Activity-Based Protein Profiling of Cysteine Proteases and Serine Hydrolases in *Plasmodium falciparum*, and Genetic Interrogation of Potential Antimalarial Targets.

Dara Davison

The Francis Crick Institute
and Imperial College London
PhD Supervisor: Dr Edgar Deu

A thesis submitted for the degree of
Doctor of Philosophy
Imperial College London
2020
**Declaration**

I, Dara Davison, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

‘The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.’
Abstract

Malaria continues to be an important global health issue. *Plasmodium falciparum*, the most virulent human-infecting malarial parasite, causes over 200 million cases per year. New drugs and drug targets are desperately needed to combat emerging resistance to frontline therapies. Two thirds of *P. falciparum* genes have predicted functions but very few have been verified. Broad-spectrum activity-based probes (ABPs) profile the activity of enzyme families based on common catalytic mechanisms. Activity-based protein profiling (ABPP) can functionally annotate many enzymes at once, as well as report on their activity in the presence of inhibitors.

In this thesis, we use ABPP to interrogate two large enzyme families in *P. falciparum*: serine hydrolases (SHs) and cysteine proteases (CPs). We describe a novel ABP, W-sCy5-VS, for the profiling of papain-like CPs (PLCPs), a family that has roles in haemoglobin degradation and red blood cell invasion in *P. falciparum*. SH activities were profiled using well-characterised fluorophosphonate (FPP) ABPs. Our first aim was to develop a medium-throughput, gel-based ABPP screen combining the two ABPs to profile both families simultaneously. The screen was performed in competition with the Malaria Box, a diverse set of anti-parasitic compounds. We identified five Malaria Box compounds that targeted SHs or PLCPs. The SH inhibitors were the most potent and we concentrated follow-up studies on this family.

*P. falciparum* has forty-eight predicted metabolic SHs, forty-three of which share a common α/β-fold. These SHs may play important roles in lipid scavenging, signalling and metabolism in *P. falciparum*, but few has verified biological function. We used FPP probes to develop chemical proteomic techniques to profile this large, uncharacterised family. We used chemical proteomics to identify possible targets of the Malaria Box SH inhibitors. By extending the approach to different life stages, we achieved unprecedented depth coverage of SHs, verifying the catalytic activity for more than half of the family. Life-stage dependent activation of human SHs was also observed, representing an exciting new area of host-targeted drug discovery. Making use of new technologies for gene editing in *P. falciparum*, we designed a conditional mutation strategy to determine if the hydrolytic activity of four SHs were important for parasite growth and replication. Contrary to data from recent large-scale genetic screens, none of the targeted genes appeared to be essential. The life-stage-dependent activation of SHs will help identify the roles that these enzymes play during the life cycle. This work illustrates how integrating chemical and genetic approaches can both efficiently identify new drug targets and chemically profile multiple enzymes. These outcomes are vital in malaria research due to the lack of functional genome annotation and the pressing need for new antimalarial targets.
Acknowledgement

I would like to sincerely thank the many people who have helped me, both professionally and personally, throughout my PhD. First to my supervisor, Dr Edgar Deu for giving me the chance to work on this exciting project and for being very generous with his time in helping me prepare my thesis. Also for his understanding during my time away and phased return. Without Edgar’s support and the wonderful Crick student team, I certainly would not have been able to return to complete my PhD after my back injury in 2017.

Thank you to the members of my thesis committee Antony Holder, Mike Blackman and Luiz de Carvalho for their insightful and kind guidance. I would like to thank all the members past and present of the Deu, Blackman and Holder labs for their help and friendship. Special thanks go to Fiona Hackett, Chris Collins, Christine Lehmann, Sophie Ridewood, Michele Tan, Emma Sherling, Anja Schlott and Abigail Perrin. I would like to thank Edward Tate for his advice throughout my PhD and for welcoming me into the Tate group at Imperial College where I gained an excellent training in proteomics. Particular thanks are due to Scott Lovell and Julia Morales Sanfrutos. I would also like to thank the Mass spectrometry STP at the Francis Crick Institute, especially Stephen Howell who worked closely with me on the development of the proteomics protocols described in this thesis.

On a personal note, I would like to thank my amazing family for their constant love and support. Thank you to Guy and Lucy for providing me with wonderful places to write my thesis. Thank you to my friends and everyone who amused and took care of me during my recovery. Special thanks go to Will J for his unsurpassed proofreading talents. To the housemates that sustained me at the start of my PhD, and to Laura, Evie, Ellie, India and Bambi for keeping me happy and relaxed during the writing process, Thank you. Last but not least to Will, my best and constant support, thank you for everything.
Table of Contents

Abstract ......................................................................................................................... 3
Acknowledgement ........................................................................................................... 4
Table of Contents ............................................................................................................. 5
List of figures .................................................................................................................. 9
List of tables ................................................................................................................... 10
Abbreviations ................................................................................................................ 11

Chapter 1. Introduction ................................................................................................. 15
  1.1 Malaria as a global health problem ...................................................................... 15
  1.1.1 The history of malaria and mankind ................................................................. 15
  1.1.2 Plasmodium species are part of the phylum Apicomplexa .............................. 16
  1.1.3 Malaria in modern times .................................................................................... 17
  1.2 The life cycle and pathogenicity of Plasmodium ............................................... 18
      1.2.1 Asymptomatic, pre-erythrocytic phase ......................................................... 20
      1.2.2 Symptomatic asexual replication in erythrocytes ........................................... 20
      1.2.3 Transmission phase ......................................................................................... 24
      1.2.4 Pathological effects during infection ............................................................... 24
  1.3 Malaria drug discovery ...................................................................................... 26
      1.3.1 Past and present strategies to prevent and eradicate malaria ..................... 26
      1.3.2 Strategic drug discovery for malaria elimination ............................................. 32
      1.3.3 Phenotypic and target based screening for drug discovery ......................... 34
      1.3.4 The Malaria Box by Medicines for Malaria Venture (MMV) ................. 35
  1.4 Enzyme targets .................................................................................................. 38
      1.4.1 Cysteine proteases ......................................................................................... 39
      1.4.2 Serine hydrolases ............................................................................................ 43
  1.5 Activity-based protein profiling (ABPP) ............................................................ 48
      1.5.1 Activity based probes (ABPs) and activity-based protein profiling (ABPP) .... 48
      1.5.2 Cysteine protease probes ............................................................................. 51
      1.5.3 Fluorophosphonate (FPP) Probes ................................................................. 54
  1.6 Chemical proteomics ......................................................................................... 55
      1.6.1 Chemical proteomics overview .................................................................... 55
      1.6.2 Chemical proteomics methods ..................................................................... 56
      1.6.3 Methods of quantification ............................................................................ 58
  1.7 Advances in genetic manipulation of Plasmodium falciparum provide novel
      opportunities to study parasite biology ................................................................. 61
      1.7.1 Conditional gene disruption systems ............................................................. 62
      1.7.2 Technologies for improving integration ......................................................... 64
      1.7.3 Conditional allelic replacement strategy ....................................................... 65
      1.7.4 High-throughput genetic screening ............................................................... 66

Chapter 2. Aims of this work ..................................................................................... 67

Chapter 3. Results I: Characterisation of W-sCy5-VS and FPP-TMR Activity-Based Probes
            (ABPs) in P. falciparum ..................................................................................... 68
  3.1 FPP probes profile P/SH activities in lysates and intact parasites ..................... 68
      3.1.1 FPP-TMR labels multiple P/SHs active at different life stages ..................... 68
      3.1.2 FPP-TMR and FPP-N3 probes have similar reactivity in P. falciparum.......... 70
  3.2 Proteomics confirms that FPP-N3 targets P/SHs .............................................. 71
  3.3 W-sCy5-VS is a broad-spectrum PLCP ABP ...................................................... 72
      3.3.1 W-sCy5-VS labels recombinant PLCPs ......................................................... 73
      3.3.2 W-sCy5-VS labels multiple unknown P/CPs at neutral pH ......................... 74
      3.3.3 W-sCy5-VS labels different P/CPs throughout the parasite life cycle .......... 75
3.4 Proteomics confirms W-sCy5-VS labelling of PLCPs in addition to multiple other reactive-Cys containing proteins .................................................. 75
3.4.1 W-dtBio-TMR-VS and W-sCy5- VS probes have the same targets in P. falciparum ................................................................. 75
3.4.2 W-dtBio-TMR- VS labels PLCPs at acidic pH and additional reactive-cysteine-containing proteins at neutral pH ............................................. 76
3.5 FPP-TMR and W-sCy5-VS can be used simultaneously to screen for reversible inhibitors ................................................................. 80
3.5.1 W-sCy5-VS and FPP-TMR do not cross-react ................................................................. 80
3.5.2 WsCy5-VS and FPP-TMR labelling can be blocked by broad-spectrum inhibitors E64 and PSMF, respectively ............................................. 82
3.5.3 Dynamic labelling conditions are required to screen for reversible inhibitors ........ 82
3.5.4 Optimisation of screening procedure for a medium-throughput ABPP screen .......... 82
3.5.5 Visual assessment of gels followed by densitometry provides quick and easy analysis ................................................................. 83
3.6 Discussion .......................................................................................................................... 84

Chapter 4. Results II: ABPP identified five compounds affecting CP or SH activity in P. falciparum ........................................................................................................... 86
4.1 ABPP at schizont and trophozoite stages identified five Malaria Box inhibitors.... 86
  4.1.1 The Malaria Box ABPP screen resulted in sixteen hits, eight of which were reproducible under the same conditions ........................................ 86
  4.1.2 Confirmation of potential hits identified five compounds for follow-up studies .... 88
4.2 The multiplexed ring screen did not produce any original hit compounds for follow-up studies ................................................................. 92
4.3 Two compounds increase W-sCy5-VS labelling of multiple targets including FP3 93
  4.3.1 M941 (but not M738)-dependent W-sCy5-VS labelling is E64 sensitive ................. 94
  4.3.2 M941 increases W-sCy5-VS labelling of rFP3 at acidic pH ................................. 94
  4.3.3 M941 inhibits recombinant FP3 in a fluorogenic substrate assay .................. 95
4.4 M563 is a broad-spectrum SH inhibitor ................................................................. 96
  4.4.1 M563 blocks parasite development ................................................................. 97
  4.4.2 M563 is likely to inhibit by a reversible tight-binding mechanism ......... 98
4.5 M438 and M914 appear to target the same SH but may have other distinct targets ................................................................. 101
  4.5.1 M438 and M914 block parasite development .................................................. 101
  4.5.2 M438 inhibits the Coenzyme A (CoA) biosynthesis pathway but this is not the FPP-labelled SH target .................................................. 103
4.6 Discussion .......................................................................................................................... 104

Chapter 5. Results III: Chemical proteomics prioritises PfSHs for genetic validation .... 108
5.1 Optimisation of chemical proteomics protocols achieved increased depth of coverage ........................................................................ 108
  5.1.1 FPP-N3 was a better SH probe for chemical proteomics than FPP-Biotin ....... 108
  5.1.2 Dilution of neutravidin linked agarose beads identified ideal pull-down conditions .......................................................................... 110
5.2 Comparative chemical proteomics for Malaria Box target deconvolution .... 112
5.3 ABPP of SHs during the P. falciparum blood-stage asexual life cycle ............... 115
  5.3.1 Optimised chemical proteomics methods achieve good depth of coverage of PfSH’s ................................................................. 115
  5.3.2 Levels of active SHs vary throughout the asexual life cycle ....................... 117
5.4 Four SHs are prioritised for genetic validation .............................................. 119
5.5 Discussion .................................................................................................................. 120

Chapter 6. Results IV: Human SHs may be potential antimalarial targets ........ 124
6.1 ABPP showed changing levels of active human SHs throughout the P. falciparum asexual blood-stage life cycle ............................................. 124
Chemical Proteomics

Molecular biology techniques for the generation of plasmids for genetic modification.

Transfection, cloning and testing of P. falciparum conditional allelic replacement lines.

Chapter 10 Appendix

Organisation of lysis conditions for ABPP

Malaria Box ABPP screening data

Malaria Box screening gels

Data for unconfirmed Malaria Box screen hits.

Inhibition of labelling by Malaria Box compounds was not seen in intact parasites.

Analouges of the M563 and M438 did not inhibit SHs.

Discussion

Generation of conditional mutant parasite lines.

A conditional allelic replacement strategy to test the essentiality of active Psta1, Psta2, abH112 and abH114.

Clonal conditional mutant and WT control lines were generated.

Expression of catalytically-dead mutant SHs has no effect on parasite replication.

PCR's confirm DiCre mediated excision in conditional parasite lines.

Loss of HA-tagged WT protein after excision is detectable by western blot.

None of the conditional mutation lines caused an effect in parasite replication.

Measurement of parasitaemia by flow cytometry.

The Malaria Box ABPP competition screen

The Malaria Box Screening conditions.

Confirmation and retesting of Malaria Box hits.

Fluorogenic assay to measure FP3 activity and inhibition.

Standard replication assays to measure inhibition of parasite growth.

Measurement of parasitaemia by flow cytometry.

Human SHs

Monitoring of parasite replication in the presence of HsSH inhibitors.

Genetic validation of SHs.

Molecular biology techniques for the generation of plasmids for genetic modification.

Transfection, cloning and testing of P. falciparum conditional allelic replacement lines.

Inhibitors of human SH's block parasite replication.

Human SH inhibitors block parasite replication.

LYPLA1 and APEH inhibitors block parasite replication in the first cycle.

Discussion

Results V: Genetic interrogation of SHs in P. falciparum

General replication assays to measure inhibition of parasite growth.

Fluorogenic assay to measure FP3 activity and inhibition.

Confirmation and retesting of Malaria Box hits.

The Malaria Box ABPP competition screen.

Chemical Proteomics

Protein labelling by ABPs.

Tandem-tagging by CuAAC-chemistry.

Affinity purification of labelled proteins.

Trypsin digestion of proteins and peptide sample clean-up by stage tipping.

TMTsix/tenplex™ labelling and high pH fractionation.

Proteomics method.

Proteomics analysis.

Interpretation of data.

Human SHs

Monitoring of parasite replication in the presence of HsSH inhibitors.

Genetic validation of SHs.

Molecular biology techniques for the generation of plasmids for genetic modification.

Transfection, cloning and testing of P. falciparum conditional allelic replacement lines.
List of figures

**Figure 1.2.1:** The *Plasmodium falciparum* life cycle. 19
**Figure 1.2.2:** Architecture of a mature schizont and merozoite. 23
**Figure 1.3.1:** The chemical structures of common antimalarial drugs. 30
**Figure 1.3.2:** The MMV Malaria Box. 36
**Figure 1.4.1:** Basic common structure of a PLCP active site. 41
**Figure 1.4.2:** The $\alpha/\beta$-H fold. 45
**Figure 1.5.1:** The structures of inhibitors and ABPs that target PLCPs. 52
**Figure 1.5.2:** The structure of FPP ABPs that target SHs. 55
**Figure 1.7:** Selection-linked integration. 65
**Figure 3.1:** Characterisation of broad spectrum FPP probes. 69
**Figure 3.2:** Characterisation of broad spectrum CP probe W-sCy5-VS. 74
**Figure 3.3:** Characterisation of broad spectrum CP probe W-sCy5-VS. 74
**Figure 3.4.1:** Confirmation of W-dtBio-TMR-VS reactivity. 76
**Figure 3.4.2:** Chemical proteomics protocol for target identification of W-sCy5-VS/W-dtBio-TMR-VS. 78
**Figure 3.5.1:** Optimisation of FPP-TMR and W-sCy5-VS labelling conditions for ABPP screening. 81
**Figure 3.5.2:** Medium-throughput gel-based ABPP screening procedure. 83
**Figure 4.1.1:** The Malaria Box ABPP screen results. 87
**Figure 4.1.2:** Final hits from the Malaria Box screen. 89
**Figure 4.1.3:** Dose dependent inhibition/activation of probe labelling by the Malaria Box hit compounds. 90
**Figure 4.2:** The Malaria Box ABPP ring screen results. 93
**Figure 4.3.1:** W-sCy5-VS labelling in the presence of Malaria Box compounds M941 and M738. 95
**Figure 4.3.2:** Catalytic activity of FP3 in the presence of compound M941. 96
**Figure 4.4.1:** Standard replication assay and phenotypic effects with M563. 98
**Figure 4.4.2:** Investigation into the inhibition mode of M563. 100
**Figure 4.5.1:** Standard replication assay and phenotypic effects of M438 and M914. 102
**Figure 4.5.2:** Standard replication assays with CoA rescue. 104
**Figure 5.1.1:** Competition of broad spectrum SH FPP probes. 109
**Figure 5.1.2:** Optimisation of affinity purification for chemical proteomics. 111
Figure 5.2: Comparative proteomics to identify Malaria Box compound targets. 114
Figure 5.3.1: ABPP protocol for the profiling of SHs throughout the *P. falciparum* asexual-blood stage life-cycle. 116
Figure 5.3.2: ABPP profiling of SHs throughout the *P. falciparum* asexual-blood stage life-cycle. 118
Figure 6.1: Asexual life cycle proteomics results, *Homo sapiens.* 125
Figure 6.2.1: Standard replication assay with human SH inhibitors. 128
Figure 6.2.2: Flow cytometry parasite life stage grouping. 130
Figure 6.2.3: Replication assay to monitor the effect of human SH inhibitors on parasite growth and replication. 132
Figure 6.2.4: Quantification of replication assay with HsSH inhibitors and controls. 133
Figure 7.1.1: Strategy for conditional allelic replacement. 139
Figure 7.1.2: Integration of conditional mutants. 141
Figure 7.2.1: PCRs and western blots confirm DiCre-mediated excision. 144
Figure 7.2.2: Replication assay with conditional mutation lines over multiple cycles. 145
Figure 9.6: The generation of pT2A-GOI-cMut/WT constructs for transfection 175

List of tables

Table 1  
\(\alpha/\beta\)Hs of *P. falciparum.* 46
Table 3.2  
Proteins enriched in FPP-N\(_3\) compared to DMSO 72
Table 3.4.1  
Proteins enriched in W-dtBio-TMR-VS compared to DMSO, pH 5.5 79
Table 3.4.2  
Proteins enriched in W-dtBio-TMR-VS compared to DMSO, pH 7.2 79
Table 4  
Comparison of IC\(_{50}\) and EC\(_{50}\) values of Malaria Box compounds. 91
Table 5.1.1  
Proteins enriched in FP-Biotin compared to DMSO 110
Table 5.1.2  
Proteins identified in samples after different affinity purification conditions 112
Table 5.2  
Comparative proteomics with Malaria box inhibitors 115
Table 6  
Human SHs and their inhibitors 126
Table 9.1  
Human SH inhibitors 170
Table 9.2  
PCRs for the construction of transfection plasmids 173
Table 9.3  
PCRs for the interrogation of parasite lines 173
Table 9.4  
List of oligonucleotide primers 174
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>abH112</td>
<td>PF3D7_1120400</td>
</tr>
<tr>
<td>abH114</td>
<td>PF3D7_1143000</td>
</tr>
<tr>
<td>ABPP</td>
<td>activity-based protein profiling</td>
</tr>
<tr>
<td>ABPs</td>
<td>activity based probes</td>
</tr>
<tr>
<td>AChE</td>
<td>acetylcholine esterase</td>
</tr>
<tr>
<td>ACTs</td>
<td>artemisinin combination therapies</td>
</tr>
<tr>
<td>ALAD</td>
<td>sigma-aminolevulinate dehydratase</td>
</tr>
<tr>
<td>APEH</td>
<td>acyl-amino acid releasing enzyme</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BZTs</td>
<td>benzothiazepines</td>
</tr>
<tr>
<td>C2</td>
<td>compound 2</td>
</tr>
<tr>
<td>Cas9</td>
<td>CRISPR-associated protein 9</td>
</tr>
<tr>
<td>CatC</td>
<td>cathepsin C</td>
</tr>
<tr>
<td>CESTA</td>
<td>cellular thermal shift assays</td>
</tr>
<tr>
<td>cKO</td>
<td>conditional knock-out</td>
</tr>
<tr>
<td>cMUT</td>
<td>conditional mutant</td>
</tr>
<tr>
<td>CoA</td>
<td>coenzyme A</td>
</tr>
<tr>
<td>CP</td>
<td>cysteine protease</td>
</tr>
<tr>
<td>CQ</td>
<td>chloroquine</td>
</tr>
<tr>
<td>CRISPR</td>
<td>clustered regularly interspaced short palindromic repeats</td>
</tr>
<tr>
<td>CSP</td>
<td>circumsporozoite rotein</td>
</tr>
<tr>
<td>CuAAC</td>
<td>copper(I)-catalysed alkyne-azide cycloaddition</td>
</tr>
<tr>
<td>cWT</td>
<td>conditional wild-type</td>
</tr>
<tr>
<td>DDT</td>
<td>dichlorodiphenyltrichloroethane</td>
</tr>
<tr>
<td>DiCre</td>
<td>dimerisable cre recombinase</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulphoxide</td>
</tr>
<tr>
<td>DPAP</td>
<td>dipeptidyl aminopeptidase</td>
</tr>
<tr>
<td>DSB</td>
<td>double strand break</td>
</tr>
<tr>
<td>EC$_{50}$</td>
<td>half maximal effective concentration</td>
</tr>
<tr>
<td>EETs</td>
<td>epoxyeicosatrienoic acids</td>
</tr>
<tr>
<td>ESTHER</td>
<td>esterases and a/b-hydrolase enzymes and relatives</td>
</tr>
<tr>
<td>ETC</td>
<td>electron transport chain</td>
</tr>
<tr>
<td>etramps</td>
<td>early transcribed membrane protein genes</td>
</tr>
<tr>
<td>Ex/Em</td>
<td>excitation/emission</td>
</tr>
<tr>
<td>FASN</td>
<td>fatty acid synthase</td>
</tr>
<tr>
<td>FDR</td>
<td>false discovery rate</td>
</tr>
<tr>
<td>FPPs</td>
<td>fluorophosphonates</td>
</tr>
<tr>
<td>FPs</td>
<td>falcipains</td>
</tr>
<tr>
<td>FQ</td>
<td>ferroquine</td>
</tr>
<tr>
<td>FSC</td>
<td>forward scatter</td>
</tr>
<tr>
<td>G6PD</td>
<td>glucose-6-phosphate dehydrogenase</td>
</tr>
<tr>
<td>GFP</td>
<td>green fluorescent protein</td>
</tr>
<tr>
<td>GOI</td>
<td>gene of interest</td>
</tr>
<tr>
<td>GSK</td>
<td>GlaxoSmithKline</td>
</tr>
<tr>
<td>h</td>
<td>hour</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Hb</td>
<td>haemoglobin</td>
</tr>
<tr>
<td>hDHFR</td>
<td>human dihydrofolate reductase</td>
</tr>
<tr>
<td>HDP</td>
<td>haem detoxification protein</td>
</tr>
<tr>
<td>HLPC</td>
<td>high performance liquid chromatography</td>
</tr>
<tr>
<td>hpi</td>
<td>hours post invasion</td>
</tr>
<tr>
<td>hPrx-2</td>
<td>peroxin 2</td>
</tr>
<tr>
<td>HR</td>
<td>homology region</td>
</tr>
<tr>
<td>HTS</td>
<td>high-throughput screen</td>
</tr>
<tr>
<td>iBAQ</td>
<td>intensity-based absolute quantification</td>
</tr>
<tr>
<td>IC₅₀</td>
<td>half maximal inhibitory concentration</td>
</tr>
<tr>
<td>ID</td>
<td>identification</td>
</tr>
<tr>
<td>IFA</td>
<td>immunofluorescence analysis</td>
</tr>
<tr>
<td>iRBC</td>
<td>infected red blood cell</td>
</tr>
<tr>
<td>IRS</td>
<td>indoor residual spraying</td>
</tr>
<tr>
<td>ITNs</td>
<td>insecticide treated bed nets</td>
</tr>
<tr>
<td>iTRAQ</td>
<td>isobaric tags for relative and absolute quantification</td>
</tr>
<tr>
<td>IVEWGA</td>
<td>in vitro evolution and whole-genome analysis</td>
</tr>
<tr>
<td>K13</td>
<td>K13 Kelch protein</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>kDa</td>
<td>kilodalton</td>
</tr>
<tr>
<td>KO</td>
<td>knockout</td>
</tr>
<tr>
<td>LB</td>
<td>loading buffer</td>
</tr>
<tr>
<td>LC-MS/MS</td>
<td>liquid chromatography and tandem mass spectrometry</td>
</tr>
<tr>
<td>LCAT</td>
<td>PF3D7_0629300</td>
</tr>
<tr>
<td>LFQ</td>
<td>label-free quantification</td>
</tr>
<tr>
<td>LLIINs</td>
<td>lasting insecticide treated bed nets</td>
</tr>
<tr>
<td>LYPLA1/2</td>
<td>acyl protein thioesterases</td>
</tr>
<tr>
<td>M127</td>
<td>MMV007127</td>
</tr>
<tr>
<td>M389</td>
<td>MMV006389</td>
</tr>
<tr>
<td>M438</td>
<td>MMV011438</td>
</tr>
<tr>
<td>M563</td>
<td>MMV000563</td>
</tr>
<tr>
<td>M672</td>
<td>MMV645672</td>
</tr>
<tr>
<td>M738</td>
<td>MMV019738</td>
</tr>
<tr>
<td>M914</td>
<td>MMV665914</td>
</tr>
<tr>
<td>M941</td>
<td>MMV665941</td>
</tr>
<tr>
<td>malERA</td>
<td>Malaria Eradication Agency</td>
</tr>
<tr>
<td>MC</td>
<td>Maurer’s clefts</td>
</tr>
<tr>
<td>MF</td>
<td>mefloquine</td>
</tr>
<tr>
<td>MLPL</td>
<td>PF3D7_1328500</td>
</tr>
<tr>
<td>MMV</td>
<td>Medicines for Malaria Venture</td>
</tr>
<tr>
<td>MoA</td>
<td>mechanisms of action</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometer</td>
</tr>
<tr>
<td>MSP</td>
<td>merozoite surface protein</td>
</tr>
<tr>
<td>NBs</td>
<td>neutravidin beads</td>
</tr>
<tr>
<td>NCEH1</td>
<td>neutral cholesterol esterase</td>
</tr>
<tr>
<td>NPT</td>
<td>neomycin phosphotransferase</td>
</tr>
<tr>
<td>OP</td>
<td>organophosphophate</td>
</tr>
</tbody>
</table>
PAFAH1B3  platelet activating factor acetyl hydrolase 1B gamma
PAGE  polyacrylamide gel electrophoresis
PBS  phosphate-buffered saline
PBS  phosphate buffered saline
PBS-T  0.01% tween in PBS
PBSa  phosphate buffered saline buffer A
PC  phosphatidylcholine
PCR  polymerase chain reaction
PEP  posterior error probabilities
PEXEL  *Plasmodium* export element
PFA  paraformaldehyde
PKG  cyclic-GMP (cGMP) dependent protein kinase
PLA2  phospholipase A2
*PlasmoGEM*  *Plasmodium* Genetic Modification Project
PLCPs  papain-like clan CA CPs
PNPLA6  neuropathy target esterase
PPI  protein-protein interaction
PSM  peptide-spectrum match
Psta1  PF3D7_1476700
Psta2  PF3D7_1476800
PV  parasitophorous vacuole
PVM  PV membrane
PyBr  pyridostigmine bromide
RAP  rapamycin
RBC  red blood cell
ROM4  PF3D7_0506900
RON  rhoptry neck protein
RR1/2  recodonised region 1/2
RT  room temperature
S9C  PF3D7_0403800
SAR  structure-activity relationship
SDS  sodium dodecyl sulphate
SDS-PAGE  sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC  single exposure chemoprotection
SENP1/2  small ubiquitin-related modifier(SUMO)-specific proteases
SERA6  serine-repeat antigen 6
SERA6s  serine repeat antigens
SERCaP  single encounter radical cure and prophylaxis
SHs  serine hydrolases
SILAC  stable isotope labelling with amino acids in cell culture
SIT  sterile insect technique
SLI  selection-linked integration
SP  sulfadoxine and pyrimethamine
SSC  side scatter
SUB1  subtilistin-like serine protease 1
SUMO  small ubiquitin-related modifier
T2A  2A element from the virus *thosea asigna*
T2A  
TCPs  target candidate profiles
TMT  tandem-mass tagging
TPPs  target product profiles
Trop1Q  MMV183
TVN  tubovesicular network
uRBC  uninfected RBC
UTR  untranslated region
WHO  World Health Organisation
WT  wild-type

*those asigna*
Chapter 1. Introduction

1.1 Malaria as a global health problem

1.1.1 The history of malaria and mankind

Mosquito vectors have been spreading malaria in tropical regions across the globe for far longer than human existence. The earliest evidence of malaria was found preserved inside mosquitoes fossilised in amber, from around 20 million years ago (Poinar, 2005). The history of the relationship between humans and malaria extends into antiquity. Malarial DNA has been detected in the tombs of ancient Egypt from 1500-500 BC (Nerlich et al., 2008). The characteristic cycles of fever caused by the disease were described as early as 2700 BC in ancient Chinese medical texts (Chinese Canon of Medicine, Nei Ching) and later in the records of Sumerian, Assyrian, Babylonian, Egyptian, Indian, Greek and Roman scholars and physicians. Homer (circa 750 BC) even mentioned malaria in The Iliad, linking its fever and misery with the appearance of the dog star Sirius in late summer/autumn. Historians have speculated that malaria epidemics contributed to the fall of Rome (Sallares, 2002). At that time, malarial fevers were linked to swamp miasma, and it is thought that the word malaria comes from the Italian “mala aria” or “bad air”. The discovery of the real cause of malaria required many stepping stones: the development of germ theory (Louis Pasteur and Robert Koch, 1878-1879); discovery of the Plasmodium parasites by Charles Louis Alphonse Laveran in 1880; and finally, identification of mosquitoes as the vector (Sir Ronald Ross 1897, Giovanni Battista Grassi 1898).

The close ancient relationship between malaria and humans is exhibited in the prevalence of certain gene alleles in endemic populations. For example, the mutant β-globin allele that causes sickle cell anaemia in homozygous individuals is pervasive in regions with high malaria incidence because despite the serious (and often life-threatening) disease, carriers of the allele have a tenfold decrease risk of serious malaria (Allison, 1954; Piel et al., 2010). There is a dynamic balance between the negative selection of the disease state and the positive selection of heterozygous protection in the population. Glucose-6-phosphate dehydrogenase (G6PD) deficiency is another example of this kind of polymorphism (Tishkoff et al., 2001). This close connection of host and parasite in the genetic code often has implications for drug development (Section 1.3.1.3.1).
1.1.2 *Plasmodium* species are part of the phylum Apicomplexa

*Plasmodium* species are a part of the phylum Apicomplexa: single cellular eukaryotic organisms characterised by their complex apical structures and parasitic lifestyle (apart from the marine endosymbiotic *Nephromyces*; Saffo et al., 2010). Most Apicomplexa, including *Plasmodium*, also have an unusual organelle called the apicoplast thought to be a relic of a photosynthetic endosymbiote (Kalanon and McFadden, 2010). Apicoplast function is essential *Plasmodium* and is thought involve the synthesis of crucial molecules, such as isoprenoids (Guggisberg et al., 2014). The phylum contains many parasites important for human and animal health including *Toxoplasma gondii* and *Babesia* species.

*Plasmodium* species have been causing malaria in diverse organisms for millions of years. Currently over one hundred modern species have been identified affecting reptiles, birds and mammals (Martinsen et al., 2008). All are transmitted by mosquito vectors, which is an essential step in their lifecycles. There are five known species that cause malaria in humans: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium knowlesi*, *Plasmodium malariae*, *Plasmodium ovale curtisi* and *Plasmodium ovale wallikeri*. *P. falciparum*, the most virulent species, predominates in sub-Saharan Africa (99.7% of cases) and *P. vivax* is most prevalent in South America (75% of cases; WHO, 2019). *P. falciparum* is responsible for the majority of malaria mortality. The study of *P. vivax* is neglected in comparison to *P. falciparum*, even though it still causes significant morbidity and high risk of relapse (Battle et al., 2014; Mendis et al., 2001). Both *P. falciparum* and *P. vivax* have been shown to have evolved in Africa from parasites that originally infected apes (Liu et al., 2014, 2010). *P. falciparum* is considered part of the *Laverania* sub-genus of *Plasmodium*, all the other members of which, including *P. reichenowi* and *P. gaboni*, infect African great apes (Bray, 1958). *P. knowlesi* illustrates the zoonotic potential of *Plasmodium* species: it primarily infects macaques in South East Asia, but is able to switch hosts and infect humans (Cox-Singh et al., 2008; Singh et al., 2004).

*P. knowlesi*, *P. malariae* and *P. ovale* are generally considered to be at a low level of prevalence, however new research suggests that their contribution to malaria burden may be more serious than previously thought via mixed infections and misdiagnosis as *P. falciparum* (Bichara et al., 2017; Bruce et al., 2000; Mayxay et al., 2004). *P. vivax*, *P. malariae* and *P. ovale* cannot be continuously cultured *in vivo* which hampers efforts to understand their biology.
1.1.3 Malaria in modern times

The WHO (World Health Organisation) still classifies malaria as a global health problem. In 2018, there were an estimated 228 million cases worldwide. Although this is significantly less than in 2010 (251 million cases), the WHO predicts that unless change is accelerated, milestones set out in The Global Technical Strategy for Malaria 2016-2030 will not be met (WHO, 2019). Malaria incidence is confined to areas that are hospitable for the vectors, the *Anopheles* mosquito. These are in broad tropical regions around the equator including much of Africa, South America, South-East Asia, and India. The decrease in incidence over recent years is, in part, due to prevention efforts. In 2018, 50% of the population in sub-Saharan Africa slept under insecticide-treated bed nets (Section 1.3.1.1.2), compared to 29% in 2010 (WHO, 2019). However, rising average global temperatures have sparked concerns that in the future larger areas could be at risk. In 2018, 93% of cases were in Africa with just six countries in Sub-Saharan Africa carrying more than half of the malaria burden (Nigeria, Democratic Republic of Congo, Uganda, Côte d’Ivoire, Mozambique and Niger (WHO, 2019). The mortality is broadly the same with 94% of the 405,000 malaria-related deaths in 2018 coming from the Sub-Saharan African region (WHO, 2019).

The disconnect between the morbidity and mortality of malaria has caused public perception challenges. Most malaria cases are mild. Healthy adults rarely die after a malaria infection and may suffer from multiple acute presentations in their lifetime (Section 1.2.4.1). This creates an expectation of health; where adults are used to living in the midst of a disease, they are less likely to spend time, money and effort travelling to clinics to pick up preventative drugs. The adult morbidity caused by malaria reduces the effective size of the workforce putting a huge economic burden on endemic-countries (Gallup and Sachs, 2001; Malaney et al., 2004). In endemic areas children are the biggest sufferers, the loss of time in education for the child and in work for caring parents has long-term socio-economic consequences.

As well as strong geographical factors, the burden of malaria on a country is linked to the development of its infrastructure. Malaria is often considered as a disease of poverty, as it disproportionally affects underdeveloped countries with low GDP (Gallup and Sachs, 2001). Before even considering public healthcare resources, infrastructure such as road coverings, more enclosed housing and air conditioning can have a huge effect on mosquito numbers and biting. There have been malaria elimination successes after drainage programs of mosquito breeding grounds such as swamps and other still bodies of water (Section 1.3.1.1.1). However, these measures alone cannot eliminate the vectors. Mosquitos can lay eggs in the smallest bodies of stagnant water, for instance rainfall collected in the hoof impression of an animal. In many low
GDP countries, there are also inherent difficulties arising from supply of interventions, medicines and treatment. Many communities are isolated, making it difficult both for people to access treatment and for healthcare professionals to monitor incidence numbers, treatment compliance and patient outcome. Any new diagnostic tool, treatment or preventive measure must take into account the feasibility of being implemented within the available infrastructure. All interventions must also be thoughtfully considered in the context of the countries and communities they target. In a stark illustration of foreign aid failing to understand the communities they wished to help, there have been many anecdotal cases of bed nets being used as fishing nets. Recently reviewed on a global level, this practise has long-term consequences such as overfishing of lakes due to too many juvenile fish being caught in the fine nets (Short et al., 2018). Banning fishing with bed nets however has ethical implications in communities affected by extreme poverty. Despite the economic burden of malaria, 30% of funding for malaria still comes from governments of endemic counties, totalling US$ 0.9 billion in 2018 (WHO, 2019).

1.2 The life cycle and pathogenicity of *Plasmodium*

*Plasmodium* parasites have a complex life cycle that involves both sexual and asexual replication and requires transmission between humans and mosquitos (Fig. 1.2.1). We commonly refer to the human “host” and mosquito “vector”. As both parasite development and sexual reproduction takes place in the mosquito however, according to the strictest definitions both humans and mosquitos are hosts of malaria. In the human host, there is first an asymptomatic phase in the liver, and an intra-erythrocytic phase responsible for the majority of symptoms. Below is described the basic life cycle for the five human infecting *Plasmodium* species, with a particular focus on *P. falciparum*. 
Figure 1.2.1: The *Plasmodium falciparum* life cycle. Three important phases in the *P. falciparum* life cycle are shown. *P. falciparum* parasites are transmitted to humans by the bite of a mosquito. Parasites then undergo an asexual-asymptomatic phase in the liver (hepatic stage) where thousands of merozoites are produced inside hepatic cells. Upon egress these merozoites enter the circulatory system and the asexual-erythrocytic stage. This stage causes symptoms in human patients. Some parasites leave the erythrocytic cycle and enter the sexual phase via gametocytogenesis. Gametocytes are taken up by mosquitoes if they bite an infected host, and the sexual cycle completes inside the mosquito. Reproduced with permission from (Philip J. Rosenthal, April 24, 2008, N Engl J Med 2008; 358:1829-1836 DOI: 10.1056/NEJMct0709050), Copyright Massachusetts Medical Society.
1.2.1 Asymptomatic, pre-erythrocytic phase

*Plasmodium* infected female *Anopheles* mosquitos have haploid, motile sporozoites in their salivary glands (Fig. 1.2.1). When a mosquito bites a human, around one hundred sporozoites are transferred to the human dermis via the mosquito saliva (Frischknecht et al., 2004; Medica and Sinnis, 2005; Rosenberg et al., 1990). The sporozoites migrate through the epidermis into the circulatory system (Amino et al., 2006). When they encounter the liver sinusoids, sporozoites enter and travel through hepatocytes via an unusual sequence of breach and rapid repair of the hepatocyte plasma membrane (Mota et al., 2001). At some point, the parasite surface circumsporozoite protein (CSP) binds to heparin sulphate proteoglycans on a hepatocyte (Coppi et al., 2007; Frevert et al., 1993). The sporozoites invades the hepatocyte, creating itself a parasitophorous vacuole (PV) in which it resides (Mota et al., 2001; Risco-Castillo et al., 2015). Inside the hepatocyte schizogony occurs, whereby the parasite goes through multiple rounds of nuclear division without cytokinesis to give a multinuclear schizont. The parasite plasma membrane then invaginates around each nucleus to enclose thousands of hepatic merozoites, the infective agents in the next stage of the parasite life cycle (Sturm et al., 2009). When merozoites egress, the PV membrane (PVM) ruptures, the hepatocyte is destroyed and merozoites are released in protective merosomes derived from the hepatocyte plasma membrane (Tarun et al., 2006; Vaughan et al., 2012). Merosomes travel through hepatic blood vessels to reach the circulatory system, protecting the merozoites from the host immune system (Sturm et al., 2006). Merosomes are thought to rupture in the pulmonary capillary system, where they can invade red blood cells (RBCs) and begin the asexual symptomatic part of the life cycle (Baer et al., 2007). In *P. vivax* and *P. ovale* infections some sporozoites do not go through schizogony and stay quiescent inside hepatocytes. These forms, known as hypnozoites, can be activated by unknown mechanisms up to two years after initial infection, causing relapse in disease (White, 2011).

1.2.2 Symptomatic asexual replication in erythrocytes

The description of the asexual stage of development is made easier by splitting it into stages according to the morphology of the parasite: ring, trophozoite, schizont and merozoite (Fig. 1.2.1). *P. falciparum* parasites have a 48-h life cycle in which genes are expressed in cyclic manner, 80% of genes have peak mRNA levels at a single point over the 48 h (Bozdech et al., 2003).
1.2.2.1 Invasion and remodelling of the RBC

Invasion of RBCs by merozoites happens in multiple, highly-regulated steps: 1. Attachment of merozoite to the RBC surface; 2. Deformation of RBC membrane; 3. Orientation of the merozoite apical end to RBC surface; 4. Formation of a moving junction and entry into RBC by actin-myosin motor action; 5. Resealing of the RBC membrane; and 6. Formation of the PVM (Dvorak et al., 1975; Gilson and Crabb, 2009; Riglar et al., 2011; Weiss et al., 2015).

There is a vast amount of remodelling of the host RBC upon invasion. An extensive exomembrane system is generated inside the infected RBC (iRBC) (Sherling and van Ooij, 2016). The parasites reside in a PV, as described in hepatocytes, but there is a network of other membranous structures including Maurer’s clefts (MC), the cavelo-vesicle complex, J dots and the tubovesicular network (TVN). This network is thought to be crucial in facilitating the export of parasite proteins into the RBC lumen and membrane. Hundreds of proteins are exported into the RBC cytosol performing multiple functions: virulence, immune invasion, uptake of haemoglobin (Hb), and import of nutrients from serum (Spielmann and Gilberger, 2015). For example, *P. falciparum* erythrocyte protein 1 (*Pf*EMP1) is exported to the iRBC membrane and plays a key role in RBC adhesion, preserving mature stages from circulation and preventing their recognition by the immune system (Baruch et al., 1995; Crabb et al., 1997). There are changes in RBC shape and deformability by proteins that bind to cytoskeleton (Maier et al., 2009, 2008). Exported parasite proteins also mediate the increase in influx of glucose and other nutrients into the RBC lumen (Nguitragool et al., 2011). The parasite also co-opts RBC proteins for its own benefit. Several host enzymes are now confirmed to be imported into the parasite such as sigma-aminolevulinate dehydratase (ALAD; Bonday et al., 2000) and peroxidin 2 (hPrx-2) for peroxide detoxification (Koncarevic et al., 2009).

There is still much debate on the formation mechanism of the PVM and the main source of its lipids. Initial studies using fluorescent lipophilic probes suggested that the majority of lipids were derived from the RBC membrane during invasion (Haldar and Uyetake, 1992; Pouvelle et al., 1994; Ward et al., 1993). However, further experiments refuted these results (Dluzewski et al., 1995) and it was found that no significant change in RBC surface area takes place during invasion (Mikkelsen et al., 1988). One proposed parasite-derived source of PVM phospholipids are the rhoptries, apical merozoite secretory organelles that discharge their contents during invasion concomitantly with PVM formation (Aikawa et al., 1981; Bannister and Mitchell, 1989). Opinions remain divided (Sherling and van Ooij, 2016); however, it is likely that PVM formation
involves both host- and parasite-derived material, the rhoptries being the probable source of the latter.

1.2.2.2  Parasite growth and metabolism

The ring stage of the parasite, so called for its distinctive shape in Giemsa-stained thin blood films, lasts for the first 20 h of the cycle. The PVM must continue to expand to accommodate the growing intracellular parasite. As the RBC has little to no de novo lipid biosynthesis, it is assumed that the lipids required are parasite derived or scavenged from plasma (Marks et al., 1960). It has been shown that *P. falciparum* parasites in vitro culture require fatty acids, but not phospholipids, from the media (Asahi et al., 2005). The parasite has the capacity to make its own phospholipids, although the remarkable complexity of this process and the difference between human and rodent malaria models have somewhat confused this issue (Déchamps et al., 2010a, 2010b). The parasite eventually develops into a trophozoite with a more rounded morphology. This stage is characterised by the digestion of Hb from the RBC lumen. Hb digestion provides a source of amino acids for protein synthesis and creates more space for the parasite to grow (Elliott et al., 2008). Hb digestion takes place in the highly specialised digestive vacuole (Goldberg et al., 1990) where liberated haem is detoxified and crystallises into haemoglobin (Coronado et al., 2014; Egan et al., 2002). A complex of numerous cysteine and aspartyl and metallo-proteases are involved in the degradation of Hb (Section Error! Reference source not found.) and haem detoxification protein (HDP) catalyses haem biocrystallisation (Jani et al., 2008). This process is highly important for drug activity and resistance, both quinoline and artemisinin-based antimalarials target the process via different mechanisms, and mutations in Hb-processing proteases has been linked to reduced drug sensitivity (Section 1.3.1.3.3; Section 1.4.1.3.2).

1.2.2.3  Schizogony

In *P. falciparum*, schizogony commences approximately 36 hours after host cell invasion. As with hepatic schizogony, this involves multiple rounds of asynchronous nuclear division without cytokinesis, followed by invagination of the parasite membrane to enclose daughter merozoites (White and Kilbey, 1996). Mature erythrocytic schizonts have only between 8 and 32 nuclei. The merozoites are highly organised structures with specialised secretory organelles at their apical end: exonemes for egress; micronemes and rhoptries for invasion, and dense granules (Fig. 1.2.2).
1.2.2.4 Merozoite egress

The egress of merozoites is a highly-regulated and tightly-controlled event (Blackman, 2008). A master trigger has yet to be discovered but the identification of various key players has led to the development of small molecule tools that allow us to study the process in great detail. For example, the cyclic-GMP dependent kinase (PKG) has been shown to regulate the release of the contents of secretory organelles: exonemes and micronemes. This initiates a protease cascade that ultimately results in egress (Collins et al., 2013b; Thomas et al., 2018). A reversible inhibitor of PKG, Compound 2 (C2), can be used to arrest mature schizonts just before egress (Collins et al.,

Figure 1.2.2: Architecture of a mature schizont and merozoite. Illustration depicting a mature-stage schizont in the asexual-blood stage of *P. falciparum*. One of the merozoites inside the parasitophorous vacuole membrane (PVM) is coloured and labelled to show the main features and organelles of the parasite. Also labelled are haemoglobin crystals and features of the infected RBC including knobs on the RBC membrane surface and the tubovesicular network. Artwork is an original zinc plate etching by the author.
This is an important tool as it provides a checkpoint, helping to order other egress events. Practically, C2 can be used to achieve highly synchronous egress of a culture of schizonts by stalling the population for 2 to 4 h and then washing the inhibitor away. The subtilisin-like protease 1 (SUB1) is released from exonemes and cleaves a number of substrates including cysteine protease Serine-repeat antigen 6 (SERA6) (Thomas et al., 2018). SERA6 is thought to be one of the targets of cysteine protease inhibitor, E64. E64 halts egress downstream of PKG activation where PVM, but not RBC membrane, rupture has occurred (Thomas et al., 2018; Section 1.4.1.3.1). Downstream of SERA6 activation therefore, the RBC membrane is compromised and merozoites are released, closing the asexual replication cycle by invading a new RBC.

1.2.3 Transmission phase

A subset of asexual-stage parasites enter the sexual cycle and undergo gametocytogenesis. Initiated by environmental cues, merozoites in a mature schizont commit to the sexual path. After egress newly-invaded rings become male or female gametocytes (Bruce et al., 2000; Ngotho et al., 2019; Silvestrini et al., 2000). After 9 to 12 days of maturation (in *P. falciparum*) the gametocytes are transmission-competent, meaning that if a mosquito now feeds on the infected person, the cells are taken up and will continue sexual development (Hawking et al., 1971). Once in the mosquito mid-gut, the different environment causes male and female gametes to develop into microgametes and macrogametes, respectively (Billker et al., 1997; Garcia et al., 1998; Janse et al., 1989, 1986; Kawamoto et al., 1991). The macrogametes are then fertilised by the microgametes to form a diploid zygote. The zygote develops into a motile ookinete which embeds itself in the mosquito gut wall, transversing the midgut epithelial, and becomes an oocyst. Eventually the oocyst goes through sporogony to release approximately 3000 haploid sporozoites (Rosenberg and Rungsiwongse, 1991). These are the infective agents that end up in the mosquito salivary glands ready to be transmitted to a new host.

1.2.4 Pathological effects during infection

The pathological effects of malaria are caused by the asexual-erythrocytic phase of the parasite life cycle. The transient epithelial and liver stages are asymptomatic. Clinical incidences of malaria may present in mild symptoms, severe life-threatening disease or be completely asymptomatic. The reasons why some patients are affected more severely are not fully understood but there are many parasitic, host and environmental factors at play (L. H. Miller et al., 2002; Wassmer et al., 2015).
1.2.4.1 Uncomplicated malaria

Mild or uncomplicated malaria symptoms arise one to two weeks after infection and can involve fever, chills, headache, nausea, diarrhoea and vomiting. Symptoms are caused by the innate immune system response to endotoxins such as haemozoin and foreign parasite DNA that are released when schizonts rupture (Gowda and Wu, 2018; Pichyangkul et al., 1994; Sharma et al., 2011). These symptoms usefully resolve within weeks, and prompt treatment helps to ensure cases do not become severe (WHO, 2019).

1.2.4.2 Complicated Malaria

In some cases, however, the symptoms worsen to a severe (complicated) form of the disease with a higher risk of death. Severe malaria is most common in *P. falciparum* infections but can also develop from *P. vivax* and *P. knowlesi* (Cox-Singh et al., 2008; Price et al., 2007). Higher risk of complications exists for immunocompromised people such as those with bacterial or viral co-infections (Cohen et al., 2005; Kwan et al., 2018), children under the age of 5 whose immune systems are not yet fully developed (Barry and Hansen, 2016; Gonçalves et al., 2014). People in low transmission areas are also at higher risk of severe malaria as it takes multiple episodes of malaria to develop acquired immunity (Barry and Hansen, 2016). Pregnant individuals and their babies are also particularly at risk as malaria can target the placenta (Ayres Pereira et al., 2016; Black et al., 2017).

There are a number of pathophysiology mechanisms associated with severe disease. Malarial anaemia is caused by the depletion of RBCs. Apart from egress-related lysis of iRBCs, there is bystander RBC lysis (Jakeman et al., 1999), reduced RBC production and increased clearing of RBCs by the spleen (Abdalla et al., 1980; Buffet et al., 2009; Totino et al., 2016). Another major cause of pathology is the sequestration of infected erythrocytes. Surface antigens such as PfEMP1 help mediate cytoadhesion to vascular endothelial cells. The endothelial cells are damaged by reactive oxygen species from haemolysis and disruption of blood flow causes tissue hypoxia and acidosis (L. H. Miller et al., 2002; Schaer et al., 2013). Clusters of infected and uninfected RBCs (rosettes) can also contribute to blood vessel obstruction (Kaul et al., 1991). These effects can occur in the haematopoietic system, the brain, the lungs, the kidneys or the placenta. Each location is associated with its own symptoms and risk factors. 15 to 20% of malaria-related deaths are caused by cerebral malaria (Miller et al., 2013). Symptoms of cerebral malaria include respiratory distress and seizures, and it can lead to coma, or long-term neurological sequelae in 25% of sufferers (John et al., 2008). Acute respiratory distress syndrome, defined as dyspnoea,
respiratory failure an pulmonary oedema is a rare but often-lethal complication of severe malaria (Taylor et al., 2006). Organ failure of the kidneys and liver also occurs in some severe cases (Barsoum, 2000; Shoukier et al., 2006). Placental malaria in pregnancy is a large part of the disease burden in malaria endemic regions, and leads to poor outcomes for both mother and foetus (WHO, 2004).

1.3 Malaria drug discovery

1.3.1 Past and present strategies to prevent and eradicate malaria

The control and eradication of malaria requires a concerted multi-approach strategy. The reduction in malaria since 2010 (mostly in Africa), is attributed to three key interventions: insecticide treated bed nets (ITNs); Artemisin combination therapies (ACTs); and indoor residual spraying (IRS) with insecticides (Bhatt et al., 2015). There are many considerations that make developing control and elimination strategies a complicated business. The development of new tools is hampered by the adaptability of both the parasite and vector. Malaria affects many different countries and communities, and measures must be assessed for their sustainability and feasibility in the field.

1.3.1.1 Vector control

An important and historically-successful way of controlling malaria incidence has been to target the mosquito vectors that transfer the disease from person to person. This can be done either by decreasing mosquito numbers or by preventing mosquitos from biting humans. However, the large number of Anopheles species that transmit malaria and the ability of mosquitos to quickly evolve resistance and/or behavioural changes means that these techniques are not always straightforward. As previously mentioned, increases in mosquito territories due to rising global temperatures could make it more difficult to control mosquitos on a population level in the future.

1.3.1.1.1 Indirect methods of vector control

Elimination of mosquito breeding grounds, e.g. swampland drainage, has reduced the incidence of multiple mosquito-borne diseases. However, modulation of the physical environment in malaria endemic regions is not always possible— wetlands are often important areas for biodiversity, and reservoirs may be crucial for urban water supply. Economic development is often inextricably linked to a country’s malaria burden. Urbanisation locally reduces both the number and size of Anopheles breeding habitats; however, deforestation, improved irrigation and
dam construction create new ones. The increase in population density caused by urbanisation also increases transmission rates. The improvement of building materials and housing, and the installation of electricity and air conditioning reduces inside mosquito feeding. However, this selects for mosquitoes that feed outside and electric lighting results in people staying outside and awake later into the night when they may be bitten more. Evidently these complex and counteracting factors must be thoroughly surveyed and taken into consideration in order to maximise malaria elimination efforts. In some countries, such as Mauritius, the elimination effort included larval control through the introduction of lavivorous fish and larvicides such as the organophosphate temephos. In 1985 nearly 16,000 potential larval habitats in Mauritius were treated with temephos (Smith Gueye et al., 2016) and the fish are still pointed out to visitors by proud tour guides of the botanical gardens to this day.

1.3.1.1.2 Insecticide-based vector control methods

Direct methods of reducing malaria transmission by vector control have shown historical and recent successes. The two main strategies are IRS (application of stable insecticides to indoor walls and rooves of houses) and distribution of ITNs, including long lasting insecticide treated bed nets (LLINs). The discovery of the organochlorine insecticide dichlorodiphenyltrichloroethane (DDT) in the 1940s led to great successes in malaria elimination in Taiwan, much of the Caribbean, the Balkans, parts of northern Africa and northern Australia. It also reduced deaths considerably in Sri Lanka and India (“Mosquitoes, malaria and man,” 1979). DDT acts by opening sodium ion channels in the neurons of insects, but soon resistant insects with mutations in their ion channels were seen and this led to a catastrophic reversal of initial successes (Chapin and Wasserstrom, 1981). There were also concerns about DDT affecting human and animal health. It is highly stable and hydrophobic and it rapidly accumulates in food chains, although the exact risks to health are not well understood. Currently the four main classes of insecticides are: organochlorides like DDT, organophosphates, carbamates and pyrethroids. All are used in IRSs and pyrethroids are also used in LLINs. Although resistance has been reported for each of these classes of insecticides (WHO, 2018), their hugely significant contribution to the reduction in malaria cases since 2010 is undeniable (Section 1.1.3; WHO, 2019).

1.3.1.1.3 Genetic methods of vector control: sterile insect technique

In recent years there have been inventive genetic methods of vector control. The sterile insect technique (SIT) uses laboratory-bred sterile male mosquitos to compete with wild mosquitos for mating partners in order to drastically decrease the population size. In the past SIT has been shown to be successful and is one of the only interventions that increases in effectiveness as the target
population decreases (Knipling, 1955). However, STI faces challenges both social, such as public understanding and acceptance, and technical, including accounting for immigrating male mosquitos (Oliva et al., 2014). Modern genome editing techniques such as CRISPR (See Section 1.7.2.1) have revolutionised this field allowing for gene drives (Gantz and Bier, 2016). A gene drive is when a genetic modification is designed to spread through a population at much higher than Mendelian inheritance rates. Gene drives have shown to be successful in introducing sterility in Anopheles gambiae and parasite resistance in Anopheles stephensi, two malaria carrying mosquito species (Gantz et al., 2015; Hammond et al., 2016). It should also be noted that there are over forty other species of Anopheles that transmit malaria, and these methods affect one at a time. Complicated governance and ethics arguments surround the area of gene drive technologies (Adelman et al., 2017; Rudenko et al., 2018). Ultimately, multiple vector control methods must be intelligently designed and coordinated to ensure the continued decrease in malaria incidence worldwide. Factors such as cost, feasibility, public perception and implementation must all be taken into account.

1.3.1.2 Vaccines

The ability to confer protection to a person for life would be hugely beneficial to the eradication effort. Protection would eliminate the logistical and financial challenges of administering courses of drug treatment. Although protective immunity to malaria is seen in populations, individuals still act as reservoirs of the disease and continued exposure to the disease is required to maintain it. This partial immunity is due to low immunogenic antigens and the high antigenic variability of parasites that enables remarkable evasion of the immune system (Doolan et al., 2009; Portugal et al., 2015). Despite these challenges, scientists have invested much time into vaccine research and development leading to a number of candidates currently in clinical trials. The most advanced example is RTS,S/AS01, (Mosquirix) a vaccine developed by GlaxoSmithKline (GSK; Casares et al., 2010). It contains parts of the P. falciparum CSP. The immunity it confers is partial and short-lived, protecting only 40 to 50% of phase 2 trail participants (Kester et al., 2009; Polhemus et al., 2009). Phase 3 trails in children in Africa prevented a substantial number of clinical malaria cases despite low and short-lived efficacy (RTS,S Clinical Trials Partnership, 2015). Phase 4 pilot schemes are currently underway in Ghana, Kenya and Malawi (Laurens, 2019). In the meantime, we must continue the discovery and development of antimalarial treatments.
1.3.1.3 Multiple and single drug therapies

1.3.1.3.1 Quinoline-based antimalarials

The race between antimalarial drugs and resistance has been running since the first drugs were introduced. A single patient may be infected by more than $10^{12}$ asexual blood-stage parasites. Even at slow random mutation rates, genetic changes that confer resistance can occur within a few replication cycles. The first modern antimalarial was derived from the bark of the Peruvian *Cinchona* tree, a bitter traditional medicine that had been used to treat fever for hundreds of years. The alkaloid quinine, one of the active ingredients, was isolated in 1820 and quickly became the most important antimalarial treatment of the time (Fig. 1.2A). Quinine was widely and effectively used until more effective synthetic derivatives such as chloroquine (CQ), developed in 1934, became available. Quinine and CQ are both 4-aminoquinolines (Fig. 1.2A), containing the quinoline group. They have similar mechanisms of action (MoA), primarily interfering with parasite metabolism of haemoglobin in the digestive vacuole leading to the build-up of toxic haem complexes (Section 1.2.2.2; Banyal and Fitch, 1982; Slater, 1993). CQ was the gold-standard for malaria treatment post World War II, but after unprecedented widespread use CQ resistance was seen after just fifteen years (Eyles et al., 1963). Parasite resistance to CQ soon spread to over forty countries causing a catastrophic increase in deaths, especially of children in African countries (Payne, 1987; Trape, 2001). Resistance to CQ and other quinolines has been linked to multiple different point mutations and/or amplification of transporters that promote efflux of drugs from the digestive vacuole (Sanchez et al., 2010). Despite its higher toxicity, quinine remains to this day an important anti-malarial in the treatment of severe malaria, especially in areas affected by resistance to front-line drugs (Yeka et al., 2009). In response to the CQ resistance crisis, other quinoline-based drugs such as mefloquine (MF) were developed (Fig.1.2A). MF acts by disrupting protein synthesis, a different MoA to quinine and CQ (Wong et al., 2017). However, like CQ and other derivatives it was only a matter of time before resistance was reported. In fact the long half-life of MF leads to long exposure to low levels of the drug, driving the selection of resistant parasites (Mockenhaupt, 1995). Another 4-aminoquinoline Ferroquine (FQ, Sanofi) is currently in phase 2 clinical trials as part of a new two-drug antimalarial treatment (Phyo et al., 2016). Unlike all other quinine derivatives, FQ is an organometallic containing a ferrocene group (Fig. 1.3.1A). It is a more potent antimalarial than CQ and, crucially, is active in drug-resistant parasite strains (Held et al., 2015; Wells and Hooft van Huijsduijnen, 2015).
8-aminoquinolines are another class of quinoline-containing anti-malarial drugs such as primaquine and tafenoquine (Fig. 1.3.1B). These drugs act differently from the 4-

Figure 1.3.1: The chemical structures of common antimalarial drugs. (A-B) The 4-aminoquinolines including quinine and its synthetic derivatives and the 8-aminoquinolines all contain a common quinoline structure (shown in pink). (C) The antifolate drug structures. (D) Artemisinin and its semi-synthetic derivatives all contain an endoperoxide group important for their activity (shown in blue). 8-aminoquinolines are another class of quinoline-containing anti-malarial drugs such as primaquine and tafenoquine (Fig. 1.3.1B). These drugs act differently from the 4-
aminoquinolines, targeting the hypnozoites formed by *P. vivax* and combat the latent infections that are a problem with non-*falciparum* species. However in many malaria-endemic populations these drugs are only cautiously used, as they can cause acute haemolytic anaemia in G6PD-deficient patients (Ashley et al., 2014).

1.3.1.3.2 Antifolate antimalarials

Combination therapy of antifolate drugs, sulfadoxine and pyrimethamine (SP therapy), was developed in 1966 (Richards, 1966) and SP quickly became an important treatment in high CQ-resistance areas (Fig. 1.3.1C). Both sulfadoxine and pyrimethamine inhibit *de novo* folate synthesis in *Plasmodium* parasites by synergistic mechanisms. Folate is required for pyrimidine synthesis at all stages of the asexual life cycle and early gametogenesis. However, in a now familiar pattern, resistance to SP quickly emerged in the form of point mutations or copy number polymorphism in the folate biosynthesis pathway (Nair et al., 2008; Peterson et al., 1988; Sibley et al., 2001; Triglia et al., 1997). SP therapy is now only used for intermittent preventative therapy in pregnant people (ter Kuile et al., 2007).

1.3.1.3.3 Artemisinin-based malaria therapies

The current gold-standard treatment for malaria is ACT. Like quinine, artemisinin has its roots in traditional medicine. It was identified as the active ingredient from the sweet wormwood plant *Artemisia annua* which had historically been used to treat fevers in Chinese traditional medicine (White et al., 2015). The endoperoxide bridge is responsible for the activity of artemisinin and its more bioavailable semi-synthetic derivatives: dihydroartemisinin, artemether and artesunate (Fig. 1.3.1D). Haem-mediated cleavage leads to the generation of free radicals. These cause widespread destructive intercellular damage and parasite death (Kamchonwongpaisan and Meshnick, 1996; Meshnick et al., 1991). This unique reactivity means artemisinin-derivatives are highly efficacious but have short therapeutic half-lives (White, 1997). The combination of artemisinin-derivatives with other partner drugs such as MF in ACTs gives a longer elimination time and, it was predicted, a longer time before parasite resistance develops. However, yet again artemisinin resistance has been documented in South-East Asia leading to a reduction of parasite clearance rates (Dondorp et al., 2009). Whole genome sequencing helped associate resistance with mutations in the propeller domain of K13 Kelch protein (K13; Ariey et al., 2014). It is thought that these mutations confer resistance via a complicated chain of events that causes upregulation of protein degradation by the proteasome, reducing free-radical cell damage (Suresh and Haldar, 2018). The timeline of ACT resistance development or spread in Africa remains an unclear and
concerning problem (Tacoli et al., 2016). The modification of multidrug therapies such as ACT and carefully-planned cycling of drugs may help to stave off another drug-resistance crisis as was seen in the past with CQ. However, in the long-term new drugs with novel modes of action are desperately needed.

1.3.2 Strategic drug discovery for malaria elimination

There are many complex external factors affecting malaria drug discovery including: the development of resistance, the feasibility of implementing complicated treatment regimens, and the need for drugs suitable for patients coinfected with other pathogens. Funding for malaria research has historically been low as it affects mostly low GDP counties. Internal problems facing the malaria community are the high demands that are made on the drugs to further eradication. Ideal new drugs should be transmission blocking, prevent *P. vivax*-like relapses, act as post-treatment prophylaxis and chemoprotection for vulnerable groups such as pregnant or co-infected patients. It is unlikely that a single treatment will cover these aims.

In 2018 US $663 million was invested in basic research and product development for malaria. This is the third year in a row that investment has increased, and the highest annual total since the peak of 2009 (US$ 672 million; WHO, 2019). Funding for drug R&D and basic research both increased in 2018, driven by increased investment by the private sector. Private funding now roughly equals that of the philanthropic sector, although just over half of investments still come from the public sector (WHO, 2019). The increase in investment from the private sector is largely due to several phase 2 trials of new drugs with the potential of single-exposure radical cure such as Ganaplacide/Lumefantrine (Novartis; Koller et al., 2018).

The Malaria Eradication Agency (malERA) described the new strategy needed to work towards malaria eradication. The strategy is a reorientation back to focusing on eliminating infection and transmission rather than control of the disease (Drugs, 2011). For example, treatments to reduce clinical symptoms are not relevant to eradication unless they have transmission-blocking (eg. anti-gametocytocidal) activity. Among other recommendations they described an ideal treatment of Single Encounter Radical Cure and Prophylaxis (SERCaP) (Drugs, 2011), consisting of multiple compounds. SERCaP means completely eradicating all asexual blood-stage forms and gametocytes in a patient (and hypnozoites in *P. vivax*). This process eliminates the human reservoir of disease, while also conferring protection against reinfection for the patient after cure. They also described the need for a new generation of casual prophylaxis, Single Exposure
Chemoprotection (SEC), to target pre-erythrocytic liver stages. This would be useful in preventing resurgence of the disease after elimination in areas.

The lead-time for antimalarial treatments is longer than other medicines due to the complexities of developing combination therapies. To fully exploit funding for malaria it is important to have clear but flexible goals in mind when pursuing a new product. Target Product Profiles (TPPs) describe a desired treatment with a clear hypothesis of how it will meet the current or predicted medical need (Burrows et al., 2013). TPPs are made up of multiple different Target Candidate Profiles (TCPs) for specific molecules (Burrows et al., 2013). Some issues are universal and for every TPP, researchers and funders must consider clinical safety and efficacy, resistance and affordability. Medicines for Malaria Venture (MMV) is a foundation set up in 1999 with the mission to reduce the burden of malaria in endemic countries by facilitating the discovery, development and delivery of new antimalarial drugs. In the context of malERA’s eradication-focused strategy MMV set out its TCPs in 2014. For the SERCaP TPP, three TCPs were described: fast parasite clearance, long duration, and transmission-blocking/relapse-prevention (Burrows et al., 2013). The TCP for SEC is to be a liver schizontocide with a high safety profile for extended use (one week - one month). Other desirable but non-critical attributes are vector-stage activity, gametocytocidal activity and an orthogonal MoA to all other TCPs (Burrows et al., 2013). The TPPs and TCPs have been recently reviewed and updated to reflect changing understanding of the challenges of malaria drug discovery (Burrows et al., 2017). For example, “long duration” is no longer discussed as a TCP, as all compounds in development have been selected for long duration. Instead more sophisticated measures such as Parasite Reduction Ratios are encouraged (Hastings et al., 2016). There has also been a recent focus on the safety profiles of drugs in development to be able to serve vulnerable populations such as people experiencing co-infections, malnutrition or pregnancy. MMV presents a constantly-updated Global Malaria Portfolio describing the molecules currently in pre-clinical development (Wells et al., 2015; http://www.mmv.org/research-development/rd-portfolio). The common platform that MMV has facilitated for the discussion of goals, priorities and developments will surely continue to boost efforts towards malaria control, elimination and ultimately eradication. Before compounds can be assessed and prioritised in the context of TCPs and TPPs however, the antimalarial drug pipeline needs to be populated with novel chemotypes and modes of action.
1.3.3 **Phenotypic and target based screening for drug discovery**

Unfortunately, the availability of *Plasmodium* genetic information and expansion of our knowledge of basic biology has not lead to an immediate increase in new drugs. This is possibly due to the overreliance on target-based drug discovery: many novel hypothetical targets are identified, but there is high attrition of inhibitory compounds due to a lack of whole-cell activity (Cameron et al., 2004; Tasdemir et al., 2006). The lack of conditional genetic tools for target validation in *P. falciparum* (until relatively recently) has also contributed to this attrition. Many drug and vaccine discovery studies were focused on targets assumed to be essential because knockout parasites could not be obtained. However using new techniques, many targets were show to be dispensable, such as Merozoite surface protein 1 (MSP1; Das et al., 2015), the falcipains (Sijwali et al., 2006) and the plasmepsins (Liu et al., 2006). Over the last decade, most MMV antimalarial compounds have been identified by phenotypic high-throughput screening (HTS; Hovlid and Winzeler, 2016). In the last fifty years, despite increases in R&D expenditure across most disease areas, the rate of new drug approvals has remained relatively constant (an average of one new chemical entity per company, every 6 years; (Munos, 2009). This reflects difficulties in target validation and subsequent increased stringency from regulatory authorities. Whole-parasite screening means the success rate for pre-clinical candidates becoming launched is often higher for neglected diseases as non-valid chemotypes are rejected quickly. High-throughput phenotypic screens with large, highly-diverse compound collections have been successful in finding new chemical series.

Two of the main classes of antimalarial compounds are derived from natural products, quinine and artemisinin (Section 1.3.1.3). The natural products were used in traditional medicine for hundreds of years, followed by their successful application to malarial treatment, with no MoA known. However, with resistance concerns and the need for combinatorial treatments to fit ideal TPPs, knowledge of a compound’s MoA is more important now than ever before. MMV’s Project of the Year 2018 was awarded to the team at the Drug Discovery Centre in Dundee for the identification of a novel chemical series targeting a novel biological target lysyl-tRNA synthase (PfKRS1). A search for the target of antimalarial natural product cladosporin identified PfKRS1, which is essential to protein synthesis in pre-erythrocytic liver stages, asexual and sexual blood-stages (Baragaña et al., 2019). Compound screening and structure-based drug design followed to optimise small-molecule hits (Baragaña et al., 2019). The MMV supported discovery and validation of an antimalarial target, and subsequent hit screening shows the restoration of openness to target-based drug discovery.
Many new antimalarials still have no known MoA. Tafenoquine (Fig.1.1B) was developed by GSK in collaboration with MMV and was approved in the USA in 2018 for both SERCaP and prophylaxis. The targets of Tafenoquine are still unclear, although its action may involve the mitochondria, haemazoin polymerisation and/or cryptosis (Al Mamun Bhuyan et al., 2016; Carvalho et al., 2010; Vennerstrom et al., 1999). However, there are multiple compounds in the late-stage antimalarial pipeline that do have known MoAs. Two such are DSM265, which inhibits plasmodial dihydroorotate dehydrogenase (phase 2, Takeda), and P218 which inhibits Plasmodium dihydrofolate reductase (phase 1, Janssen), both developed in partnership with MMV (Abbat et al., 2015; McCarthy et al., 2017).

Advances in proteomics and genomics have helped drive the deconvolution of targets from large phenotypic screens. Affinity chromatography can be used to identify targets, pulling down proteins that bind to the compound of interest. In vitro evolution and whole-genome analysis (IVEWGA) can identify targets by generating compound-resistant parasites and using whole genome sequencing to find mutations (Flannery et al., 2013). Although it is not always straightforward, resistance can be indirect as a result of drug influx mediation or prodrug-activating enzymes. This approach has had great successes in target deconvolution. For example, translation elongation factor 2 was identified as the target of M5717 (Marck KGaA) a compound now in phase 1 trials for SERC (Baragaña et al., 2015). Proteomics-based deconvolution techniques are also highly useful. Specially designed strategies can be very powerful to confirm on-target selectivity if there is some pre-existing knowledge, although unbiased fishing for targets can be difficult to interpret. Proteomics helped identify Plasmodium phosphatidylinositol 4-kinase as the target of a high-throughput screening compound MMV048, now in phase 1 clinical trials for SERC (Paquet et al., 2017). The future of malaria drug discovery undoubtedly lies in the integration of all these techniques, utilising target research, large compound screens and sophisticated MoA confirmation in the next generation of antimalarial medicines.

1.3.4 The Malaria Box by Medicines for Malaria Venture (MMV)

One example of the successful application of high-throughput data is the MMV Malaria Box project. This aimed to prioritise and make available compounds from multiple large-scale private sector screens to the academic community. Many pharmaceutical companies already have huge compound libraries developed for screens against other infectious and non-communicable diseases. These could be easily directed towards antimalarial screening. Pharmaceutical companies (GSK (Gamo et al., 2010; Garcia-Bustos and Gamo, 2011), Novartis (Meister et al.,
2011), and St Jude’s Children’s Hospital (Guigue et al., 2010) have screened in excess of six million compounds against the erythrocytic life cycle of *P. falciparum* and collectively identified 20,000 compounds that block parasite replication below 5 μM (Fig. 1.3.2). Computational analysis of the full target space has predicted an enrichment in kinase and protease inhibitors (Spitzmüller and Mestres, 2013), reflecting the essential nature of these enzyme classes and their importance as targets to the pharmaceutical industry. An advantage of phenotypic screens is that one can find unbiased chemotypes without any *a priori* assumptions. However as the phenotypic screens come from previously assembled libraries they may be naturally biased to traditional targets such as kinases.

![Figure 1.3.2: The MMV Malaria Box.](image)

The wealth of data generated from these screens has already informed multiple hit-to-lead drug discovery campaigns, but the majority of the 20,000 antimalarial hits have no known MoA. In an unprecedented move considering the scale and nature of the data, the structure and potency of all these hits were made available to the academic community. However, barriers remain to the optimum exploitation of these results. First the compounds themselves are not freely available, and few groups would have the means to purchase or synthesise a significant number. Secondly there are simply too many compounds to be able to apply these results to screens of different *Plasmodium* life stages or other Apicomplexa, as few systems are suitable for such large screens. MMV undertook a project to simplify this large collection of new antimalarial compounds. The
aim was to make it financially and technically feasible for academic groups around the world to
test a diverse subset of compounds in a variety of assays and organisms.

MMV compiled a box of 400 of the above-mentioned HTS hits into a freely-available resource.
The “Malaria Box” is a set of potent and commercially-available compounds that were chosen
for maximum chemical diversity (Spangenberg et al., 2013; Fig. 1.3.2). The first selection criteria
applied was that they be commercially available, this gave a subset of around 5000 compounds.
These were subjected to multiple filters including searches for known toxicophores and lipinski’s
rule-of-five, a computational way to predict physiochemical properties and oral absorption
(Lipinski et al., 2001). Other liabilities were identified using Rapid Elimination of Swill and Pan
Assay Interference Compounds filters. Around half of the compounds complied with these
conditions and were termed “drug-like”. The other half were labelled “probe-like” reflecting the
hope that they could be useful molecules to probe parasite biology, even if they were unsuitable
as hit-to-lead compounds. Each set was now separately optimised for the structural diversity and
potency, while retaining some structural analogues to allow for easy structure-activity
relationship (SAR) studies of hit compounds. After a final round of verification of antimalarial
activity (EC$_{50} < 4$ μM), 200 drug-like and 200 probe-like compounds were chosen by “wisdom
of the crowd” selection to make up the final Malaria Box. The box was made freely available as
solid compound stocks split across five 96-well plates labelled A-E. The plates were divided up
into A: 80 most potent compounds, B-C: 160 “Drug-like” compounds”, D-E: 160 “Probe-like”
compounds.

The Malaria Box has been screened many times against specific enzyme targets, pathways and
other life stages of $P. falciparum$ as well as other apicomplexa. The results of these studies and
many others are deposited into an online resource on CHEMBL (Mendez et al., 2019). These
compounds have the potential to be useful tools in studying parasite biology as well as pointing
to new essential genes. A meta-analysis of 291 Malaria Box screens, carried out by 55 different
research groups showed only 135 (34%) of the compounds has been assigned a likely MoA (Van
Voorhis et al., 2016).

By using the focused Malaria Box library, diverse chemical entities can be screened for activity
against proteins identified as potential drug targets in highly diverse assays (target-based drug
discovery). Screening of the Malaria Box in a surface plasmon resonance-based competition assay
identified a series of compounds that disrupted the protein-protein interaction (PPI) between
PfAtg8 and PfAtg3, which is important in the essential $Plasmodium$ autophagy machinery
Duszenko et al., 2011; Hain et al., 2014, p. 8). The series was then stripped back to a chemical scaffold and optimised. Deoxyhypusine hydrolase is a metalloprotein essential for hypusine biosynthesis, predicted to be a “druggable” target due to the essential nature of amino acid hypusine and the low similarity (26% sequence identity) to the human orthologue (Mittal et al., 2013). The Plasmodium hypusine biosynthesis pathway was reconstituted recombinantly and gas chromatography mass spectrometry was used to measure metabolites in the presence of Malaria Box compounds, identifying three inhibitory compounds (Koschitzky and Kaiser, 2013; von Koschitzky et al., 2015).

Malaria Box studies have also helped to identify new antimalarial targets. The Coenzyme A (CoA) biosynthesis pathway is essential in Plasmodium. A chemical rescue approach was designed to identify novel CoA synthesis inhibitors (Fletcher and Avery, 2014). The compounds identified were then used as tools to begin to identify which enzyme in the pathway was affected, thus providing a new antimalarial drug target and lead compounds (Fletcher et al., 2016). IVEWGA has also been used to find the targets of the Malaria Box compounds, by generating resistant strains. For example, IVIEWGA identified resistance-conferring mutations in the P. falciparum farnesyl/geranylgeranyl diphosphase synthase in the presence of Malaria Box compound MMV019313 and subsequently validated it as a target (Gisselberg et al., 2018).

High throughput screens of the scale that went into the Malaria Box are currently out of the question for other P. falciparum life stages. However many screens of gametocytocidal activity have now been carried out using the more focused Malaria Box (Bowman et al., 2013; Lucantoni et al., 2013; Ruecker et al., 2014). By a similar argument, more challenging in vivo culture systems make large-scale screens difficult for other organisms. Repurposing of the Malaria Box has identified inhibitors of Apicomplexa Toxoplasma gondii and Cryptosporidium parvum as well as other pathogens like Schistosoma mansoni and Entamoeba histolytica (Bessoff et al., 2014; Boyom et al., 2014; Ingram-Sieber et al., 2014). After the success of the Malaria Box, MMV have put together the Pathogen Box, now available free of charge in its stead. This box of 400 diverse drug-like compounds is active against a range of neglected tropical diseases, including malaria.

### 1.4 Enzyme targets

Proteins are by far the easiest biological macromolecules to target with small molecules. Enzyme catalysis is essential in all life processes and is dependent on highly specific ligand binding. This makes enzymes particularly vulnerable to modulation of activity by the binding of small
molecules (Copeland et al., 2007). Almost half of all orally-dosed drugs have their pharmacologic effect through the inhibition of enzymes (Hopkins and Groom, 2002). The “druggable genome” of *P. falciparum* is constantly expanding as we discover new pathways and begin to assign functions to the hundreds of predicted protein-coding genes in the genome. Currently 5268 genes are annotated on the *P. falciparum* genome over fourteen chromosomes, 66% of genes have a predicted function (Böhme et al., 2019). Around 57% of the proteome are unique proteins, and the median global identity between related proteins is 13%, a strikingly low level of redundancy (Ludin et al., 2012). There are a huge number of distinct enzyme families, classed according to properties such as their reaction, active site residues or cofactor binding. The enzyme families concerned in this thesis are serine hydrolases and cysteine proteases.

### 1.4.1 Cysteine proteases

Cysteine proteases (CPs) are proteolytic enzymes that share a common catalytic mechanism. The mechanism involves a Cys-His dyad or a Cys-His-Asn/Asp triad in their active site. The active site cysteine residue has an acidic pKa, modulated by its interaction with histidine, which renders it reactive as a nucleophile. The nucleophillic cysteine residue attacks the electron-deficient carbon centre in an amide bond, releasing the amine moiety via a tetrahedral intermediate stabilised by hydrogen bonding. Hydrolysis of the enzyme-intermediate thioester bond via a water molecule regenerates the cysteine nucleophile and releases the carboxylic acid product. CPs are generally synthesised as zymogens, inhibited by a prodomain that is cleaved to give the active mature enzyme.

The MEROPS peptidase database reports 171 peptidases in *P. falciparum*, with 83 of these being CPs (Rawlings et al., 2018). Only around one third of CPs have been studied and very few have been fully characterised (Deu, 2017). However some of the previously studied proteases have already been shown to play essential roles in all stages of parasite development. If applied to schizonts before egress, the broad-spectrum CP inhibitor E64 blocks rupture of the host red cell membrane (Section 1.2.2.4) suggesting that this process is CP dependent (Glushakova et al., 2009). Inhibitor studies originally implicated the human CP calpain in parasite invasion but it was later shown to be dispensable, implicating another essential role for parasite CPs (Hanspal et al., 2002; Olaya and Wasserman, 1991). CPs are subdivided into clans based on sequence and structural similarities. The clans represented in *Plasmodium* are CA, CD and CE (Barrett et al., 2012; Barrett and Rawlings, 2001; Rosenthal, 2004).
1.4.1.1 Clan CD proteases

Clan CD CPs have a His-Cys dyad and appear to have a much tighter substrate specificity than Clan CAs. This is a promising characteristic for drug development (Mottram et al., 2003). Clan CD CPs are not inhibited by the CP inhibitor E64. At least five clan CD enzymes have been predicted in *P. falciparum* of families C13, C14 and C50. Thus far none has been characterised or validated as an active protease in *Plasmodium* although all five appear to have interesting roles in other protozoa (Mottram et al., 2003).

1.4.1.2 Clan CE proteases

There are two clan CE CPs represented in the *P. falciparum* genome. These are described as the SUMO-specific proteases (SENP1 & SENP2). They have a His-Glu/Asp-Cys triad and are thought to regulate post-transcriptional modification of proteins by small ubiquitin-related modifier (SUMO; Ponder et al., 2011). The role of SUMO-lation in *P. falciparum* is not well understood but inhibitors of SENP1 have been shown to block parasite growth and replication in asexual stages (Ponder et al., 2011).

1.4.1.3 Clan CA proteases

One of the most abundant families of CPs are lysosomal cathepsins (clan CAs), including both cathpepsin-like (family C2) and papain-like (family C1) proteases. Papain-like Clan CA CPs (PLCPs) have a Cys-His-Acid catalytic triad and perform a diverse range of functions in all living organisms (Vasiljeva et al., 2007; Verma et al., 2016). Their role in inflammation and cancer has led to their consideration as potential drug targets (Kramer et al., 2017; Olson and Joyce, 2015). They have also been linked to key roles in parasite processes, and the structural distinctness to human proteins provides exciting opportunities for chemotherapy (Sajid and McKerrow, 2002). The best characterised CPs in *Plasmodium* species are the PLCP family. There are ten members of this family belonging to three distinct subfamilies: the serine repeat antigens (SERAs), falcipains (FPs), and dipeptidyl aminopeptidases (DPAPs). Multiple members of these groups have been suggested as interesting drug targets (Deu, 2017). They contain a common ‘papain’ fold with two subdomains: Left and Right (Fig. 4.1). There is also *Pf*calpain which is a clan CA, C2 family, Ca\(^{2+}\)-dependent protease that is significantly different from vertebrate calpains, making it an interesting candidate for drug discovery (Mitchell and Bell, 2003; Soh et al., 2013). Its biological function is not well understood but it may be involved in essential Ca\(^{2+}\) modulation in the parasite (Glushakova et al., 2013; Gomes et al., 2015).
Serine repeat antigens (SERAs)

There are nine SERA homologues in *P. falciparum*. Eight are clustered on Chromosome 2 (SERA1-8) and one is on chromosome 9 (SERA9). They are expressed at different levels in the late asexual blood-stages (S. K. Miller et al., 2002). SERA5 (previously known as SERA/SERP) has long been considered a potential drug/vaccine target (Delplace et al., 1987; Knapp et al., 1989; Li et al., 1989). Only three SERA (SERA6-8) have a catalytic cysteine, all the others have an unusual cysteine to serine substitution and may not retain proteolytic activity (S. K. Miller et al., 2002). The SERAs are secreted into the PV as zymogens. SERA6 in particular has been shown to be essential for parasite egress. It is activated by the subtilisin-like serine protease 1 (SUB1) 15 minutes before egress, illustrating the fine temporal control of these processes (Section 1.2.2.4.

---

**Figure 1.4.1: Basic common structure of a PLCP active site.** Schematic representation of a generic substrate in the active site of a Clan CA, family C1 papain fold protease. The active site lies in between the L(left)-domain and R(right)-domains. The conserved positions of the catalytic cysteine and histidine are shown, as are the four main pockets for substrate binding. Electrophilic attack by the catalytic cysteine is shown in blue, the position of amide bond cleavage is shown in red and hydrogen bonding between the histidine and amide carbonyl is shown in purple. The exclusion domain found in DPAPs is shown in grey. Figure adapted from (Cordara et al., 2016).
Falcipains (FPs)

Four FPs are expressed in *P. falciparum* asexual blood stages: FP1, FP2a, FP2b and FP3. FP1 is the most distinct, sharing only 39% sequence similarity with the others and localising to granules in newly-formed merozoites. FP1 was originally ascribed a role in invasion after epoxide inhibitor studies, however disruption of FP1 presented no phenotype, suggesting there are other undiscovered CP targets involved (Greenbaum et al., 2002; Sijwali et al., 2004). Further targeted disruption studies have since implicated FP1 in oocyst production in the mosquito-hosted life stage (Eksi et al., 2004). FP2a and FP2b only differ by one amino acid so they are generally discussed together as “FP2”. FP2a/FP2b and FP3 are homologous sharing 68% identity and are both food vacuolar haemoglobinases (Drew et al., 2008; Shenai et al., 2000; Sijwali et al., 2001; Section 1.2.2.2). FP2a-null parasites show reduced Hb degradation, characterised by a swollen food vacuole in trophozoites, but recover and replicate as normal (Sijwali et al., 2006; Sijwali and Rosenthal, 2004). FP2b knockout (KO) has no effect, suggesting that these enzymes are perhaps redundant (FP2b expression may only significant in the absence of FP2a; Sijwali et al., 2006). In one study disruption of FP3 was not possible, suggesting it may be essential (Sijwali et al., 2006). However further studies did achieve FP3 KO, and showed this PLCP to be non-essential in asexual-blood stages (Dr Daniel Goldberg, personal communication). Interestingly, the swollen food vacuole phenotype of FP2a KO parasites is associated with reduced susceptibility to artemisinin (Klonis et al., 2011; Xie et al., 2016). Recently mutations in FP2a have been found (in addition to mutations in K13 Kelch protein) in artemisinin-resistant parasites suggesting a role for FP2a in resistance (Section 1.3.1.3.3; Ariey et al., 2014; Siddiqui et al., 2018). Hb degradation also involves multiple plasmepsin aspartyl proteases that may be activated by FPs (Bonilla et al., 2007). The plasmepsins were originally considered to be potential drug targets but the redundancy among them led to a shift in focus onto the FPs (Liu et al., 2006). Other proteases involved in Hb degradation include metalloproteases (Eggleson et al., 1999), aminopeptidases (Dalal and Klemba, 2007) and the DPAPS (Section 1.4.1.3.3).
1.4.1.3.3 Dipeptidyl aminopeptidases (DPAPs)

The *P. falciparum* DPAPs (DPAP1, DPAP2 and DPAP3) cleave dipeptides from the N-terminus of protein substrates, in homology with Cathepsin C (CatC, also known as DPPI). This exopeptidase activity is due to an additional ‘exclusion’ domain that interacts with the free N-termini of protein substrates and prevents binding beyond the S2 binding site (Fig. 1.4; Turk et al., 2001). Studies using the small molecule inhibitor SAK1 suggested a role for DPAP3 in egress (Section 1.5.2; Arastu-Kapur et al., 2008). The SAK1-related phenotype was later attributed to off-target effects after multiple approaches showed that DPAP3 localised to novel apical organelles in merozoites and was dispensable in egress (Ghosh et al., 2018; Lehmann et al., 2018). In trophozoite stages DPAP1 localises to the food vacuole where it participates in Hb degradation, but is non-essential (Dalal and Klembo, 2007; Deu et al., 2010; Klembo et al., 2004). In mature schizonts DPAP1 re-localises to the PV (Klembo et al., 2004). Studies in our lab have now shown that DPAP1 and DPAP3 cKOs showed no defects in parasite growth or egress but a defect in invasion, which was additive in a DPAP3/DPAP1 double cKO (Rahbari et al., in preparation). These studies suggest DPAP1 and DPAP3 play complementary roles in RBC invasion (Rahbari et al., in preparation). A dual inhibitor of DPAP1 and DPAP3 also has gametocytocidal activity suggesting an extra role for at least one of them in sexual stage development (Tanaka et al., 2013). DPAP2 meanwhile is only expressed in gametocyte life stages. It was shown to be dispensable in gametogenesis in *P. falciparum* and sporozoite development in *P. berghei* (Le Roch et al., 2003; Tanaka et al., 2013). However a more recent study report a 50% decrease in gametocyte egress in DPAP2-null mutants (Suárez-Cortés et al., 2016).

1.4.2 Serine hydrolases

Serine hydrolases (SHs) are a huge superfamily including all enzymes that use a nucleophilic serine to catalyse the hydrolysis of amide, ester, and thioester bonds. SHs are involved in a variety of biological processes such as metabolism, post-translational modification and proteolysis. They have been shown to be targetable with small molecule inhibitors. Serine proteases, like cysteine proteases have been extensively studied and pursued as drug targets. They too are classed into families, such as subtilisin, trypsin and chymotrypsin, based on structural similarities. *P. falciparum* has many examples of serine proteases, for example the aforementioned SUB1 that initiates the protease cascade required for egress in *P. falciparum* (Section 1.2.2.4; Blackman, 2008). Rhomboid serine proteases also have roles in parasite invasion (Baker et al., 2006a; Singh et al., 2007; Srinivasan et al., 2009) and host-cell interaction (Vera et al., 2011).
1.4.2.1 Metabolic SHs

This thesis focuses more specifically on the metabolic SHs, which may cleave ester, amide or thioester bonds and remain largely uncharacterized in \textit{P. falciparum}. All of these bond cleavages proceed by a conserved catalytic mechanism via an acyl-enzyme intermediate. The intermediate then undergoes water-induced hydrolysis to release the product, and regenerate the serine nucleophile. The study of metabolic serine hydrolases has been enhanced by the development of activity-based probes such as fluorophosphonates (FPPs; Section 1.5.3). The ubiquity of metabolic SHs is such that they are important in almost all mammalian disease states: acetylcholinesterase in neurotransmission (Lane et al., 2006); phospholipase A2 in inflammation (Bonventre et al., 1997); and lipase activity in bacterial infection (Shao et al., 2007) to name a few.

Metabolic serine hydrolases can broadly be classed into three families: small molecule hydrolases, protein/polysaccharide-modifying hydrolases and peptidases (Long and Cravatt, 2011). Small molecule hydrolases consist of: neutral lipases that hydrolyse triglycerides and cholesterol esters resulting in fatty acids; phospholipase and lysophospholipases enzymes that cleave diverse fatty acid chains from phospholipids and lysophospholipids, respectively; small molecule amidases; and acyl-CoA hydrolases that release free fatty acids from acyl-CoA molecules and cholinesterases such as acetylcholinesterase (Long and Cravatt, 2011). The protein-modifying hydrolases include proteins that remove post-translational modifications such as acyl-protein thioesterases (Hirano et al., 2009), depalmitoylases, and unusual methylesterification (Ogris et al., 1999). Glycan hydrolases are diverse but hydrolyse fatty acids from molecules with polysaccharide cores such as hydrolysis of fatty acids from glycosylphosphatidylinositol (GPI) protein anchors (Ueda et al., 2007). Metabolic SH peptidases include the Clan SC serine peptidases which, unlike other serine proteases, have an \( \alpha/\beta \)-fold (Page and Di Cera, 2008).

60\% of mammalian metabolic serine hydrolases have an \( \alpha/\beta \)-fold (\( \beta \)-sheet surrounded by \( \alpha \)-helices), usually accompanied by a Ser-His-Asp catalytic triad (Fig. 1.4.2; Holmquist, 2000). Other classes include amidase signature enzymes (Ser-Ser-Lys triad; Shin et al., 2002), patatin domain-containing lipases (Ser-Asp dyad; Kienesberger et al., 2009; Six and Dennis, 2000) and dipeptidylpeptidases (Yu et al., 2010). \( \alpha/\beta \)-fold-containing hydrolases (\( \alpha/\beta \)Hs) are an example of divergent evolution. In contrast to enzymes that evolved different structures independently to perform similar functions such as trypsin and subtilisin families, the \( \alpha/\beta \)H family have evolved
from a common protein fold to give a highly diverse class of enzyme (Ollis et al., 1992). Though diverse, α/βHs have a very well conserved arrangement of catalytic triad residues and a common “nucleophile elbow” that allows efficient presentation of the nucleophile (Heikinheimo et al., 1999). The ESTHER (Esterases and α/β-hydrolase enzymes and relatives) database is a sever dedicated to the analysis of the α/βH superfamily (Hotelier et al., 2004). Evolutionary analysis in humans shows that both well-known and uncharacterised SHs have low homology between each other but are well conserved with mouse orthologues (Simon and Cravatt, 2010). This suggests the majority of serine hydrolases have distinct and unique roles with strong evolutionary conservation. Most predicted malarial SHs show little homology to well-characterised enzymes in other organisms and might therefore be good antimalarial targets.

1.4.2.2 SHs in P. falciparum

There are twenty serine proteases in P. falciparum: two chymotrypsin-like proteases, three subtilisin-like proteases (including SUB1), ten predicted rhomboid proteases (clan S54) and five others of the S26, S16 and S14 families. Besides the well characterised subtilisin proteases, most of these have not been validated as active, let alone significant proteases. There are forty-three α/βHs in P. falciparum according to structural predictions, most of these are totally uncharacterised putative proteins (Table 1; Hotelier et al., 2004). Other metabolic serine hydrolases in P. falciparum include five phospholipase A2 enzymes, (all papatin-like phospholipases (PLPs; Wilson and Knoll, 2018)), and a unique amidase glutamyl-tRNA amidotransferase subunit 1 (Mailu et al., 2015).
Table 1 α/β-Hs of P. falciparum.

<table>
<thead>
<tr>
<th>PF3D7 Gene ID</th>
<th>Name</th>
<th>Sub-family</th>
<th>Putative function</th>
<th>TM domain</th>
<th>signal peptide</th>
<th>MW / kDa</th>
<th>PBANKA Gene ID</th>
<th>PlasmoGEM</th>
<th>piggybac screen</th>
</tr>
</thead>
<tbody>
<tr>
<td>0702200</td>
<td>Pst-a lysophospholipase</td>
<td>0 n</td>
<td></td>
<td>49.2</td>
<td></td>
<td></td>
<td></td>
<td>dispensable</td>
<td></td>
</tr>
<tr>
<td>0709700</td>
<td>PIPARE esterase</td>
<td>0 n</td>
<td></td>
<td>42.4</td>
<td></td>
<td>1220300</td>
<td>dispensable</td>
<td>dispensable</td>
<td></td>
</tr>
<tr>
<td>1038900</td>
<td>Pst-a phospholipase</td>
<td>0 n</td>
<td></td>
<td>41.8</td>
<td></td>
<td>0</td>
<td></td>
<td>essential</td>
<td></td>
</tr>
<tr>
<td>1476700</td>
<td>Psta1 lysophospholipase</td>
<td>0 n</td>
<td></td>
<td>41.3</td>
<td></td>
<td>low fitness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1476800</td>
<td>Psta2 lysophospholipase</td>
<td>0 n</td>
<td></td>
<td>42.7</td>
<td></td>
<td>essential</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0937200</td>
<td>Pst-a enzyme</td>
<td>0 n</td>
<td></td>
<td>40.8</td>
<td></td>
<td></td>
<td></td>
<td>low fitness</td>
<td></td>
</tr>
<tr>
<td>0936700</td>
<td>PfPARE Pst-a esterase</td>
<td>0 n</td>
<td></td>
<td>49.6</td>
<td></td>
<td></td>
<td></td>
<td>essential</td>
<td></td>
</tr>
<tr>
<td>1252600</td>
<td>Pst-a esterase</td>
<td>0 n</td>
<td></td>
<td>52.9</td>
<td></td>
<td></td>
<td></td>
<td>essential</td>
<td></td>
</tr>
<tr>
<td>1401500</td>
<td>Pst-a esterase</td>
<td>0 n</td>
<td></td>
<td>43.7</td>
<td></td>
<td></td>
<td></td>
<td>essential</td>
<td></td>
</tr>
<tr>
<td>0102400</td>
<td>S33H Pst-a / S33 α/β-hydrolase</td>
<td>0 n</td>
<td></td>
<td>44.4</td>
<td></td>
<td></td>
<td></td>
<td>essential</td>
<td></td>
</tr>
<tr>
<td>0629300</td>
<td>LCAT PC-sterol acyltransferase</td>
<td>0 y</td>
<td></td>
<td>99.2</td>
<td></td>
<td>1128100</td>
<td>dispensable</td>
<td>low fitness</td>
<td></td>
</tr>
<tr>
<td>0731800</td>
<td>S33G S33 α/β-hydrolase</td>
<td>1 y</td>
<td></td>
<td>78.4</td>
<td></td>
<td>low fitness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1001400</td>
<td>PXIL1 / S33F exported lipase 1</td>
<td>1 y</td>
<td></td>
<td>107.7</td>
<td></td>
<td></td>
<td>low fitness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1001600</td>
<td>PXIL2 exported lipase 2</td>
<td>0 n</td>
<td></td>
<td>88.6</td>
<td></td>
<td></td>
<td></td>
<td>essential</td>
<td></td>
</tr>
<tr>
<td>1328500</td>
<td>MLPL α/β-hydrolase</td>
<td>0 n</td>
<td></td>
<td>115.9</td>
<td></td>
<td></td>
<td></td>
<td>dispensable</td>
<td></td>
</tr>
<tr>
<td>1441600</td>
<td>α/β-hydrolase</td>
<td>0 n</td>
<td></td>
<td>67.6</td>
<td></td>
<td>1305500</td>
<td>dispensable</td>
<td>dispensable</td>
<td></td>
</tr>
<tr>
<td>1427100</td>
<td>Lipase_3 α/β-hydrolase</td>
<td>1 y</td>
<td></td>
<td>155.7</td>
<td></td>
<td>1017500</td>
<td>dispensable</td>
<td>dispensable</td>
<td></td>
</tr>
<tr>
<td>0709900</td>
<td>Duf_900 α/β-hydrolase</td>
<td>4 n</td>
<td></td>
<td>274.0</td>
<td></td>
<td>1220500</td>
<td>essential</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1359900</td>
<td>Duf_900 α/β-hydrolase</td>
<td>4 n</td>
<td></td>
<td>249.8</td>
<td></td>
<td>1136100</td>
<td>dispensable</td>
<td>dispensable</td>
<td></td>
</tr>
<tr>
<td>1205900</td>
<td>Duf_726 membrane protein</td>
<td>5 n</td>
<td></td>
<td>124.2</td>
<td></td>
<td>0604600</td>
<td>essential</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1116000</td>
<td>RON4 Duf_676 serine esterase</td>
<td>0 y</td>
<td></td>
<td>135.6</td>
<td></td>
<td>0932000</td>
<td>essential</td>
<td>low fitness</td>
<td></td>
</tr>
<tr>
<td>0823400</td>
<td>α/β-hydrolase</td>
<td>0 n</td>
<td></td>
<td>58.6</td>
<td></td>
<td>0707200</td>
<td>essential</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1142900</td>
<td>S33C S33 α/β-hydrolase</td>
<td>0 n</td>
<td></td>
<td>61.3</td>
<td></td>
<td>0906100</td>
<td>slow</td>
<td>essential</td>
<td></td>
</tr>
<tr>
<td>0301300</td>
<td>EH1 / S33E S33 α/β-hydrolase</td>
<td>1 y</td>
<td></td>
<td>50.5</td>
<td></td>
<td></td>
<td>low fitness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0826200</td>
<td>S33B S33 α/β-hydrolase</td>
<td>0 n</td>
<td></td>
<td>70.7</td>
<td></td>
<td>0704400</td>
<td>dispensable</td>
<td>essential</td>
<td></td>
</tr>
<tr>
<td>1401300</td>
<td>EH2 / S33D S33 aminopeptidase</td>
<td>0 n</td>
<td></td>
<td>55.2</td>
<td></td>
<td>low fitness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1410100</td>
<td>S33A S33 α/β-hydrolase</td>
<td>1 n</td>
<td></td>
<td>41.5</td>
<td></td>
<td>1032400</td>
<td>dispensable</td>
<td>essential</td>
<td></td>
</tr>
<tr>
<td>0728700</td>
<td>α/β-hydrolase</td>
<td>0 n</td>
<td></td>
<td>84.8</td>
<td></td>
<td>0212800</td>
<td>essential</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1120400</td>
<td>abH112 α/β-hydrolase</td>
<td>0 n</td>
<td></td>
<td>44.7</td>
<td></td>
<td></td>
<td>essential</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0403800</td>
<td>S9C S9 α/β-hydrolase</td>
<td>0 n</td>
<td></td>
<td>83.4</td>
<td></td>
<td>1001500</td>
<td>sig slow</td>
<td>essential</td>
<td></td>
</tr>
<tr>
<td>0805000</td>
<td>S9B S9 α/β-hydrolase</td>
<td>0 n</td>
<td></td>
<td>28.5</td>
<td></td>
<td>1225600</td>
<td>low fitness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0321500</td>
<td>S9A acylaminoacyl-peptidase</td>
<td>0 n</td>
<td></td>
<td>125.9</td>
<td></td>
<td>1217000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1126600</td>
<td>sterol-ester hydrolase</td>
<td>0 n</td>
<td></td>
<td>81.1</td>
<td></td>
<td>0921800</td>
<td>dispensable</td>
<td>dispensable</td>
<td></td>
</tr>
<tr>
<td>0818600</td>
<td>PBLP BEM46-like α/β-hydrolase</td>
<td>1 y</td>
<td></td>
<td>34.9</td>
<td></td>
<td>0712200</td>
<td>essential</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1129300</td>
<td>α/β-hydrolase</td>
<td>0 n</td>
<td></td>
<td>221.7</td>
<td></td>
<td>0918900</td>
<td>dispensable</td>
<td>dispensable</td>
<td></td>
</tr>
<tr>
<td>1134500</td>
<td>PGAP1-like α/β-hydrolase</td>
<td>2 n</td>
<td></td>
<td>210.5</td>
<td></td>
<td>0913900</td>
<td>essential</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1143000</td>
<td>abH114 α/β-hydrolase</td>
<td>2 y</td>
<td></td>
<td>44.8</td>
<td></td>
<td>0906000</td>
<td>dispensable</td>
<td>essential</td>
<td></td>
</tr>
<tr>
<td>1116100</td>
<td>serine esterase</td>
<td>0 n</td>
<td></td>
<td>217.1</td>
<td></td>
<td>0931900</td>
<td>dispensable</td>
<td>dispensable</td>
<td></td>
</tr>
<tr>
<td>0630100</td>
<td>α/β-hydrolase</td>
<td>0 n</td>
<td></td>
<td>79.4</td>
<td></td>
<td>1128900</td>
<td>dispensable</td>
<td>dispensable</td>
<td></td>
</tr>
<tr>
<td>0814400</td>
<td>phospholipase DODH1 α/β-hydrolase</td>
<td>0 n</td>
<td></td>
<td>80.8</td>
<td></td>
<td>1423100</td>
<td>slow</td>
<td>essential</td>
<td></td>
</tr>
<tr>
<td>1306200</td>
<td>α/β-hydrolase</td>
<td>0 y</td>
<td></td>
<td>56.8</td>
<td></td>
<td>1404700</td>
<td>essential</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1458300</td>
<td>α/β-hydrolase</td>
<td>2 n</td>
<td></td>
<td>183.5</td>
<td></td>
<td>1322000</td>
<td>dispensable</td>
<td>dispensable</td>
<td></td>
</tr>
<tr>
<td>0808000</td>
<td>α/β-hydrolase</td>
<td>0 n</td>
<td></td>
<td>163.9</td>
<td></td>
<td>1222700</td>
<td>slow</td>
<td>essential</td>
<td></td>
</tr>
</tbody>
</table>

Colours indicate paralogue groups. PlasmoGEM (Schwach et al., 2015), piggybac screen (Zhang et al., 2018).
Some Pfα/βHs are similar enough to α/βHs in other organisms such that putative functions can be assigned, such as PF3D7_0629300 (PfLCAT), a predicted phosphatidyl choline-sterol acyltransferase. The P. berghei analogue of PfLCAT (PBANKA_1128100) has been shown to be involved in disruption of the liver-stage PV (Bhanot et al., 2005; Burda et al., 2015). However the vast majority have only been annotated with an α/β-fold or general lysophospholipases/esterase activity, which does not narrow down their predicted molecular function.

The few Pfα/βHs that have been studied display diverse activities and functions in the parasite. Four exported α/βHs were identified by their Plasmodium export element (PEXEL) motifs: two predicted lipases PF3D7_1001400 (PfXL1) and PF3D7_1001600 (PfXL2) and two predicted expoxide hydrolases PF3D7_0301300 (PfEH1) and PF3D7_1401300 (PfEH2/PfPAP; Spillman et al., 2016). PEXEL motif refers to a sequence near the N-terminus, which results in trafficking of a protein from the PV into the host cell (Hiller et al., 2004; Marti et al., 2004). The PfEH1 and PfEH2 were validated as expoxide hydrolases of epoxyeicosatrienoic acids (EETs), lipid-containing signalling molecules present in RBCs that are important in vasodilation and anti-inflammation in the blood. Other reports suggested PfEH2 is a proline aminopeptidase associated with changes in erythrocyte deformability (da Silva et al., 2016). The function of the two lipases is still unknown, but these four enzymes are only present in the Laverania subgenus, possibly suggesting a specific role in host immune response modulation (Spillman et al., 2016a). The most abundant α/βH in P. falciparum blood stages is a pro-drug resistance esterase (PfPARE, PF3D7_0709700). This is a non-essential esterase that has been shown to activate certain antimalarial compounds, and mutations in its sequence are associated with resistance (Istvan et al., 2017). Another annotated α/βH is the rhoptry neck protein PfRON4, although it’s potential enzymatic function has not been studied (PF3D7_1116000). It’s homologue TgRON4 was first characterised in T. gondii and has now been shown to be conserved in Plasmodium species and is important for tight junction formation during RBC invasion (Lebrun et al., 2005; Morahan et al., 2009). The Plasmodium Bud Emergence 46 protein (BEM46, PF3D7_0818600) is also annotated. BEM46 α/βHs are conserved in eukaryotes, little is known about their function but in other organisms they have been implicated in cell polarity and signal transduction (Kumar et al., 2013; Mercker et al., 2009).
The rest of the Pfα/βHs are largely unannotated. They can be assigned orthologue groups based on structural similarities (Table. 1). There are a number of closely associated pairs such as PfXL1 and PfXL2 and PF3D7_1476700 (Psta1) and PF3D7_1476800 (Psta2) that are in adjacent gene loci, possibly pointing to gene duplications. At least ten of the α/βH’s are part of a *Plasmodium* sub-telomeric family called the *pst-a* family (Table. 1; Carlton et al., 2002; Templeton, 2009). Others appear to be associated with members of another sub-telomeric family, the early transcribed membrane protein genes (*etramps*) (Spielmann and Beck, 2000). Both PF3D7_1001400 and PF3D7_1001600, and PF3D7_1401300 and PF3D7_1401500 flank *etramp* genes in sub-telomeric regions. Multicopy sub-telomeric gene families have been be the subject of much study in *Plasmodium*. These include the *var* gene family that encodes variant surface antigens, important determinants of parasite virulence. Sub-telomeric regions are characterised by a number of features: repeat regions; segmental duplications; considerable homology between chromosomes; and large multi-gene families such as the *var* genes (Fischer et al., 1997; Hernandez-Rivas et al., 1997; Templeton, 2009).

1.5 *Activity-based protein profiling (ABPP)*

1.5.1 *Activity based probes (ABPs) and activity-based protein profiling (ABPP)*

Traditional profiling techniques such as transcriptomics and proteomics measure the expression and abundance of proteins respectively. But these techniques fall short for proteins whose activities are highly post-translationally regulated such as proteases, kinases, and other hydrolases. Expression levels do not always correspond with protein abundance, as in the case of translationally regulated gametocyte proteins in *Plasmodium* (Mair et al., 2006). For 30% of *Plasmodium* genes, there is a delay between peak mRNA and protein levels (Foth et al., 2011; Le Roch et al., 2003). Enzymes are often expressed as zymogens that require cleavage before they can be active (Khan and James, 1998). Enzymes can also be activated by other post-translational modifications, such as phosphorylation (Cohen, 2000), or by interactions with small molecules or other proteins (Ngounou Wetie et al., 2014; Zorn and Wells, 2010).

ABPs are small molecule tools that specifically and covalently react with the key catalytic nucleophillic residue (e.g. Cys/Ser) of an enzyme family via an electrophilic warhead. The spacer/recognition element structure of ABPs can be tuned to confer specificity for a single enzyme or an entire class of enzymes. Crucially, all ABPs possess a reporter tag so that their
targets are labelled and thus can be visualised or measured. ABPs covalently modify enzymes via their intrinsic catalytic activity, requiring the catalytic dyad/triad to be properly formed and at the optimal pH. ABPs therefore discriminate between the active and inactive forms of the same enzyme and are useful tools in measuring enzymatic activity. In ABPP, broad spectrum ABPs are used to profile global the activity of whole enzyme families. They can therefore report on different patterns of enzyme activation between different biological samples or during biological processes such as cell fate commitment, life cycle progression, or responses to external stimuli. The ability to profile enzymes according to a common catalytic mechanism allows us to characterise proteins into functional classes. ABPP can also be used in competition with inhibitor libraries to determine the specificity of small molecules against all members of an enzyme family.

Affinity-based probes are also useful tools, but these do not label proteins using the intrinsic catalytic mechanism. They may still bind tightly to enzymes in the active site and tools such as photocrosslinker elements can be introduced into the probe so that covalent bonds are made upon irradiation, most likely to non-catalytic residues. An affinity-based probe is capable of reporting activity of an enzyme, for example if its specific binding is dependent on the enzyme adopting an active structure. The wide variety of ABPs have a huge utility in profiling enzyme activities both in vivo and in vitro.

Since the development of the first ABPs the scope and variety of studies using them has grown every year. There are a number of excellent reviews on ABPs and their uses in ABPP (Cravatt et al., 2008; Liu et al., 1999; Niphakis and Cravatt, 2014; Willems et al., 2014). Chemical probes have great utility in target validation for drug development (Cravatt and Sorensen, 2000; Drewes and Knapp, 2018; Wang et al., 2012; Ziegler et al., 2013). This field is only limited by the probes available, but the development of new chemical tools is not swift. The great usefulness of ABPP in target deconvolution relies on its use in the context of chemical proteomics. This is described in more detail in Section 1.6. Below are described some of the most common features of ABPs.

### 1.5.1.1 Visualisation of labelled proteins

There are many different and creative visualisation tools. The most common are isotope tags, fluorophores and biotin. Radioisotopes such as $^{125}$I, and $^3$H are useful in ABPs because they can be readily incorporated into any molecule without changing the structure. In this way any tight binding inhibitor could become a probe (Bogyo et al., 2000). However, there are drawbacks in terms of stability and the extra handling care that radiolabels require in the laboratory.

Fluorophores are the most diverse class of tags. They allow extremely swift and easy visualisation
of labelled proteins, both in gel and in cell. Their hydrophobicity means they readily penetrate cell membranes and can therefore be used in vivo to label proteins and, in some cases for imaging purposes by fluorescent microscopy. They are many to choose from, all with different properties, bleaching levels and price, allowing scientists to choose the best fit for their applications. Furthermore, probes with orthogonal fluorophores can be multiplexed in ABPP. Two commonly used fluorophores are TAMRA (TMR, also Rhodamine/Rho) and Cy-dyes. TMR is a relatively inexpensive fluorophore that emits in the red wavelength (Patricelli et al., 2001). It has some photo-bleaching drawbacks making it less desirable for imaging purposes. The Cy-dyes (eg Cy5, sulpho-Cy5, Cy3 etc.) are highly stable and amenable to many applications but are more expensive to incorporate into probes (Chan et al., 2004). Biotin can be used as a visualisation tag via affinity blotting methods but the incredible utility of this kind of tag lies in affinity-purification.

1.5.1.2 Affinity tags

Biotin is by far the most common affinity tag used in ABPs. Its small size and strong binding to the avidin has been exploited to give a powerful tool for the isolation of tagged proteins. Tagged proteins may be separated from complex mixtures by using avidin-coated (or avidin derivative-coated) resin/beads. The purified subset of relevant tagged enzymes can then be interrogated with powerful identification methods such as chemical proteomics (Section 1.6). There are two major drawbacks to biotin-avidin-based affinity purification systems. First harsh conditions such as heat and strong detergents, are required to disrupt the strong biotin-avidin interaction in order to released captured proteins. This means that elution of labelled proteins from resin is not simple and denatures or damages proteins. A number of methods have been developed to work around this problem including probes with readily cleavable linkers (Yang et al., 2013) or the elution step can be bypassed altogether by digesting proteins for proteomics while still attached to the resin (Section 1.6.2). The second problem is that biotin has low cell permeability, limiting its use to in vitro applications. Tandem tagging approaches fix this problem.

1.5.1.3 Post-labelling functionalisation

For intact cell and in vivo uses, tandem-tagging allows tags such as biotin to be added after labelling by highly specific biorthogonal ligation chemistry. A common strategy uses ‘click’ chemistry, the copper(I)-catalysed alkyne-azide cycloaddition (CuAAC) of functional groups (Rostovtsev et al., 2002; Speers et al., 2003; Speers and Cravatt, 2005). If a probe possesses an azide group, a labelled protein sample can be treated with standard reagents to conjugate an alkyne-reporter tag (or vice versa). Both the azide and alkyne groups are very small so their
incorporation into probes is unlikely to affect protein binding. Using this highly versatile method, an affinity purification step can be added to probe labelling protocols in any system, greatly expanding the utility of ABPPs and chemical proteomics.

1.5.2 Cysteine protease probes

The PLCPs are some of the most numerous and well-studied CPs in nature (Section 1.4.1.3). Their enzyme mechanism is highly conserved and well understood and many inhibitors have been developed based on electrophiles attached to a peptide specificity moiety. These include vinyl sulphones (Palmer et al., 1995) and epoxides (Barrett et al., 1982). Here we will discuss some of the probes that have been developed for PLCPs and used in *P. falciparum*.

1.5.2.1 E64 and DCG-04

The natural product E64 was first described in 1982 as a broad-spectrum clan CA inhibitor (Barrett et al., 1982). It has a leucine side chain and agmatine moiety that bind in the S2 and S3 positions respectively. The epoxide electrophile reacts irreversibly with the catalytic cysteine (Fig. 1.4.1; Fig. 1.5.1; Matsumoto et al., 1999). DCG-04 is a probe developed based on the structure of E64 with a tyramine moiety instead of the agmatine, which had been previously shown to have similar reactivity (Fig. 1.5.1; Greenbaum et al., 2000; Meara and Rich, 1996; Shi et al., 1992). Elaboration of the structure did not affect its reactivity, as predicted by the fact that the P2 substrate position is the main specificity determinant for PLCPs (Greenbaum et al., 2000). Therefore, a biotin affinity tag was added to turn this E64 derivative into a useful probe for profiling CP activities. DCG-04 has since been used as an ABP in a myriad of different assays, systems and pathogen screens including study of novel SARs and Ebola virus inhibitors (Shah et al., 2010); studying protease activity in *Arabidopsis* autophagy (Havé et al., 2018); and shedding of extracellular cathepsins in cancer (Sobotič et al., 2015).

In *P. falciparum*, these CP probes have made a huge contribution to the characterisation of the PLCPs. After the characterisation of the distinctive phenotype of E64 inhibition (Section 1.2.2.4; Section 1.4.1.3.1; Greenbaum et al., 2002; Salmon et al., 2001), DCG-04 was used to profile the proteases involved: FP1, FP2, FP3, DPAP1, and human calpain I (Greenbaum et al., 2002). This led to link between FP1 and invasion/oocyst production (Section 1.4.1.3.2) as well as the identification of FP1 specific inhibitors by competitive ABPP (Eksi et al., 2004; Greenbaum et al., 2002).
Figure 1.5.1: The structures of inhibitors and ABPs that target PLCPs. (A) The natural product E64 and the synthetic derivative DCG04 contain an epoxide electrophile, shown in purple. The attachment site of a TMR fluorophore on DCG04 is indicated. (B) The probes FY01, L-WSAK-Alk, W-sCy5-VS and W-DthioBio-TMR-VS and the inhibitors SAK1, SAK2 and L-WSAK all contain a vinyl sulphone warhead, shown in orange. The attachment sites of a TMR/sCy5 fluorophores or (desthio)Biotin are indicated.
Given the interest in PLCPs as antimalarial drug targets, DCG-04 has since been used in a wide range of studies to profile them, for example studies into the link between haemoglobin degradation and artemisinin resistance (Xie et al., 2016). DCG-04 has also been used with success in multiple other parasite systems including *T. gondii*, *Schistosoma* and *C. parvum* (Dou et al., 2013; Dvorák et al., 2008; Ndao et al., 2013).

**1.5.2.2 FY01**

FY01 was developed as a more selective probe than DCG-04 that could be used specifically to monitor the activity of a single protease: the exopeptidase CatC (Yuan et al., 2006). A cysteine-reactive vinyl-sulphone warhead was chosen, attached at the C-terminus to a dipeptide designed to bind to CatC, with an alkyl linker customisable tagging element distal to the warhead (Fig. 1.5.1; Yuan et al., 2006). In rat liver extracts FY01-BODIPY was shown to label specifically CatC, in comparison to broader labelling by DCG-04 (Yuan et al., 2006). FY01 was used in *P. falciparum* alongside DCG-04 to characterise the DPAPs (Section 1.4.1.3.3; Arastu-Kapur et al., 2008). Competition experiments led to the identification of inhibitors thought to be specific for DPAP1 and DPAP3, SAK2 and SAK1 respectively (Fig.1.5.1; Arastu-Kapur et al., 2008). It was the use of SAK1 that led to the linking of DPAP3 to egress, although it was later shown that this was due to off-target effects and that DPAP3 is only involved in invasion (Section 1.4.1.3.3; Arastu-Kapur et al., 2008; Lehmann et al., 2018).

**1.5.2.3 W-sCy5-VS**

Further efforts by our group to develop ever-more specific probes for PLCP members, led both to the identification of DPAP3-specific probes (de Vries et al., 2019; Lehmann et al., 2018) and a novel broad-spectrum PLCP probe, W-sCy5-VS-VS that labels more than either DCG-04 and FY01 (Section 3.3; Tan and Davison et al., 2020).

Based on the original screening results that identified SAK1, a more selective DPAP3 inhibitor was synthesised by replacing the N-terminal tyrosine with a tryptophan to give L-WSAK (Fig. 1.5.1). L-WSAK was more selective for DPAP3 over FP1-3 than SAK1. However like SAK1 these compounds blocked egress in DPAP3-null mutants showing there were still other off-target effects (Lehmann et al., 2018). DPAP3 investigation continued using libraries of peptide-based substrates and inhibitors to determine the optimal amino acids for the P1 and P2 position, as had already been done for DPAP1 and CatC (Fig. 1.4.1; de Vries et al., 2019; Poreba et al., 2014). Surprisingly, there were discrepancies between the activity of certain P2 residues in substrates.
compared to inhibitors. Aromatic residues in the P2 position gave potent DPAP3 inhibition, whereas in substrates P2 Phe/Tyr or Trp showed relatively poor turnover (de Vries et al., 2019). An alkyne version of L-WSAK, L-WSAK-Alk (also called L-Trp-hPG-VS) was tested and, although it had 5- to 100-fold less potency against recombinant DPAP3, showed the most potent inhibition of DPAP3 in intact parasites. This suggested an advantage in stability, cell permeability or higher effective concentration due to sequestration (Fig. 1.5.1; de Vries et al., 2019).

The alkyne of L-WSAK-Alk is in the S1 pocket, which is generally solvent exposed. It was postulated that the addition of fluorescent/affinity tags in this position would create a useful DPAP3-selective ABP while retaining the demonstrated activity (Tan and Davison et al., 2020). Using click chemistry, different fluorophores and tags were conjugated to the L-WSAK-Alk scaffold (Cy5, sulpho-Cy5, Cy3, sulpho-Cy3, desthiobiotin-TMR) and the activities and specificities of these probes were interrogated in whole parasite lysates (Tan and Davison et al., 2020). Surprisingly, W-sCy5-VS showed unprecedented broad-spectrum activity against the malarial PLCPs, targeting FP1-3 and DPAP1 & 3 at acidic pH making it a highly useful probe to profile all family members asexual life stages (Section 3.3; Tan and Davison et al., 2020). Unlike DCG-04 and FY01 this probe could be used to examine inhibitor specificities for all PLCPs in parallel. In addition, at neutral pH this probe labelled numerous other supposed CPs (Section 3.3).

1.5.3 Fluorophosphonate (FPP) Probes

It has long been known that serine-type hydrolases are inhibited by FPPs, and that binding of FPPs is dependent on the SH in question being in its active state (Piñeiro-Sánchez et al., 1997). Cysteine-, aspartyl- and metallo-hydrolases are mostly unaffected by such agents. Liu et al postulated that an FPP warhead attached to a reporter would give a highly useful tool to profile the activities of the huge number of SHs based on their activity (Fig. 1.5.2; Liu et al., 1999).

FPP probes were quickly widely adopted to profile and identify serine hydrolases in complex proteomes (Jessani et al., 2005; Liu et al., 1999; Okerberg et al., 2005; Patricelli et al., 2001) as well as to assess their activity and sensitivity towards inhibitors (Bachovchin et al., 2010; Kidd et al., 2001). For example, the identification and subsequent inhibitor development of SHs that have increased activity in cancer cells (Chiang et al., 2006). However, this probe may have even more scope and functionality than first thought. One study of human proteomes also suggested that some threonine hydrolases may be targeted as various proteasome sub-units were identified (Jessani et al., 2005). More recently the FPP probe helped characterise a novel non-typical class
of hydrolytic enzymes (Parsons et al., 2016). Now, various FPP probes functionalised with different tags are commercially available (Fig. 1.5.2).

![Figure 1.5.2: The structure of FPP ABPs that target SHs. FPP probes contain a fluorophosphonate warhead, shown in pink. The probes shown are all functionalised with a different tag. Top to bottom: Azide for tandem tagging, Biotin for affinity purification or TMR fluorophore for visualisation.]

1.6 Chemical proteomics

1.6.1 Chemical proteomics overview

After the influx of information from whole genome sequencing there has been a paradigm shift in the way we view biological processes. This change has been referred to as the post-genomic era, moving away from interrogation of single genes/proteins/processes and towards the development of other “omics” techniques to profile and classify biological macromolecules/processes on a whole cell level. Proteomics can identify and quantify individual proteins in complex samples using liquid chromatography and tandem mass spectrometry (LC-MS/MS). Proteomic techniques have been advancing rapidly and can achieve more than just quantification, including subcellular location (Zhu et al., 2019), protein-protein interactions (Turriziani et al., 2016) and temporal changes (Woo et al., 2017). Cell surface capture technology
can even directly profile extracellular proteins by mass spectrometry (Bausch-Fluck et al., 2015; Roux et al., 2012; Schnider et al., 2018). In modern proteomics, proteins are digested to peptides. Peptides have a better solubility than whole proteins and are easier to mass tag if needed. Peptides are quantified and fragmented before the sequence of amino acids is identified. A peptide of more than six amino acids is likely to be unique, so can be assigned to a protein. Very few peptide matches are needed to identify most proteins in a complex mixture, although it is always difficult to differentiate highly similar proteins that differ in only a few amino acids (e.g. \( PfFP2a \) and \( PfFP2b \)). The development of tools such as ABPs allow us to investigate protein activity separately from expression or abundance (Section 1.5). In addition, only relatively recently it is possible to perform routine whole proteome analysis with good depth of coverage (Corthals et al., 2000). Previously, it was desirable to find ways to reduce the complexity of proteomic samples, for example affinity purification of enzyme families using ABPs. The field of chemical proteomics refers to the MS-aided profiling of proteins in the context of their interaction with small molecules. Chemical proteomics is highly useful for target-deconvolution, and identification of off-target interactions (Wright and Sieber, 2016). ABPs are not the only small molecule tools used in chemical proteomics. Metabolic tagging approaches are also useful, using click chemistry-equipped metabolites such as a myristic acid analogue containing an alkyne to profile all myristoylated proteins, as has been achieved in \( P. falciparum \) (Wright et al., 2014). Probe-free methods also have much use such as cellular thermal shift assays (CESTA) which rely on the measurable change in protein stability upon ligand binding (Savitski et al., 2014).

### 1.6.2 Chemical proteomics methods

Sample preparation for proteomics is highly important. A very robust workflow is needed to ensure agreement between replicates. Each step of the protocol can and should be tailored depending on: the aim of the experiment; the controls and replicates required; the method of quantification; and the type of ABP being used.

In order to extend ABPP to a chemical proteomics workflow, a biotin affinity tag must either be present on the ABP, or be added later via click chemistry (Section 1.5.1.3). Labelled targets are purified from a complex mixture using avidin immobilised on agarose beads. In this thesis, neutravidin was used, a synthetic avidin derivative, which is inexpensive and retains tight binding with biotin. The complex protein sample is incubated with the beads to allow labelled proteins to bind. The beads are then washed to remove non-specifically bound proteins. Alternatively, the ABP itself can be immobilised on beads and contain a cleavable linker to aid with elution. While
target proteins are immobilised on the beads, the buffer can be exchanged and proteins treated
with reagents for the reduction and alkylation of free cysteines with minimal loss in material. The
reduction/alkylation reagents used are often important, depending on how labile the probe-protein
bond is. Elution methods to disrupt biotin/avidin interactions are harsh and many alternative
solutions have been described above (Section 1.5.1.2). One simple method is ‘on-bead digestion’; the
digestive enzyme (such as trypsin) is applied directly to washed beads to give a mixture of
tryptic peptides. These peptides can then be prepared directly for LCMS/MS analysis. Using on-
bead trypsin digestion has some problems with background noise, as the immobilised neutravidin
is also digested, but far less material is lost in sample preparation compared to elution techniques.
On-bead trypsin-digestion therefore maximises sample amounts and helps to standardise between
replicates and experiments. De-salting and buffer exchange is necessary before peptides can be
vacuumed to dryness ready for reconstitution in the LC-MS/MS running buffer. Quick protocols
to do this with minimal sample loss have been developed such as the “stage tipping” method
(Kulak et al., 2014). Fractionation procedures (such as high pH fractionation) can also be used to
further process the samples in order to reduce complexity and therefore increase coverage (Gilar
et al., 2005). The fractions are run separately on the LC-MS/MS and then the combined data is
assembled by analysis software to give a full picture of the complex sample.

1.6.2.1 LC-MS/MS

The mixture of peptides is fractionated by high performance liquid chromatography (HLPC) to
separate and concentrate peptides before they reach the tandem mass spectrometer (MS). In MS1
peptides are detected and their abundance may be measured. Exclusion windows are set to pick
the most abundant peptides, for example the top ten most abundant ions eluted in a one-second
time frame. These are subjected to fragmentation by high energy collision dissociation. This is
achieved by colliding ions with a stream of inert gas. Bond breakage occurs through the lowest
energy pathways (amide bond cleavage). The resulting fragment ions go on to MS2 and the
fragment spectra are collected for each peptide. From these MS2 spectra, the identity of the
peptide can be assigned.

1.6.2.2 Protein identification

The raw MS/MS data must be filtered and processed to identify what peptides (and therefore what
proteins) are present. In this thesis, we have used the open access software MaxQuant (Tyanova
et al., 2016a). MaxQuant starts with feature detection from the MS1: it finds peptide peaks from
the raw data and thus distinguishes them from background noise; local maxima are then found
and fitted to a Gaussian distribution. Peaks from different time points are used to create a three-
dimensional feature for each peptide detected, which will correspond to relative abundance. Iterative filtering processes are employed to maximise information and minimize background noise. If samples contain labelled peptide pairs such as heavy/light isotopes from SILAC (Section 1.6.3.2.1), this is when they would be identified.

MaxQuant has an integrated search engine called Andromeda (Cox et al., 2011). This uses a proteome database to match experimental MS2 spectra to known peptide sequences by a spectral comparison method. The database takes an input proteome and performs in silico digests to predict the tryptic peptides. The algorithm then generates virtual MS2 spectra for those peptides. The virtual and experimental spectra are then compared to match and assign peptides. The search also delivers a score that aims to limit erroneous protein IDs, called the false discovery rate (FDR). The FDR is found using posterior error probabilities (PEP) and decoy databases. PEP calculates the probability that a peptide-spectrum match (PSM) is incorrect (Käll et al., 2008a). Peptides are scored based on many characteristics such as missed cleavages, modifications, length, charge and other properties. A protein FDR is generated by multiplying its constituent peptide FDRs. FDR is calculated for each protein to ensure there is only a 1% chance of matches being made by chance. Some proteins are classed as ‘only identified by site’ which means the individual peptides pass PEP but the protein does not. Decoy databases are ‘false positive factories’. An easy way to generate one is to reverse the peptides predicted by the real database, meaning that the decoys have similar physiochemical properties (Käll et al., 2008b; Storey and Tibshirani, 2003). The FDR here is calculated as the number of decoy hits divided by the number of forward hits.

1.6.3 Methods of quantification

MS is not inherently quantitative: peptides ‘fly’ in the MS better or worse depending on their different physiochemical properties. Absolute quantification methods can measure copy numbers or concentration of a protein within a sample using internal standards. Relative quantification methods are experimentally easier. These calculate a quantitative ratio or relative change of protein between samples, relying on well-thought-out experimental controls and good agreement between replicates. All of these methods can be successfully combined with ABPP to give quantitative chemical proteomics workflows.

1.6.3.1 Label-free quantification

There are many computational approaches to quantify proteins from MS data (Mueller et al., 2008; Nahnsen et al., 2013). Label-free methods do not require any extra preparation steps and
can be applied to any types of samples, although they do require MS platforms with high resolution power and good mass precision such as OrbiTrap (Thermo) mass analysers. Although only one sample is analysed per MS run, any number of samples can be analysed by the same methods and compared. There is a high risk of systematic error though, so a robust workflow and consistent replicates are required.

1.6.3.1.1 MaxQuant label-free quantification (LFQ)

MaxQuant contains a label-free quantification algorithm (LFQ, (Cox et al., 2014)). This is a widely-applicable approach with a high dynamic range. It has two important features. First a “delayed normalisation” algorithm helps to compare samples that may have been handled in slightly different ways. For this to work there must be a dominant population of proteins that are unchanged between samples. In practice, this means that experiments need to be designed carefully with appropriate controls—samples for comparison that are too different will not be quantified to high precision. Secondly, peptide ratios from multiple samples are used to extract the maximum information for accurate quantification. This LFQ algorithm has been shown to perform well when compared to other available software (Weisser et al., 2013).

1.6.3.1.2 Intensity-based absolute quantification (iBAQ)

iBAQ is another label-free intensity quantification carried out in MaxQuant. It is used to estimate the relative abundance of proteins within samples. iBAQ values are quantified by scaling the raw intensities according to the number of theoretical peptides present. This provides accurate relative protein quantification (Arike et al., 2012; Krey et al., 2014).

1.6.3.2 Label-based quantification

1.6.3.2.1 Stable Isotope Labelling with Amino acids in Cell culture (SILAC)

SILAC is a metabolic labelling method. Amino acids containing stable heavy isotopes are introduced into cell culture and incorporated into proteins during normal cell protein synthesis (Ong and Mann, 2005). Typically, labelled Arginine and Lysine are used; trypsin cleaves after Arg or Lys so after trypsin digestion of proteins there is likely to be one labelled residue in each peptide. Up to three conditions can be labelled with different-weight Arg and Lys. These are mixed, processed for proteomics and analysed by LC-MS/MS together. Heavy/medium/light versions of the same peptides are identified in clusters during MS1 and relative quantification is achieved by calculating the intensity ratios between them. This increases sample throughput but having three states increases the complexity of the proteomic sample, so less coverage is achieved.
compared to label-free techniques. This method has high accuracy. The labelling is done early in the proteomic workflow to control for any systematic errors accumulated in sample preparation. Many specialised SILAC workflows now exist, for example pulse SILAC measures the changes in protein expression over time by taking protein samples before and after cells are switched to heavy Arg/Lys-containing growth media (Schwanhäusser et al., 2009). Super SILAC techniques allow this type of quantification even in cases where cell culture is not possible. An example is human tumour samples, where labelling is decoupled from the biological experiment by ‘spiking-in’ an internal standard (Geiger et al., 2010).

1.6.3.2.2 Stable isotope dimethyl labelling

The easiest and most inexpensive example of chemical labelling relies on reductive amination. This labels all free N-termini and lysines with isotopically tagged dimethyl groups (Hsu et al., 2003; Yang et al., 2018). All primary amines are converted to dimethylamines and, using heavy (2C13D3), medium (2CD3) and light (2CH3) methyl groups, three samples can be analysed at once. The samples can be combined after labelling and the mix is subjected to affinity purification and further proteomic preparation. This method has similar advantages and disadvantages to SILAC, but it is easier to apply to any sample type. As labelling is performed later, there is more risk of systematic error between samples.

1.6.3.2.3 Isobaric labelling

Isobaric mass tags are used to label tryptic peptides at free amines (Evans et al., 2012). The two most common labelling systems are isobaric tags for relative and absolute quantification (iTRAQ; Ross et al., 2004) and tandem-mass tagging (TMT; Dayon et al., 2008). This thesis focuses on TMT, which allows up to ten samples to be labelled and mixed for simultaneous LC-MS/MS. As the TMT labels are isobaric, there is no mass difference in MS1 for peptides from different samples. The sample complexity is not increased, so high coverage can be achieved (Mertins et al., 2018). This lack of mass difference also means that the signal for individual peptides is boosted by the summation of the samples, aiding identification, especially of lowly-abundant proteins. Unlike the other methods described above, quantification takes place in the MS2 rather than MS1. After fragmentation, the isobaric labels have distinct masses, and the quantification of these reporter ions gives accurate relative quantification of peptides between samples. Each identified peptide is quantified across all samples. This means that there are very few missing values, which can be a problem in other quantification systems. In 6-plex TMT there is an easily detectable 1 Da difference between the reporter ions. 10-plex labelling relies on the 6.32 mDa difference between 13C and 15N, so this can only be used with appropriately sensitive
instrumentation. TMT labelling gives very good coverage of complex mixtures and precise quantification, allowing much higher throughput of samples than other techniques. However, TMT reagent kits are expensive and require careful handling as they are solvent and pH sensitive. Labelling is done on the peptide level, so there must be robust methods in place to process samples before trypsin digestion to reduce systematic error between samples. Another problem that can arise with both TMT and iTRAQ is ratio compression. This occurs when similar contaminating peptides are isolated together with targeted peptides leading to underestimation of ratio differences. To combat this one can adjust precursor selection on the MS, or reduce the complexity of a sample by fractionation. The problem is completely avoided if MS3 is used for reporter quantification, but obviously this requires specialist instrumentation (Ting et al., 2011).

1.7 Advances in genetic manipulation of *Plasmodium falciparum*

provide novel opportunities to study parasite biology

*P. falciparum* is responsible for the greatest malaria burden of all *Plasmodium* spp.. The ease of *in vitro* culture has meant *P. falciparum* is more widely studied than any other human malaria species (Trager and Jensen, 1976). When the genome sequence of *P. falciparum* was published in 2002 (Gardner et al., 2002), only 61% of the 5268 predicted genes could be assigned functional roles based on similarity to known proteins in other organisms. This is much greater uniqueness than generally seen in eukaryotes. The lack of similarity is probably related to the remarkable AT-rich nature of the genome, at 80% A/T the most ever sequenced at that time. Seventeen years later, the improved genome annotation now reports 5438 genes of which only 33% have no predicted function (Böhme et al., 2019). The web-based open resource *PlasmoDB* is an amazing repository for genomic, transcriptomic and proteomic data helping to integrate and further our understanding of the *P. falciparum* genome as a community (Aurrecoechea et al., 2009).

Although genetic manipulation of *P. falciparum* has been achievable since 1995 (Wu et al., 1996, 1995), the AT-richness of the genome and the very low transfection efficiency have both meant it remained challenging (O’Donnell et al., 2002). *P. falciparum* lacks essential components of the canonical non-end homologous joining pathway that is usually responsible for repairing double strand breaks (DSB) in eukaryotes (Gardner et al., 2002). Therefore, DSB repair can only take place via homologous recombination. This can be exploited by transfecting parasites with plasmids containing an integration construct with a region homologous to the target locus. When a DSB occurs the parasite will repair the damage using the integration plasmid as a template,
therefore integrating it into the genome sequence. DSB events are relatively rare. Plasmids now commonly contain selectable markers, such as the human dihydrofolate reductase (hDHFR) (Fidock and Wellems, 1997) but for a long time few selectable markers were available. It can take months of drug cycling before chromosomal integration is observed because the episomal targeting plasmid persists in parasites conferring drug resistance in the absence of integration. In the case of gene disruption, a negative result is ambiguous, as it is not clear if KO is not possible due to gene essentiality or technical issues. The asexual stages are haploid, so any gene disruption is a full KO, and straightforward gene disruption of essential genes is lethal. However after transfection there will be a mixture of integrant and wild-type (WT) parasites. Even if a gene disruption gives a modest fitness cost, integrant parasites may be out-competed by the WT parasites and never detected. It is also important to note that given time parasites could compensate for the loss of an important, but not essential, gene by upregulating others. In this case KO parasites would be detected with no apparent change in growth or replication. The development of conditional KO (cKO) systems in \textit{Plasmodium} has been hugely beneficial both in confirming (or debunking) previously-assigned essential genes. cKO systems can also help to probe the molecular function of proteins, as one can monitor the associated mutant phenotype immediately after cKO.

1.7.1 Conditional gene disruption systems

1.7.1.1 Conditional knockdown systems

Conditional knockdown systems allow the knockdown (rather than KO) of a gene to be controlled by the addition of a small molecule trigger. One can target any one of: the gene, the transcribed RNA, translation of RNA, or the protein product. RNA-level knockdown approaches have been hampered by the fact that RNA interference approaches are not functional in \textit{Plasmodium} (Baum et al., 2009). However a ribozyme-based system has been described, whereby an inducible self-cleaving glmS ribozyme sequence is inserted downstream of the Gene Of Interest (GOI) and subsequently expressed in the gene mRNA (Prommana et al., 2013). Another system interferes with the translation of mRNA: TetR-binding aptamers are inserted up/downstream the GOI such that, in the absence of anhydrotetracycline, TetR protein binds to the aptamers and prevents mRNA translation by the ribosome (Goldfless et al., 2014). A variety of methods affect the protein product itself. In these methods sequences are inserted into the GOI locus to result in a fusion protein that can be controlled by small molecules. Protein mislocalisation knocks a protein “sideways”, by fusing an inducible localisation signal to the protein (Oberli et al., 2016). A
number of different protein degradation methods have been described where an unstable protein domain is fused to the protein of interest. A stabilising ligand maintains the system but, when it is removed, the unstable protein is rapidly degraded by the proteasome (Banaszynski et al., 2006; Muralidharan et al., 2011). While these are all powerful techniques, they are “knockdown” rather than “knockouts” and the level to which the gene product is depleted may vary between different parasites and depending on the GOI. As with all fusion proteins, there is also the risk that the fusion does not behave exactly as the WT protein in its localisation and interactions with other proteins.

1.7.1.2 Inducible, Dimerisable Cre recombinase (DiCre) conditional knockout (cKO)

True cKO requires the GOI to be excised or disrupted at a genome level, giving a complete ablation of expression. An inducible DiCre system has been adapted for use in *P. falciparum* and allows for rapid and efficient conditional gene disruption (Collins et al., 2013a). Cre recombinase is a protein from the bacteriophage P1 that catalyses site specific recombination between two sites called *loxP* sites (Abremski and Hoess, 1984). This process was adapted into the DiCre system in 2003, which gave excellent temporal control over Cre recombinase function (Jullien et al., 2003). The Cre recombinase is expressed in two fragments, each of which is fused to a different rapamycin (RAP) binding domain. Separately the fragments are inactive but addition of RAP promotes dimerization and restores Cre recombinase activity. If *loxP* sites are inserted to flank the gene or DNA sequence of interest, then excision of that gene can be rapidly induced by the addition of RAP. Futhermore, *loxP* sites can be introduced within introns to minimise disruption of the coding sequence. Sequences are integrated into the genome such that the WT GOI is expressed until the addition of RAP, this is the conditional switch that triggers DiCre mediated excision to give the desired mutant expression, such as cKO.

This system was successfully adapted for use in *P. falciparum* in 2013 and has been incredibly powerful for cKOs ever since (Collins et al., 2013a). The AT-rich *Plasmodium* genome continued to present some challenges however, as it was difficult to target *loxP* sites to intergenic regions and there are relatively few introns. Recently, artificial introns have been used to introduce *loxP* sites, called the *loxPint* module (Jones et al., 2016). Both the DiCre and *loxPint* systems are used in this thesis (Chapter 7).
1.7.2 Technologies for improving integration

1.7.2.1 CRISPR Cas9

Recently, a highly-efficient gene editing technique has been adapted for use in *Plasmodium*. Clustered regularly interspaced short palindromic repeats (CRISPR) and the CRISPR-associated protein 9 (Cas9) form the CRISPR-Cas9 system, a prokaryotic adaptive immune mechanism (Bhya et al., 2011). CRISPR-Cas9 significantly decreases the long time it takes to achieve stable integrants when relying on stochastic DBS and homologous recombination. This system has been exploited by molecular biology to allow for highly-efficient and targeted genome editing (Doudna and Charpentier, 2014). The Cas9 endonuclease makes DSBs targeted towards a locus of interest using a single guide RNA. The adaptation of CRISPR-Cas9 to *P. falciparum* (Ghorbal et al., 2014) has made genome editing much quicker and more efficient.

1.7.2.2 Selection-linked integration

Other strategies can be employed to improve the efficiency of single homologous recombination-based techniques. In this thesis we employ selection-linked integration (SLI, Fig. 1.7). SLI is a method whereby a resistance marker selects for chromosomal integration of an plasmid (Birnbaum et al., 2017a). SLI relies on a 2A element from the virus *Thosea asigna* (T2A). T2A is a linking peptide that causes a ribosome to skip a bond formation during mRNA translation (Kim et al., 2011). When the T2A sequence is fused between two coding regions, two proteins are translated from the same mRNA. In cKO systems, T2A can be used to link the region expressing the WT version of the GOI (before conditional switch) to a selectable marker. In this thesis we use the neomycin phosphatase gene (*npt*), an infrequently used selection marker in *P. falciparum*, which confers resistance to the drug G418 (Wang et al., 2002).

When a plasmid, targeted to the GOI locus, is integrated into the genome, the coding sequence on the plasmid will be expressed via the endogenous promoter (Fig. 1.7). In cKO systems, a recodonsis version of the WT GOI is present on the integrating plasmid and will be expressed before RAP is added. T2A allows us to link the WT GOI to *npt* so that both proteins are expressed by the endogenous promoter. Only when the plasmid is integrated will *npt* be expressed and the parasites become resistant to G418. In this way, chromosomal integration can be selected for by treating transfected parasites with G418. Drug cycling is not required so integrants are detected far quicker than traditional methods, on a similar timescale to CRISPR-Cas9 technology. A caveat of SLI is that *npt* expression relies on the endogenous promoter, if expression levels of the GOI are not high enough then the parasite will never exhibit G418 resistance.
1.7.3 Conditional allelic replacement strategy

The conditional KO/knockdown systems described above allow us to investigate the effect on parasite growth and replication when a protein is eliminated. Thus one can make inferences about the protein’s role in parasite cellular biology. However, determining the molecular and cellular function of a protein using these methods is not always straightforward. Effects on parasite growth and replication may be a complex result of changes caused by many reasons: the loss of protein/protein interactions; mislocalisation of members of a protein complex; loss of downstream processes in a reaction cascade to name a few. If some information about a protein target is known from its sequence or predicted structure/homology, targeted conditional mutation strategies may be able to narrow down the search for specific KO outcomes. For example, mutating the active site of a suspected enzyme. This allows one to specifically test if the activity of that enzyme is essential for parasite growth and replication, without affecting any non-enzymatic functions of the protein. The molecular function may still be difficult to assign due to downstream reaction cascades but in this case, there are fewer variables to consider. Conditional allelic replacement is
one way of achieving conditional mutation of an active site. In this strategy two copies of the GOI are integrated into the genome at the endogenous gene locus: one copy is wild type and flanked by `loxPint` sites (Section 1.7.1.2) and the other copy is a catalytically dead mutant sitting out of frame. After rapamycin-induced DiCre-mediated excision of the WT copy, the mutant allele comes into frame and is now expressed by the endogenous promoter. Thus, the effect of the ablated enzyme activity can be assessed while controlling for the presence and localisation of the protein.

**1.7.4 High-throughput genetic screening**

Recent high-throughput genetic screens in *Plasmodium* have generated genome-wide essentiality data. *P. berghei* has long been the most genetically tractable *Plasmodium* species. A new class of low copy number linear plasmids allowed for the generation of a genomic library in *E. coli,* covering the whole of the *P. berghei* genome (Godiska et al., 2010; Pfander et al., 2011). The library was transformed into barcoded genetic modification vectors and made into a community resource: the *Plasmodium* genetic modification project (*PlasmoGEM*; Schwach et al., 2015; Wang et al., 2006; Zhang et al., 1998). Using *PlasmoGEM*, large genetic screens have been done in mice, with more data being deposited into the *PlasmoGEM* database continuously (Bushell et al., 2017). As *P. falciparum* shares a core of around 4500 genes (85%) with *P. berghei*, this data has become an important resource in predicting essentiality of genes. More recently however, the first whole genome-wide screen has been performed in *P. falciparum* (Zhang et al., 2018). This study used high-throughput piggyBac transposon mutagenesis to saturate the genome. *P. falciparum* mutants in excess of 38,000 were quantified by quantitative insertion site sequencing to identify 2680 genes likely essential for *in vitro* asexual stage parasite growth (Zhang et al., 2018). It is extremely important to note that both of these screens measure mutant fitness in competition with other mutants, and care must be taken in assuming that “essential” or “dispensable” labels are definitive proof of either. There is no denying that these resources are a huge boon in prioritising potential drug targets, but in many cases cKO is the only way of definitively validating essentiality.
Chapter 2. **Aims of this work**

- To profile the activity of *Plasmodium falciparum* serine hydrolases and cysteine proteases in competition with the Malaria Box (MMV) in order to identify inhibitor targets.

- To use chemical proteomics to profile the activity of *Plasmodium falciparum* serine hydrolases throughout the asexual lifecycle of the parasite in order to learn more about this large enzyme family and identify interesting targets for genetic validation.

- To generate conditional allelic replacement lines of selected targets in order to test the essentiality of serine hydrolase catalytic activity on parasite growth.
Chapter 3. **Results I: Characterisation of W-sCy5-VS and FPP-TMR Activity-Based Probes (ABPs) in *P. falciparum***.

ABPs are highly useful tools for profiling the activity either of a specific enzyme or of an enzyme family with a conserved catalytic mechanism (Section 1.5). This thesis aims to profile two enzyme families in *P. falciparum*: CPs and SHs. Here we discuss the characterisation of ABPs for each enzyme family. The intention being, to combine the ABPs in a competition screen to identify targets of MMV antimalarial compounds from the Malaria Box (Section 1.3.4; Spangenberg et al., 2013). It has been shown that CPs are important in numerous human diseases and pathogens, including *P. falciparum*. The malarial PLCPs particularly have been investigated as potential drug targets and much of their characterisation has relied on ABPs (Section 1.5.2). In this chapter we describe the characterisation of a new broad-spectrum probe (W-sCy5-VS) to target the PLCPs and other potential CPs. We also use the well-characterised FPP-TMR probe to profile metabolic SHs, a family largely uncharacterised in *P. falciparum* (Section 1.5.3, Section 1.4.2.2). The FPP probes were later used to profile the activity of SHs across the erythrocytic cycle by chemical proteomics (Chapter 5).

### 3.1 FPP probes profile PfSH activities in lysates and intact parasites.

FPP probes (Fig. 1.5.2) have been extensively used to profile and identify SHs in complex proteomes (Liu et al., 1999; Okerberg et al., 2005; Simon and Cravatt, 2010) and to assess their activity and sensitivity towards inhibitors (Section 1.5.3; Bachovchin et al., 2010; Kidd et al., 2001). In this thesis they were applied to study SHs in *Plasmodium* parasites in an effort to chemically annotate this large enzyme family.

#### 3.1.1 FPP-TMR labels multiple PfSHs active at different life stages.

During the intra-erythrocytic stages of *P. falciparum*, 80% of genes are expressed in a cyclic manner (Bozdech et al., 2003). Lysates from different parasite life stages were labelled with the commercially-available fluorescent FPP-TMR (Fig 1.5.2) to observe changes in the activity profile of SHs throughout the parasite life cycle.

Different life stages were collected from cultures of synchronous parasites (Materials and Methods 9.2.1). Ring and trophozoite stage parasites were collected by treating cultures with 0.15% saponin (w/v) and washed to remove RBC debris. Saponin treatment causes the RBC and PV membranes to lyse due to osmotic pressure. This step removes the majority of haemoglobin which interferes with probe labelling, in-gel fluorescence and proteomics sample preparation. However, it is important to
note that proteins that are exported into the PV or RBC are lost. Schizonts were purified from uninfected RBCs by density centrifugation before saponin lysis. Purified mature schizonts were collected after a 4-h treatment with 1 μM C2, which inhibits PKG, arresting schizont development before egress (Section 1.2.2.4). Merozoites were purified from egressing culture by centrifugation.

Frozen pellets of each parasite life stage were lysed by suspending them in 0.1% TritonX in PBSa buffer on ice, before the soluble protein fraction was isolated by centrifugation. These lysis conditions were found to result in the most efficient labelling compared to freeze-thaw and NP40 lysis (Supplementary Fig. 10.1). The soluble fraction was collected and diluted to 6.67 mg/ml before labelling with 1 μM FPP-TMR in PBSa (pH 7.2) at room temperature (RT). The use of TritonX for protein extraction ensured that the majority of bands labelled in the insoluble fraction were observed in the soluble fraction (data not shown). Samples were run on SDS-PAGE gels and labelling was visualised using a fluorescence scanner. Across all life stages, approximately 22 different proteins were labelled and a

Figure 3.1: Characterisation of broad-spectrum FPP probes. (A) To profile SH activities throughout the asexual life cycle, lysates from rings, trophozoites, early and mature schizonts, and merozoites were labelled for 30 min with 50 or 100nM FPP-TMR. (B) The labelling of FPP-TMR and FPP-N3 were compared in schizonts using lysates and intact parasites. For intact labelling, schizonts were purified and diluted 1/10 in media, 0.5 μM probe was added and incubated for 30 min at room temperature (RT) with gentle shaking. Schizonts were saponin lysed and the soluble protein extracted. Schizont lysates were labelled with 1 μM probe for 30 min. Samples labelled with FPP-N3 were treated with CuAAC-chemistry reagents to attach an alkyne-sCy5 fluorophore for visualisation. (C) Schizont lysates were labelled with 0.1 or 1 μM of FPP-N3 for 30 min before labelling with 0.1 μM FPP-TMR for 30 min. (A-C) All labelling reactions were quenched by addition of loading buffer (LB) and immediate boiling. 12.5 % SDS-PAGE gels were run and a fluorescence scanner was used to visualise TMR (red channel Ex/Em 552/572nm).
different activity pattern was seen at each life stage (Fig. 3.1A). Thus FPP-TMR labels multiple SHs that are expressed or activated in a cyclic manner throughout the \textit{P. falciparum} asexual life cycle.

### 3.1.2 FPP-TMR and FPP-N\textsubscript{3} probes have similar reactivity in \textit{P. falciparum}.

The FPP-containing probes have been used to profile SHs for over twenty years. Different versions are now commercially available with distinct functionalisation. Probes such as FPP-Biotin or FPP-N\textsubscript{3}, which can be functionalised with biotin via CuAAC-chemistry, can be used to affinity purify targets (Fig 1.5.2). We compared FPP-TMR vs FPP-N\textsubscript{3} labelling in both schizont lysates and intact schizonts. This would verify that the different probes has the majority of their targets in common. The labelling in schizont lysates was achieved as described above (Section 3.1.1). To label intact parasites, purified schizonts were diluted 1/10 in RPMI media and treated with 0.5 μM probe for 30 min at RT with gentle shaking. The schizonts were washed and saponin lysed before the soluble protein was extracted. FPP-N\textsubscript{3} labelled lysates were then treated with CuAAC-chemistry reagents to attach an alkyne-sCy5 fluorophore for visualisation of labelling.

The probes exhibited very similar activity profiles in lysates with few exceptions. This indicates that the FPP-containing probes have most of their targets in common, although there are some differences (Fig. 3.1B). The activity profiles in intact schizonts were distinct from those in lysates and differed between the two probes. Differences in labelling between lysate and intact cells is to be expected. On one hand proteins may be sequestered in organelles and membrane compartments in intact parasites and therefore might not be accessible to the probes. On the other hand, probes may also concentrate in sub-cellular compartments, and therefore label some proteins more efficiently in intact parasites than in lysates. Additionally, proteins that do not retain their activity in lysates may be labelled only in their native environment. FPP-N\textsubscript{3} labelled more proteins in intact parasites than FPP-TMR. FPP-N\textsubscript{3} is smaller and therefore likely to access more targets in a cellular environment.

To further verify that the probes had the same targets, we competed FPP-N\textsubscript{3} against FPP-TMR. Schizont lysates were pre-treated with 0.1 or 1 μM of FPP-N\textsubscript{3} for 30 min before labelling with 0.1 μM FPP-TMR for 30 min. At 0.1 μM FPP-N\textsubscript{3}, all FPP-TMR labelled bands were outcompeted apart from 2 bands, which merely decreased in intensity. At 1 μM FPP-N\textsubscript{3} all FPP-TMR labelled bands disappeared (Fig. 3.1C). Thus, at saturating conditions FPP-N\textsubscript{3} targets all the proteins that FPP-TMR targets. This is useful as we can perform chemical proteomics with FPP-N\textsubscript{3} and the results will infer the targets of FPP-TMR.
3.2 Proteomics confirms that FPP-N₃ targets PfSHs.

Chemical proteomics was used to both identify the FPP-N₃ targets at schizont stage and confirm that they were indeed SHs. Schizont lysates were treated with DMSO or with 1 μM FPP-N₃ in triplicate (Fig. 3.2A). An alkyne-biotin tag was added to modified proteins via CuAAC-chemistry, and the excess small molecule reagents were removed by protein precipitation. The labelled proteins were resolubilised and affinity-purified using neutravidin-agarose beads. After washing steps, free cysteines were reduced and alkylated. Proteins were digested by the addition of trypsin directly to the beads. The resulting tryptic peptides were prepared for proteomics by stage tipping, and analysed by LCMS/MS. For further details on the chemical proteomics method see Materials and Methods (Section 9.4).

**Figure 3.2:** Chemical proteomics protocol for the identification of the targets of FP-N₃.  
A) Schematic to illustrate the main steps in the chemical proteomics protocol. Soluble protein lysate was treated with FP-N₃ probe or DMSO. CuAAC-chemistry reactions attached the biotin-alkyne affinity tag. Labelled proteins were purified with neutravidin-agarose beads. Trypsin was applied to the beads to digest captured proteins. After clean-up steps the resultant peptides were analysed by LC-MS/MS.  
B) Volcano plot showing the enrichment of proteins in FP-N₃-labelled samples, compared to DMSO-treated controls. After a student t-test (s0 = 0.5, FDR = 0.01), the negative Log (p-value) was plotted against the Log2 Fold difference between the FP-N₃ sample compared to DMSO to give a visual representation of the proteins that were significantly enriched. The significantly enriched proteins are indicated in red and labelled with their gene ID.
Proteins were identified by MaxQuant analysis and quantified by LFQ (Section 1.6.3.1.1). The list of identified proteins was filtered to remove contaminants, proteins only identified by site (Section 1.6.2.2), proteins identified from decoy peptides (Section 1.6.2.2), and proteins identified from less than two peptides. After filtering, an average of 146 proteins were identified across three replicates. Fifteen proteins were significantly enriched in FPP-N\textsubscript{3} vs DMSO-treated samples with a fold change of at least 3 (student t-test, s\textsubscript{0} = 0.5, FDR = 0.01; Fig. 3.2B). Of these, eleven were predicted SHs, one patatin-like phospholipase and ten α/βHs (Table 3.2). Few of these SHs have known functions, but some are putative metabolic enzymes such as esterases and lysophospholipases (Section 1.4.2). This represents over one fifth of the 43-predicted α/βHs in \textit{P. falciparum} (Table 3, Table 3.2). We have shown that FPP probes label SHs in \textit{P. falciparum}. They appear principally to target α/βHs, a large family that is mostly uncharacterised in \textit{Plasmodium} species. FPP-TMR can therefore be used in ABPP screens in \textit{P. falciparum} to identify metabolic SH inhibitors.

### Table 3.2 Proteins enriched in FPP-N\textsubscript{3} compared to DMSO

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Protein name</th>
<th>Mol. weight [kDa]</th>
<th>Fold-enrichment (log\textsubscript{2})*</th>
<th>-Log (t-test p value)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF3D7_0709700</td>
<td>PARE</td>
<td>42.4</td>
<td>8.59</td>
<td>3.35</td>
<td>Prodrug activation and resistance</td>
</tr>
<tr>
<td>PF3D7_1001600</td>
<td>XL2</td>
<td>88.6</td>
<td>5.40</td>
<td>3.32</td>
<td>Exported lipase 2</td>
</tr>
<tr>
<td>PF3D7_1038900</td>
<td>Esterase</td>
<td>41.8</td>
<td>5.28</td>
<td>2.51</td>
<td>Putative protein</td>
</tr>
<tr>
<td>PF3D7_0403800</td>
<td>u/fi hydrolase</td>
<td>83.4</td>
<td>4.78</td>
<td>3.02</td>
<td>Putative protein</td>
</tr>
<tr>
<td>PF3D7_0805000</td>
<td>u/fi hydrolase</td>
<td>28.5</td>
<td>4.59</td>
<td>2.07</td>
<td>Putative protein</td>
</tr>
<tr>
<td>PF3D7_1143000</td>
<td>u/fi hydrolase</td>
<td>44.8</td>
<td>3.97</td>
<td>3.34</td>
<td>Putative protein</td>
</tr>
<tr>
<td>PF3D7_1328500</td>
<td>u/fi hydrolase</td>
<td>115.9</td>
<td>3.80</td>
<td>2.22</td>
<td>Putative protein</td>
</tr>
<tr>
<td>PF3D7_0818600</td>
<td>PBLP</td>
<td>34.9</td>
<td>3.75</td>
<td>1.81</td>
<td>Putative BE46-like protein</td>
</tr>
<tr>
<td>PF3D7_0209100</td>
<td>Phospholipase A2</td>
<td>78.3</td>
<td>3.39</td>
<td>2.10</td>
<td>Putative patatin-like phospholipase</td>
</tr>
<tr>
<td>PF3D7_1476800</td>
<td>Lyso phospholipase</td>
<td>42.7</td>
<td>2.91</td>
<td>1.96</td>
<td>Putative protein</td>
</tr>
<tr>
<td>PF3D7_0929400</td>
<td>RHOPH2</td>
<td>162.7</td>
<td>2.91</td>
<td>2.32</td>
<td>high molecular weight rhoptry</td>
</tr>
<tr>
<td>PF3D7_0321500</td>
<td>SSA</td>
<td>125.9</td>
<td>2.78</td>
<td>2.35</td>
<td>Putative peptidase</td>
</tr>
<tr>
<td>PF3D7_0917900</td>
<td>BIP</td>
<td>72.4</td>
<td>2.30</td>
<td>1.28</td>
<td>heat shock protein 70</td>
</tr>
<tr>
<td>PF3D7_0827900</td>
<td>PDH8</td>
<td>55.5</td>
<td>2.20</td>
<td>1.51</td>
<td>protein disulfide isomerase</td>
</tr>
<tr>
<td>PF3D7_0930300</td>
<td>MSP1</td>
<td>195.7</td>
<td>1.69</td>
<td>2.53</td>
<td>merozoite surface protein 1</td>
</tr>
</tbody>
</table>

*Fold enrichment (log\textsubscript{2}) = 1.58 is equivalent to Fold change = 3

| α/β-fold SHs |

### 3.3 W-sCy5-VS is a broad-spectrum PLCP ABP.

Members of the PLCP family have been found to play roles in essential processes throughout the \textit{Plasmodium} life cycle (Section \textbf{Error! Reference source not found.}). They can be split into three subfamilies: the FPs (FP1, FP2a, FP2b and FP3); the DPAPs (DPAP1, DPAP2 and DPAP3); and the SERAs (SERA1-9). Of the SERAs, only SERA6-8 actually have a catalytic cysteine. Due to the short activation window of SERAs during the asexual-blood stage, ABP-based characterisation has focused on the FPs and DPAPs (Section 1.5.2). Previously published CP probes do not label all the PLCPs
simultaneously: DCG04 labels the FPs but not all DPAPs (Greenbaum et al., 2000), whereas FY01 labels the DPAPs efficiently but not all FPs (Yuan et al., 2006).

W-sCy5-VS was initially designed to be a DPAP3-selective ABP. An initial screen of vinyl sulphone-containing DPAP inhibitors suggested that a Trp in the P2 position would be selective for DPAP3 (Arastu-Kapur et al., 2008). Subsequently the compound L-WSAK-Alk was shown in our lab to be a potent inhibitor of DPAP3 in intact parasites (Fig. 1.5.1; de Vries et al., 2019; Lehmann et al., 2018). L-WSAK-Alk has an alkyne in the P1 position to which a fluorophore was added via CuAAC-chemistry to yield the probe W-sCy5-VS (Fig. 1.5.1).

W-sCy5-VS was tested in parasite lysates at acidic pH. Unexpectedly, this probe was only selective for DPAP3 at 1 nM, at higher concentrations many PLCPs were labelled (FP1, FP2, FP3, DPAP1 and DPAP3; Tan and Davison et al., 2020). L-WSAK-Alk was only able to outcompete W-sCy5-VS for DPAP1 and DPAP3, even if used at ten times the concentration (Tan and Davison et al., 2020). The broader specificity of W-sCy5-VS compared to previous probes meant that potential inhibitors could now be screened for activity against both DPAPs and FPs simultaneously. This would not only increase the efficiency of inhibitor screening but also make it easier to identify off-target effects when designing specific inhibitors. A drawback is that the sCy5 fluorophore precludes the probe from in-cell applications. The following experiments were performed both to confirm targets of W-sCy5-VS and to investigate its reactivity and specificity at neutral pH. The hypothesis was that under neutral conditions other CPs, that do not reside in acidic organelles, may be labelled. If PLCPs and other CPs could be labelled at neutral pH, then W-sCy5-VS could be used in combination with the FPP SH probes (Section 3.1) to perform multiplexed ABPP for both enzyme families.

### 3.3.1 W-sCy5-VS labels recombinant PLCPs.

W-sCy5-VS was tested against available recombinant proteases to confirm the PLCP labelling that was observed in parasite lysates (Fig. 1.5.1; Tan and Davison et al., 2020). Bovine cathepsin C (CatC) and three recombinant malarial PLCPs (FP2a, FP3 and DPAP3) were labelled with 1 μM W-sCy5-VS in acidic pH for 30 min. This pH is optimal for their reactivity (Lehmann et al., 2018; Shenai et al., 2000, p. 2; Sijwali et al., 2001). All the PLCPs were labelled as expected (Fig. 3.3A). It was noted that rFP2 and rFP3 are the same size and will likely run concomitantly in gels of labelled parasite lysates (Fig. 3.3A). Two bands are labelled in rDPAP3 samples, as expected from previous studies with FY01 (Lehmann et al., 2018).
3.3.2 W-sCy5-VS labels multiple unknown *Pf*CPs at neutral pH.

W-sCy5-VS was then compared to previously characterised probes, FY01 and DCG04 at neutral pH. Probe labelling was also tested in competition with the broad-spectrum CP inhibitor, E64 (Section 1.5.2.1). Lysates were pre-incubated for 30 min with E64 (10 μM) or DMSO followed by 1 h labelling with 1 μM of W-sCy5-VS, FY01-Cy5 or DCG04-TMR.

**Figure 3.3: Characterisation of broad-spectrum CP probe W-sCy5-VS.** (A) Recombinant proteins were labelled with 1 μM W-sCy5-VS for 30 min in Acetate buffer at pH 5.5. (B) Schizont lysates were labelled with W-sCy5-VS, FY01-Cy5 or DCG04-TMR (1 μM) for 30 min at pH 7.2. One sample was pre-treated with E64 (10 μM) for 30 min followed by W-sCy5-VS labelling. (C) Lysates from rings, trophozoites, early and mature PKGI blocked schizonts, and merozoites were labelled with W-sCy5-VS (0.5 or 1 μM). (A-C) All labelling reactions were quenched by addition of LB and immediate boiling. 12.5% SDS-PAGE gels were run and a fluorescence scanner was used to visualise TMR (red channel Ex/Em 552/572nm) and sCy5 (green channel Ex/Em 647/663nm) labelling. Arrows indicate the likely bands corresponding to known PLCPs.
DCG04-TMR labelled FP2/3 and DPAP1, whereas FY01-Cy5 only labelled DPAP1. W-sCy5-VS labelled FP1, FP2/3 and DPAP1 in addition to approximately twenty-five other E64 sensitive bands (Fig. 3.3B). These results show that, at neutral pH, W-sCy5-VS labels many E64-sensitive proteins that may be CPs, confirming it is a better tool than previously-published ABPs to profile the activity of PLCPs in *P. falciparum*.

### 3.3.3 W-sCy5-VS labels different *Pf*CPs throughout the parasite life cycle.

PLCPs are expressed/active at different stages of parasite development whenever functionally required (Section 1.4.1.3). This is seen throughout the *Plasmodium* proteome. To observe how the activity profile of PLCPs changes over time, parasite lysates collected at different asexual stages (rings, trophozoites, early and late schizonts, and merozoites) were labelled with 1 μM W-sCy5-VS for 1h at neutral pH.

The changes in labelling at each life stage reflect stage-dependent expression and/or activation of PLCPs (Fig. 3.3C). For example, the putative FP2/3 and DPAP1 bands are most strongly labelled at trophozoite stage, as expected from their role in haemoglobin degradation (Klemba et al., 2004; Sijwali et al., 2006, p. 3; Sijwali and Rosenthal, 2004). DPAP3 labelling however is most clear in merozoites (Lehmann et al., 2018). The depth of coverage of this broad-spectrum probe is therefore increased by resolving the separate activity patterns at different asexual stages. The fact that all the PLCPs, which are active at acidic pH (Klemba et al., 2004; Lehmann et al., 2018; Shenai et al., 2000; Sijwali et al., 2001), can be labelled by W-sCy5-VS will be highly important for using this probe in ABPP experiments.

### 3.4 Proteomics confirms W-sCy5-VS labelling of PLCPs in addition to multiple other reactive-Cys containing proteins.

### 3.4.1 W-dtBio-TMR-VS and W-sCy5-VS probes have the same targets in *P. falciparum*.

The unknown targets of W-sCy5-VS were hypothesised to be CPs due to their sensitivity to E64. However other reactive cysteines, for example in redox-related enzymes or non-protease active sites, could also attack the vinyl sulphone warhead. To identify and quantify more of the targets we used a bifunctional version of the probe, W-dtBio-TMR-VS, to carry out chemical proteomics. W-dtBio-TMR-VS is also derived from L-WSAK-Alk. A desthiobiotin moiety, attached via a linker to a TAMRA fluorophore, was added to the P1 position via CuAAC-chemistry (Fig. 1.5.1 ; Tan and Davison et al.,
W-dtBio-TMR-VS was first tested to verify that it had the same specificity as W-sCy5-VS. W-dtBio-TMR-VS was then used to pull-down target proteins for proteomics analysis by LCMS/MS.

Schizont lysates were labelled with W-sCy5-VS (0.5 μM), W-dtBio-TMR-VS (0.5 μM) or a mixture of both probes (both 0.5 μM) at neutral pH. The resulting TMR and Cy5 gel images were combined to give a composite image (Fig. 3.4.1A). The single probe labelling was quantified by densitometry (Fig. 3.4.1B). Both the images and the graph clearly show the similarity in labelling profiles, although there are some differences in band intensities. When the probes were competed against each other, both showed a decrease in fluorescence labelling for the majority of bands. Therefore, these probes have similar overall reactivity with some differences, and W-dtBio-TMR-VS can be used in proteomics experiments to infer the targets of W-sCy5-VS. The coomassie-stained gel indicated equal protein loading in each lane (Fig. 3.4.1A).

Figure 3.4.1: Confirmation of W-dtBio-TMR-VS reactivity. (A) Schizont lysates were labelled with W-sCy5-VS (1 μM), W-dtBio-TMR-VS (1 μM), or a mixture of both, at pH 7.2. Photoshop was used to create a composite image of the gel in two colours. Coomassie blue protein stain confirms even protein loading on the SDS PAGE gel. Labelling reactions were quenched by addition of LB and immediate boiling. 12.5% SDS-PAGE gels were run and a fluorescence scanner was used to visualise TMR (red channel Ex/Em 552/572nm) and sCy5 (green channel Ex/Em 647/663nm) labelling. (B) Image J was used to extract densitometry data from the single probe labelling lanes in (A) in order to plot a graph of the labelling patterns. PLCPs are labelled on the graph.

3.4.2 W-dtBio-TMR-VS labels PLCPs at acidic pH and additional reactive-cysteine-containing proteins at neutral pH.

To identify the targets of W-dtBio-TMR-VS, we carried out chemical proteomics. We compared a W-dtBio-TMR-VS-treated sample to two different controls: no probe (DMSO); and pre-treating with W-sCy5-VS, which should block any subsequent labelling.
Schizont lysates were labelled in triplicate, at both neutral and acidic pH, with DMSO, W-dtBio-TMR-VS (1 μM) alone, or W-dtBio-TMR-VS (1 μM) after reaction with W-sCy5-VS (10 μM). The labelled proteins were affinity-purified using neutravidin-agarose beads. After washing, trypsin was added directly to the beads to digest the proteins and release tryptic peptides. Each condition was labelled separately with different TMT-6plex tags and then pooled together (Fig.3.4.2). This resulted in three TMR-6plex mixture samples as the labelling had been done in triplicate. The resulting mixture was separated by high pH fractionation and run on tandem LCMS/MS for detection and quantification. The isobaric TMT tags have the same molecular weight in MS1 but fragment differently in MS2. These fragments are quantified to give a relative measurement of the identified peptides between samples (Section 1.6.3.2.3). During analysis, a student t-test was done to identify proteins that were significantly enriched (p = 0.05) in the W-dtBio-TMR-VS samples compared to the two controls.

The results of the proteomics reflected the gel labelling results 3.3. Under acidic conditions fewer proteins were significantly enriched than at neutral pH, but the probe was more specific for PLCPs. At low pH, nine proteins were significantly enriched with a fold change of at least 3, including DPAP1, DPAP3, FP3 and FP2a/b (Table 3.4). Note that “FP2” enrichment actually represents two proteins FP2a and FP2b, which differ only in a single amino acid. Surprisingly, only DPAP3 was also significantly enriched in the W-dtBio-TMR-VS-treated samples compared to the W-sCy5-VS pre-treated control (data not shown). DPAP3 is the first protease to be labelled at low W-sCy5-VS concentration in acidic pH (Tan and Davison et al., 2020). It follows that this is the most efficiently blocked by pre-treatment. However, the lack of enrichment of other proteases suggests these conditions require further optimisation. None of the SERAs were detected, but this is expected since they are secreted into the PV and would be lost after saponin lysis (Thomas et al., 2018).

At neutral pH, 38 proteins were significantly enriched with a fold change of at least 3 compared to the DMSO control, but only two Cys proteases were present, DPAP1 and FP3 (Table 3.4) with DPAP3 notably absent. This could be because DPAP1 and FP3 are the most abundant PLCPs in schizont stage. In addition to PLCPs, enriched proteins include redox enzymes such as thioredoxin and disulphide isomerases. These are known to contain highly nucleophilic cysteine residues which could react with the probe’s vinyl sulphone warhead. The remainder of the significantly-enriched proteins may either be labelled non-specifically or contain other reactive cysteines for example cysteines involved in metal ion binding or disulphide bridges. We have shown that the main targets of W-dtBio-TMR-VS (and hence of W-sCy5-VS) are PLCPs in acidic conditions, and that at neutral pH they can label other diverse proteins, presumably containing highly reactive cysteines. This proteomics method requires further optimisation. Increasing the concentration of W-dtBio-TMR-VS would hopefully improve results.
Figure 3.4.2: Chemical proteomics protocol for target identification of W-sCy5-VS/W-dtBio-TMR-VS. (A) Schematic to illustrate the main steps in the chemical proteomics protocol. Soluble protein lysate in PBSa (pH 7.2) or sodium acetate buffer (pH 5.5) were treated with W-dtBio-TMR-VS probe with or without pre-treatment with W-sCy5-VS. A DMSO-only treated control was also included. Labelled proteins were purified by incubation with neutravidin agarose beads. Trypsin was applied to the beads to digest captured proteins and the resultant peptides processed through clean-up steps. (B) Peptides from different samples were labelled with isobaric TMT-sixplex labels and combined. The mixture was fractioned by high pH fractionation and analysed by LC-MS/MS.
*Fold enrichment (log2) = 1.58 is equivalent to Fold change = 3

Table 3.4.1 Proteins enriched in W-dtBio-TMR-VS compared to DMSO, pH 5.5

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Protein name</th>
<th>Mol. Weight [kDa]</th>
<th>Fold-enrichment (log2)</th>
<th>-Log (t-test p value)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF3D7_1116700</td>
<td>DPAP1</td>
<td>80.4</td>
<td>3.15</td>
<td>5.01</td>
<td>Cysteine protease</td>
</tr>
<tr>
<td>PF3D7_0827900</td>
<td>PD18</td>
<td>55.5</td>
<td>2.67</td>
<td>3.79</td>
<td>Protein disulphide isomerase</td>
</tr>
<tr>
<td>PF3D7_1115400</td>
<td>FP3</td>
<td>56.7</td>
<td>2.16</td>
<td>2.19</td>
<td>Cysteine protease</td>
</tr>
<tr>
<td>PF3D7_1121300/700</td>
<td>FP2a/b</td>
<td>55.8</td>
<td>1.96</td>
<td>4.00</td>
<td>Cysteine protease</td>
</tr>
<tr>
<td>PF3D7_1205400</td>
<td>Tetrapeptide repeat protein</td>
<td>39.0</td>
<td>1.93</td>
<td>1.96</td>
<td>Putative protein, unknown function</td>
</tr>
<tr>
<td>PF3D7_1302100</td>
<td>G27/25</td>
<td>26.0</td>
<td>1.73</td>
<td>2.68</td>
<td>Gamete antigen</td>
</tr>
<tr>
<td>PF3D7_1352500</td>
<td>Thioredoxin-related protein</td>
<td>24.0</td>
<td>1.70</td>
<td>4.38</td>
<td>Putative thioredoxin</td>
</tr>
<tr>
<td>PF3D7_1108600</td>
<td>ERC</td>
<td>39.4</td>
<td>1.61</td>
<td>3.85</td>
<td>ER-resident calcium binding protein</td>
</tr>
<tr>
<td>PF3D7_0409700</td>
<td>DPAP3</td>
<td>110.5</td>
<td>1.59</td>
<td>5.76</td>
<td>Cysteine protease</td>
</tr>
</tbody>
</table>

Table 3.4.2 Proteins enriched in W-dtBio-TMR-VS compared to DMSO, pH 7.2

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Protein name</th>
<th>Mol. Weight [kDa]</th>
<th>Fold-enrichment (log2)</th>
<th>-Log (t-test p value)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF3D7_1205600</td>
<td>Tetrapeptide repeat protein</td>
<td>39.0</td>
<td>4.12</td>
<td>2.86</td>
<td>Putative protein, unknown function</td>
</tr>
<tr>
<td>PF3D7_0212900</td>
<td>ATEL1</td>
<td>42.9</td>
<td>3.93</td>
<td>1.72</td>
<td>Arginyl-IRNA-protein transferase</td>
</tr>
<tr>
<td>PF3D7_0718500</td>
<td>Prefoldin subunit 3</td>
<td>22.6</td>
<td>3.49</td>
<td>2.38</td>
<td>Putative protein</td>
</tr>
<tr>
<td>PF3D7_1252300</td>
<td>HAD2</td>
<td>33.3</td>
<td>3.12</td>
<td>1.65</td>
<td>Putative halocid dehalogenase-like hydrolase</td>
</tr>
<tr>
<td>PF3D7_0609700</td>
<td>CCT6</td>
<td>61.5</td>
<td>3.00</td>
<td>2.61</td>
<td>T-complex protein 1 subunit zeta</td>
</tr>
<tr>
<td>PF3D7_0179700</td>
<td>RPS0</td>
<td>16.5</td>
<td>2.96</td>
<td>0.91</td>
<td>Putative 40S ribosomal protein S10</td>
</tr>
<tr>
<td>PF3D7_1405600</td>
<td>RNR</td>
<td>40.6</td>
<td>2.52</td>
<td>2.00</td>
<td>Putative ribonucleoside-diphosphatase reductase small chain</td>
</tr>
<tr>
<td>PF3D7_0516200</td>
<td>RPS1</td>
<td>16.1</td>
<td>2.45</td>
<td>0.90</td>
<td>40S ribosomal protein S11</td>
</tr>
<tr>
<td>PF3D7_1438900</td>
<td>Trx-Px1</td>
<td>21.8</td>
<td>2.31</td>
<td>5.85</td>
<td>Thioredoxin peroxidase 1</td>
</tr>
<tr>
<td>PF3D7_1147800</td>
<td>MCM2</td>
<td>111.5</td>
<td>2.20</td>
<td>0.86</td>
<td>DNA replication licensing factor</td>
</tr>
<tr>
<td>PF3D7_1429800</td>
<td>Coatsomer subunit beta</td>
<td>160.0</td>
<td>2.20</td>
<td>1.33</td>
<td>Putative protein</td>
</tr>
<tr>
<td>PF3D7_0810800</td>
<td>PPK-DHPS</td>
<td>83.4</td>
<td>2.12</td>
<td>0.89</td>
<td>Hydroxymethylhydroperoxidase pyrophosphokinase-DHPS</td>
</tr>
<tr>
<td>PF3D7_0715300</td>
<td>Calcium-dependent kinase</td>
<td>61.3</td>
<td>2.11</td>
<td>2.15</td>
<td>Putative protein</td>
</tr>
<tr>
<td>PF3D7_1127100</td>
<td>dUTPase</td>
<td>19.6</td>
<td>2.11</td>
<td>3.16</td>
<td>Deoxyuridine 5’-triphosphate nucleotidohydrolase</td>
</tr>
<tr>
<td>PF3D7_1357100/000</td>
<td>Elongation factor 1-alpha</td>
<td>49.0</td>
<td>2.07</td>
<td>3.02</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1027300</td>
<td>nPxi</td>
<td>43.9</td>
<td>2.07</td>
<td>1.64</td>
<td>Peroxiredoxin</td>
</tr>
<tr>
<td>PF3D7_1414800</td>
<td>SMB1</td>
<td>22.3</td>
<td>2.04</td>
<td>0.68</td>
<td>Putative small nuclear ribonucleoprotein-associated</td>
</tr>
<tr>
<td>PF3D7_1423300</td>
<td>PP7</td>
<td>112.6</td>
<td>2.02</td>
<td>0.75</td>
<td>Serine/threonine protein phosphatase 7</td>
</tr>
<tr>
<td>PF3D7_1338300</td>
<td>EF-1gamma</td>
<td>47.8</td>
<td>1.98</td>
<td>3.18</td>
<td>Putative elongation factor 1-gamma</td>
</tr>
<tr>
<td>PF3D7_0930300</td>
<td>MSP1</td>
<td>195.7</td>
<td>1.95</td>
<td>2.92</td>
<td>Merozoite surface protein 1</td>
</tr>
<tr>
<td>PF3D7_1115400</td>
<td>FP3</td>
<td>56.7</td>
<td>1.94</td>
<td>2.32</td>
<td>Cysteine protease</td>
</tr>
<tr>
<td>PF3D7_1246200</td>
<td>ACT1</td>
<td>41.9</td>
<td>1.93</td>
<td>4.27</td>
<td>Actin I</td>
</tr>
<tr>
<td>PF3D7_1402300</td>
<td>RPN6</td>
<td>78.4</td>
<td>1.92</td>
<td>0.60</td>
<td>265 proteasome regulatory subunit</td>
</tr>
<tr>
<td>PF3D7_1011400</td>
<td>Proteasome beta type-S</td>
<td>30.6</td>
<td>1.82</td>
<td>2.99</td>
<td>Threonine-type endopeptidase activity</td>
</tr>
<tr>
<td>PF3D7_0827900</td>
<td>PDB1</td>
<td>55.5</td>
<td>1.82</td>
<td>2.70</td>
<td>Protein disulphide isomerase</td>
</tr>
<tr>
<td>PF3D7_1229400</td>
<td>MIF</td>
<td>12.8</td>
<td>1.80</td>
<td>0.69</td>
<td>Macrophage migration inhibitory factor</td>
</tr>
<tr>
<td>PF3D7_0109800</td>
<td>cPhfR5</td>
<td>67.7</td>
<td>1.79</td>
<td>4.28</td>
<td>Phenylation-IRNA ligase alpha subunit</td>
</tr>
<tr>
<td>PF3D7_0627800</td>
<td>ACS</td>
<td>113.9</td>
<td>1.79</td>
<td>1.01</td>
<td>Putative actetyl-CoA synthetase</td>
</tr>
<tr>
<td>PF3D7_0309300</td>
<td>N2227-like protein</td>
<td>112.3</td>
<td>1.76</td>
<td>1.30</td>
<td>Putative SAM methyltransferase, cysteine rich repeats</td>
</tr>
<tr>
<td>PF3D7_1116700</td>
<td>DPAP1</td>
<td>80.4</td>
<td>1.74</td>
<td>3.51</td>
<td>Dipeptidyl aminopeptidase 1</td>
</tr>
<tr>
<td>PF3D7_0528200</td>
<td>EEF3E</td>
<td>61.4</td>
<td>1.73</td>
<td>1.88</td>
<td>Putative eukaryotic translation initiation factor 3 subunit E</td>
</tr>
<tr>
<td>PF3D7_1213900</td>
<td>Conserved Plasmodium protein</td>
<td>134.6</td>
<td>1.66</td>
<td>2.66</td>
<td>Unknown function</td>
</tr>
<tr>
<td>PF3D7_1420400</td>
<td>GlyR5</td>
<td>104.3</td>
<td>1.66</td>
<td>2.98</td>
<td>Glycin-IRNA ligase</td>
</tr>
<tr>
<td>PF3D7_1447000</td>
<td>40S ribosomal protein S5</td>
<td>30.0</td>
<td>1.64</td>
<td>1.79</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1334200</td>
<td>Chaperone binding protein</td>
<td>62.7</td>
<td>1.61</td>
<td>0.90</td>
<td>Putative protein</td>
</tr>
<tr>
<td>PF3D7_0802200</td>
<td>T-CysPxi</td>
<td>25.2</td>
<td>1.60</td>
<td>1.03</td>
<td>1-cys peroxiredoxin</td>
</tr>
<tr>
<td>PF3D7_1352500</td>
<td>Thioredoxin-related protein</td>
<td>24.0</td>
<td>1.59</td>
<td>2.51</td>
<td>Putative protein</td>
</tr>
<tr>
<td>PF3D7_1311900</td>
<td>vapaA</td>
<td>68.6</td>
<td>1.59</td>
<td>1.55</td>
<td>V-type proton ATPase catalytic subunit A</td>
</tr>
</tbody>
</table>

*Fold enrichment (log2) = 1.58 is equivalent to Fold change = 3

Redox related enzymes

Papain-like cysteine protease

V-type proton ATPase catalytic subunit A
3.5 FPP-TMR and W-sCy5-VS can be used simultaneously to screen for reversible inhibitors.

We have shown that FPP-TMR and W-sCy5-VS can be used to profile active SHs and PLCPs at neutral pH in *P. falciparum* (Chapter 3). The probes have orthogonal reactivity and different fluorophores so they can be used simultaneously. We therefore designed an ABPP screening strategy to use both probes in parallel so that we can concurrently monitor the activity of these two families in competition with the Malaria Box antimalarial compounds. Our aim was to identify specific inhibitors of CP or SH members. Generally, pharmaceutical drug discovery favours inhibitors with a reversible rather than irreversible mechanism as they are usually less toxic. The process of selection for the Malaria Box included steps to eliminate highly-reactive chemical moieties (Section 1.3.4). We wanted to find screening conditions where it would be possible to visualise competition between reversible inhibitors and the irreversible ABPs. As FPP-TMR and W-sCy5-VS both labelled different profiles of active enzymes at each life stage, we knew that we should do the screening at multiple stages to maximise depth of coverage.

In order to profile the probes targets efficiently in competition with many inhibitors at multiple life stages a medium-throughput screening procedure was designed and optimised. A simple analysis workflow was possible as the fluorescent gel-based results could be visually assessed and confirmed via densitometry.

3.5.1 W-sCy5-VS and FPP-TMR do not cross-react.

In order to confirm that the two probes did not cross-react or interfere with each other’s labelling, schizont lysates were treated with 1 μM WsCy5-VS, 1 μM FPP-TMR, or both probes at once for 30 min. The single-channel and composite SDS-PAGE gel images show that both probes label the same targets alone or when mixed together, with no increase in fluorescent background when both probes are combined (Fig. 3.5.1A). Thus, they could be used simultaneously.
Figure 3.5.1: Optimisation of FPP-TMR and W-sCy5-VS labelling conditions for ABPP screening. A) Schizont lysates were treated with 1 μM WsCy5-VS, or 1 μM FP-TMR or both at once. The single channels were combined in Adobe Photoshop to give the composite gel image. B) Schizont lysates were pretreated for 30 minutes with either 10 μM E64, 2mM PMSF or both before being labelled with 1 μM W-sCy5-VS and 0.1 μM FP-TMR. C) Schizont lysates were treated with 500nM W-sCy5-VS and 50 nM FP-TMR. Labelling reactions were quenched after different amounts of time. Bands were chosen at random to quantify and are indicated by their molecular weight and coloured arrows corresponding to the graphs (left). The intensity of band labelling over time was quantified using Image Lab™, the background was subtracted and intensities were normalised to the most intense signal quantified. Intensities were fit to an association kinetics model in Graphpad Prism®. The kinetics of labelling of representative bands (indicated by coloured arrows) is shown on the right panel. (A-C) Labelling reactions were quenched by addition of LB and immediate boiling. 12.5% SDS-PAGE gels were run and a fluorescence scanner was used to visualise TMR (red channel Ex/Em 552/572nm) and sCy5 (green channel Ex/Em 647/663nm) labelling.
3.5.2 WsCy5-VS and FPP-TMR labelling can be blocked by broad-spectrum inhibitors E64 and PSMF, respectively.

We have shown that the broad-spectrum covalent CP inhibitor E64 blocks the majority of W-sCy5-VS labelling (Section 3.3). E64 can therefore be used as a positive control. We used the general SH covalent inhibitor PMSF as the positive control for inhibition of SHs.

Schizont lysates were pre-treated for 30 min with either 10 μM E64, 2mM PMSF, or both, followed by labelling with 1 μM W-sCy5-VS and 0.1 μM FPP-TMR for 30 min. The results showed that the majority of FPP-TMR labelling was blocked by PMSF and that it could be used efficiently in combination with E64 (Fig. 3.5.1B). This allowed us to use a single positive control condition for the inhibitor screen.

3.5.3 Dynamic labelling conditions are required to screen for reversible inhibitors.

As the probe modifies enzymes covalently, we needed to find sub-saturating conditions so that competition with reversible inhibitors would be apparent. To obtain a time course of labelling, multiple lysate samples were treated with the probes FPP-TMR (50 nM) and W-sCy5-VS (500 nM) for different durations from 5 to 120 min. Image Lab™ gel analysis software was used to quantify the intensity of eleven FPP-TMR and twenty W-sCy5-VS labelled bands. The labelling over time was plotted and association kinetics were used to model the binding of the probes (Equation 1).

\[ Y = Y_{SAT} \cdot (1 - e^{-kx}) \quad \text{Equation 1} \]

The majority of bands showed time-dependent saturation above 15 min (Fig. 3.5.1C). Therefore, we chose 10 min labelling time for the screen, to ensure conditions in which all bands are within the dynamic linear phase of labelling kinetics.

3.5.4 Optimisation of screening procedure for a medium-throughput ABPP screen.

We have identified labelling conditions such that both probes can be used simultaneously to screen for reversible inhibitors of SHs and Cys proteases (Section 3.5.3). Now we wanted to streamline the experimental procedure so that the screen could be carried out efficiently by one scientist in a medium-throughput fashion (Fig. 3.5.2). Assay conditions were scaled to a 96-well round-bottom plate format to test eighty compounds at once (the Malaria Box format has eighty compounds per plate). The use of
multichannel pipettes to set up the plate and load triple-wide SDS-PAGE gels significantly decreased assay time.

Parasite lysates were pre-treated with antimalarial compounds (50 μM) for 30 min before the addition of a mix of FPP-TMR (50nM) and W-sCy5-VS (500nM). The labelling reaction was quenched after ten minutes by addition of loading buffer (LB) and boiling. Electronic multichannel pipettes allowed rapid addition of probes and LB across a plate, achieving the precise 10 min incubation time tested above. This, combined with immediate boiling of the entire plate using a sand-bath, ensured consistent labelling across a plate (Fig. 3.5.2). Evaporation during boiling was minimal under these conditions (data not shown). In this way, 400 compounds could be screened in a single day, ensuring consistent treatment and labelling conditions with the same parasite lysate. The samples could be frozen in-plate and later defrosted to run on SDS-PAGE gels to give an output of twenty triple-wide gels that could be easily assessed both visually (after fluorescent scanning) and using ImageLab™ (Fig. 3.5.2). The screening conditions are described in detail in the Materials and methods (Chapter 9).

3.5.5 Visual assessment of gels followed by densitometry provides quick and easy analysis.

The advantage of the gel-based read-out was that results could be generated quickly and cheaply, and hits retested with ease. Proteomics could be used to quantify changes in probe labelling in the presence
of Malaria box compounds. This method would likely return a greater number of hits and would have
the advantage of identifying protein targets. However, for 400 compounds, sample preparation would
be extremely time consuming and mass spectrometry analysis costly. The aim of the project was to
identify a small set of hypothetically essential drug targets feasible for genetic validation. This means
that maximal depth of coverage is not critical. The choice of gel-based results also means that we are
selecting for protein labelling and inhibition that is easily visualised on a fluorescent SDS-PAGE gel,
and therefore targets whose abundance will make follow-up studies easier to perform.

3.6 Discussion

We have characterised two broad-spectrum ABPs targeting SHs and PLCPs in *P. falciparum* in
preparation for simultaneous ABPP of these enzyme families. The previously-characterised FPP probes
(Section 1.5.3) were shown to label at least twenty-two SHs in *P. falciparum*. Different profiles of SH
activity were observed at different life stages. It is important to note that, while the FPP-labelling gels
specifically report enzyme activity, they will also reflect changes in expression or localisation. If an
active SH is exported into the PV or RBC lumen at a specific time in the life cycle it will no longer be
detected by FPP-TMR due to the saponin lysis step. We have shown that FPP-TMR and FPP-N₃ probes
have common SH targets and that FPP-N₃ targets more SHs in intact parasites. Chemical proteomics
using FPP-N₃ identified nine *P. falciparum* SHs in schizonts, seven of which were completely
uncharacterised. The ubiquitous presence of the α/β-fold is striking, suggesting that the FPP probes
may preferentially target enzymes of this family. α/β-fold enzymes have a conserved fold and catalytic
site architecture, but can have very diverse substrates and cellular functions (Section 1.4.2). There are
43 predicted α/β-fold hydrolases in *P. falciparum* (Table 1). Later in this thesis we describe the
optimisation of the chemical-proteomic method employed in order to increase the depth of coverage of
this large family (Chapter 5). We investigated another biotin-containing FPP probe, optimised the
method for proteomic sample preparation, and employed more precise quantification methods (Chapter
5). Most of the *Pf*SHs have low homology to human SHs. Therefore, we hypothesise that the
characterisation of this family will help to discover new antimalarial targets, as well as new aspects of
*P. falciparum* biology. Many of the *Pf*SHs are predicted to be lysophospholipases and may play roles
in the elusive processes of parasite lipid synthesis, scavenging and metabolism (Table 1). At least three
of the *Pf*SHs detected above are part of an expanded subtelomeric family *Pst-a* (Fischer et al., 2003).
This is interesting because other multicopy sub-telomeric gene families have vital roles in virulence,
such as the *var* genes (Fischer et al., 1997). However, multiplicity of members may also indicate gene
expansion and redundancy. New genetic screens in *P. berghei* (Schwach et al., 2015) and *P. falciparum*
(Zhang et al., 2018) help to prioritise SHs with essential functions in the asexual-blood stages. If SHs
are identified as novel drug targets, the FPP probes are likely to be highly useful in the screening of
specific inhibitors and confirmation of on-target activity. The FPP probes are therefore excellent tools for unbiased ABPP. They allow us to profile enzymes based on common catalytic features, while covering diverse enzyme functions.

We also characterised the novel ABP, W-sCy5-VS. W-sCy5-VS was confirmed to label the well-characterised malarial PLCPs with broader reactivity than previously published probes DCG-04 or FY01. The discovery of W-sCy5-VS was serendipitous, arising from an effort to identify more specific DPAP3 inhibitors. But the utility of being able to profile the activities of multiple PLCPs concurrently is undeniable, particularly in the context of ABPP. It was interesting that W-sCy5-VS labelled multiple other enzymes at neutral pH, expanding the ABPP functionality of this probes beyond the well-characterised PLCPs. The neutral conditions were also essential in the design of the forthcoming ABPP screen, as they allowed the probe to be used in parallel with FPP-TMR. Labelling at different life stages allowed us to observe the expected patterns of activation of the *Plasmodium* PLCPs (FP2, FP3, DPAP3 and DPAP1), with the exception of the SERAs. Detection of the cysteine-type SERA activity is precluded by the loss of the PV lumen contents upon saponin lysis, but would be unlikely anyway due to their short activation window before egress (Thomas et al., 2018). The increased activity of this probe compared to its alkyne precursor and analogues with different fluorophores (Tan and Davison et al., 2020) is curious, and suggests that the electron dense sCy5 fluorophore contributes to PLCP targeting. Fortunately, the bifunctional probe, W-dtBiotin-TMR-VS, showed a similar reactivity and therefore could be used to infer W-sCy5-VS targets by chemical proteomics. Much of the W-sCy5-VS labelling at neutral pH was sensitive to the CP inhibitor E64. However, W-dtBiotin-TMR-VS proteomics did not identify any non-PLCPs, only reactive-cysteine-containing redox-related proteins. In general, the chemical proteomics results showed much less specificity for PLCPs and other CPs than expected. Further optimisation of the proteomic method is required to provide a full characterisation of these probes. We did not repeat the W-dtbiotin-TMR-VS proteomics in the course of this thesis as after the Malaria Box ABPP screen, SH hits were prioritised.

We used the orthogonal reactivity of the two probes to design an ABPP screen at neutral pH in which approximately 50 SHs and reactive-cysteine-containing enzymes could be profiled in parallel. This medium throughput screen was suitable to test hundreds of reversible inhibitors per day. The Malaria Box is an open-access resource containing 400 compounds with antimalarial activity in phenotypic screens (Section 1.3.4). We aimed to use the ABPP screen as an efficient and unbiased method to discover novel molecular targets of the Malaria Box compounds.
Chapter 4. **Results II: ABPP identified five compounds affecting CP or SH activity in *P. falciparum*.**

The Malaria Box was constructed by MMV using open-access data from phenotypic screens of thousands of small molecules against asexual stage *P. falciparum* (Section 1.3.4). Of all the compounds that blocked parasite replication (EC$_{50}$ greater than 5 μM), 400 structures were chosen for the Malaria Box based on their potency, commercial availability and maximum chemical diversity (Spangenberg et al., 2013). Half of the chosen compounds had passed toxicophore and bioavailability screening and were termed “drug-like”, with the rest labelled “probe-like”.

Many groups have used the Malaria Box to screen for inhibitors against specific enzymes/pathways, non-blood stage of *P. falciparum*, and other Apicomplexa (Section 1.3.4). The results of these studies and many others are deposited into an online resource on CHEMBL, however the mode of action of the majority of hits are still unknown. Discovering the compounds’ mechanisms of action (MoA) will not only further drug discovery aims but lead to better understanding of basic parasite biology and potentially identify new druggable targets.

An ABPP screen was developed (Section 3.5) to compete Malaria Box compounds with FPP-TMR and W-sCy5-VS to profile inhibition of metabolic SHs and PLCPs, respectively. As the ABPs only label active enzymes, any inhibited enzymes can be identified as a decrease in labelling intensity of specific fluorescent bands visualised in SDS-PAGE gels. Because the ABPs have different fluorophores, they can be used in combination to profile both families simultaneously. This approach allows screening of the Malaria Box against more than 40 targets simultaneously. The screen was optimised to a medium-throughput format to efficiently test all 400 compounds against multiple life stages (Section 3.5).

### 4.1 ABPP at schizont and trophozoite stages identified five Malaria Box inhibitors.

#### 4.1.1 The Malaria Box ABPP screen resulted in sixteen hits, eight of which were reproducible under the same conditions.

Using a gel-based ABPP screen, we were able generate results quickly and easily. The 400 Malaria Box compounds were competed with the probes in either schizont or trophozoite lysates, alongside a control plate containing DMSO or positive control inhibitors E64 and PMSF (10 μM / 2 mM). The samples were run on 20 triple wide gels in a single day. The labelling was then visualised using a fluorescence
scanner. The screening procedure is described in more detail in the Materials and Methods (Section 9.3).

Each gel lane was visually compared to the DMSO control to see if any bands had disappeared or decreased in labelling intensity in the presence of compound (Fig. 4.1.2B). Subsequently, the densitometry profiles of each lane (generated by ImageJ™) were compared in order to confirm the visual inspection or identify more subtle changes in labelling (Fig. 4.1.2C). Using this assay, the trophozoite and schizont lysate screens showed consistent labelling in the majority of samples. Visual inspection identified 28 potential hits, 16 of which were confirmed by densitometry analysis (Fig. 4.1.1A, examples in Fig. 4.1.2B-C). These 16 hits were first re-tested with either schizont or trophozoite lysate under the same conditions using the same compound stocks used in the screens. Only eight compounds displayed reproducible inhibition results when re-tested (Fig. 4.1.1A, data shown for final five only). The highly reproducible labelling and relatively low number of hits (8/400 = 2%) was a good indication that the results were real. However, further retesting with freshly purchased compound was required before performing follow up studies.

![Figure 4.1.1: The Malaria Box ABPP screen results. (A) The attrition of screening hits. Multiple rounds of hit confirmation were performed: visual inspection of gels; densitometry analysis of lane profiles; retesting of hits from the Malaria Box; and retesting with purchased compounds. At each step, the total number of hits is given, followed by the number of W-sCy5-Vs hits in blue and FP-TMR hits in pink. (B) Venn diagram to show the compounds that were purchased for retesting and the life stage at which they were detected. The final hits confirmed by retesting are indicated in bold.](image_url)
4.1.2 Confirmation of potential hits identified five compounds for follow-up studies.

4.1.2.1 Five of out eight Malaria Box ABPP screen hits were confirmed with fresh compound stocks.

The Malaria Box consisted of 400 frozen DMSO stocks in 96-well plates, which were defrosted and diluted for the screen. Some compounds may have degraded or changed in concentration over time, so we wished to re-test the eight hit compounds using fresh stocks. These compounds were MMV665914, MMV665798, MMV655941, MMV019313, MMV000563, MMV019738, MMV000986, and MMV011438 (Fig. 4.1.1B). The compounds were purchased commercially and tested in competition with the relevant probe. We used ring, trophozoite and schizont lysates to see if the different targets of these compounds were present and active in different life stages.

Five out of the remaining eight hits were confirmed with freshly purchased stocks (Fig. 4.1.1A-B, Supplementary Fig. 10.2.1). Surprisingly, two compounds induced increased labelling of specific bands by W-sCy5-VS in schizont lysates. Retesting confirmed that MMV665941 (M941) and MMV019738 (M738) both increased labelling of a 28 kDa band in schizonts only, and M738 of two bands at 65 and 26 kDa in schizonts, trophozoites and rings (Fig. 4.1.2D). We hypothesised that the 28 kDa could be FP3 as indicated by its size and presence in trophozoite and schizont lysates.

Three compounds inhibited FPP-TMR SH labelling of specific bands (MMV00563, MMV665914 and MMV011438). MMV000563 (M563) (Fig 4.1.2A), which was identified from the trophozoite screen, was confirmed to block labelling of two bands in trophozoite lysates at 80 and 56 kDa (Fig. 4.1.2D). M563 also inhibited multiple bands in schizont and ring lysates, which suggested that the fresh M563 stock was more potent than the Malaria Box stock (Fig. 4.1.2D). We observed that stocks of M563 would lose their orange colour and potency in assays if not kept at -20°C and anhydrous (data not shown). MMV011438 (M438) and MMV665914 (M914) (Fig. 4.1.2A) were found in the schizont screen, and confirmed to inhibit the same 49 kDa band in schizont lysates (Fig. 4.1.2D). Neither M438 or M914 inhibited any bands in other life stages (Fig. 4.1.2D). Note that compound M914 was also previously identified in a pilot screen in schizont lysates by then masters student Anja Schlott (Anja Schlott and Edgar Deu, unpublished). Five compounds therefore showed reliable effects on labelling, as confirmed by their fresh stocks (Fig. 4.1.1B). M563, M914 and M738 were all in the “drug-like” set of compounds, with M438 and M914 in the “probe-like” set. Although further compounds may have been identified if the screen had been repeated in triplicate and on merozoites derived lysates, this was an ideal number of compounds to perform follow up studies.
We now aimed to assess the compounds’ specificity and potency by measuring the IC\textsubscript{50}s of labelling inhibition or activation. As ABPs only react with active enzymes, a decrease in labelling intensity of a protein on a gel can be taken as a measure of inhibition of that enzyme. IC\textsubscript{50}s were also calculated for

4.1.2.2 All five final hit compounds affect probe labelling in a dose-dependent manner.

We now aimed to assess the compounds’ specificity and potency by measuring the IC\textsubscript{50}s of labelling inhibition or activation. As ABPs only react with active enzymes, a decrease in labelling intensity of a protein on a gel can be taken as a measure of inhibition of that enzyme. IC\textsubscript{50}s were also calculated for
the W-sCy5-VS activators using the increase instead of decrease in labelling to give a measure of their potency. However the effect of these compounds on normal enzyme activity was yet unknown.

Figure 4.1.3: Dose dependent inhibition/activation of probe labelling by the Malaria Box hit compounds. Trophozoite (M563) or schizont (rest) lysates were pre-treated for 30 min with different concentrations of compounds before treatment with 500 nM W-sCy5-VS for 30 min or 50 nM FP-TMR for 10 min. Saturation of labelling (SAT) was achieved using 1 μM W-sCy5-VS or 100nM FP-TMR for 30 min. Labelling reactions were quenched by addition of LB and immediate boiling. 12.5% SDS-PAGE gels were run and a fluorescence scanner was used to visualise TMR (red channel Ex/Em 552/572nm) and sCy5 (green channel Ex/Em 647/663nm) labelling. IC50s were calculated by quantifying band labelling in Image Lab™ and fitting to a standard dose response in Graphpad Prism®; the calculated standard error is given.
Trophozoite (for M563) or schizont (for M438, M914, M738 and M941) lysates were pre-treated with different concentrations of compounds before the addition of either 50 nM FPP-TMR for 10 min or 500 nM W-sCy5-VS for 30 min (as M738 and M941 induce rather than block labelling, sub-saturating conditions were not required). After running the reactions on SDS-PAGE gels and imaging using a fluorescence scanner, the inhibited/activated band intensities were quantified using ImageLab™. The data were fit to a standard dose response curve, and the IC_{50} and standard errors calculated.

The hits that activate W-sCy5-VS labelling (M4941, M738) were one to two orders of magnitude less potent than the FPP-TMR inhibitors. They also had IC_{50}s at least 10-fold higher than the reported EC_{50}. This discrepancy may suggest that enzyme activation might not be the primary cause of these compounds’ anti-parasitic activity (Table 4). The SH inhibitor IC_{50}s were in the micromolar (M563) or sub-micromolar (M438, M914) range (Fig.4.1.3). For M438 and M914 the IC_{50} was less than the reported EC_{50} for inhibition of parasite replication (ChEMBL) (Table 4), so inhibition of the 49 kDa target could be the source of the compounds’ anti-parasitic activity. Dose dependent inhibition of two bands by M563 gave IC_{50}s 30-fold higher than the reported EC_{50} (Table 4). Note that at high concentrations this compound inhibits additional bands, which suggests that it may be a broad-spectrum inhibitor that affects parasite growth via cumulative inhibition of multiple targets.

<table>
<thead>
<tr>
<th>Compound ID</th>
<th>Mol. Weight / kDa</th>
<th>IC_{50} / μM</th>
<th>EC_{50} / μM</th>
<th>ChEMBL EC_{50} / μM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMV665941</td>
<td>31</td>
<td>32 ± 5*</td>
<td>-</td>
<td>0.620</td>
</tr>
<tr>
<td>MMV019738</td>
<td>65</td>
<td>67 ± 4*</td>
<td>63 ± 5*</td>
<td>1.130</td>
</tr>
<tr>
<td>MMV000563</td>
<td>80</td>
<td>6.0 ± 0.5</td>
<td>0.40 ± 0.18</td>
<td>0.238</td>
</tr>
<tr>
<td>MMV011438</td>
<td>49</td>
<td>0.08 ± 0.013</td>
<td>0.26 ± 0.025</td>
<td>0.332</td>
</tr>
<tr>
<td>MMV665914</td>
<td>49</td>
<td>0.3 ± 0.1</td>
<td>1.20 ± 0.080</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Overall, five compounds have been shown to have clear and consistent dose-dependent effects on ABP labelling. We now wanted to investigate the SH and CP inhibitors in further detail to try and identify the targets and mode of inhibition of the compounds. The FPP-TMR hits were given higher priority as they blocked labelling and they are more potent than the W-sCy5-VS hits (Table 4). The W-sCy5-VS labelling activators were briefly investigated to determine if the 28k Da band was FP3 as predicted.

4.1.2.3 Clear inhibition of labelling was not observed in live parasites.

As previously shown, intact parasites can be labelled with FPP-TMR before lysis and a different profile of active enzymes is seen (Section 3.1). We wanted to test the inhibitors under these conditions to see
if the enzyme targets could be labelled in intact parasites and whether or not we could observe inhibition by the hits compounds. Parasites at trophozoite stage were treated with 50 μM of M563, and parasites at schizont stage were treated with 50 μM of M438 or M914 for 30 min, followed by labelling with 0.5 μM of FPP-TMR for 30 min as previously described (Section 3.1). W-sCy5-VS is not cell permeable so the equivalent experiment was not performed with M941 or M738. Unfortunately the cell permeable versions of this probe, W-Cy5-VS and L-WSAK-Alk, do not have the same reactivity as W-sCy5-VS so they cannot be used instead (Tan Davison et al., 2020).

Labelling inhibition by M563 was less clear in intact parasites compared to lysates while no inhibition was seen for M438 or M914 (Supplementary Fig. 10.2.2). Many combinations of probe/inhibitor concentrations and reaction times were tested but it was not possible to find conditions to match the dynamic labelling that was optimised in lysates (Supplementary Fig. 10.2.2).

4.2 The multiplexed ring screen did not produce any original hit compounds for follow-up studies.

We next wanted to perform the screen on ring stage lysates. Generation of ring lysates is challenging, as much less parasite protein can be collected compared to equal numbers of trophozoites or schizonts. For this reason, Malaria Box compounds were multiplexed to reduce the scale of the ring screen. Compounds from Malaria Box plates B, C, D and E were mixed in a separate 96-well plate, with each well then containing 4 inhibitors. Plate A was prioritised to be screened alone, as this contained the most potent drug- and probe-like chemotypes (Spangenberg et al., 2013).

The labelling in the ring screen was less consistent across plates than the schizont or trophozoite screens (data not shown). This led to many more compounds being considered potential hits initially, most of which we were unable to confirm when retested (Fig. 4.2). Two BCDE reaction mixes were confirmed during re-retesting. One appeared to induce labelling of bands in W-sCy5-VS and contained the compound M738. The other BCDE mix appeared to inhibit multiple FPP-TMR labelled bands and contained M563. It was assumed that M738 and M563 were the effector compounds in their reaction mixes, which confirms their previously observed activity (Fig. 4.1.2). Therefore, the ring screen results were not pursued any further.
4.3 Two compounds increase W-sCy5-VS labelling of multiple targets including FP3.

Neither M941 nor M738 had been implicated in previously published Malaria Box screens. However, M941 is commonly known as methylrosanilide, which is the basic form of the dye “crystal violet” (also known as gentian violet or methyl violet). It has a bright violet colour in solution and has anti-bacterial and anti-fungal properties (Maley and Arbiser, 2013). It has been used historically as a topical antiseptic and histological stain. Although both compounds induce labelling of the same 28 kDa band, M738 is not structurally similar to M941. However, both compounds have an extended conjugated system of pi bonding. M738 also induces labelling of two further bands at 65 and 26 kDa. The purpose of using W-sCy5-VS in the ABPP screen was to monitor inhibition of the multiple PLCPs which it labels (Section...
3.3). Although no Malaria Box compounds were observed to inhibit PLCPs, we hypothesised that the induced 28 kDa band could be FP3. Interestingly neither M941 or M738 had any effect on FY01 or DCG04 labelling in schizont lysates (data not shown). We wanted to test the FP3 hypothesis, and also discover if the increase in labelling was accompanied by inhibition or activation of the enzyme activity using recombinant FP3.

4.3.1 M941 (but not M738)-dependent W-sCy5-VS labelling is E64 sensitive.

We previously showed that W-sCy5-VS labelling of the FP3/FP2 doublet is blocked by pre-incubation with the E64 general CP inhibitor (Section 3.3). We wanted to test if this was also the case for the M941/M738-dependent labelling to help assign the identity of the 28 kDa band. Schizont lysates were first treated with 10 $\mu$M E64 for 30 min, followed by 50 $\mu$M M941 or M738 for 30 min, and finally labelling with 500 nM W-sCy5-VS for 15 min.

The unknown 65 and 26 kDa proteins affected by M738 were not blocked by E64, suggesting that either they are not E64-sensitive CPs, or that M738 does not bind in their active site and therefore is unaffected by E64 binding (Fig. 4.3.1A). E64 successfully blocked the M941/M738-dependent labelling of the hypothetical FP3 band (Fig. 4.3.1A). Therefore, the induced W-sCy5-VS labelling, in this case, is dependent on a functional active site. As these results supported the FP3 hypothesis we decided to investigate the effect of the compounds on recombinant FP3 (rFP3).

4.3.2 M941 increases W-sCy5-VS labelling of rFP3 at acidic pH.

We have shown that rFP3 is labelled by W-sCy5-VS (Section 3.3.1). We wanted to repeat this experiment after M941/M738 treatment to observe the changes in labelling intensity. rFP3 (approximately 0.1 $\mu$M) was labelled with 1 $\mu$M W-sCy5-VS for 30 min in acetate buffer (pH 5.5), in which it has maximal activity (Sijwali et al., 2001). Labelling of rFP3 was enhanced in the presence of M941, but not in the presence of M738 (Fig. 4.3.1B). This may reflect different mechanisms of activation/inhibition of the two compounds. If M738 interacts with FP3 distal to the active site, these interactions may be different in the recombinant, compared to the native, enzyme. There is evidence to suggest that FPs can be inhibited allosterically (Marques et al., 2015). We chose to focus on the more specific effect of M941 on FP3 enzymatic activity.
4.3.3 M941 inhibits recombinant FP3 in a fluorogenic substrate assay.

To further confirm that FP3 was the target, we used a fluorogenic substrate assay to measure rFP3 activity in the presence of M941. rFP3 was added to different concentrations of Z-LR-AMC substrate in the presence of different M941 concentrations. Hydrolysis over time was measured using a MSE spectrofluorimeter (excitation 355 nm, emission 460 nm). The initial rates of hydrolysis at each substrate concentration were calculated ($V_0$).

A plot of activity against M941 concentration at 50 μM Z-LR-AMC shows clear dose-dependent inhibition with an IC$_{50}$ of 46 ± 5.5 μM, close to the calculated AC$_{50}$ (32 ± 5 μM) measured in parasite lysates (Fig. 4.3.2A, Table 4). Substrate turnover completely was ablated at 200 μM M941 (Fig. 4.3.2A). To determine the type of inhibition the $V_0$ was plotted against substrate concentration. However, prediction of the inhibition model was complicated by substrate inhibition over 100 μM Z-LR-AMC. The raw data was fit to a substrate inhibition model in Prism and $K_m$ and $V_{max}$ were calculated and plotted against inhibitor concentration (Fig. 4.3.2B-C). $V_{max}$ appeared to decrease with increasing...
M941 concentration while \( K_M \) appeared constant, suggesting that inhibition was non-competitive with the substrate (Fig. 4.3.2C). However, there were large standard errors in both data sets and this experiment was only done once.

Thus, M941 inhibits rFP3 proteolytic activity \textit{in vitro}. Non-competitive binding provided a hypothesis for the inhibition/probe activation effect. A non-competitive inhibitor binds to the enzyme-substrate complex as well as the free enzyme, slowing conversion to the enzyme-product complex. If the inhibitor binds to and stabilizes the enzyme-substrate tetrahedral intermediate, it could also stabilize the irreversible formation of the enzyme-probe bond, explaining the increase in probe labelling. However, to verify the mode of inhibition multiple replicates of this experiment are required in order to more accurately calculate \( V_{\text{max}} \) and \( K_M \) values.

4.4 M563 is a broad-spectrum SH inhibitor.

The further investigation of hits was focused on three Malaria Box compounds that were found to ablate FPP-TMR labelling of specific bands, M563, M438, M914. M563 primarily affected SH bands at 80 and 56 kDa in trophozoite and schizont lysates but also inhibited further bands in ring lysates, and at higher concentration it appeared to broadly inhibit most labelled SHs. We now aimed to discover how M563 affected live parasites and the phenotype associated with inhibition, as well as probe the binding mechanism of M563.
4.4.1 M563 blocks parasite development.

We first wanted to confirm the anti-parasitic activity of M563 in live parasite cultures. We used a standard 72-h replication assay to calculate the EC\textsubscript{50}, which could then be compared to the previously reported EC\textsubscript{50} values and the IC\textsubscript{50} values calculated above. Synchronous ring stage parasites at 1% parasitaemia and 0.5% haematocrit were treated with DMSO or different concentrations of M563 in triplicate. After 72 h in standard growing conditions, the parasites were fixed and the parasitaemia determined by flow cytometry. For more details on parasite synchronisation (Section 9.1) and flow cytometry (Section 9.3.5), see Materials and Methods. The parasitaemia was plotted against inhibitor concentration and the data were fitted to a standard dose response curve to calculate EC\textsubscript{50}s and standard errors. M563 completely blocked parasite replication at 5 μM and had a sub-micromolar EC\textsubscript{50} of 0.4 μM, comparable to the ChEMBL reported EC\textsubscript{50} (Fig. 4.4.1A, Table 4.3). These EC\textsubscript{50}s are ten times more potent than the IC\textsubscript{50}s obtained from the dose-dependent inhibition of two ABP-labelled bands (Fig. 4.1.3; Table 4.3). Perhaps in parasites, as in lysates, this inhibitor targets multiple SHs that have a cumulative effect on parasite growth. It could also be that M563 accumulates in parasites leading to a higher effective concentration, or that M563’s anti-parasitic MoA is through a non-SH target.

To gain more information on the mechanism of M563-mediated anti-parasitic activity we used microscopy to observe the life stage at which development is blocked. The standard dose response curves were used to determine the EC\textsubscript{90} value of 3.1 μM, this was to ensure that the majority of target proteins would be inhibited. Synchronised parasites at ring or trophozoite stage were treated with 3.1 μM M563 or DMSO and incubated under standard growing conditions. Thin blood smears were made every 24 h and stained with Giemsa to observe parasite development and morphology. The smears shown here are representative of three replicates (Fig. 4.4.1B).

When treated with M563 at ring stage, parasites did not progress to trophozoite stage and by 48 h were visible only as shrunken, dead parasites (Fig. 4.4.1B). After trophozoite stage treatment the parasites did not develop any further or enter schizogony (Fig. 4.4.1B). We may therefore conclude that the targets of M536 are active and essential for parasite growth at both ring and trophozoite stage. In order to find out how M563 engages with its targets we undertook experiments to test the mechanism of inhibition.
4.4.2 M563 is likely to inhibit by a reversible tight-binding mechanism.

In order to determine the mode of inhibition of M563 we first wanted to verify if it acted as a reversible inhibitor. We increased the probe concentration to observe if M563 could still outcompete labelling. Trophozoite lysates were pre-treated with 50 μM M563 or DMSO, followed by labelling with FPP-TMR at 50 nM or 500 nM, for 10 min. We observed that M563 inhibited labelling even at 500 nM, which was not expected for a reversible inhibitor (Fig. 4.4.2A). From the structure of M563 we formulated a model for covalent inhibition. We can draw the structure of M563 tautomerised to contain an α,β-unsaturated ketone (Fig. 4.4.2B). In reality, the compound exists in an in-between of these two states, with the electron density spread out over the nitrogen, oxygen and the three carbons in-between, with the electronegative oxygen and nitrogen being the most negatively charged. This means that the...
carbon atom indicated with an arrow (Fig. 4.4.2B), is an electrophilic centre which could react with a catalytic serine.

To determine the mode of M563 inhibition we wished to observe the time-dependence of inhibition. Schizont lysates were incubated for 15 or 60 min with a dose response of M563 prior to labelling with 1 μM FPP-TMR for 1 h (Fig. 4.4.1C). Dose dependent inhibition of multiple FPP-labelled bands was observed at both pre-incubation times, further confirming M563 as a broad-spectrum SH inhibitor (Fig. 4.4.1C). The IC50 was calculated for the two most intense bands 80 and 56 kDa. The IC50s were the same at both M563 pre-treatment times (Fig. 4.4.2D). This lack of change suggested a time-independent reversible mechanism, although it could still be a tight binding or covalent reversible inhibitor. Previous quantification of band labelling showed that the 80 kDa band was labelled slowly and did not reach saturation after 60 min treatment with FPP-TMR. This slow rate of labelling explains why M563 is able to inhibit the 80 kDa, even at high probe concentrations, with a reversible mechanism. Lack of stability has been noted for M563, one limitation of this experiment is that the compound may degrade during the pre-treatment time and over the time course. We can conclude that M563 is a broad-spectrum SH inhibitor that acts via a reversible inhibition mechanism.
Figure 4.4.2: Investigation into the inhibition mode of M563. (A) Trophozoite or schizont lysates were pre-treated with DMSO or the Malaria Box compounds (50 μM) for 30 min before 10 min labelling with 50 nM or 500 nM FP-TMR at pH 7.2 (B) Tautomerism of M563, with the arrow showing possible site of nucleophilic attack. (C) Trophozoite lysates were pre-treated for 30 min with different concentrations of M563 before treatment with 1 μM FP for 15 or 60 min. Inhibited bands are labelled with an arrow. (D) IC₅₀s for different molecular weight bands were calculated by quantifying band labelling in Image Lab™ and fitting to a standard dose response in Graphpad Prism®, the calculated standard error is given. (A&C) Labelling reactions were quenched by addition of LB and immediate boiling. 12.5% SDS-PAGE gels were run and a fluorescence scanner was used to visualise TMR (red channel Ex/Em 552/572nm) labelling.
4.5 M438 and M914 appear to target the same SH but may have other distinct targets.

4.5.1 M438 and M914 block parasite development.

M438 and M914 both inhibited the same 49 kDa SH present in schizont lysates and are therefore considered together. We wanted to confirm and observe the effect of the M438 and M914 on live parasite cultures using a standard 72 h replication assay as described above (Section 4.4.1). The inhibition and EC$_{50}$s calculated for both inhibitors were consistent with the ChEMBL reported EC$_{50}$s, with M438 being more potent as expected, although both fully blocked parasite replication at high concentrations (Fig. 4.5.1A, Table. 4.3). The standard dose response curves were used to estimate EC$_{90}$ values of 0.75 μM and 3.5 μM for M438 and M914, respectively. We used light microscopy to determine the life stage-specific inhibition by the two compounds as described above (Section 4.4.1). M438 (like M563; Fig.4.4.1) gave a complete block in parasite development whether treated at ring or trophozoite stage (Fig. 4.5.1B). M914 blocked any further parasite development when treated at ring stage. In contrast, after trophozoite treatment parasites progressed to early schizont stage, although development was slowed compared to the DMSO control (Fig. 4.5.1B). M438 and M914 were assumed to target the same 49kDa protein but the difference in chemical inhibition phenotype here suggests this may not be the case, although it could be due to a difference in effective concentration of inhibitor due to stability/solubility/penetration issues. Another explanation is that the concentration of M914 used in these assays, based on the EC90 of ring-stage parasites at 72 h, is not enough to ensure full inhibition in the metabolically more active trophozoite stage.

Our hypothesis is that the targets of these Malaria Box compounds are metabolic SHs. They all start to block parasite development at trophozoite stage when the parasite is most metabolically active (Section 1.2.2.2). We continued to investigate the targets of M438 and M914 by exploring a previously reported target of M438.
Figure 4.5.1: Standard replication assay and phenotypic effects of M438 and M914. (A) A standard 72-h growth assay was carried out in the presence of different concentrations of inhibitor. After 72 hours post invasion (hpi), the parasites were fixed and stained with Hoechst, and the infected/uninfected RBCs counted by flow cytometry. The parasitaemia was plotted against inhibitor concentration and fit to a standard dose response curve in Graphpad Prism© to determine EC$_{50}$ values. Error bars represent the SEM (standard mean error) of three replicates. (B) Giemsa-stained thin blood smears were made of parasites at 0 hpi (ring stage), 24 hpi (trophozoite stage) and 48 hpi (schizont stage) after treatment with approximately EC$_{90}$ concentration of inhibitors (or DMSO) at ring or trophozoite stage.
4.5.2 M438 inhibits the Coenzyme A (CoA) biosynthesis pathway but this is not the FPP-labelled SH target

M438 was previously identified in a Malaria Box screen designed to find inhibitors of the *P. falciparum* CoA synthesis pathway (Fletcher and Avery, 2014). The study used a chemical rescue approach to find compounds where the inhibition phenotype could be rescued by supplementing the media with CoA. They performed HTS on asexual-stage parasites using the Malaria Box alongside another focused compound library in the presence and absence of CoA. M438 was found to have an IC$_{50}$ of 0.43 ± 0.05 μM and IC$_{80}$ inhibition could be rescued by CoA to 80% (Fletcher and Avery, 2014).

SHs are not involved in CoA synthesis, the enzymatic steps of which are highly conserved between eukaryotes (Huthmacher et al., 2010; Leonardi et al., 2005). However, we know that many metabolic SHs are annotated as esterases, lipases and lysophospholipases and, as the main acyl carrier, CoA is intimately involved in lipid metabolism. The Malaria Box structural analogues were not identified in this screen and neither was M914, which we had expected to inhibit the same target. We wanted to repeat the experiments described in this study for M438 and M914. Panthenol is a specific inhibitor of the CoA synthesis via the first pathway enzyme, pantothenate kinase, and was used as a positive control (Fletcher and Avery, 2014; Saliba et al., 2005).

A standard 72-h replication assay was carried out, in triplicate, in the presence of different inhibitor concentrations and with or without the addition of 1 mM CoA. After 72 h, the parasitaemia was measured by flow cytometry as previously described.

As expected, the addition of CoA gave close to 100% rescue from Panthenol growth inhibition. Although M438 antiplasmodial activity was reduced after CoA treatment, as observed previously (Fletcher and Avery, 2014), there was no effect on the M914 inhibition phenotype (Fig. 4.5.2). These results, combined with the microscopy observations (Section 4.5.1) imply that M438 has multiple enzyme targets including one related to the CoA biosynthesis pathway and one SH that is also targeted by FPP-TMR and M914.
4.6 Discussion

We aimed to develop and use an ABPP screen in competition with the Malaria Box to identify new essential enzymes and potential drug targets. We developed a medium-throughput screen that could test 400 compounds in one day and has potential to be expanded further. ABPP screens often use libraries of irreversible inhibitors focused towards an enzyme family of interest. This Malaria Box screen is an unbiased screen of a diverse set of reversible compounds. Combining two probes with orthogonal fluorophores allowed ABPP to be multiplexed, simultaneously screening against fifty SH and CP targets. Three compounds were identified that blocked FPP-TMR labelling of SHs (M438, M914 and M563) and two compounds were found that increased W-sCy5-VS of potential CPs (M941 and M738).

The biggest difficulty in the Malaria Box screen was the collection of parasite material, which took months to accumulate. For this reason, the screen was not performed in triplicate, as would have been ideal. The more time-consuming process of merozoite collection meant we did not attempt the screen at this stage. We reasoned that although repeating the screen may have led to more hits, five hits after multiple round of confirmation was an ideal number to investigate further. With limitless resources, the screen could have been extended to a high throughput proteomics platform to test the competition for targets between probes and inhibitors. However, the gel-based screen was simple and inexpensive to perform, with no specialist instrumentation required. The gel-based method selects for the highest abundance proteins that can be easily visualised, ensuring straight-forward follow-up studies.
The induction of labelling caused by the two W-sCy5-VS hits was unexpected and unprecedented. Neither has been identified in other *P. falciparum* Malaria Box screens, but M941 can interconvert to gentian violet, a bright blue dye with reported antibacterial, antifungal, antiparasitic and antitumor properties (Maley and Arbiser, 2013). M941 and M738 both caused an increase in E64-sensitive labelling of the PLCP FP3. M941, but not M738, had the same effect on recombinant FP3. M941 was subsequently shown to inhibit FP3 substrate turnover. Modelling of the mode of inhibition was complicated by pronounced substrate inhibition, and further repetitions are necessary to make formal conclusions. However, from our preliminary data we speculate that M941 binds to FP3 allosterically: inhibiting the enzyme while leaving the active site open to interact with the probe, or E64. The non-competitive binding may inhibit FP3 such that nucleophilic attack of a substrate can still occur, but is not followed by substrate hydrolysis. By raising the activation energy of the latter steps of the hydrolysis reaction, the forward reaction is disfavoured, which would not hinder the irreversible mechanism of probe binding. This could be due to more efficient stabilisation of the tetrahedral intermediate formed in the course of the reaction by His hydrogen-bonding, that is also required in the reaction with E64 and W-sCy5-VS. There is evidence of allosteric inhibition by haem and suramin of the closely related FP2 in fluorogenic assays using similar substrates (Marques et al., 2015, 2013). The prediction of the molecular basis of the inhibition mechanism has been studied, building towards the design of novel allosteric inhibitors against both FP2 and FP3 (Hernández González et al., 2019). The results we report here support this aim for FP3. Binding of haem to FP2 is adjacent to the active site, probably functioning to regulate the accumulation of free toxic haem by negative feedback (Marques et al., 2015). In the 2015 FP2 study, they also report that haem binding induces a regulatory substrate binding site that allows for substrate inhibition (Marques et al., 2015). They showed that the allosteric haem binding results in changes around the FP2 “arm” region, although further studies are required to determine how this affects the catalytic site (Marques et al., 2015). Another study using chalcone derivatives suggested a more classical non-competitive mechanism (Bertoldo et al., 2015). M738 and M941 may interact allosterically with FP3 via two different binding modes, similar to those described for FP2. M738 did not affect recombinant FP3 labelling, this could be due to a difference (for example, in structure or pH dependence) between the recombinant and endogenous (lysate) FP3 enzyme. M738 also induced labelling of two other targets in lysates. This labelling was E64-insensitive, and the identity of the bands remains a mystery. No compound-dependent labelling effects were observed with W-dtBio-TMR-VS (data not shown), so this cannot be used to identify the targets by proteomics. It would be very interesting to identify them in order to employ non-probe methods of chemical proteomics such as detection of binding by thermal stability changes, CESTA.

The SH inhibitor M563 primarily inhibited two high intensity bands, but showed broad-spectrum inhibition at higher concentration. M563, along with other tetracyclic benzothiazepines (BZTs), has
previously been shown to inhibit complex III (cytochrome bc₁) of the *P. falciparum* electron transport chain (ETC; Dong et al., 2011). The ETC is well-established as the target of the atovaquone/proguanil anti-malarial drug combination (Malarone™; Kremsner et al., 1999). In the BZT study, they showed that the inhibition phenotype could be rescued in a transgenic *P. falciparum* Dd2 strain that circumvents complexes I-III in the ETC. Further *in vitro* assays identified cytochrome bc₁ as the target (Dong et al., 2011). However, in their SAR studies M563 was one of the least potent BZTs tested against *P. falciparum*, having an EC₅₀ of 2 μM, 100-fold higher than the majority of the series. However, it is worth noting that both the EC₅₀ reported in this thesis and by ChEMBL were lower than this (Table 4.5.1). Therefore, cytochrome bc₁ may be among the targets of M563, but we can still make no inferences as to any principal cause of its anti-parasitic activity. The Malaria Box was intelligently designed to assist with SAR follow-up studies and includes sets of structural analogues. MMV006389, MMV007127 and MMV645672 are tetracyclic benzothiazepines (BZTs) similar to M563. These were not identified in the initial screens, or in subsequent retesting (Supplementary Fig. 10.2.3). But as we know that M563 degrades over time, it is possible that fresh stocks of the analogues would also show inhibition. It would be interesting to test if fresh stocks of BZTs showed the same pan-SH inhibition as M563.

Lastly, two structurally distinct compounds (M438 and M914) blocked FPP-TMR labelling of the same 49 kDa SH band. As noted for M563 and the BZTs, similar fused-ring structures to M438 in the Malaria Box (MM085583 and MMV011436) showed no inhibition in the screen or in resting (Supplementary Fig. 10.2.3). Previous target-based Malaria Box screens have generally identified hits in the micromolar range, the IC₅₀s of M438 and M914 calculated above show that these hits are particularly potent (Fong et al., 2015; Paiardini et al., 2015; Tiwari et al., 2016). Both inhibitors blocked parasite growth at ring stages, although M914 caused a less pronounced phenotype than M438 when applied at trophozoite stages. M914 is less potent than M438, so it is possible that not enough M914 was used to observe the same effect in the highly metabolically active trophozoite stages.

In previous Malaria Box studies, M438 was identified as an inhibitor of the *P. falciparum* CoA synthesis pathway, but M914 was not (Fletcher and Avery, 2014). This pathway has been validated as an antimalarial drug target, and the acetyl-coA synthetase inhibitor MMV183 (TropIQ) is in preclinical development. We were able to confirm these results showing that addition of exogenous CoA modulated M438 inhibition, but not M914 inhibition. These results combined with the microscopy observations suggest at least two targets for M438: a SH target in rings shared with M914; and a CoA-synthesis target in trophozoite stages. M438 has also undergone target deconvolution studies via IVEWGA. In one study, no M438 resistant parasites were obtained, teasing a resistance-proof chemotype (Corey et al., 2016). However, in another study, resistance-conferring mutations were found in one of the α/β-fold SHs, *PfPARE* (Istvan et al., 2017). *PfPARE* was shown to be a non-essential
esterase that was not the target of M438, but was responsible for activation of the chemotype by hydrolysis of the ester group on M438 (Fig. 4.1.2; Istvan et al., 2017). The contradiction of the two IVEWGA studies is confusing, but the interaction with P/PARE suggests that there are easy mechanisms towards parasites acquiring M438 resistance.

Our aims for further target deconvolution focus on the SH inhibitors M563, M438 and M914. We hope to use chemical proteomics to identify the SH target(s) of each compound. The chemical proteomic methods that we have already described using FPP-N₃, identified 9 out of a predicted 43 α/β-fold SHs (Section 3.2). We therefore first aimed to optimise our proteomics methods to improve the depth of coverage of this large enzyme family. One of the challenges was to achieve detectable changes from the reversible inhibitors via proteomics, as we had found for the gel-based screen. Towards these objectives, we compared different FPP probes, optimised the affinity purification procedure, and used more precise methods of quantification (Chapter 5).
Chapter 5. **Results III: Chemical proteomics prioritises PfSHs for genetic validation.**

We demonstrated that FPP-TMR could be used to profile active SHs across multiple life stages of *P. falciparum* (Section 3.1.1) and identified three antimalarial compounds from the Malaria box that inhibit FPP-TMR labelling of a specific SH (M914 and M438) or multiple SHs (M563) (Chapter 4). We also showed that FPP-N$_3$ could be used to identify SHs via chemical proteomics (Section 3.2). Nine SHs were identified as probe targets in schizont lysates, however structural database searches predicted at least 43 hydrolases in the α/β-fold superfamily alone (Section 3.2, Table 1). We wanted to increase the depth of coverage and quality of quantification of the chemical proteomics method in preparation for the profiling of SHs across multiple life cycles, and to identify the targets of the Malaria box compounds.

### 5.1 Optimisation of chemical proteomics protocols achieved increased depth of coverage.

#### 5.1.1 FPP-N$_3$ was a better SH probe for chemical proteomics than FPP-Biotin.

It has been shown that different FPP probes can have slightly different reactivity profiles, and some studies have combined the probes for optimum coverage (Kidd et al., 2001). We have successfully demonstrated the use of FPP-N$_3$ in chemical proteomics to identify SH targets, but we wanted to see how using different FPP probes in the chemical proteomics method would affect the results. Using a probe such as FPP-Biotin would eliminate the CuAAC-chemistry step used to biotinylate FPP-N$_3$ targets, thus reducing background noise and limiting loss of material during precipitation. However, FPP-N$_3$ unlike FPP-Biotin is cell-permeable so can be used in intact cell conditions; it’s small size could also confer additional reactivity.

We had previously performed competition labelling to verify that FPP-N$_3$ had the same targets as FPP-TMR; when pre-treated with 0.1 µM FPP-N$_3$ almost all subsequent labelling with 0.1 µM FPP-TMR was blocked and it was completely blocked with 1 µM FPP-N$_3$ (Section 3.1.2). We now repeated this experiment using FPP-Biotin in competition with FPP-TMR. Although every band decreased in FPP-TMR labelling intensity after 0.1 µM FPP-Biotin treatment, no bands were completely blocked and faint FPP-TMR labelling of multiple bands could be seen even following 1 µM FPP-Biotin (Fig. 5.1.1).
More FPP-Biotin was required to block FPP-TMR labelling than FPP-N₃, this suggest that although the probes share common targets, FPP-Biotin is less reactive than the other two.

We now wished to test FPP-Biotin reactivity more quantitatively by proteomics and compare this to previous results using FPP-N₃ (Section 3.2). Schizont lysates were labelled with 1 μM FPP-Biotin or DMSO and processed for proteomics as previously described (Section 3.2). *P. falciparum* proteins only were considered. Compared to FPP-N₃ under the same conditions (Section 3.2), 16% fewer proteins were quantified by LFQ. Eleven proteins were significantly enriched in the FPP-Biotin compared to DMSO sample with fold change in enrichment of at least 3 (Table 5.1.1). Of these, only three were SHs compared to 11 SHs in 15 enriched proteins with FPP-N₃ (Table 3.2).

Thus, FPP-N₃ appeared to be a more specific probe for SHs than FPP-Biotin in lysates. The experiment could not be extended to live parasites as FPP-Biotin is not cell permeable. The CuAAC-chemistry step may actually facilitate LFQ of proteins due to a more consistent background between samples, therefore improving the signal to noise ratio of protein activation changes. However, there is a balance to be struck between a robust background and low background noise to achieve high resolution, especially for low abundance proteins. We therefore considered other methods of improving the quality of proteomic data by reducing contamination and improving quantification strategies.

**Figure 5.1.1: Competition of broad spectrum SH FPP probes.** A) Schizont lysates were labelled with 0.1 or 1 μM of FPP-Biotin for 30 minutes before labelling with 0.1 μM FPP-TMR for 30 minutes. Labelling reactions were quenched by addition of LB and immediate boiling. 12.5% SDS-PAGE gels were run and a fluorescence scanner was used to visualise TMR (red channel Ex/Em 552/572nm) labelling (left). The gel was transferred to a blot and biotin labelling was visualised streptavidin conjugated HRP-chemiluminescent substrate (right).
5.1.2 Dilution of neutravidin linked agarose beads identified ideal pull-down conditions.

Contamination from tryptic neutravidin peptides is inherent in the proteomics method. Peptides from trypsin and other common contaminants such as keratin are also present in lower amounts. The MS employs dynamic exclusion windows in order to prevent highly abundant peptides from being repeatedly detected over time, but this can mask low abundance peptides at a similar mass. As a limited amount of material can be loaded on the LC-MS/MS, we proposed to reduce the ratio of neutravidin to non-neutravidin peptides in order to boost the signal of lower abundant peptides and ensure their detection. One solution would be to elute target proteins from the neutravidin beads (NBs) before trypsin digestion by chemical/thermal denaturation or chemical cleavage of a special probe linker. However, all elution techniques incur losses in material. We therefore decided to proceed with on-bead digestion but to optimise the volume of NBs used. The aim was to achieve an optimal engagement of labelled proteins using the least volume of NBs. The protocol requires the scientist to be able to see the bead pellet clearly, so we supplemented the NBs with blank agarose beads in different ratios and observed the effect on protein identifications (Fig. 5.1.2A). This work was undertaken in collaboration with the proteomics facility at The Francis Crick Institute (Steve Howell, Bram Snijders).

Schizont lysates were labelled in triplicate with 1 μM FPP-N₃ or DMSO and treated with CuAAC-chemistry reagents (Section 1.5.1.3). Labelled proteins were affinity purified using six different agarose bead mix suspensions containing different ratios of NB:blank beads (Table. 5.1.2). The proteins were digested with trypsin, prepared and run on the LC-MS/MS (Section 1.6.2). For more details on proteomics methods see Materials and Methods (Section 9.4). The raw intensities (before LFQ) of specific peptides from neutravidin, trypsin and SHs were assessed using Skyline. In probe-treated samples neutravidin peptides increased in intensity linearly with NB, but the increase in SH peptide

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Protein name</th>
<th>Mol. weight [kDa]</th>
<th>Fold-enrichment (log)</th>
<th>-Log (t-test p value)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF3D7_0709700</td>
<td>PARE</td>
<td>42.4</td>
<td>9.05</td>
<td>4.33</td>
<td>Prodrug activation and resistance esterase</td>
</tr>
<tr>
<td>PF3D7_1341300</td>
<td>L71-2</td>
<td>21.5</td>
<td>3.37</td>
<td>1.09</td>
<td>60s ribosomal</td>
</tr>
<tr>
<td>PF3D7_1343000</td>
<td>PMT</td>
<td>31.0</td>
<td>2.82</td>
<td>1.18</td>
<td>Phosphoethanolamine N-methyltransferase</td>
</tr>
<tr>
<td>PF3D7_0922500</td>
<td>PGK</td>
<td>45.4</td>
<td>2.81</td>
<td>2.18</td>
<td>Phosphoglycerate kinase</td>
</tr>
<tr>
<td>PF3D7_1143000</td>
<td>α/βH</td>
<td>44.8</td>
<td>2.52</td>
<td>1.80</td>
<td>Putative protein</td>
</tr>
<tr>
<td>PF3D7_1324900</td>
<td>LDH</td>
<td>34.1</td>
<td>2.50</td>
<td>2.36</td>
<td>L-lactate dehydrogenase</td>
</tr>
<tr>
<td>PF3D7_0929400</td>
<td>RhFH</td>
<td>162.7</td>
<td>2.45</td>
<td>1.67</td>
<td>Rhoptry protein</td>
</tr>
<tr>
<td>PF3D7_1462800</td>
<td>GAPDH</td>
<td>36.6</td>
<td>2.00</td>
<td>1.90</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>PF3D7_10155900</td>
<td>ENO</td>
<td>48.7</td>
<td>1.88</td>
<td>2.57</td>
<td>Enolase</td>
</tr>
<tr>
<td>PF3D7_1476800</td>
<td>LysoPL</td>
<td>42.7</td>
<td>1.88</td>
<td>1.44</td>
<td>Putative protein</td>
</tr>
<tr>
<td>PF3D7_0708400</td>
<td>HSP90</td>
<td>86.2</td>
<td>1.66</td>
<td>1.75</td>
<td>Heat shock protein 90</td>
</tr>
</tbody>
</table>

*Fold enrichment (log) = 1.58 is equivalent to Fold change = 3 α/β-fold SHs
intensity slowed over 10 \mu l NB. (Fig. 5.1.2C, shows representative example). Trypsin peptide intensities were not as constant as expected but showed no clear trend (Fig. 5.1.2C shows representative example). Proteins were then quantified using iBAQ (Section 1.6.3.1.2) and filtered as previously described (Section 3.2) (for this dataset iBAQ identified more many proteins than LFQ). The number of \textit{P. falciparum} proteins and SHs identified was considered, as an average of the three replicates. In probe-treated samples the highest number of total proteins and SHs were identified using 5 \mu l NB, followed by 10 \mu l NB (Table. 5.1.2). The next aim was to calculate how many SHs were significantly enriched in FPP-N3 vs DMSO-treated samples. The protein list was filtered to remove proteins that were not identified in 3/3 replicates in either FPP-N3- or DMSO-treated samples, and a student t-test was done (FDR = 0.05, \sigma_0 = 0.5). For each NB volume, the number of significantly enriched SHs was constrained by the number of SHs that were identified in DMSO-treated samples (Table. 5.1.2). The most SHs were identified in the DMSO samples using 30 \mu l NB, presumably due to a higher background protein binding to the NBs.

In order to circumvent the problem of low SH identifications in the control samples, we turned to TMT labels as a quantification method. TMT-labelling techniques allow samples to be multiplexed. Proteins are identified globally in MS1 followed by relative quantification of the protein for each sample.

\textbf{Figure 5.1.2: Optimisation of affinity purification for chemical proteomics.} (A) Pictorial representation of probe and DMSO treated samples for proteomics showing the amounts of blank/neutravidin beads. (B) Pictorial representation of probe and DMSO treated samples for proteomics, showing the amounts of blank/neutravidin beads and labelling for TMT quantification. TMT labelled samples are subsequently combined and the mixture is fractionated by high pH before LCMC-MS. (C) Representative examples of precursor ion intensities in each NB condition, extracted from Skyline.
separately in MS2. Therefore, every protein that is identified has a relative quantification value for all probe-treated and control samples. TMT labels were used to re-quantify the 1, 5 and 10 μl NB samples as these had given high total protein IDs, but low raw intensities of Neutravidin peptides. The FPP-N3- and DMSO-treated 1, 5 and 10 μl NB samples were labelled with TMT-sixplex tags as indicated (Fig. 5.1.2B). The labelled samples were combined and the resulting mix was analysed by LCMS/MS (Fig. 5.1.2B). Fewer total proteins were identified in the TMT samples, but the number of proteins identified was consistent over all FPP-N3- and DMSO-treated samples as expected (Table 5.1.2). The highest number of SHs significantly enriched in FPP-N3- vs DMSO-treated samples (student t-test, FDR = 0.05, s0 = 0.5) was found with 10 μl NB. Thus, we reasoned that 10 μl NB was the optimal pulldown condition to allow for maximal SH capture and identification, and this was used in the subsequent large-scale proteomic experiments.

<table>
<thead>
<tr>
<th>Neutravidin agarose / µl</th>
<th>Blank agarose / µl</th>
<th>Number of proteins, iBAQ quantification</th>
<th>Number of proteins, sixplex-TMT quantification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>FPP-N3* DMSO*</td>
<td>SHs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>308</td>
<td>25</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>356</td>
<td>26</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>363</td>
<td>26</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>365</td>
<td>26</td>
</tr>
<tr>
<td>1</td>
<td>29</td>
<td>326</td>
<td>25</td>
</tr>
<tr>
<td>0</td>
<td>30</td>
<td>144</td>
<td>2</td>
</tr>
</tbody>
</table>

*Numbers represent the average of three replicates, **significantly enriched FPP-N3 vs DMSO (s0=0.5, FDR = 0.01)

5.2 Comparative chemical proteomics for Malaria Box target deconvolution.

The antimalarial compounds M563, M438 and M914 were all found to inhibit FPP-TMR labelling of one or more SHs (Chapter 4). We wanted to utilise the optimised proteomics method to perform a comparative study of SH activities in the presence of the Malaria box compounds and identify their targets (Fig. 5.2A). Each inhibitor was used at four different concentrations alongside DMSO and no probe controls (Fig.5.2B). Comparative proteomics detects more subtle changes than the gel-based method, but more labelled material is needed to input into the MS. In the Malaria Box screen 50 nM FPP-TMR was used for 10 min in an effort to remain in a dynamic range of labelling. For the proteomics method, we hypothesised that treating samples with higher probe concentration (1 µM, 10 min) would provide enough protein material to effectively identify targets on the MS. We hoped that the chemical proteomics methods would be sensitive enough to still detect reversible binding of M563, M914 or M438 in the active site of their targets. Inhibitor concentration ranges were chosen to include the IC50.
for each compound calculated based on probe labelling as an activity readout (Fig. 4.1.3). For M563 and M914 the IC\textsubscript{50}s were 6.5 μM and 0.3 μM, respectively, inhibitor concentrations from 0.15 μM - 12.5 μM were used (Table 4). For M438, the IC\textsubscript{50} was 0.08 μM, and we used concentrations from 0.015-1.25 μM (Table 4).

Schizont lysates were pre-treated with inhibitor (or DMSO) for 30 min before labelling with 1 μM FPP-N\(_3\) (or DMSO) for 10 min in triplicate. The sample preparation for proteomics was carried out as previously described (Section 3.2) using 10 μl NB (Section 5.1.2, Fig. 5.2A-B). In each inhibitor replicate set, samples were labelled with 6-plex TMT tags for quantification and combined (Fig. 5.2B), giving a total of nine 6-plex TMT mixes. The resulting mixtures were separated into 8 fractions (by high pH fractionation) before LC-MS/MS analysis in an effort to increase the depth of coverage of the method. Across all samples, 1093 proteins were identified, 561 proteins were left after filtering as previously described (Section 3.2). The log\textsubscript{2}(x) intensities were normalised by subtracting the median intensities of each replicate across all samples, followed by the median intensities of each protein within replicate groups.

Twenty-One \textit{P. falciparum} SHs were significantly enriched in probe-treated compared to the DMSO-treated samples, of which twenty were predicted SHs (Student t-test, FDR = 0.05, s\(_0\) = 0.5). For each protein, the fold change enrichments were transformed into percentage activity relative to the probe only condition and plot against inhibitor concentration. From the Malaria Box screening results, we expected that M914 and M438 may share a 48kDa target, and that M563, while primarily inhibited two targets of 56 and 63 kDa, may target multiple other SHs (Fig. 4.1.2). Three of the enriched proteins showed a general trend of decreasing activity with increasing inhibitor concentration in at least 2 out of 3 replicates (Table 5.2): PF3D7\_0629300 (LCAT) decreased with all three compounds; PF3D7\_1120400 (abH112) decreased with M563; and PF3D7\_1143000 (abH114) decreased with both M914 and M438, and (Fig. 5.2C-E). These results supported our previous observations, that M563 inhibits multiple SHs and M914 and M438 shared a common target (Section 4.5). Furthermore, abH114 is the correct size to be the common target of M914 and M438 (target \(\approx\) 49kDa, abH114 = 44.8kDa, Fig. 4.1.2) although LCAT is not (LCAT = 99.2kDa). However, multiple student t tests showed only M438-induced LCAT activity changes were statistically significant (student t-test, s\(_0\) = 0.5, FDR = 0.01, FPP-N\(_3\)-treated vs. 1.25 μM M438, fold difference = 1.13, p = 0.003). abH114 was only quantified in two out of the three replicates, due to missing values for this protein in the third replicate. We decided to keep all three proteins in consideration for genetic interrogation. Five human SHs were also detected, none of these showed changes in enrichment levels in the inhibitor-treated samples (data not shown).
Figure 5.2: Comparative proteomics to identify Malaria Box compound targets. (A) Schematic of the comparative proteomic method. Soluble protein lysates are pre-treated with inhibitor (or DMSO) before being labelled with probe (or DMSO), target proteins that have been inhibited will not be labelled. CuAAC-chemistry is used to attach biotin to labelled proteins for pull down on neutravidin agarose beads. On-bead digestion liberates peptides from beads for subsequent processing for proteomics. (B) Schematic of comparative proteomics method, continued from (A). Peptide samples are labelled with different 6-plex TMT tags. Sets of 6-plex labelled samples are combined and the mixture is fractionated by high pH into 8 fractions for LCMSMS. (C-E) The normalised log₂ enrichment of proteins PF3D7_0629300 (LCAT), PF3D7_1120400 (abH112) and PF3D7_1143000 (abH114) was used to calculate the percentage activity relative to zero inhibitor concentration. Each replicate is plotted separately. abH114 could only be quantified in 2/3 replicates.
5.3 ABPP of SHs during the *P. falciparum* blood-stage asexual life cycle.

We have shown that the profile of active SHs changes throughout the life cycle (Section 3.1.1) and that FPP detection of SHs can provide the first evidence of a catalytically active serine residue in proteins that were previously unannotated (Section 3.2). We now wanted to extend the in-depth chemical proteomics profiling of SHs to the whole asexual life cycle of *P. falciparum* with two main aims. First, to have the best chance of achieving saturation of the all SHs active in the sexual stages. And secondly, we hypothesised that SHs that were active at specific time points in the life cycle may play an essential role at that time. Therefore, identifying SHs specifically activated at certain time points, for example around merozoite egress, may highlight interesting candidates for genetic interrogation.

5.3.1 Optimised chemical proteomics methods achieve good depth of coverage of *PfSH’s*.

We showed that FPP-N₃ labels SHs in intact schizonts as well as lysates (Section 3.1.2). In order to construct the most accurate profile of stage-dependent SH activity we aimed to label intact rings, trophozoites and schizonts with FPP-N₃ (Fig. 5.3.1A). However, merozoites are time consuming to collect and have a very short lifespan outside the RBC, so these were collected and frozen prior to probe labelling.

To ensure full life cycle coverage, parasites were roughly synchronised to achieve large time windows, (rings 0-20 hpi, trophozoites 20-36 hpi, early schizonts 36-44 hpi). Although mature (C2-treated) schizonts and merozoites represent discrete time points before and after invasion (48hpi), Life stages were synchronised and collected (Section 9.4.1.2). The intact early and mature schizonts were purified and labelled with 1 μM FPP-N₃ (or DMSO), and rings and trophozoites were labelled in the presence of RBCs (Section 9.4.1.2). Labelled parasites were washed, saponin lysed and frozen. Merozoites were collected, purified and frozen (Section 9.4.1.2). Frozen parasites from different days and RBC batches were pooled, and soluble protein was extracted using 0.1% Triton X in PBS at 4°C. Merozoite lysates were labelled with 1 μM FPP-N₃ (or DMSO) for 1 h. The samples were split into three to give technical replicates and processed for proteomics as previously described using the 10 μl NB conditions (Section 3.2, Section 5.1.2).
For each replicate set there were ten conditions, the five life stages (rings, trophozoites, early schizonts, mature schizonts, merozoites) with FPP-N3 or DMSO treatment. Peptides from each condition were labelled with 10-plex TMT tags and combined (Fig. 5.3.1B). As each sample was prepared in triplicate, this gave three 10-plex TMT mixes. These were each split into 8 fractions by high pH fractionation and submitted for LCMS/MS analysis.

In total 1024 proteins were identified. Proteins were initially filtered as previously described (Section 3.2), then filtered again to consider only proteins identified in 3/3 replicates in at least one life stage, and from greater than 4 peptides, giving a final of 405 proteins. The log2 intensities were normalised by subtracting the median intensities of each replicate. Thirty-six proteins were significantly enriched in FPP-N3 vs DMSO-treated samples in at least one life stage (Fig. 5.3.2, student t-test, FDR = 0.05, s0
Eight were human proteins and are discussed in Chapter 6. Only two *P. falciparum* proteins were not SHs. A proteasome subunit (beta type-5, PF3D7_1011400) and an endonuclease (MUS81, PF3D7_1449400), were identified in all FPP-N3-dependent proteomic experiments and are probably covalently modified by the probe. Proteomic sub-units have been previously identified with FPP probes (Jessani et al., 2005). But the interaction with the endonuclease MUS81 is unknown. One unknown transmembrane protein, PF3D7_1117300, was actually significantly depleted in probe-treated mature schizonts, likely an artefact of the proteomic protocol. Of the other twenty-five proteins, three were PLPs (PF3D7_0209100, PF3D7_1358000 and PF3D7_0218600) and the rest were α/β-fold containing lysophospholipases, lipases, peptidases or unknown hydrolases.

By using intact parasites from multiple stages, we achieved the most in-depth profiling of active SHs thus far. We were surprised to detect PfXL2 (PF3D7_1001600), an exported lipase which we expected to be lost in the PV fraction-after saponin lysis, this has been shown to be non-essential in blood stages (Spillman et al., 2016a), the related PfXL1 was not detected (PF3D7_1001400). It is essential to note that this experiment was quantified using technical replicates, it would be desirable to repeat the experiment with biological replicates and this is discussed further below (Section 5.5). We now wanted to observe how individual protein activities change during the asexual cycle.

### 5.3.2 Levels of active SHs vary throughout the asexual life cycle.

The log2 enrichment of active protein at each life stage was used to generate a heat map of hydrolase activity over time (Fig. 5.3.2). The SHs were grouped by Euclidean clustering based on their temporal profile. The majority of SHs had 2- to 6-fold greater activity at some life stages than at others, with the lowest levels of active protein in ring stages. Notably PF3D7_1129300, PF3D7_1476800, LCAT and PF3D7_1134500 had most activity in schizont stages, with PF3D7_1458300, PF3D7_1358000 and abH112 being specifically highly active in early schizonts. SHs whose activity increased dramatically between mature schizont and merozoite stages were of particular interest such as PF3D7_1476700 (Psta1), PF3D7_1328500 (MLPL) and PF3D7_0403800 (S9C). The tightly regulated process of merozoite egress occurs in just 10-15 minutes after the C2 stall point (Collins et al., 2017). We hypothesised that these SHs that increased in enrichment in this time period may be specifically at the time of parasite egress and may play important roles in egress processes.
Figure 5.3.2: ABPP profiling of SHs throughout the *P. falciparum* asexual-blood stage life-cycle. 

(A) For each protein, at each life stage, Log$_2$ fold enrichment was calculated (Log$_2$(Intensity(FPP-N$_3$/DMSO))) and used to create a heat map showing Log$_2$ fold enrichment on a colour scale from -1 (green) to 6 (bright red). Proteins were clustered by Euclidean clustering. Proteins of interest discussed in the text are highlighted. * denotes the non-SHs, ** denotes SHs of PLP-type, all others are $\alpha/\beta$-fold hydrolases. 

(B) The log$_2$ enrichment of candidate proteins are plotted against life cycle. The bars represent the mean of three replicates, the replicate values are also indicated.
5.4 Four SHs are prioritised for genetic validation.

We now wished to integrate the data from both the proteomics studies with literature search results in order to identify candidates for genetic interrogation. We considered six proteins that had not been previously studied. LCAT, abH112 and abH114 were taken from the Malaria Box comparative proteomics screen, as potential targets of the Malaria Box compounds (Section 5.2). MLPL and Psta1 were selected for genetic interrogation from the life cycle ABPP proteomics due to their activation at merozoite stages. A close paralogue of Psta1, PF3D7_1476800 (Psta2), which showed most activity at schizont stages was also chosen. S9C was not included as it was already under investigation as a putative protease in our lab (Sophie Ridewood, personal communication). Two large genetic screening data sets were examined to predict essentiality: the \textit{P. berghei} based PlasmoGEM screen (Schwach et al., 2015); and the \textit{P. falciparum} piggyback transposon screen (Section 1.7.4; Zhang et al., 2018). Only LCAT and abH114 had orthologues in \textit{P. berghei}, (PBANKA\_1128100 and PBANKA\_0906000). \textit{Pb}abH114 was labelled as dispensable on PlasmoGEM however \textit{P. falciparum} abH114 was found to be essential in the transposon screen. abH114 activity peaked in early schizont stage with a second increase in merozoites (Fig. 5.3.2B). LCAT had a predicted function of a Phosphatidylcholine-sterol O-acyltransferase and had its greatest activity towards the end of the life cycle in schizonts/merozoites (Fig. 5.3.2B). \textit{Pb}LCAT was annotated as dispensable in the PlasmoGEM screen, however previous studies showed that knockout caused disruption of parasite egress from liver stage hepatocytes due to inefficient rupture of the parasitophorous vacuole (Burda et al., 2015). Although LCAT mutants had reduced fitness in the transposon screen, we considered it likely that LCAT had its main function in hepatocytes. As our focus was on asexual stage parasites, LCAT was not investigated further.

Proteins that did not have \textit{P. berghei} orthologues were not discounted as there are examples of differences in lipid metabolism pathways between rodent and human malarial parasites (Section 1.2.2.2). MLPL and abH112 were conserved among human parasites, and Psta1 and Psta2 were conserved just among other \textit{laverania} sub species, \textit{P. reichenowi} and \textit{P. gaboni}. abH112 was annotated as essential in the transposon screen, and its activity peaked in early schizont stage (Fig. 5.3.2B). Although Psta1 KO was only found to slow parasite replication, Psta2 was essential in the transposon screen. MLPL was found to be dispensable and was therefore discarded for follow up studies. Among this small selection of genes, it is possible to see how one must be careful when assuming the essentiality of proteins. For example, abH114 could have been discounted from further investigation due to \textit{Pb}abH114 being dispensable in the PlasmoGEM screen and abH112 or Psta2 may have been discounted due to their lack of conservation in other \textit{Plasmodium} species. However, all three proteins were found essential in the transposon screen. Therefore, we concluded that 4 SHs, Psta1, Psta2, abH112 and abH114 would be good targets for genetic interrogation.
5.5 Discussion

We have successfully developed a proteomic method in *P. falciparum* for the in-depth profiling of metabolic SHs. The results here described illustrate the utility of optimising standard proteomics protocols to maximise on one’s aims with the given targets, tools and instrumentation. First, we showed that FPP-N₃ specifically labelled more SHs than FPP-Biotin. We hypothesise that this is due not only to the higher reactivity of FPP-N₃, as indicated by the in-gel competition experiments, but also to the better quantification achieved when the CuAAC-chemistry step is included. A caveat of the MaxQuant LFQ algorithm is that relative changes in a protein’s intensity are quantified assuming that the majority of proteins are constant between different samples. A low level of unspecific labelling arising from the CuAAC-chemistry step may provide a common background between FPP-N₃- and DMSO-treated replicates that is not present in FPP-Biotin- and DMSO-treated replicates. There is no background of endogenously biotinylated proteins in *P. falciparum*. Acetyl-CoA carboxylase is the only protein predicted to be biotinylated, and it is not biotinylated in the asexual-blood stages (Dellibovi-Ragheb et al., 2018; Müller and Kappes, 2007). Therefore, although any extra step in the protocol introduces systematic errors, in this case the tandem tagging step contributes overall to the success of the approach. This illustrates the phenomenon that in chemical proteomics, “cleaner” sample preparation workflows are not always better. It could be that the previous proteomics experiments with W-dtBio-TMR-VS suffered from this same problem (Section 3.4). In summary, FPP-N₃ proved a better probe for chemical proteomics than FPP-Biotin. FPP-N₃ was shown to have the highest reactivity, and its tandem tagging functionality proved indirectly useful for protein quantification. FPP-N₃ is also suited for intact cell labelling, where FPP-Biotin is not. All of these considerations meant that FPP-N₃ was preferentially used in all further experiments.

The next step that was optimised was the affinity purification of labelled SHs. There is a technical limit to the amount of peptide material that can be analysed by the LC-MS/MS in each run. Our aim was to limit the proportion of neutravidin-derived peptides in the samples to be analysed by LC-MS/MS, while maximising the number of target proteins (SHs) identified. We therefore performed the affinity purification and subsequent on-bead digestion with different amounts of NBs, diluted with blank-agarose beads. We saw that while the raw intensities of neutravidin-derived peptides increased linearly with NB bead volume, SH-derived peptides did not. This suggested that the high proportion of neutravidin-derived peptides was masking the detection of SH-derived peptides. Using both iBAQ and TMT-based quantification, we looked at the total number of proteins and SHs identified, as well as the number of SHs significantly enriched in FPP-N₃- vs DMSO-treated samples. We showed that the highest number of targeted SHs was identified using NBs diluted 1:2 with blank-agarose, quantified with TMT labels (10 μl NB, 20 μl blank-agarose beads). Nineteen SHs were identified under these conditions. Seventeen of these were α/β-fold SHs, eight more than previously identified using FPP-N₃.
(Section 3.2) and now representing 40% of the predicted α/β-fold SHs in *P. falciparum*. The remaining two were PLPs, that had not been previously identified. In addition to increased SH identifications, the use of isobaric TMT-labelling allows samples to be multiplexed, allowing ABPP to be done in a higher-throughput format. Here we showed that by understanding the technical limits of the instrumentation, we could refine the workflow accordingly to achieve increased depth of coverage. Optimising the method in this way has advantages over the development of other elution methods. One alternative would be to use a probe with an integrated cleavage site, thus allowing proteins can be chemically cleaved from the NBs and trypsin digested without neutravidin contamination. However efficient elution methods like this mean the low-level background of proteins binding un-specifically to neutravidin is lost. This can have an impact on protein quantification in a similar way to as discussed above for FPP-Biotin. Since these experiments took place, other methods have been trailed such as derivatisation of NBs prior to affinity purification to limit digestion. NB derivatisation has proved partially successful and further optimisations are underway (Steven Howell, personal communication).

We then applied the optimised methods in comparative chemical proteomics to identify the targets of the Malaria box compounds, M438, M914 and M563. The clear and quantifiable inhibition of labelling observed using in-gel fluorescence suggested that detection by LC-MS/MS would be simple (Chapter 4). In the gel-based study we used 50 nM FPP-TMR for 10 min to achieve sub-saturating labelling conditions. We reasoned that more protein was required for MS-based quantification than direct in-gel fluorescence, so we used a higher probe concentration, 1 μM FPP-N3 for 10 min. In hindsight, this concentration was probably too high. The inhibitor concentrations used in this experiment were based on IC50 values calculated from competition with 50 nM FPP-TMR, much higher concentrations would probably be required to observe competition with 1 μM FPP-N3. With the addition of a high-pH fractionation step before LC-MS/MS analysis, twenty SHs were identified. Although the depth of coverage of the metabolic SHs was the best so far reported, the results for the target deconvolution of M438, M914 and M563 were disappointing. A number of proteins visually appeared to decrease in enrichment with increasing inhibitor concentration: abH112 with M563; abH114 with M914 and M438; and LCAT with all three compounds. However only changes in active LCAT levels with the highest concentration of M438 were statistically significant. Further optimisation is required therefore to find conditions where the reversible inhibitors can be studied by ABPP-chemical proteomics methods. Other target deconvolution methods could also be employed. IVEWGA approaches have already been performed on the Malaria Box with varying results (Corey et al., 2016; Istvan et al., 2017). Label-free methods such as CESTA could also be used, this would predict target engagement without bias, so if non-SH targets exist they could also be detected.
Finally, we aimed to profile the activity of SHs throughout the erythrocytic cycle. Five time points were used representing rings, trophozoites, early schizonts, mature schizonts (C2-arrested), and merozoites. Six additional SHs were identified compared to the previous experiments, totalling twenty-three α/β-fold SHs and three PLPs. This now represents over half of the predicted α/β-fold SHs and 3 out of 4 predicted PLPs. The remaining hydrolases may be reasonably accounted for. First, not all predicted hydrolases may be expressed and active in the asexual-blood stages. Secondly, all SHs exported into the PV or RBC lumens will likely be lost after saponin lysis. Although these proteins may be visualised by running a labelled saponin fraction on a gel, the high levels of Hb present would make them difficult to identified by proteomics. This is also the reason why it is difficult to include an uninfected RBC control sample. Other methods of RBC lysis could be considered to retain exported PV proteins. The bacterial pore-forming protein Streptolysin-O (Bhakdi et al., 1985) can be used as an alternative to detergents to selectively permeabilise the RBC membrane while keeping the PVM intact (Ansorge et al., 1997). As well as increasing the number of SHs identified, access to the exported proteins would allow us to profile changes in SH activity while controlling for (or tracking) changes in sub-cellular localisation. Although it is sensible to assume that a SH is only active when it is at its final location, currently we cannot ensure that a decrease in SH activity is not due to the time-dependent export of that protein.

We used the enrichment of proteins in probe-treated samples from each life stage to produce a heat map showing the time-dependent activity of all the identified SHs. The main limitation of this study is the lack of biological replicates. In order to represent the full *Plasmodium* life cycle, many different parasite preparations were pooled. For each life stage, parasites that were synchronised and labelled on different days using different batches of blood were pooled. The combined parasite material for each life stage was split into three immediately after soluble protein was extracted, providing replicates for proteomics. Ideally, we would repeat the entire experiment at least twice more. This aim is hampered, even more so than with the Malaria Box screen, by the time and labour required to generate the parasite material. To partly counteract the lack of biological replicates, we have been stringent in the filtering methods, requiring proteins to be identified from at least four peptides.

There are a number of caveats in comparing the proteomics results from different parasite life stages. Although parasites are washed with PBSa after saponin-lysis of RBC and PV membranes, some membrane debris is probably present in all samples. At ring-stage when parasites are small compared to the RBC, parasite preparations probably contain a lower ratio of parasite-derived material to membrane debris than at other stages. This may skew the results from this life stage, leading to underestimation of SH intensities. The heat map shows that all proteins are least active in ring stages, supporting this idea. Every SH showed some change in activity across the life cycle. A group of proteins...
are equally enriched in trophozoite-merozoite stages (PF3D7_1401500, PfPARE (PF3D7_0709700), PF3D7_080500 and PF3D7_0301300). If ring enrichment values are underestimated, these SHs may be equally active in all life stages. The heat map should be of aid to any future studies of *P. falciparum* metabolic SHs. The activity profiles could also be integrated with transcriptomics and whole proteomics data to help build a full picture of the expression, translation and activation of SHs. The development of further genetic, chemical and immunochemical tools directed towards specific SHs will be able to corroborate the results reported here. For example, S9C has been confirmed by genetic and probe methods to be specifically expressed and active in merozoites as was predicted here (personal communication, Dr Sophie Ridewood). Recently, another group has also attempted to profile SHs in *P. falciparum*. They used FPP-desthiobiotin to label lysates from saponin-lysed schizonts and analysed both the soluble and insoluble fractions after protein extraction (Elahi et al., 2019b). Twenty-One SHs were significantly enriched in FPP-Biotin- vs DMSO-treated samples. This included two extra proteins: the exported epoxide hydrolase PfEH2; and PF3D7_1126600. Both of these were present in the FPP-desthiobiotin insoluble fraction and not identified in our FPP-N3 ABPP experiments.

We have used the combined results of the Malaria Box and life-cycle proteomics studies to prioritise four genes for genetic knockout: abH114, which may be the common target of M438 and M914; abH112, which may be one of the targets of M563; Psta1, which is activated in merozoite stages; and Psta2, a paralogue of Psta1. The genetic interrogation of these SHs is described below (Chapter 7). In addition to the *Pf*SHs discussed we also identified and quantified eight human enzymes. The differences in the presence of active *Hs*SHs enzymes throughout the parasite life cycle is intriguing. If these host SHs are co-opted by the parasite, they may present novel resistant-proof targets for malaria drug-discovery.
Chapter 6. **Results IV: Human SHs may be potential antimalarial targets.**

Metabolic SHs in *P. falciparum* are mostly uncharacterised and many have low sequence similarity to SHs in other organisms. We hypothesise that this family may be a source of novel antimalarial targets. The annotation of the metabolic SHs will also contribute to the understanding of different aspects of parasite cell biology such as lipid metabolism. We have reported the first large-scale, functional profiling of *P. falciparum* metabolic SH activity throughout the asexual blood-stages (Section 5.3). Twenty-Five *P. falciparum* SHs were identified in the 5-point life stage proteomics study. Many were confirmed as SHs for the first time, and we were able to profile their life-stage specific activity (Section 5.3). Eight human SHs were also identified in this study. Much more is known about these HsSHs, and many play important roles in human health and disease (Long and Cravatt, 2011). But it is unknown what role, if any, they play in a *Plasmodium*-infected RBC. When RBCs are infected by *P. falciparum* asexual stage parasites, the RBC is remodelled and some host enzymes and pathways are co-opted by the parasite (Section 1.2.2.1). Exploring the interaction between parasite and host derived SHs could provide a new source of antimalarial targets. Human antimalarial targets would be more impervious to drug resistance, as mutation of the drug target is not possible. Other drug resistance mechanisms would still be possible however. Existing HsSH inhibitors could also be easily repurposed. In this chapter, we use the results of the large-scale SH profiling to identify and interrogate human SHs that are active in iRBCs.

### 6.1 ABPP showed changing levels of active human SHs throughout the *P. falciparum* asexual blood-stage life cycle.

We wanted to examine the human SHs detected in the asexual life cycle proteomics study to discover if they play a role during *P. falciparum* infection. In addition to the twenty-five *Pf*SHs (Section 5.3), eight human proteins were significantly enriched in FPP-N₃- vs DMSO-treated samples (student t-test, s₀=0.5, FDR=0.01). All eight were known SHs (Table 6). The saponin lysis of iRBCs during sample preparation should have precluded the detection of RBC SHs (Section 5.3). Some RBC membrane debris is expected to remain in the parasite pellet after saponin lysis and there may be debris in merozoite preparations also. This could explain the presence of membrane associated SHs such as Acetylcholine esterase (AChE) (Ott, 1985). It is possible that cytoplasmic SHs could be imported into the parasite as has been observed for other RBC metabolic enzymes such as ALAD and hPrx-2 (Section 1.2.2.1; Bonday et al., 2000; Koncarevic et al., 2009).
In order to determine the relationship of the HsSHs to P. falciparum-infection we examined their activities throughout the asexual-blood stages. The RBC performs no de novo protein synthesis. Therefore any change in human SH activation levels reflects either a change in active state, or a change in localisation of an active protein (for example, import into the parasite).

The human SH activities were processed in the same way as the plasmodium proteins (Section 5.3) and a heat map of life stage activities was produced (Fig. 6.1). All the SHs showed some life stage dependent activation (Fig. 6.1). Platelet activating factor acetyl hydrolase 1B gamma (PAFAH1B3) and Acyl-amino acid releasing enzyme (APEH) were activated at merozoite stage (Fig. 6.1). Both Acyl protein thioesterases (LYPLA1 and LYPLA2) and fatty acid synthase (FASN) were most active at early life stages in rings/trophozoites (Fig. 6.1). AChE, Neut ral cholesterol esterase (NCEH1) and Neuropathy target esterase (PNPLA6) were most active in trophozoite and schizont stages (Fig. 6.1). These results suggest that the enzymes may have specific roles to play in parasite growth and replication. Many of the human SHs has been extensively characterised. This meant existing inhibitors for each could be used to test the importance of their activity in asexual-stage parasites.

![Figure 6.1: Asexual life cycle proteomics results, Homo sapiens. A) For each protein, at each life stage, Log2 fold enrichment was calculated (Log2(Intensity(FP-N3/DMSO))) and used to create a heat map showing Log2 fold enrichment on a colour scale from -1 (green) to 6 (bright red). Proteins were clustered by Euclidian clustering.](image-url)
## Table 6 Human SHs and their inhibitors

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Name</th>
<th>MW/KDa</th>
<th>Function</th>
<th>Notes</th>
<th>Plasmodium</th>
<th>Inhibitor</th>
<th>I&lt;sub&gt;IC50&lt;/sub&gt; / nM</th>
<th>PEC&lt;sub&gt;50&lt;/sub&gt; / μM</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAFAH1B3</td>
<td>Platelet-activating factor acetyl hydrolase 1B gamma</td>
<td>25.7</td>
<td>Phospholipase A2 activity (Hattori and Arai, 2015)</td>
<td>Hydrolyses an acetyl group from the glycerol backbone of platelet-activating factor and circulating aspirin in erythrocytes (Hattori and Arai, 2015; Kohnz et al., 2015). 30% homology to PF3D7_0321500 peptidase.</td>
<td>P11 (Chang et al., 2015) Reversible</td>
<td>880</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>FASN</td>
<td>Fatty acid synthase</td>
<td>273</td>
<td>Multi-domain fatty acid synthase (Chang and Hammes, 1990)</td>
<td>The malonyl-CoA-/acetyl-CoA-ACP-transacylase (MAT) and thioesterase (TE) domains use nucleophilic serines. (Bunkoczi et al., 2009; Pemble et al., 2007). No FASN1, FASN2 not essential in blood stages (Vaughan et al., 2009).</td>
<td>TBV-3166 (Heuer et al., 2016) Reversible</td>
<td>42 (Ventura et al., 2015)</td>
<td>14 ± 5</td>
<td></td>
</tr>
<tr>
<td>PNPLA6</td>
<td>Neuropathy target esterase</td>
<td>143-150</td>
<td>Patatin-like phospholipase, (Kenescherger et al., 2009)</td>
<td>Membrane associated (Tienhoven et al., 2002). Plays a critical role in the mammalian brain (Richardson et al., 2013) via hydrolysis of phosphotidyl choline (Qustad et al., 2003).</td>
<td>-</td>
<td>TOCP (Davis and Richardson, 1980; Richardson et al., 2013) Irreversible</td>
<td>150 (Veronesi et al., 1991)</td>
<td>7 ± 1</td>
</tr>
<tr>
<td>LYPLA1 (APT1)</td>
<td>Acyl protein thioesterase</td>
<td>26</td>
<td>Depalmitoylase (Duncan and Gilman, 1998) / lysophospholipases (Sugimoto et al., 1996)</td>
<td>APT1/APT2 share 60% identity and some substrates.</td>
<td>ML348 (Won et al., 2016) Reversible</td>
<td>840 (Adibekian et al., 2010a)</td>
<td>1.7 ± 0.2</td>
<td></td>
</tr>
<tr>
<td>LYPLA2 (APT2)</td>
<td>Acyl protein thioesterase</td>
<td>24.7</td>
<td></td>
<td></td>
<td>ML349 (Won et al., 2016) Reversible</td>
<td>510 (Adibekian et al., 2010b)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>NCEH1 (KIAA1363/AADAACL1)</td>
<td>Neutral cholesterol esterase</td>
<td>31-46.6</td>
<td>Esterase</td>
<td>Membrane glycoprotein, Hydrolysis of cholesterol esters (Sekiya et al., 2011) and 2-acetyl-monoa{llyl}glycerol ethers (Buchebner et al., 2010; Chiang et al., 2006).</td>
<td>-</td>
<td>JW480 (Chang et al., 2011) Irreversible</td>
<td>12</td>
<td>-</td>
</tr>
<tr>
<td>APEH</td>
<td>Acyl-amino acid releasing enzyme</td>
<td>81.2</td>
<td>Hydrolyses acyl-amino acids from peptides (Tsunasawa et al, 1975)</td>
<td>Adherent to oxidized erythrocyte membranes and preferentially degrades oxidatively damaged proteins.(Fujino et al., 2000). 30% identity to plasmodium peptidases.</td>
<td>AA74-1 (Adibekian et al., 2011) Irreversible</td>
<td>5 (Adibekian et al., 2011)</td>
<td>0.20 ± 0.01</td>
<td></td>
</tr>
</tbody>
</table>
6.2 Inhibitors of human SH’s block parasite replication

6.2.1 Human SH inhibitors block parasite replication

To test our hypothesis that human SHs play a role in the asexual blood-stage of *P. falciparum* we wanted to see if parasite growth and replication could be disrupted by their inhibition. Previously characterised, commercially-available inhibitors were obtained for each *HsSH* (Table 6). The majority of inhibitors were described as highly specific and had a mix of reversible and irreversible reaction mechanisms.

Five of the inhibitors have been shown to be highly selective in mammalian systems using FPP probe-based ABPP studies: P11 selectively and reversibly inhibits PAFAH1B3 and also the related PAFAH1B2 (Chang et al., 2015); ML438 and ML439 selectively and reversibly inhibit LYPLA1 and LYPLA2 respectively (Adibekian et al., 2012, 2010a, 2010b); JW480 is a highly potent and selective carbamate inhibitor of NCEH1, presumed irreversible (Chang et al., 2011); and AA74-1 selectively inhibits APEH (Adibekian et al., 2011).

TVB-3166 was developed by 3-V Biosciences (Huesca et al., 2005) reversibly inhibits FASN via the keto-reductase unit. This disrupts palmitate synthesis *in vivo* and on target selectively was demonstrated by rescue with exogenous palmitate (Heuer et al., 2016; Ventura et al., 2015). TOCP is a organophosphosphosphate (OP), this toxic class of compounds is responsible for human OP neuropathy. Human OP neuropathy is caused by OP inhibition of SHs in the brain. TOCP has been shown to be selective for PNPLA6 over AChE (Johnson, 1975, 1969; Veronesi et al., 1991). Other OPs target AChE specifically. The carbamate Pyridostigmine bromide (PyBr) can protect from AChE-dependent OP poisoning by reversible inhibition of AChE and has been shown to target the erythrocyte variant (Dawson, 1994; Herkert et al., 2011; Koster, 1946). Though their primary human targets are known, TOCP and PyBr have the least evidence of selectivity. It is important to note that despite the reported on-target specificities of these inhibitors in mammalian systems, their selectivity in *P. falciparum* is unknown.

Synchronous, ring stage parasites at 1% parasitaemia and 0.5% haematocrit, were treated with DMSO or different concentrations of the eight *HsSH* inhibitors in a 1/4 serial dilution. After 72 h in standard growing conditions, the parasites were fixed, and parasite nuclei in iRBC stained with Hoechst. The number of iRBC/uRBC (infected/uninfected RBCs) were counted by flow cytometry, and the parasitaemia was calculated. This was done in triplicate. The parasitaemia was plotted against inhibitor concentration and the data was fitted to a standard dose response curve in order to calculate EC_{50} and
standard errors. For more details on parasite synchronisation (Section 9.1) and flow cytometry (Section 9.3.5), see Materials and Methods.

Figure 6.2.1: Standard replication assay with human SH inhibitors. A standard 72-h growth assay was carried out in the presence of different concentrations of inhibitor. The title of each graph indicated the target human SH and the inhibitor used is indicated on the X-axis. After 72 h, the parasites were fixed and stained with Hoechst. The parasitaemia was quantified by flow cytometry and fitted to a dose response curve to determine EC$_{50}$. Error bars represent the SEM (standard mean error) of three replicates.

The inhibitors of LYPLA1, APEH, PNPLA6, FASN and AChE were all found to affect parasite replication in a dose dependent manner after 72 h (Fig. 6.2.1, Table 6). In all cases the EC$_{50}$ was less than the reported IC$_{50}$ for each compound (Table 6). The inhibitors of the other three human SHs had no effect, including (surprisingly) the LYPLA2 inhibitor M349 (data not shown). LYPLA2 and LYPLA1 have been shown to have the same function and share 60% identity, although different
substrates have been reported *in vitro* (Tomatis et al., 2010). These results supported our initial hypothesis that some of these host SHs may be co-opted by the parasite and are essential for growth and replication. We decided to focus on the inhibitors of APEH and LYPLA1, which gave the lowest EC₅₀s and were also the only inhibitors to completely block replication at high concentrations (Table 6, Fig. 6.2.1).

**6.2.2 LYPLA1 and APEH inhibitors block parasite replication in the first cycle**

In order to better understand the role of LYPLA1 and APEH in parasite replication, we wanted to determine at which life stage the parasite is affected by their inhibition (Table 6). A new flow cytometry-based assay had been developed in our lab in order to closely monitor the effect of small molecules on parasite growth and replication (Section 9.5.1; Bell et al, in preparation). In addition to Hoechst staining of DNA, fixed parasites are stained with an RNA dye (132A) (Cervantes et al., 2009). Infected RBCs can be classed into life stages based on the DNA/RNA content measured by flow cytometry. Ring-stage parasite populations have a signal from DNA and little RNA content, but as parasites progress into trophozoite stage, the RNA signal increases. Schizont populations have high levels of RNA and DNA resulting from the 8-32 nuclei after DNA replication. For this experiment, growth assays are set up in the presence of inhibitors (and controls) and time points are taken at different intervals after treatment. Analysis of each time point allows the monitoring of parasite growth and replication in comparison to a DMSO-treated control. In this way, the stage at which parasite growth is affected can be ascertained. Two control compounds were used to benchmark specific inhibition phenotypes: CQ should kills parasites immediately if treated at ring stage; and C2 blocks egress, arresting parasites in schizogony.

A standard growth assay was set up, as described (Section 6.2.1). Parasites were treated at ring stage (0 hpi), in triplicate, with 5 μM ML-348 or 0.625 μM AA71-1 (approximately 3 times the EC₅₀). Three controls were used: DMSO; C2 at 1 μM, sufficient for stall of egress (Collins et al., 2013b); or an excess of CQ (2 μM). Samples were taken and fixed at four different time points: 24, 41, 48 or 72 hpi. Fixed parasites from each time point were stained with Hoechst and 132A dye at 2 μM. Using flow cytometry, the UVA fluorescence (355 nm, filter 450/50) was used to determine DNA content and the blue fluorescence (488 nm, 610/20 nm) was used to determine RNA content. The life stage of populations on the flow cytometric plots could be assigned based on the DNA/RNA content: uRBC have low RNA and low DNA content, rings have low RNA and low DNA (one nuclei), trophozoites have higher RNA and low DNA (one nuclei), and schizonts have high RNA and higher DNA (greater than one nuclei). This strategy is illustrated by a growth assay monitoring parasite DNA and RNA and every 4 hours (Fig. 6.2.2). Using these assignments, parasites from the first cycle (0-48 hpi) could be distinguished from parasites from the invasion of new RBCs (48-72 hpi) (Fig. 6.2.3). The mean parasitaemia, RNA
and DNA content from each parasite population (cycle 1 or cycle 2) was quantified and the triplicate values plotted for each time point (Fig. 6.2.4, Supplementary Table 10.4).

**Figure 6.2.2: Flow cytometry parasite life stage grouping.** A growth assay was carried out and every 4 hours parasites were fixed parasites were fixed and stained with Hoechst and the RNA dye 132-A. The infected/uninfected RBCs counted by flow cytometry and their DNA and RNA content detected. On a graph of RNA over DNA content the parasite populations are gated to assign their life stage and can be followed throughout a cycle. Uninfected RBCs (uRBCs) are labelled in green, cycle 1 parasites in blue and cycle 2 parasites in pink.
DMSO-treated samples replicated as expected - some parasites developed faster than others, so some cycle 1 schizonts had already eggressed by 41 hpi, and cycle 2 rings could be observed. CQ-treated samples did not progress past ring stage, and C2-treated samples progressed to schizont stage where they were arrested and started to deteriorate (Fig. 6.2.3).

ML348-treated parasites appeared to slow midway through the life cycle, at 41 hpi parasites were still trophozoite/early schizont stage compared to the late schizonts seen in the DMSO (Fig. 6.2.3). At 48 hpi, almost all DMSO schizonts had eggressed and new rings could be seen, but ML348 parasites were stalled in schizogony similar to C2-treated parasites (Fig. 6.2.3). At 72 hpi, the mature schizonts looked like they started to deteriorate like C2-treated parasites; there is a small population of ring-type parasites which could be from schizonts which did manage to egress, or simply dead parasites (Fig. 6.2.3). We previously showed that LYPLA1 was most active in ring and trophozoites (Fig. 6.1), taken together these data suggest that if LYPLA1 is the main target of ML438 then it may be important in metabolism in trophozoite stage that is required for later successful schizogony and egress. AA74-1-treated parasites appeared to be blocked before 41 hpi, in the same way as CQ-treated parasites (Fig. 6.2.3). APEH was found to be most active in merozoites (Fig. 6.1), this may suggest that it has a role in the same way as *P. falciparum* S9C, being packaged into merozoites ready for an essential role in the metabolism of the growing ring parasite (Ridewood et al, in preparation). Thus, our hypothesis that human SHs are essential in asexual replication of *P. falciparum* may be true for at least two of the SHs detected by our unbiased SH profiling.
Figure 6.2.3: Replication assay to monitor the effect of human SH inhibitors on parasite growth and replication. A growth assay was carried out in the presence of different concentrations of inhibitor or DMSO. Parasites were treated at ring stage (0 hpi). After 24, 41, 48 and 72 hpi, the parasites were fixed and stained with Hoechst and the RNA dye 132-A. The infected/uninfected RBCs counted by flow cytometry and their DNA and RNA content detected. On a graph of RNA over DNA content the parasite populations are gated to assign their life stage. Uninfected RBCs (uRBCs) are labelled in green, cycle 1 parasites in blue and cycle 2 parasites in pink. This experiment was performed in triplicate, each graph shown was representative of three replicates, apart from for AA74-1 which only had two replicates.
Eight human SHs were quantified by the ABPP proteomics screen across the asexual parasite life cycle. Unlike the *Pf* SHs, much is known about these mammalian enzymes. We were surprised to see the changes in enrichment of active *Hs* SHs throughout the life cycle. It is important to consider that

**Figure 6.2.4: Quantification of replication assay with *HsSH* inhibitors and controls.** Newly invaded rings (0 hrs) were treated with human SH inhibitors (ML-348 or AA74-2) or controls (DMSO, CQ, C2). Parasites were fixed 24, 41, 48 and 72 h after treatment. Fixed parasites were incubated with Hoechst and 132A to stain DNA and RNA respectively. Flow cytometry was used to determine the parasitaemia, RNA and DNA content of parasites and assign them as cycle 1 (rings-schizonts from cycle of drug treatment in blue) or cycle 2 (newly invaded rings in pink). The mean RNA and DNA of parasite populations was normalised by dividing the RNA or DNA content of uRBCs. The data for each time point were plot using prism. Bar graphs represent the median of three replicates and error bars show the standard error. Statistical significance was determined by unpaired one-way ANOVA. Significance levels are indicated: p < 0.0001, ****; p ≤ 0.001, ***; p ≤ 0.01, **; p ≤ 0.05, *; p < 0.05, ns. Significance values are given in Supplementary Table 10.4.

### 6.3 Discussion

Eight human SHs were quantified by the ABPP proteomics screen across the asexual parasite life cycle. Unlike the *Pf* SHs, much is known about these mammalian enzymes. We were surprised to see the changes in enrichment of active *Hs* SHs throughout the life cycle. It is important to consider that
different ratios of parasite-derived material to RBC debris may be present in lysates from each life stage. However, if the HsSHs enrichment only varied for this reason then the activity profiles shown in the heat map would be the same for each life stage, which they are not. PNPLA6, AChE and NCEH1 are associated with the RBC membrane, so their presence may be due to RBC debris only (Ott, 1985; Sekiya et al., 2011; Tienhoven et al., 2002). In support of this, PNPLA6, AChE and NCEH1 all have similar activity profiles, although they are most enriched in trophozoite and schizont-stage parasites, not ring-stages as might be expected. The different activity profiles in the other HsSHs suggest that these are activated or internalised by the parasites in some way during the life cycle.

*Plasmodium* species are parasitic and depend on their host for survival throughout the life complex cycle. In asexual-blood stages the residency of parasites inside RBCs provides resources essential for growth and replication. Erythrocytic disorders can confer protection from malaria to the host (Section 1.1.1). The study of human enzymes and how they interact with, and are co-opted by, the parasite is essential to building a full picture of asexual-stage parasite biology. There are numerous examples of the parasite’s coevolution with its host, adapting to employ human enzymes for its own purposes. This practice reduces the time and energy the parasite has to spend in *de novo* protein synthesis. Human sigma-aminolevulinate dehydratase (HsALAD) has been shown to be imported into the parasite, by a hypothesised phagocytosis mechanism (Bonday et al., 2000). It has been hypothesised that *HsALAD* and other host enzymes are hijacked by the parasite for haem biosynthesis in asexual-blood stages (Bonday et al., 1997; Nagaraj et al., 2013; van Dooren et al., 2012). The parasite-encoded haem biosynthesis pathway has been shown to be dispensable (Nagaraj et al., 2013; van Dooren et al., 2012). Microscopy methods have shown human peridoxin 2 (HshPrx-2) to be imported into the blood-stage parasite cytosol and also in Maurer’s clefts (Koncarevic et al., 2009). *HshPrx-2* accounts for around 50% of overall thioredoxin peroxidase activity in the parasite, suggesting that it is essential in this pathway (Koncarevic et al., 2009). Other internalised host enzymes include ferrochelatase (Varadharajan et al., 2004), catalase (Clarebout et al., 1998), and superoxide dismutase (Fairfield et al., 1983) (also this is disputed (Dive et al., 2003)).

If host enzymes are viable antimalarial drug targets this would provide exciting opportunities for drug discovery. If a drug targets a human enzyme, the parasite cannot directly evolve resistance to that drug. Although resistance can be conferred by other mechanisms, for example via mutations in parasite transporters (Sanchez et al., 2010; Valderramos and Fidock, 2006), it is hoped that the time taken for resistance to develop would be longer. Human enzymes are much more likely than parasite enzymes to have crystal structures and inhibitory chemotypes already available. This was demonstrated by the commercial availability of inhibitors to the *HsSHs*. Some of these inhibitors have already been considered as drugs for other conditions. TBV-3166 inhibits palmitate synthesis in cancer cell lines (Ventura et al., 2015), it is orally available and has been investigated as an anti-cancer agent (Heuer et
al., 2016). PyBr has already been used as OP poisoning prophylaxis (Dawson, 1994; Koster, 1946). The repurposing of drugs with previously approved toxicology profiles towards malaria would greatly decrease drug development costs.

The two most promising HsHS targets were APEH and LYPLA1. As well as inhibitors, antibodies are available for the HsSHs. The next priority would be to confirm the presence and location of the HsSHs by western blot and Immuno-fluorescence assay (IFA). The use of antibodies in combination with FPP-TMR could identify the bands corresponding to APEH and LYPLA1. The selective inhibition of HsSHs could then be confirmed by using the inhibitors in a gel-based ABPP assay with FPP-TMR. By combining the results of this thesis with previously reported studies of the HsSHs, we can start to hypothesise what their role in parasites may be. Active LYPLA1 was shown to be most enriched in ring and trophozoite stages, the profile for LYPLA2 was similar although it was less enriched overall. The acyl protein thioesterases LYPLA1 and LYPLA2 (sometimes called APT1 and APT2), share 60% structural identity. Although they share functionality and some substrates they display distinct reactivity (Long and Cravatt, 2011). The inhibition of parasite growth in the presence of ML-348 but not ML-349 suggests that LYPLA1 but not LYPLA2 is important in iRBCs. They are involved in palmitoylation/depalmitoylation, although they were originally thought to be lysosphospholipases (Sugimoto et al., 1996). Dual inhibition by palmostatin B causes disruption of Ras signalling in cells leading to APT1/2 being investigated as targets for cancer therapy (Dekker et al., 2010, p. 1). Post-translational palmitoylation (both dynamic and stable) of proteins in *P. falciparum* has been shown to be essential in asexual-blood stage growth and replication (Corvi et al., 2012; Corvi and Turowski, 2019; Jones et al., 2012). There are 12 predicted palmitoyltransferases encoded in *P. falciparum* but no validated parasite depalmitoylases. Perhaps, LYPLA1 plays this role in iRBC. We have shown that ML-348 causes a slow in parasite growth and block in egress similar to inhibition by C2. This is surprising if LYPLA1 has its effect in ring/trophozoite stages. It would be very interesting to repeat the experiment, washing off the ML-348 to see if, like C2, parasites are then able to egress. This would show if the block in egress is a halt or if inhibition throughout the early life stages has led to an aberrant schizont that is incapable of efficient egress. Further experiments are required to demonstrate that ML-438 is inhibiting LYPLA1 and that it does not have other parasite derived targets. To investigate the role of LYPLA1 in palmitoylation we could employ metabolic-tagging chemical proteomics to profile the parasite palmitoylome in the presence of LYPLA inhibitors (Jones et al., 2012).

APEH (previously known as ACPH/OHP) catalyses the hydrolysis of N-terminal acyl-amino acids from small peptides (Tsunasawa et al., 1975). APEH has been shown to respond to oxidative stress in erythrocytes, adhering to the membrane and preferentially degrading oxidatively damaged proteins (Fujino et al., 2000). In this study, active APEH was most enriched in merozoite stages, although it was present throughout the parasite life cycle. When treated with the APEH inhibitor, AA74-1, parasite
growth and replication was blocked from ring stages, phenocopying CQ treatment. These results suggest that APEH may be important throughout the parasite life cycle, including in the initial stages of ring development. Endocytic import processes are thought to be initiated in ring-stage parasites 12-24 hrs after invasion (Bakar et al., 2010), although they are most active in trophozoites (Hanssen et al., 2012). We hypothesise that the parasite may have adapted to package APEH in merozoites so that it is present when needed immediately after new RBC invasion. After the completion of our experiments it was shown that the internalisation of APEH is essential for replication of *P. falciparum* in asexual-blood stages (Elahi et al., 2019a). They observed that the effect of AA74-1 on parasites in the first cycle was moderate (EC$_{50} > 1$ μM), with a much more striking block in parasite growth occurring in the second cycle (EC$_{50} = 0.1$ μM) (Elahi et al., 2019a). This is consistent with the EC$_{50} = 0.2$ μM after 72-h treatment reported here, although we were also able to show that 0.625 μM AA74-1 was sufficient to completely block parasite growth in the first cycle. They showed that APEH was present in the parasite ring stages and that late treatment with AA74-1 in the first cycle was sufficient to cause growth defects in the second cycle (Elahi et al., 2019a). They predict that the increase in potency observed in the second cycle is due to the inhibition of APEH in RBCs before invasion. This could be tested by pre-inhibiting RBCs prior to infection with *P. falciparum* and observing the effect on invasion and parasite growth. They hypothesise that APEH is internalised by endocytosis into the food vacuole, where it contributes to erythrocyte protein digestion by cleaving N-acylated amino acids from peptides generated by endopeptidases. However, their results are also consistent with our hypothesis that APEH has an essential role early in parasite development that is provided for by protein already packaged into merozoites. APEH activity is likely to be essential for *P. falciparum* asexual life stage growth, but further work is required, especially in merozoite and early-ring stages, to determine its functional role in the parasite. These results present novel and exciting avenues for anti-parasitic target discovery among host SHs.
Chapter 7. Results V: Genetic interrogation of SHs in *P. falciparum*

Although the *P. falciparum* genome sequence was published in 2002 (Gardner et al., 2002), many protein-coding genes are still unannotated. Metabolic SHs have diverse roles in cell biology and robust chemical tools are available to study them (Section 1.5.3). In *P. falciparum*, forty-three proposed metabolic SHs can be identified via their common α/β-fold (Table 1). We have confirmed the identity of many of these using FPP probes, however their roles in the parasite are still unknown (Chapter 5). We have used ABPP to identify four SHs as interesting candidates for further study (Chapter 5). Two of these candidates were identified as the possible targets of Malaria box anti-parasitic compounds (Section 5.2) and two were chosen from the life cycle proteomics study (Section 5.3). They have all been predicted to be essential (Psta2, abH112, abH114) or important (Psta1) for parasite growth and replication in a recent *P. falciparum* genome-wide screen (Zhang et al., 2018).

In this chapter, we outline a conditional allelic replacement system designed to test the essentiality of the genes of interest (GOIs) and describe the effect of catalytic site mutation on parasite growth and replication. Conditional knock out (cKO) strategies involve the tight temporal control of excision of a gene, induced by a chemical stimulus. One may then monitor in real-time the effect of the loss of the gene product on the cell or organism in question. Drawbacks to this technique include the difficulty of teasing out the catalytic function of a protein from the effects caused by loss of protein-protein interactions. The GOIs had all been labelled as SHs via their reaction with the FPP probe, although the proteomics study did not allow for confirmation of the active site serine as the probe modification site (Chapter 5). We wished to determine the function of their enzymatic activity in the parasite by mutating the reactive serine in the active site of our SHs of interest to an alanine. We chose to use a system of conditional allelic replacement. In this approach, chemically activated excision of the wild-type (WT) gene causes a second, mutated version of the gene to be expressed. By this method we would not detect essentiality of a protein if their role is unrelated to their predicted SH function, as the catalytically dead mutant would be still be able to perform non-hydrolytic functions. This method also enables an elegant internal control by conditionally replacing the WT gene with a second WT copy.
7.1 Generation of conditional mutant parasite lines.

7.1.1 A conditional allelic replacement strategy to test the essentiality of active Psta1, Psta2, abH112 and abH114.

In order to test the essentiality of the GOIs we designed a conditional system of allelic replacement. Recent advances in the genetic manipulation of *P. falciparum* have allowed the modification of genes to be performed in weeks rather than months (Section 1.7.2). We aimed to generate the conditional lines quickly and efficiently by employing single crossover recombination and two methods of selection.

The integration construct contained a human dihydrofolate reductase (*hdhfr*) selection cassette on the plasmid backbone that confers resistance to the drug WR99210 (WR). This allows selection for parasites that retain the plasmid. We also used selection-linked integration (SLI, Section 1.7.2.2), achieved using the viral T2A peptide (Kim et al., 2011) to attach the neomycin phosphotransferase gene (*npt*) to the GOI. When the integration construct is integrated into the chromosome the GOI and npt will be transcribed as one long mRNA, linked by T2A. The T2A causes ribosome skipping during translation so that two proteins are transcribed from the same mRNA: the protein of interest and NPT. *npt* confers resistance to the drug G418, only when the plasmid has been integrated into the genome. In this way chromosomal integration can be selected for by G418 treatment (Section 1.7.2.2; Birnbaum et al., 2017). SLI allows for integration positive parasites to be detected within weeks, which is quicker than using cycles of WR alone and comparable to other methods of integration, such as CRISPR/Cas9 (Section 1.7.2.1). Another benefit of the SLI is that the pre-clonal parasite lines should be close to 100% integrated. This allows the functionality of the GOI to be tested before clonal lines are established for follow-up studies. NPT is expressed under the control of the endogenous promoter, so the success of SLI relies on sufficient expression of the GOI. However, as all our SHs have been detected by chemical proteomics, we expected that it should be sufficient to achieve G418 resistance.
The DiCre cKO system allows rapid excision of chromosomal DNA between two \textit{loxP} sites upon the addition of rapamycin (Collins et al., 2013b; Jullien et al., 2003). A \textit{loxP} flanked version of the GOI can be integrated into the endogenous locus by single crossover recombination. To limit disruption of the expressed protein, \textit{loxP} sites are placed within an artificial intron (loxPint; Matera and Wang, 2014; Pieperhoff et al., 2014). A DiCre-expressing line of \textit{P.falciparum}, “B11-DiCre” was used (Perrin et al., 2018). The construct backbone contained a 3’ UTR from the \textit{P. berghei} DHFR to allow for efficient expression, as the endogenous 3’ UTR is displaced after integration (Fig. 7.1.1). Homology regions

\textbf{Figure 7.1.1: Strategy for conditional allelic replacement.} The p2TA-cMUT construct has a \textit{hdhfr} cassette on the backbone. The targeting section starts with the homology region (HR) which targets the construct to the gene of interest (GOI) in the B11 \textit{P. falciparum} line containing a DiCre background. This is followed by the first loxPint, the recodonised region 1 (RR1) containing the rest of the WT GOI tagged with a triple HA (3xHA), and a T2A peptide sequence linking the RR1 with \textit{npt} selection marker. A second loxPint, downstream of the \textit{npt} stop codon, precedes the second recodonised region (RR2) containing either a catalytic Ser (cWT) or a Ser to Ala mutation (cMUT). This is followed by a T2A peptide linked to \textit{gfp}. The construct is integrated into the genome via homologous recombination. In the pre-excision locus the HA-tagged WT SH is expressed, along with \textit{npt}, which confers resistance to G418. The rest of the targeting construct is after the stop codon and out of frame. Upon the addition of rapamycin, DiCre mediates the excision of DNA between the two \textit{loxP} sites. The post-excision locus expresses RR2 encoding the Ser to Ala mutant SH, and GFP.

The DiCre cKO system allows rapid excision of chromosomal DNA between two \textit{loxP} sites upon the addition of rapamycin (Collins et al., 2013b; Jullien et al., 2003). A \textit{loxP} flanked version of the GOI can be integrated into the endogenous locus by single crossover recombination. To limit disruption of the expressed protein, \textit{loxP} sites are placed within an artificial intron (loxPint; Matera and Wang, 2014; Pieperhoff et al., 2014). A DiCre-expressing line of \textit{P.falciparum}, “B11-DiCre” was used (Perrin et al., 2018). The construct backbone contained a 3’UTR from the \textit{P. berghei} DHFR to allow for efficient expression, as the endogenous 3’ UTR is displaced after integration (Fig. 7.1.1). Homology regions
(HR) of approximately 400-500 base pairs in the start of each GOI were chosen to target constructs to
the correct locus (Fig. 7.1.1). Integration of the plasmid inserted two new copies of the GOI downstream
of the endogenous promoter (Fig. 7.1.1). The first copy was the recodonised WT gene (RR1), tagged
with a triple HA peptide and linked to the npt gene via a T2A peptide. This region including the RR1
and npt was flanked by loxPints (Fig. 7.1.1). The second copy was differentially recodonised (RR2)
and contained a mutation of Ser to Ala in the active site of the GOI (or Ser in the control constructs
pT2A-cWT) (Fig. 7.1.1). The RR2 is linked to the green fluorescent protein (GFP) gene by T2A. Both
the RR2 and gfp regions are out of frame and preceded by the npt stop codon, so they are not expressed
(Fig. 7.1.1). Upon the addition of rapamycin (RAP), the first (WT) copy of the gene of interest will be
excised along with the npt selection marker (Fig. 7.1.1). The second copy will come into frame and the
mutated gene and gfp will now be expressed. The excised DNA forms a free episome, however this will
not contain the full gene or any promoter, so continued WT expression is not a concern (Fig. 7.1.1).

7.1.2 Clonal conditional mutant and WT control lines were generated.

We now aimed to use the system described above to interrogate the catalytic function of our four GOIs,
Psta1, Psta2, abH112 and abH114. Schizonts, purified from B11-DiCre parasite culture (Perrin et al.,
2018), were transfected with the pT2A-cMUT constructs designed for each GOI and their pT2A-cWT
control versions, creating eight parasite lines (Materials and Methods). The cultures were treated with
WR to select for parasites maintaining the plasmid construct. When stable populations of WR resistant
parasites were observed, G418 treatment selected for chromosomal integration. G418-resistant
parasites were recovered for all lines after 2 to 4 weeks. All the lines were cloned by serial dilution in
flat-bottomed 96-well plates in the presence of WR and G418. Wells containing a single plaque in the
RBC layer (i.e. originating from a single parent parasite) were identified by microscopy and grown to
obtain at least two clonal lines for each cMUT (Psta1-cMUT: 1A, 1G; Psta2-cMUT: 3F4, 39B; abH112-
cMUT: 5A, 5G; abH114-cMUT: 7A, 7B, GOI-cWT lines Supplementary Fig. 10.5.2). They could be
cultured stably under the same conditions as WT parasites (with the addition of WR and G418).

The pre-clonal and clonal lines were assessed using PCRs designed to detect the endogenous gene and
the integrated product (3’ and 5’ ends) (Fig. 7.1.2A-B). Integration of the constructs into Psta2, abH112
and abH114 loci was highly efficient, as shown by the absence of an endogenous PCR product in the
mixed pre-clonal population (Fig 7.1.2B). This was as expected from SLI, however a band was still
observed for pre-clonal Psta1 (Fig 7.1.2B). In all clones, the 3’ PCR fragment was of the expected length
and no band was observed for the endogenous gene PCR (Fig 7.1.2B). Psta1-cMUT and Psta2-cMUT
clones also showed bands of the expected fragment length in the 5’ integration PCRs (Fig 7.1.2B). 5’
integration PCRs were never obtained for abH112 or abH114, although many combinations of primers
and PCR conditions were tried. As all subsequent experiments supported integration, it was assumed
that this was due only to unyielding chromatin structure. The same procedure was taken with the control
cWT lines for each GOI and integration in the clones obtained was checked in the same way
(Supplementary Fig. 10.5.2). Therefore, eight cMUT lines were obtained, two clonal lines for each GOI
cMUT, along with the corresponding eight cWT lines. The cWT lines were cryopreserved to be used as controls if any of the cMUT showed a phenotype upon DiCre mediated excision (Section 9.6.2.2).

**Figure 7.1.2: Integration of conditional mutants.** (A) Schematic showing the PCRs designed to test for integration. Primer numbers refer to Table 9.4. The endogenous gene PCR primers are located flanking the gene locus. In the integrated pre-excision locus, these sites are too far away to give a PCR product. The pre-excision locus contains two integration PCRs at the 5' and 3' ends of the locus, in each case one primer is on the integrated plasmid and the other in the endogenous gene locus. (B) Using the primers indicated in (A), endogenous and 3'/5' integration PCRs were done on genomic DNA extracted from mixed and clonal populations of Psta1-cMUT (1), Psta2-cMUT (3), abH112-cMUT (5) and abH114-cMUT (7) as well as the parent B11 line as a control. The mixed populations are denoted by a number in a circle and the clones are labelled with the same number. The expected sizes of the PCR products are indicated with an arrow.
7.2 Expression of catalytically-dead mutant SHs has no effect on parasite replication.

After confirming integration of our constructs, we aimed to verify efficient excision on a genomic and protein level. Following this, we could check the effect of mutant SHs on parasite growth and replication.

7.2.1 PCRs confirm DiCre mediated excision in conditional parasite lines.

Having obtained cMUT lines, we now wanted to use the conditional system to switch gene expression to the catalytically dead mutants of the GOIs. The parasite lines were synchronised, split and treated with DMSO or RAP at ring stage. At schizont stage (merozoites were used for Psta1, as it is most active at this stage, Section 5.3) parasites were collected, and genomic DNA was extracted for PCRs. Primers were designed to give a PCR product that changed length after excision (Fig. 7.2.1A). The same strategy did not work for abH114, so excision in this case was signalled by the loss of a PCR product (Fig. 7.2.1A). In all the cMUT lines the RAP-treated samples gave a reduction in the length (Psta1, Psta2, abH112) or disappearance (abH114) of the excision PCR product, with all fragments being of the expected length (Fig. 7.2.1B). These results demonstrated efficient excision in all lines. However we wanted to further confirm this on a protein level.

7.2.2 Loss of HA-tagged WT protein after excision is detectable by western blot.

We now aimed to confirm the loss of HA-tagged WT SHs after RAP-mediated excision. Schizonts (or merozoites for Psta1) were taken at the same time as samples for genomic DNA above (Section 7.2.1) and labelled with 500nM FPP-TMR for 30 mins. After 1 cycle of freeze-thaw, protein was extracted and run on SDS-PAGE gels. The gels were scanned for fluorescent proteins before being transferred to a membrane for WB. Duplicate blots were probed with anti-HA and anti-GFP antibodies.

The western blots confirmed the loss of expression of the HA-tagged SHs after excision (Fig. 7.2.1C). All the SHs migrated to a slightly lower MW than expected, which could be due to post-translational processing or an artefact from the gel running (Fig. 7.2.1C; Psta1 41.2kDa, Psta2 42.7kDa, abH112 44.7kDa, abH114 44.8kDa). One of the abH112 clones (5G) did not show a band for the non-excised WT abH112-HA. We only saw an observable difference in FPP-TMR labelling for the abH114 clones, for which a band of approximately 30ka was seen to decrease in intensity in the RAP-treated sample. This was probably due to the disappearance of one of multiple overlapping bands (Fig.7.2.1D). This is
smaller than expected for abH114 (44.8kDa) but agrees with the WB results. FPP-TMR labelled Psta1, Psta2 and abH112 were not observable on the fluorescent gel. They may be masked by other labelled SHs (Fig.7.2.1D).

Therefore, the HA WBs confirm the loss of the WT SHs after excision. Unfortunately, we were unable to support this with complementary gels showing the loss of FPP-TMR labelling or blots showing the appearance of GFP. It would be desirable to repeat these experiments under different conditions to provide further support for correct integration and excision and repeat the WB for abH112 5G.

7.2.3 None of the conditional mutation lines caused an effect in parasite replication

The integration and excision being sufficiently verified (Section 7.2.1-7.2.2), we now wanted to test the effect of the catalytically dead SHs on parasite growth and replication, and thus determine the essentiality of GOI SH activity. Synchronous ring-stage parasites from each of the cMUT lines were treated with rapamycin as previously described (Section 7.2.1). After 12 h, RAP was removed and samples from each culture were transferred to 96-well plates (at 0.5 % RBC, 0.1 % parasitaemia) for a growth assay over multiple cycles (Section 9.3.4). Tracking parasites over multiple cycles allows small changes in parasite replication rate to be seen as the cumulative effects become more apparent over time.

Parasites were fixed at different time points and stained for flow cytometry analysis (Section 9.3.5). The parasitaemia over time was plot alongside the DMSO control. None of the allelic replacement mutants showed any decrease in replication rate in the first three cycles (Fig. 7.2.3). This suggests that catalytically active Psta1, Psta2, abH112 and abH114 are not essential for parasite replication in culture. This is in contradiction to the large-scale genetic screening results (Zhang et al., 2018). It would be interesting to study these lines further to ascertain if there is a subtle growth defect, not detected here.
Figure 7.2.1: PCRs and western blots confirm DiCre-mediated excision. (A) Schematic showing the PCRs designed to test for Di-Cre mediated excision. The excision PCR primers are designed to amplify DNA between the homology region (HR) and the gfp. No product will be produced from the endogenous locus and the size of the product from the integrated locus changes after excision. A different set of primers was required for abH114. The forward primer is indicated with a * and gives a product that disappears upon excision. (B) Using the primers indicated in (A), excision PCRs show the change in size (*or loss) of the excision PCR product after RAP treatment. (C) 12.5% SDS gels were run and transferred to nitrocellulose membranes. Western blots were probed with Anti-HA antibody. The secondary antibody contained horseradish peroxidase. Binding was observed by adding a chemiluminescence substrate and imaging using a ChemiDoc (BioRad). (D) Parasite lysates were labelled with 1μM FPP-TMR for 1 h. Labelling reactions were quenched by addition of LB and immediate boiling. 12.5 % SDS-PAGE gels were run and a fluorescence scanner was used to visualise TMR (red channel Ex/Em 552/572nm) labelling.
Discussion

Here we have described a conditional allelic replacement strategy designed to interrogate the activity of four uncharacterised SHs in *P. falciparum*: Psta1, Psta2, abH112 and abH114. All four SHs had been shown to be present and active in asexual-blood stages by their reaction with FPP probes (Chapter 5). However, their molecular function, substrates and role in parasite growth and replication was unknown. For each SH the conserved alpha-beta hydrolase fold and catalytic residues were identified using the Pfam protein families database (El-Gebali et al., 2019). We designed a strategy to replace each GOI downstream of the promoter, with two recodonised copies of the GOI: a WT HA-tagged version, and a

Figure 7.2.3: Replication assay with conditional mutation lines over multiple cycles. Parasites from two clonal lines of each GOI-cMUT were synchronised and the rings were split and treated with rapamycin or DMSO (start of cycle 1). A growth assay was carried out under standard conditions in triplicate. Samples were taken at 48 hpi in the first cycle, and approximately at 30 hpi in each subsequent cycle (treatment 0 hpi, cycle 1 48 hpi, cycle 2 78 hpi, cycle 3 126 hpi, cycle 4 174 hpi). Parasites were fixed and stained with Hoechst and the infected/uninfected RBC populations counted by flow cytometry. The parasitaemia for each line was plot against cycle number. Two clones are represented in pink and blue, RAP treated samples by a brighter pink/blue and DMSO samples by light pink/blue. Error bars represent the SEM (standard mean error) of three replicates, where error bars are smaller than the size of the cross symbol they are not shown.

7.3 Discussion

Here we have described a conditional allelic replacement strategy designed to interrogate the activity of four uncharacterised SHs in *P. falciparum*: Psta1, Psta2, abH112 and abH114. All four SHs had been shown to be present and active in asexual-blood stages by their reaction with FPP probes (Chapter 5). However, their molecular function, substrates and role in parasite growth and replication was unknown. For each SH the conserved alpha-beta hydrolase fold and catalytic residues were identified using the Pfam protein families database (El-Gebali et al., 2019). We designed a strategy to replace each GOI downstream of the promoter, with two recodonised copies of the GOI: a WT HA-tagged version, and a
catalytically dead Ser-Ala mutant. Di-Cre mediated excision allows us to switch from WT to expression of the catalytically dead mutant upon addition of RAP. Plasmids for integration were targeted to the GOI loci using homologous regions. Integration relied on single homologous recombination, driven by SLI. The SLI-enhanced integration meant that parasites resistant to the selection drugs were detectable after two to four weeks. We were then able to obtain at least two clonal lines for each transfection. For each GOI we performed two transfections in parallel, one with a Ser-Ala and one with a control Ser-Ser conditional allelic replacement. In this way, if a growth defect was seen, we were immediately provided with an elegant control to boost confidence in our findings.

All four gene editing events were successful. The designed constructs were integrated and DiCre mediated excision took place as expected. None of the GOIs gave a phenotype after replacement with the catalytically dead mutant. Further experiments to characterise the clonal lines would be desirable. The FPP probes should be useful tools in labelling the SHs of interest and thereby detecting when the Ser residue is replaced with Ala. Optimisation of FPP probe-labelling and HA-WB combination should allow us to perform ABPP with the Malaria Box inhibitors (M914, M438 and M563) to confirm if abH112 or abH114 are targets. We should also test if cSER-ALA mutants have the same sensitivity to M914, M438 or M563 as the WT parasites. Fluorescent secondary antibodies could also be used to determine the subcellular location of the SHs by IFA. This was attempted once in schizonts with the cSER-ALA lines, but no significant HA or GFP signal was seen in either DMSO or RAP-treated culture (data not shown). Further trials of this technique would be desirable, although it was also not possible to measure cytosolic GFP signal in RAP-treated parasites at schizont-stage by flow cytometry (data not shown), suggesting that levels are too low for detection.

The seemingly non-essential nature of all four genes was unexpected. Psta2, abH112 and abH114 were unable to be disrupted in the piggybac mutagenesis screen and Psta1-mutated parasites showed a growth defect (Zhang et al., 2018). There are two explanations for these contradictory findings. First, the proteins may be essential but their serine-type hydrolase activity is not. In this case the strategy of conditionally generating catalytically dead mutants would not have any effect on parasite growth and replication. This could be tested by performing a full cKO of the GOIs and checking for any phenotype. However, it is unlikely that these SHs, three of which are unrelated, have all lost their biologically relevant hydrolase activity and perform other roles in the parasite, while still retaining the catalytic activity to react with FPP probes. Secondly, the proteins may perform important roles as hydrolases in the parasite but there are other hydrolases or pathways which are able to compensate for their loss in vitro. This option seems likely given the expanded nature of the SH family in P. falciparum, especially the multicopy psta family (Fischer et al., 2003). Here, there may be no discernible phenotype under standard culturing conditions. Large-scale genetic screens have the extra pressure of competition between different strains, so the outcome of a small growth defect may be amplified. We could perform
competition experiments to recapitulate these conditions: mixing parasites from all four RAP- and DMSO-treated lines and monitoring the proportion of each by quantitative PCR. Another option to introduce stress to the parasites would be to limit the nutrients in their media. Psta1 and Psta2 are predicted to be lysophospholipases and may play a role in lipid scavenging and biogenesis in the parasite. Under normal culture conditions there is an excess of nutrients, including lipids, in the media. We could limit the lipids in the media in a serial dilution and culture RAP- and DMSO-treated parasites to see if there are any differences in parasite replication. The replication assay used to interrogate the HsSH inhibitor phenotypes (Section 6.2.2) could also be applied here. By monitoring the DNA and RNA content of developing cSer-Ala parasites we might be able to observe effects on parasite growth that are not apparent in the extended end-point replication assay.

In conclusion, although we did not validate any novel essential genes, and therefore drug targets, we have expanded the study of the large family of SHs in P. falciparum. We have shown that, contrary to genetic screening results, the parasite can grow and replicate over three to four cycles without the SH activity of Psta1, Psta2, abH112 or abH114. The further experiments described would be desirable to confirm and expand on the results reported here. The conditional allelic replacement approach here described was shown to be successful. This approach could be easily applied to other SHs of interest and multiple knockout of related family members could also be considered.
Chapter 8. Discussion

Malaria has affected humankind for thousands of years and is still an important global health issue. Despite recent increases in funding worldwide, the elimination targets set by WHO (WHO, 2019) are not being met. Resistance has been detected for all current front-line treatments. If resistance spreads, the attendant drop in parasite clearance and successful patient cure will cause a crisis in malaria-endemic countries (Tacoli et al., 2016). New treatments are desperately needed to prevent this outcome. The malaria drug discovery community has recently changed its focus towards elimination and eradication, and has set out aims in the form of TPPs and TCPs, for treatments and chemotypes respectively (Burrows et al., 2017). These guidelines hope to both foster cohesion among the community and to help prioritise compounds before the long and costly process of clinical trials. The TPPs describe the ideal properties a new treatment would have, and the parasite life stages it should target. One of the desirable assets of a new TCP is a known, novel MoA. Knowing the MoA of a compound informs us about possible resistance mechanisms and is beneficial in the strategic grouping of compounds in combined therapies.

In recent years the discovery of novel antimalarial chemotypes has relied on large-scale, high-throughput phenotypic screens. The benefit of these screens is that hit compounds have proven activity on whole parasites. However, no information is gained on their MoA. Target deconvolution can be achieved using many methods such as IVEWGA and ABPP (Baragaña et al., 2015; Paquet et al., 2017). Potential targets can then be validated genetically as essential Plasmodium genes. This adds both to the knowledge of parasite biology, and to the known druggable proteome. Target identification helps inform inhibitor design through structural/substrate prediction and assays can be developed for hit compound optimisation.

In this thesis we have characterised the activity of two broad-spectrum ABPs in P. falciparum: the novel W-sCy5-VS probe targeting PLCPs (Tan and Davison et al., 2020); and the well-characterised FPP probes targeting SHs (Liu et al., 1999). ABPs are highly useful tools as they report on the activity of targeted enzymes. In ABPP, the activity of multiple enzymes can be measured in a complex mixture. By competing probe labelling with different compounds, inhibitor targets can be identified. The combination of ABPP with proteomics allowed us to profile specifically active proteins and not just all proteins that are present or expressed in the cell. We then aimed to use W-sCy5-VS and FPP-TMR for ABPP of PLCPs and SHs in asexual-blood stages. The metabolic SHs of P. falciparum are largely uncharacterised. The FPP-TMR and FPP-N3 probes were used to profile and identify SHs in P. falciparum (CROSSREF results). In the first instance, all of the SHs identified were alpha-beta fold containing enzymes. Forty-three alpha-beta fold-containing hydrolases have been predicted from
resources such as the ESTHER database (Hotelier et al., 2004). The majority of these SHs are completely uncharacterised. Other metabolic PfSHs include four predicted patatin-like phospholipases (PLPs, Wilson and Knoll, 2018) and an amidase (Mailu et al., 2015).

Unlike the metabolic SHs, the *Plasmodium* PLCP family (including the DPAPs, FPs and SERAs) has been studied and many of its members have been implicated as potential drug targets (Deu, 2017). The SERAs have been shown to participate in egress under tight temporal control. The active state enzymes are therefore difficult to profile using ABPs (Thomas et al., 2018). Many of the studies into the DPAPs and FPs have relied on ABPP (Arastu-Kapur et al., 2008; Deu et al., 2010). We characterised the activity of the novel probe W-sCy5-VS against *P. falciparum* PLCPs, showing that it labelled more members of this family than either of the well-established CP probes DCG-04 and FY01 (Tan and Davison et al., 2020). We also showed W-sCy5-VS was able to label multiple PLCPs (DPAP1, DPAP3, FP1, FP2 and FP3) in neutral conditions throughout the life cycle in addition to many other unknown E-64 sensitive proteins. We then used the bi-functionally tagged probe W-dtBio-TMR-VS to validate the targets in schizont-stage lysates. The proteomic results identified only DPAP1 and FP3 in neutral conditions, and many reactive-cysteine-containing redox enzymes were identified rather than other CPs as expected.

The ability of W-sCy5-VS to act as a broad-spectrum PLCP ABP in neutral pH gel-based assay means that we could use it in combination with FPP-TMR to build a multiplexed ABPP screen. We studied the kinetics of both probes to find sub-saturating labelling conditions that would compete against reversible inhibitors. A medium-throughput gel-based ABPP screening assay was designed and optimised to compete the two probes against hundreds of reversible inhibitors. A gel-based screening format naturally selects for hits that will be easy to follow-up in further gel-based assays. We used MMV’s Malaria Box, a focused library of 400 diverse and potent antimalarial chemotypes (Spangenberg et al., 2013). These were derived from large-scale phenotypic screens by the pharmaceutical industry (Gamo et al., 2010; Garcia-Bustos and Gamo, 2011; Guiguemde et al., 2010; Meister et al., 2011).

We performed the Malaria Box ABPP screen at schizont, trophozoite and ring stages to interrogate the different activity profiles at each stage, as indicated by the in-gel life-cycle experiments. None of the Malaria Box compounds competed with W-sCy5-VS for engagement of PLCP targets. However two compounds, M941 and M738, increased W-sCy5-VS labelling of FP3 when present. M738 also augmented labelling of two further unknown bands that it would be of great interest to identify. For M941 we confirmed *in vitro* inhibition of FP3 activity by fluorogenic assay. We hypothesised that M941 acts as an allosteric inhibitor of FP3, although further activity assays are needed to confirm a non-competitive inhibition model. If verified, these results would present further evidence for allosteric
inhibition of FP’s which had been previously shown for FP2 (Hernández González et al., 2019; Marques et al., 2015). M738 may inhibit FP3 non-competitively through either the same or a different binding site to M941. It would be desirable to repeat the inhibition assay with M738 as inhibitor. FP2 should also be assayed in the same way to check the selectivity of M914/M738 among FPs. W-dtBio-TMR-VS did not show the same M738/M941-dependent labelling so other methods will have to be developed in order to identify the other targets of M738. The unusual modulation of L-WSAK-Alk activity with the type of fluorophore attached (Tan and Davison et al., 2020) suggests that a sCy5-biotin bifunctional tag may be required for specific affinity purification of targets. Ongoing studies of the FPs are of high importance. FP2 and FP3 were shown to be non-essential for asexual-blood stage growth, although this may be due to redundancy (Dr Daniel Goldberg, personal communication; Sijwali et al., 2006; Sijwali and Rosenthal, 2004). Haemoglobin metabolism and haem detoxification are highly important antimalarial targets. The MoAs of the most historically successful malarial drugs target these pathways, including Quinine (Slater, 1993), CQ (Banyal and Fitch, 1982) and Artemisinin and derivatives (Kamchonwongpaisan and Meshnick, 1996; Meshnick et al., 1991). FP2 has been implicated in artemisinin-resistance mechanisms (Ariey et al., 2014; Siddiqui et al., 2018). New generation drugs still target haem detoxification, such as FQ in phase 2 trials (Phyo et al., 2016; Wells and Hooft van Huijsduijnen, 2015). The complex regulation of proteases in haemoglobin metabolism and haem detoxification is far from understood, as illustrated by the possible FP-activation of upstream plasmepsins (Bonilla et al., 2007) and complex systems of allosteric modulation of FPs (Bertoldo et al., 2015; Hernández González et al., 2019; Marques et al., 2015; Pant et al., 2018). The preliminary results described here report novel allosteric FP3 modulators and a new broad-spectrum ABP tool to monitor their interaction.

Three Malaria Box compounds were found to inhibit SHs and block FPP-TMR labelling: M914 and M438 both inhibited a 49kDa SH in schizonts, and M563 appeared to be a broad-spectrum inhibitor blocking labelling of multiple SHs. M563 blocked parasite replication from ring stage with sub-micromolar EC_{50}. M563, and other structurally related BZTs, were previously shown to inhibit cytochrome bc{1} of the ETC P. falciparum in vitro (Dong et al., 2011), but our results imply it is also a pan-SH inhibitor at 50 μM. In a previous Malaria Box screen, it was shown that parasite growth inhibition by M438 could be rescued by addition of CoA (Fletcher and Avery, 2014). We confirmed these results and showed that M914 did not show the same affect, suggesting that M438 may have an additional target involved in CoA synthesis. Overall, this screen identified three diverse SH inhibitors, one of which may be broad-spectrum, and two of which are chemically diverse potential allosteric FP3 inhibitors. The diversity of the hits and the unexpected allosteric activity of the FP3 inhibitors illustrates the power of using this unbiased approach. However, we must consider that the high-throughput screening libraries used by pharmaceutical companies that led to the Malaria Box are not in fact unbiased. They are generally targeted towards classical drug targets such as kinases and proteases, as
shown by computational analysis (Spitzmüller and Mestres, 2013). This screen could easily be extended to other libraries of reversible inhibitors.

We aimed to use chemical proteomics techniques for target deconvolution of the SH inhibitors M438, M914 and M563. First the FPP-probe chemical proteomics approach was optimised in order to increase the depth of coverage of the SH family. The conditions of affinity purification were adjusted to enhance the detection of SHs over contamination from neutravidin-agarose beads, by reducing their volume by one third. Isobaric labelling techniques were used to multiplex samples and provide precise quantification for the comparison of both probe- and DMSO-treated samples. After optimisation, twenty FPP-N3-labelled SHs were identified: 40% of the predicted *P. falciparum* alpha-beta hydrolases and three patatin-like phospholipases. By visual inspection of the replicates, three proteins appeared to change in activity in the presence of inhibitor: M563 appeared to inhibit abH112 and LCAT; and M914 and M438 appeared to inhibit abH114 and LCAT. It was encouraging that M914 and M438 inhibited the same targets and abH114 was the correct size to be the 49kDa band inhibited in the gel-based Malaria box screen. Only the M438-dependent LCAT enrichment changes were statistically significant. More optimisation is needed to detect reversible inhibition using these proteomics techniques. Otherwise probe-free techniques such as CESTA could be used for target deconvolution (Molina and Nordlund, 2016).

Metabolic SHs are ubiquitously important in cell biology in all organisms. Although they are generally well-conserved between different organisms, most human metabolic SHs are sequentially and structurally distinct from each other (Simon and Cravatt, 2010). Their uniqueness suggests distinct and important functions and is a striking example of divergent evolution: 60% of mammalian metabolic SHs share the α/β-fold ([Holmquist, 2000; Ollis et al., 1992]. The α/β-fold motif is curiously tolerant to insertions and deletions in the protein sequence, which explains the evolutionary diversity of its members (Nardini and Dijkstra, 1999). SHs represent approximately 1% of the proteome in humans (Rawlings et al., 2010). Some have been well-studied (Long and Cravatt, 2011), but many others remain uncharacterised (Simon and Cravatt, 2010). SHs often have multiple tissue-specific functions. Enzymatic inhibition of neuronal AChE has been a core focus in both research and treatment of Alzheimer’s disease (Saxena and Dubey, 2019). However minor splice variants have been ascribed non-hydrolytic functions (AChE-R and AChE-T; Zimmermann, 2013), and erythrocyte AChE has distinct and important roles of its own (Saldanha, 2017). Metabolic SHs are therefore highly promising drug targets, although their remarkable diversity both in sequence and function has made them difficult to study. The broad-spectrum FPP ABPs have been incredible tools in the study of metabolic SHs (Bachovchin et al., 2010; Jessani et al., 2005; Kidd et al., 2001; Liu et al., 1999; Okerberg et al., 2005; Patricelli et al., 2001). FPPs target the conserved catalytic site of SHs, allowing us to profile many members of this enzyme family at once, despite their diverse functions (Piñeiro-Sánchez et al., 1997).
This thesis uses FPP ABPs to report the most in-depth study of *Plasmodium* SHs at multiple points throughout the asexual-blood stage life cycle. We have identified forty-three α/βHs in *P. falciparum* by integrating results from structural databases such as ESTHER (Hotelier et al., 2004) and MEROPS (Rawlings et al., 2010), the online resource PlasmoDB (Aurrecoechea et al., 2009) and searches of sequence homology. Only twenty-nine of these α/βHs were conserved in other non-human-infecting *Plasmodium* spp. Other *P. falciparum* metabolic SHs include the four PLPs (Wilson and Knoll, 2018) and a unique amidase, glutamyl-tRNA(Gln) amidotransferase subunit 1 (Mailu et al., 2015). We report the ABPP of twenty-five metabolic *Pf* SHs across the *Plasmodium* life cycle. The remaining SHs predicted in *P. falciparum* may not be expressed in asexual-blood stages, or perhaps they are only active after export into the PV or RBC lumen. In the latter case, they would be lost after saponin lysis of parasites. FPP-N3 labelling was performed in intact parasites (apart from merozoites) to ensure that profiles reflected real cellular activity as much as possible. The activity of each SH was measured by calculating the FPP-N3-dependent enrichment in each, using three replicates. Twenty-two of the metabolic SHs detected were α/βHs and three were PLPs. Only two new SHs were identified when we extended the schizont proteomics studies to rings, trophozoites, mature (C2-arrested) schizonts and merozoites: PF3D7_1427100 and PF3D7_1116100. PF3D7_1116100 was lowly enriched in probe-treated samples but showed an increase in schizont and merozoite stages compared to ring and trophozoites. PF3D7_1116100 may be a lowly expressed α/βH that was not detected in previous schizont ABPP studies for this reason. PF3D7_1427100 was the only active α/βH to be most highly enriched in trophozoite stages. This suggests that it may play a role in metabolism at this stage. PF3D7_1427100 is a proposed triacylglycerol lipase based on domain prediction. It appears to be dispensable in asexual-blood stages (Zhang et al., 2018), as does its *P. berghei* homolog (PBANKA_1017500; Schwach et al., 2015).

We were surprised to detect the exported α/β-epoxide hydrolase *Pf*EH1. *Pf*EH1 hydrolyses epoxygenesatrienoic acids (EETs), which are lipid signalling molecules that modulate vasodilation and anti-inflammation in the bloodstream (Jiang et al., 2012; Larsen et al., 2006; Thomson et al., 2012). EETs are important regulators of vasoconstriction/dilation through complex mechanisms involving signalling molecules such as nitric oxide (NO) and ETTs. NO plasma levels are known to be reduced during malaria infection (Carvalho et al., 2014; Chertow et al., 2015; Eisenhut, 2015). This reduction in NO promotes cytoadherence of iRBCs (Serirom et al., 2003). ETTs may play a substantial role in anti-inflammatory (and therefore anti-cytoadhesion) processes in the absence of NO. *Pf*EH1 and *Pf*EH2 have been shown to deplete ETT levels in iRBCs (Spillman et al., 2016b). *Pf*EH1 and *Pf*EH2 were both observed at the periphery of the iRBC, associated with spectrin (Spillman et al., 2016b). By ABPP we have detected *Pf*EH1 as part of a group of α/βHs that are activated throughout the trophozoite to
merozoite stages (PF3D7_1401500, PfPARE, PF3D7_0805000 and PfEH1). This is surprising as the exported PfEH1 should be lost after saponin lysis. These α/βHs may be detected due to their interaction with the RBC membrane, seen as debris in purified parasite material. This was postulated for the membrane-associated HsSHs: PNPLA6, AChE an NCEH1. If this hypothesis is true then it implies that PfEH2, which was not detected, is associated with the RBC membrane in a different way to PfEH1. AChE has been implicated in the regulation of vasoconstriction/dilation and anti-inflammation via NO signalling. It is possible that there is an interplay between parasite and human proteins at the RBC membrane to regulate these signalling pathways (Saldanha, 2017). PfEH1 is not essential to the growth and replication of asexual-blood stages in vitro and is only conserved in the Laverania subgenus (Spillman et al., 2016b). However, the results reported here add to the evidence that PfEH1 may be important in Plasmodium infection in vivo.

Of the three PLPs that were profiled, only one has been studied in Plasmodium. Patatin-like phospholipase 1 (PfPNPLA1; PF3D7_0209100) has phospholipase A2 (PLA2) activity. PfPNPLA1 has been shown to be dispensable in asexual-blood stages, although cKO reduces the efficiency of egress and exflagellation in gametocytes, and oocyst formation in mosquitoes (Singh et al., 2019). These gametogenesis-specific effects implicate PfPNPLA1 as an interesting transmission-blocking drug target. Although there was no asexual-blood stage growth defect in the cKO, PfPNPLA1-null parasites were later shown to have increased levels of major phospholipids such as phosphotidylcholine (PC) (Flammersfeld et al., 2020). PfPNPLA1 appeared to be activated throughout the life cycle with modest increases at trophozoite and merozoite stages. Based on our ABPP results, PfPNPLA1 activity appeared to be very similar to the exported lipase PfXL1, but not to its parologue PfXL2.

Merozoite egress is a rapid highly-regulated process. Mature (C2-arrested) schizonts are stalled before egress. At the C2 stall point, merozoites are fully formed inside the RBC and PVM but the cascade of reactions that causes membrane rupture and merozoite egress is blocked (Blackman, 2008). If C2 is washed off, then merozoites can egress in less than 15 min (Blackman, 2008). Some SHs show a change in activity levels between mature schizont and merozoite stages. We hypothesised that these SHs may play an important role in egress-promoting reactions during this small time window. Because of the short timescales involved these processes must be controlled by changes in activation of proteins already present in parasites, rather than changes in protein expression or abundance. These are examples of changes that ABPP methods are able to detect as opposed to global proteomics methods. Three α/βHs were found to be activated in merozoite stages: S9C, MLPL and Psta1.

Recently, S9C became the first PfSH to have its ABPP-reported activity confirmed by genetic methods in our lab. Conditional disruption of S9C showed it to be non-essential but mutants had an asexual-
blood stage growth defect (Ridewood et al, in preparation). A similar conditional allelic replacement approach to the one described in this thesis showed that the growth defect phenotype was dependent on the loss of SH activity (Ridewood et al, in preparation). S9C-HA-expressing parasite lines were used to verify that S9C is active in merozoite stages by FPP-TMR labelling and WB (Ridewood et al, in preparation). S9C-null mutants showed a delay in early ring development and defect in PVM formation (Ridewood et al, in preparation). Activation in merozoites suggests that S9C-activity is important immediately after invasion, possibly participating in lipid metabolism before nutrient scavenging mechanisms are in place (Ridewood et al, in preparation).

MLPL had been suggested to be dispensable in asexual-blood stages by the piggybac mutagenesis screen (Zhang et al., 2018). We reasoned that it was unlikely to have an important role in egress if it was non-essential, so MLPL was not prioritised for genetic interrogation. However, it would still be of interest to verify if MLPL is indeed non-essential. Data from large-scale genetic screens does not always reflect the results of conditional gene disruption studies, as illustrated in this thesis. Active Psta1 was enriched at a lower level than S9C and MLPL, but also showed activation in merozoite stages. Psta1-disrupted parasites were detected in the piggybac mutagenesis screen but showed a growth defect (Zhang et al., 2018). Psta2, a close parologue of Psta1, was enriched in schizont stages and appeared to be essential (Zhang et al., 2018). Psta1 and Psta2 are predicted lysophospholipases and are members of the subtelomeric family Pst-a about which very little is known (Fischer et al., 2003).

One of the non-SHs detected, the endonuclease MUS81, was also highly activated in merozoites. In mammalian systems MUS81 is involved in DNS repair of DSB by homologous recombination, specifically via the formation of double Holliday junctions (Fu et al., 2015). The activation of MUS81 specifically in merozoites is surprising. We expect homologous recombination to be at its peak in schizont stages when asynchronous DNA replication makes chromatids available for repair (Lee et al., 2014). These unexpected results for MUS81 suggest that this protein has a stage-specific role in DNA repair. DNA repair pathways are potential targets for Plasmodium as unrepaired DSBs lead to cell death.

Surprisingly, we also found interesting results for human SHs in the ABPP-life cycle study. We hypothesised that some of these HsSHs could be internalised and co-opted by the parasite, and would therefore make good resistance-proof antimalarial drug targets. The development of a human enzyme as a drug target would be aided by the wealth of data and inhibitors already available for these SHs, which are much more studied than PfSHs (Long and Cravatt, 2011). We verified that inhibitors of APEH, LYPLA1, FASN, PNPLA6 and AChE inhibited parasite growth in culture, with the APEH inhibitor AA74-1 and the LYPLA1 inhibitor ML-348 being the most potent. APEH was shown both to be most active in merozoite stages and also to inhibit parasite growth from early ring stages. A recent paper confirmed our hypothesis of the essentiality of APEH in asexual-blood stages and import into the
parasite (Elahi et al., 2019a). The combined data led us to hypothesise a role for APEH in early ring stages, immediately after invasion, as predicted for Pfs9C (Ridewood et al., in preparation). LYPLA1, a predicted depalmitoylase, was most active in early stages, but did not affect parasite replication until schizont stage, when egress was halted. Further studies are required to determine the nature of the halt of egress and the role of LYPLA1. For example testing the effect of LYPLA1 inhibitors on the palmitoylome via metabolic labelling and chemical proteomics (Jones et al., 2012). The ABPP strategy described here can be used to verify the specificity of the HsSH inhibitors in iRBCs. These HsSHs represent an exciting new area of anti-parasitic drug targets. They join a number of other human proteins known to be co-opted by the parasite to perform functions such as haem biosynthesis (Bonday et al., 1997; Nagaraj et al., 2013; van Dooren et al., 2012). By employing host proteins in this way, the parasite saves time and energy in de novo protein synthesis. In fact it may have even lost capability to perform these reactions without host protein involvement. There are two main benefits to targeting host enzymes for anti-parasitic activity. First it is theoretically less likely that the parasite will evolve resistance to host protein-directed drugs, although of course indirect methods of drug resistance such as transporters are still possible. Secondly, many human metabolic SHs have been previously studied in regards to human health and disease. The information already gathered about LYPLA1 and APEH has helped to direct any further studies on the functional roles of these enzymes in the iRBC. We illustrate here the ease of procuring specific inhibitors for the HsSHs, although further ABPP is required to verify they have no P. falciparum off-targets. Any drug discovery based on these targets would benefit from previously discovered chemotypes, some of which have already been assessed for use in in vivo models. This a priori information would significantly decrease the cost and time of hit-to-lead studies. The targeting of human enzymes to fight pathogenic infection has been shown to be useful in other disease states. Human N-myristoyltransferases NMTs are co-opted by viruses and HsNMT inhibitors block viron assembly thereby suppressing virus replication and infectivity (Mousnier et al., 2018). This has been demonstrated in highly diverse viruses like Rhinovirus (responsible for the common cold) whose high replication and mutation rate have hampered any efforts to develop vaccines (Mousnier et al., 2018; Thibaut et al., 2012). Targeting the host enzyme is a promising new avenue in the development of treatments and vaccines for infective agents that rapidly acquire resistance to pathogen-directed drugs.

Recently, another group has sought to profile SHs in P. falciparum in an effort to start chemically annotating this uncharacterised family (Elahi et al., 2019b). In this study, published after the completion of our experiments, both soluble and insoluble fractions from schizont lysates were labelled with FPP-desthiobiotin (Elahi et al., 2019b). Twenty-one SHs were identified, including three SH that were not identified in any of our proteomic experiments (Elahi et al., 2019b). Two were α/βHs: PF3D7_1126600, predicted to be a steryl ester hydrolase; and PfEH2. Both of these were present in the insoluble fraction of schizont parasite material that was labelled (Elahi et al., 2019b). The insoluble fraction from purified
parasite pellets is mostly made up of membrane debris and membrane-associated proteins. The lysis and protein extraction methods used in this thesis must not have been sufficient to solubilise PF3D7_1126600 and PfEH2 from this environment. PfEH2 has been shown to be associated with the RBC periphery, like PfEH1 (Spillman et al., 2016b), but this is further evidence to suggest that the mode of attachment to the RBC surface are different for the two epoxide hydrolases. PfEH1 was also identified in the insoluble fraction. The other SH uniquely detected by the FPP-desthiobiotin study was the rhomboid protease ROM4 (PF3D7_0506900; Elahi et al., 2019). ROM4 is involved in invasion by shedding adhesins from the surface of invading merozoites via intramembrane proteolysis (Baker et al., 2006b; O’Donnell et al., 2006). ROM4 was also identified in the insoluble fraction. Five human SHs were identified by FPP-desthiobiotin: AChE and a prolylendopeptidase PREP in soluble fractions; PNPLA6 and NCEH1 in insoluble fractions; and APEH in both fractions (Elahi et al., 2019b). The presence of AChE in the soluble fraction is surprising as we predict it to be RBC membrane-associated (Ott, 1985; Soreq and Seidman, 2001). In our study it showed very similar enrichment to membrane-associated PNPLA6 (Tienhoven et al., 2002) and NCEH1 (Sekiya et al., 2011).

The schizont FPP-desthiobiotin report (Elahi et al., 2019b) and the FPP-N3 study described in this thesis mainly differ in their labelling and protein extraction methods. FPP-desthiobiotin probe labelling was performed in lysates. This is required for FPP-desthiobiton because it is not cell permeable. This does not appear to have materially decreased the number of identifications. A different activity profile might be seen if the FPP-N3 asexual-life cycle screen were done in this way. Intact parasite labelling may be a better reflection of true parasite activation levels although in-parasite labelling is subject to its own artefacts. For example, probe accumulation in certain subcellular compartments. In the FPP-desthiobiotin study soluble and insoluble lysate fractions were prepared in PBS by centrifugation only, without any detergents. While we used Triton X to solubilise as many SHs in the insoluble fraction as possible, the FPP-desthiobiotin study kept the two fractions separate. This gives more information about which proteins are membrane-associated. Only five out of twenty-one SHs were identified in both soluble and insoluble fractions. By comparison of the two data sets, we can infer more information about the nature of membrane associations of the SHs. For example, the differences between PfEH2 and PfEH1 and the unexpected solubility of AChE. We detected twenty-five PfSHs, including eight that were not labelled in the FPP-desthiobiotin. The greater number of PfSHs detected may be due to three main reasons. The first is probe labelling conditions, FPP-N3 is more reactive than FPP-desthiobiotin and may also label SHs in intact parasites that do not retain activity in lysates. Secondly our study may achieve a greater depth of coverage due to the method optimisation and sensitive isobaric labelling and quantification methods used. Finally, our study has extended the profiling to multiple life stages throughout the asexual-blood stages. As we know that active SH profiles change throughout the life cycle, it is necessary to label each stage to get the full complement of asexual-blood stage PfSHs.
A huge amount of information regarding the time-dependent activation, and therefore cellular function, can be inferred from ABPP throughout the life cycle.

We used the ABPP and chemical proteomics studies to identify four candidates for genetic validation in *P. falciparum*: abH1112, abH114, Psta1 and Psta2. These candidates were prioritised based on their predicted essentiality by high-throughput genetic screens (Zhang et al., 2018). abH112 and abH114 were chosen from the Malaria Box competitive proteomics study as they had been predicted essential by the *P. falciparum piggybac* mutagenesis screen (Zhang et al., 2018). Although the *P. berghei* homologue of abH114 was found to be dispensable in another screen (Schwach et al., 2015). Psta2 was also predicted to be essential, although the closely-related Psta1, which was active at merozoite stages, only showed a growth defect upon mutagenesis (Zhang et al., 2018). All four candidates chosen for genetic interrogation were serine hydrolases of the alpha-beta fold hydrolase superfamily with unknown molecular function. We therefore designed a strategy to probe the serine hydrolase activity of the GOIs that would be complementary to the ABPP carried out to identify them. Plasmids were designed for a conditional allelic replacement strategy that relied on DiCre mediated excision between *loxP* sites (Collins et al., 2013a; Jones et al., 2016). Single homologous recombination, driven by SLI, was used to obtain stable integrants quickly (Birnbaum et al., 2017a). An advantage of this method over CRISPR-Cas9 was that simple molecular biology techniques were required to make the suite of eight integration vectors (four Ser-Ala mutants, and four Ser-Ser controls). However, the timescales involved would have still been very similar if CRISPR-Cas9 had been chosen as the approach. Stable integrated parasite lines were detected, cloned and validated. Ultimately, contrary to what was expected from the *piggyBac* mutagenesis screen, none of the mutated lines showed a change in either parasite growth or replication in culture. We hypothesised that this difference was due to a lack of competition in our growth assays compared to the genetic screening. In large-scale genetic screens, mutant parasite lines are multiplexed and the results inferred by observing which parasite lines are detected after multiple cycles. Competition between healthy parasite lines and lines with a subtle growth defect may cause underrepresentation of the growth defect, leading the GOI to be labelled as essential. Parasites also have time to adapt to the loss of a GOI by upregulating alternative proteins/pathways. This adaptation process cannot be monitored in the large-scale genetic screens. Adaptation by upregulation of other genes is particularly relevant here as the metabolic *Pf*SH family has many similar paralogues that could have redundant functions, and could mean very subtle growth defect phenotypes. There are many examples of *P. falciparum* genes that have been annotated as essential in the past before being disproved by conditional gene disruption strategies, such as MSP1 (Das et al., 2015), the falcipains (Sijwali et al., 2006) and the plasmepsins (Liu et al., 2006). We have shown one important, and three presumed essential, α/βHs to be functionally dispensable in vitro. This illustrates the great caution that must be used when interpreting large-scale genetic screens (Zhang et al., 2018). It is, of course, possible the lack of essentiality seen here is due to our conditional allelic replacement strategy. We have only shown that
the α/βHs catalytic serines are dispensable. Several classes of non-catalytically active α/βHs have been identified in humans by their homology to AChE (De Jaco et al., 2012; Gibney et al., 1988; Swillens et al., 1986). These cholinesterase-like proteins lack a properly-positioned catalytic Ser residue and so would be unlikely to react with FPP (Swillens et al., 1986). Splice variants of AChE itself have proposed non-catalytic functions (Saldanha, 2017; Zimmermann, 2013). However, all of the α/βHs targeted here were reactive to FPP probes and so do have active catalytic triads. As such, they are unlikely to be non-hydrolytic. Using the FPP probes in combination with HA WBs, we could use broad-spectrum inhibitors to probe more specific SHs functions to further characterise these enzymes. For example, ABPP using FPP-TMR in competition with pan-lipase inhibitors, as shown in the first PfSH functional annotation paper, could help assign lipase activity (Elahi et al., 2019b). The monitoring of DNA/RNA content throughout the life cycle by flow cytometry may help detect subtle growth defects immediately after excision. Other methods such as lipidomics could be employed to test if lipid metabolism is affected, even if a growth defect is not obvious. This was shown with PfPNPLA1-null parasites (Flammersfeld et al., 2020).

The study of Homo and Plasmodium SHs is a highly interesting field. Most of SHs expressed in asexual-blood stages have unknown enzyme functions. Many metabolic SHs are predicted to be lysophospholipases or lipases (Table 1), so may play roles in lipid/phospholipid scavenging, processing or synthesis. Lipids are important in signalling, inflammation and membrane biogenesis but most of these pathways are only partially understood in P. falciparum. The phospholipid content of RBCs increases almost five-fold during intra-erythrocytic parasite growth (Tran et al., 2016). The source of lipids that make up the PVM is still a topic of debate (Sherling and van Ooij, 2016). Mature erythrocytes do not perform de novo fatty acid synthesis. This may be due to the lack of acetyl-CoA carboxylase which provides HsFASN with malonyl-CoA (Pittman and Martin, 1966). This multi-domain HsFASN is type I, common to eukaryotic species. Malarial parasites however have a prokaryotic type II FASN only, composed of a network of distinct proteins. This PfFASN II pathway has been shown to be non-essential to asexual-blood stage replication (Vaughan et al., 2009); the parasite may co-opt human proteins to create a full fatty acid synthesis network in RBCs. Different types of lipids are not only important in the membranes of parasites and their organelles but also play significant roles in signalling and immunogenic functions (Jiang et al., 2012). It is clear that lipid synthesis, scavenging and signalling pathways must all be viewed as complex systems involving both host and parasite proteins. In vivo, contributions from the serum are also likely to play a role. Interestingly, a set of secreted human PLA2s have been shown to exhibit anti-Plasmodium activity in vitro via the hydrolysis of lipoproteins in the media, releasing toxic polyunsaturated non-esterified fatty acids (Guillaume et al., 2015, 2006). As such, single gene disruptions may not be a viable option in the study of these processes. ABPP, metabolic chemical proteomics and other mass-spectrometry-based techniques such as lipidomics and
metabolomics all have important roles to play in answering the current questions. FPP probes therefore present an opportunity to profile the activities of diverse SHs from both humans and parasites simultaneously that may be part of a cooperative network of lipid metabolism.

In conclusion, the use of broad-spectrum ABPs allows many members of enzyme families to be profiled simultaneously in an unbiased way. There are many avenues for further studies suggested by this body of work. In terms of PLCPs we have described W-sCy5-VS as a novel tool for future ABPP studies. We have also identified two compounds that may be useful in the further studies of allosteric inhibition of FPs. We have identified three chemotypes that target SHs in *P. falciparum* and, with continued target identification, novel drug targets may be validated. The SH activity profiles we have reported throughout the asexual-blood stages will inform future attempts to further characterise this large enzyme family. The human SHs represent an exciting new area of study. Further experiments with our conditional parasite lines are warranted, given the disagreement with previous genetic screening results. FPP probes will undoubtedly remain a useful tool in the characterisation of all parasite-serving metabolic SHs both of human and *Plasmodium* origin.

This thesis illustrates the benefits of using ABPP to profile and chemically annotate a proteome. Annotation is of great importance in *Plasmodium* where only two thirds of the genes have any predicted function (Böhme et al., 2019). A small percentage of these gene predictions have been functionally validated. ABPP-chemical proteomics is an excellent tool to help functionally annotating the genome based on common catalytic mechanisms. The distinction between active and non-active enzyme forms also helps us to align enzyme activity with life-stage specific functions. For instance, the highly-regulated events around merozoite egress. We have shown how this approach is uniquely suited to the α/βH family, the remarkable diversity of which makes it difficult to study. The study of malarial PLCPs has also greatly benefitted from ABPP approaches, with important roles for both specific and broad-spectrum probes (Arastu-Kapur et al., 2008; Tan and Davison et al., 2020; de Vries et al., 2019; Greenbaum et al., 2002; Lehmann et al., 2018; Xie et al., 2016). SH- and CP-directed ABPs were among the first to be described (Greenbaum et al., 2000; Liu et al., 1999) and are certainly the most widely used. Probes have also been developed towards aspartyl proteases (Li et al., 2000), the proteasome (Nazif and Bogyo, 2001; Ovaa et al., 2003), glycosidasases (Vocadlo and Bertozzi, 2004; Wu et al., 2019), phosphatases (Kalesh et al., 2010) and PLP-dependent enzymes (Hoegl et al., 2018). Developing probes for other large enzyme families such as kinases is not as straightforward. Their as their activity does not proceed via covalent enzyme-intermediates that can be exploited, as with CP/SH ABPP. However, acyl-phosphate-containing nucleotides have been developed as ABPs. These show great promise in the ABPP of kinases and other nucleotide-binding proteins (Franks and Hsu, 2019; Patricelli et al., 2006; Rosenblum et al., 2013). Broad-spectrum ABPs have also been developed for bromo-domain containing proteins which lack enzyme activity by targeting non-catalytic conserved amino acids (D’Ascenzio et
After the success of this in-depth ABPP study of SHs we propose that ABPP should play a central role in the functional annotation of unannotated and uncharacterised *Plasmodium* enzymes of multiple other families. We have also shown the efficacy of combining ABPP with genetic interrogation for the unbiased identification of possible anti-parasitic drug targets and their subsequent validation. The genetic strategy described here has benefitted from the recent advances in *Plasmodium* gene editing technology, that now make it feasible to interrogate full enzyme families. This thesis illustrates the value of integrated and complementary approaches: biology and chemistry; target-based and phenotypic screens; genetic and biochemical strategies. These approaches are ideally suited for the post-genomic era of *Plasmodium* research, where novel anti-parasitic drug targets must be considered in the content of both their catalytic activity and their function in complex host-parasite reaction networks.
Chapter 9. Materials and Methods

9.1 Maintenance and synchronisation of parasite culture

The parasite lines used in this study were *P. falciparum* wild type clone 3D7 and the DiCre expressing line B11 (Perrin et al., 2018). All parasite lines were cultured at 37°C in 1-4% haematocrit of human erythrocytes (UK Blood and Transfusion service) under an atmosphere of 90% N₂, 5% CO₂, 5% O₂ (Trager and Jensen, 1976) and using RPMI-1640 (Gibco) based media supplemented with 0.292 g/L L-glutamine, 25 μg/ml gentamycin and 5% (w/v) Albumax II (referred to hereafter as ‘complete media’).

Light microscopy was routinely used to view blood films and determine parasitaemia, parasite life stage and morphology. Air-dried thin-blood smears for this purpose were fixed with 100% methanol and stained with 10% Giemsa (VWR international).

Synchronisation of parasite cultures was achieved by isolating mature schizont stage parasites on 63% (v/v) Percoll (GE Healthcare Life Science) by density gradient centrifugation (Kramer et al., 1982). Schizonts were then allowed to invade fresh erythrocytes at 37°C in complete medium with shaking (100 rpm) for 2 h unless otherwise stated. Remaining schizonts were removed by using a second Percoll purification step followed by treatment with 5% (w/v) D-sorbitol (Lambros and Vanderberg, 1979) to leave a synchronous culture of ring-stage parasites.

9.2 ABPP in *P. falciparum*

9.2.1 Preparation of parasite pellets from synchronous culture for ABPP

For the preparation of schizont pellets, schizonts were isolated from uninfected erythrocytes via Percoll purification as described (Section 9.1). “Mature schizonts” were collected by first treating schizonts with 4-[7-([dimethylamino)methyl]-2-(4-fluorophenyl)imidazo[1,2-a]pyridine-3-yl]pyrimidin-2-amine, (Compound 2, C2, Dr Simon Osbourne, LifeArc, SBC Open Innovation Campus, Stevenage) a *PfPKG* inhibitor at 1 μM for 3-4 hrs (Collins et al., 2013b). For the preparation of ring and trophozoite pellets, synchronous cultures were used without purification of parasite infected RBCs. Ring, trophozoite, schizont and mature schizont samples (Section 9.1) were treated with 10 volumes of 0.15% saponin for 5 min at 37°C to lyse the RBC and PV membranes. The parasite pellet was collected by centrifugation at 17000 x g for 5 min and washed with Phosphate buffered saline A (PBSa, pH 7.2, 137mM NaCl, 2.7mM KCl, 10mM PO₄³⁻) before being snap frozen in liquid N₂ and stored at -80°C. Schizont pellets were washed once and ring and trophozoite pellets thrice.
9.2.2 Activity-based protein labelling

The synthesis and characterisation of W-sCy5-VS have been previously described (Tan and Davison et al., 2020; Lehmann et al., 2018). DCG04 and FY01 were obtained as the kind gift of Professor Matthew Bogyo. The ActiveX™ SH probes FPP-N₃ and FPP-TMR were purchased from Thermo Scientific, FPP-Biotin was obtained as a kind gift from Dr Scott Lovell (Imperial College London). When required, FPP-N₃ labelled protein samples were treated with CuAAC-chemistry reagents to attach a sCy5-alkyne fluorophore for the visualisation of labelling (Section 9.4.2).

9.2.2.1 Probe labelling in lysates

Parasite pellets were resuspended in an equivalent volume of 0.1 % Triton X-100 (Promega) in PBSa and stored at 4°C for 20 min. The soluble fraction was recovered by centrifugation (17,000 x g, 5 min) and the insoluble pellet was discarded. The approximate protein concentration was measured using a NanoDrop 2000 spectrometer (Thermo Scientific), yielding a concentration of approximately 50 mg/ml. Parasite lysates were diluted to 2 mg/ml in PBSa before addition of the relevant probe. When required, FPP-N₃ labelled protein samples were treated with CuAAC-chemistry reagents to attach a sCy5-alkyne fluorophore for the visualisation of labelling (Section 9.4.2). Labelling reactions were quenched and prepared for SDS-PAGE by the addition of 4X Laemmli buffer (LB, 10% glycerol, 2% SDS, 67.5 mM Tris-HCl, 0.005% bromophenol blue, and 400 mM dithiothreitol (DTT)) and heating at 95°C for 5 min.

9.2.2.2 Probe labelling in intact parasites

Schizonts were diluted 1/10 in RMPI-1640 media. When indicated, inhibitor was pre-incubated with schizonts for 1 h at room temperature (RT) with gentle mixing. 0.5 µM FPP-TMR or FPP-N₃ was added (maximum 0.5 % final DMSO solvent) and samples incubated for a further 30 min. Schizonts were then washed twice with PBSa to remove excess probe. 0.15 % saponin was added (5 times schizont pellet volume) and incubated for 5 min at 37 °C. The pellet was washed with PBSa and both the saponin supernatant fraction and pellet were collected. Soluble protein lysate was extracted as above and diluted 10-fold in PBSa. When required, FPP-N₃ labelled protein samples were treated with CuAAC-chemistry reagents to attach a sCy5-alkyne fluorophore for the visualisation of labelling (Section 9.4.2). 4X LB DTT was added, and samples were incubated at 95 °C for 10 min before SDS-PAGE.

9.2.2.3 Probe labelling of recombinant proteins

Recombinant CatC was obtained as a kind gift of Professor Michael Mares (Czech Academy of Sciences, Prague). Recombinant FP2 and FP3 were obtained as a kind gifts of Dr Phillip Rosenthal (University of California, San Francisco). Recombinant DPAP3 was expressed and purified in our lab by Laura DeVries (Lehmann et al., 2018). The recombinant proteins were diluted to approximately 0.1
μM in 50 mM sodium acetate buffer (5 mM MgCl₂, 10 mM BME, 1mM EDTA, pH 5.5) and labelled with 1 μM W-sCy5-VS in acidic pH for 30 min. 4X LB DTT was added, and samples were incubated at 95°C for 10 min before SDS-PAGE.

9.2.3 SDS-PAGE and measurement of in-gel fluorescence

Soluble proteins were separated by Sodium dodecyl sulphate (SDS) polyacrylamide gel electrophoresis (PAGE) on homemade triple-wide 12.5 % polyacrylamide gels (C.B.S) in protein running buffer (25 mM Tris, 190 mM glycine, 0.1 % SDS) for 1 h55 min at 130V.

SDS-PAGE gels (Section 9.2.3) were scanned in between the glass running plates. In-gel fluorescence was measured using a PharosFX fluorescence scanner (BioRad). Different lasers were used to detect the red (TMR Ex/Em 552/572 nm) and green (sCy5 Ex/Em 647/663 nm) fluorophores. Gel images were adjusted for contrast in Adobe Photoshop. Densitometry graphs were made using data extracted from gel lane images by ImageJ™ and generated using Prism (Graph-Pad). When required, quantification of band intensity and molecular weight assignment was done using Image Lab™ software (Bio-Rad). After scanning, gels were removed from the glass running plates and stained with Quick Blue Coomassie (Triple Red) to check for consistent protein loading.

9.3 The Malaria Box ABPP competition screen

9.3.1 The Malaria Box Screening conditions

A round-bottom 96-well plate-format was used to screen 80 compounds at once. The 16 outside wells were filled with PBSa. 1 μl of a 2.5 mM DMSO stock of each Malaria Box inhibitor was added to 9 μl of PBSa in each well. At the same time, in a separate plate control compounds were prepared in the same way: Positive controls (0.5 μl of 1mM E64 (Sigma-Aldrich), 1 μl of 100mM PMSF (Sigma)); and two negative controls (1 μl DMSO). 35 μl of parasite lysate diluted in PBSa (6.67 mg/ml protein) was added to the malaria box inhibitors in the 96-well plates. Samples were incubated for 30 min at RT. FPP-TMR (1 μM) and W-sCy5-VS (10 μM) were combined and 5 μl probe mix of the probe mix was added to each reaction well (including positive/negative controls). The final reaction volume was 50 μl per well with a final concentration of 0.5 μM W- sCy5-Vs, 50 nM FPP-TMR and 50 μM inhibitor or 10 μM E64 plus 2mM PMSF for the positive control wells. The labelling reaction was quenched after 10 min by the addition of warm 4X LB and immediate incubation for 4 min in an 80°C sand bath. One of the negative controls was used to show saturation of labelling. These saturation controls received 5 μl of double probe mix to give a final 1 μM W- sCy5-Vs and 100 nM FPP-TMR, and labelling was
quenched after 45 min instead of 10 min. All steps in the screening procedure were done using a multichannel pipette. The reactions alongside the controls were run on 20 triple-wide SDS-PAGE gels and scanned for fluorescence (Section 9.2.3). After visual assessment of fluorescent gels, lane intensity profiles generated by Image Lab™ were used to confirm hits.

9.3.2 Confirmation and retesting of Malaria Box hits

The following compounds were purchased to make fresh stocks:
Life Chemicals: MMV665914 (M914, F0688-0079).
Vitas-M: MMV000986 (STL161345/STL316927), MMV011438 (M438, STK039514).
The fresh stocks of the compounds were tested against different parasite-life-stages lysates (Section 9.2.1) under the original screening conditions (Section 9.3.1). To calculate the IC_{50} of compounds inhibition, parasite lysates were labelled with the relevant probe in the presence of different concentrations of inhibitor. SDS-PAGE gels were run and scanned, and labelling intensity of inhibited bands was quantified using ImageLab (Section 9.2.3). The probe labelling intensity was taken as a measure of enzyme activity and plotted against inhibitor concentration. The data were fit to standard dose response curves (inhibitor vs. response, variable slope) in Prism and used to calculate IC_{50}s.

9.3.3 Fluorogenic assay to measure FP3 activity and inhibition.

FP3 activity was measured in 50 mM sodium acetate buffer, in a 96-well plate format at RT. rFP3 (1/1000, approximately 0.1 μM) was added to the fluorogenic substrate Z-LR-AMC (Sigma) (3-400 μM) with different concentrations of M941 (0-400 μM). Hydrolysis of the substrate was measured every 30 s for 10 min with a MSE spectrofluorometer (excitation 355 nm, emission 460 nm).

The data were analysed using Prism. The initial rates (V_0) at 50 μM Z-LR-AMC were plotted against M941 concentration. The data was fit to a standard dose response curve (inhibitor vs. response, variable slope) to calculate the IC_{50}. To model the mode of inhibition V_{0S} were plotted as a function of Z-LR-AMC concentration ([S]). The raw data were fitted to a substrate inhibition model (equation 2), which was used to calculate V_{max} and K_M values at each inhibitor concentration. Calculated V_{max} and K_M values were plot against M941 concentration to try and ascertain the mode of inhibition.
9.3.4 Standard replication assays to measure inhibition of parasite growth

96-well round-bottom plates were prepared in triplicate with serial dilutions of Malaria Box compounds. Synchronous ring-stage parasite cultures were then added to final parasitaemia of 0.1% and haematocrit of 0.5% in complete media. For the CoA rescue replication assays D-panthenol (Sigma) or the Malaria box compounds were used, with or without the addition of Coenzyme-A (Sigma) to a final concentration of 1mM.

The parasites were cultured under standard conditions (Section 9.1) Samples were taken for fixation after 72 h (unless otherwise stated), therefore parasitaemia was measured at trophozoite stage in the second cycle. Parasite cultures were fixed with 4% (v/v) paraformaldehyde (Sigma) and 0.02 (v/v) glutaraldehyde (Sigma) in PBSa for 30 min at RT before being diluted 5-fold with PBSa and stored at 4°C. Parasites were analysed by flow cytometry to measure parasitaemia (Section 9.3.5).

9.3.5 Measurement of parasitaemia by flow cytometry

Fixed cells were stained with 1 μg/ml DNA dye Hoechst 33342 (Thermo Fisher Scientific) in PBSa for 30 min at 37°C. Flow cytometry was performed using a high-throughput Fortessa analyser (BDBiosciences) using FACSDiva software v8.0.1. The forward scatter (FSC) and side scatter (SSC) height and width of cells were measured for 5000-10,000 cells per sample. The Hoechst signal was excited by the 355 nm UVA laser and detected with the 450/50 filter. Flow cytometry data were analysed using FlowJo LLC 2006-2015. Multiple gates were used to identify iRBCs and uRBCs. First, the RBC population was identified using plots of FSC/SSC area. Single RBCs were discriminated by plots of FSC and SSC height vs width. Infected RBC populations were identified by higher DNA content indicated by UVA fluorescence. Parasitaemia was calculated by the ratio of iRBC to total RBCs (uRBCs + iRBCs). The data were plotted using Prism and fit to standard dose response curves (inhibitor vs. response, variable slope) to calculate EC$_{50}$ of inhibition of parasite replication.

$$V_0 = \frac{V_{\text{max}} [S]}{K + [S]^{1 + \frac{[I]}{K}} + [S]}$$

Equation 2
9.4 Chemical Proteomics

9.4.1 Protein labelling by ABPs

9.4.1.1 Labelling of lysates for chemical proteomics

Schizont lysates were obtained as previously described (Section 9.2.1). Protein-concentration measurements were made before probe labelling using a Pierce™ BCA Protein Assay Kit (23225, ThermoFisher Scientific) per the manufacturer’s instructions.

For the CP chemical proteomics study, schizont lysates were diluted to 4 mg/ml in PBSa or sodium acetate buffer, with a total protein content of 600 μg. Three sets of labelling conditions were used: 10 μM W-sCy5-VS pre-treatment for 30 min followed by 1 μM W-dtBio-TMR-VS for 10 min; pre-treatment with DMSO for 30 min followed by 1 μM W-dtBio-TMR-VS for 10 min; treatment with DMSO only for 30 min. Affinity purification steps were then performed (Section 9.4.3).

For the initial label-free SH chemical proteomics experiments, bead optimisation studies and Malaria Box comparative proteomics studies, schizont lysates were used (Section 9.2.1). Schizont lysate was diluted to 4mg/ml protein in PBS (pH 7.4, Invitrogen, supplied as 10X and diluted to 1X in ddH₂O Pierce LC-MS grade, 600 μg protein per sample). In triplicate, diluted lysates were labelled with 1 μM FPP-N₃/FPP-Biotin or with DMSO for 1 h at RT. Labelling reactions were quenched by the addition of Triton X-100 and SDS (10% in ddH₂O, Sigma) to a final concentration of 2 and 0.2%, respectively. FPP-N₃ labelled samples were treated with CuAAC-chemistry reagents before affinity purification (Section 9.4.2), whereas FPP-Biotin were directly affinity purified (Section 9.4.3).

9.4.1.2 Labelling of live parasites and lysates for the life cycle proteomics study

Parasites at ring, trophozoite and early schizont stages were roughly synchronised by increasing the invasion time from 2 to 5 hrs (Section 9.1). For ring stage parasites, a time window of 0-20 hpi was achieved by mixing two ring populations that had been synchronised 10 hrs apart. This was not deemed necessary to obtain trophozoite (20-36 hpi) or early schizont (36-44 hpi), as parasites diverge from synchronicity slightly throughout the life cycle. Parasite life stages were monitored using Giemsa-stained thin blood smears. Highly mature schizonts were obtained following C2 treatment (Section 9.2.1). Live parasites were diluted 1/10 in RMPI-1640 media. 1 μM FPP-N₃ was added, and samples incubated for 1 h at RT with gentle shaking. Parasites were then washed twice (1 ml PBSa) to remove excess probe. 0.15 % saponin was added and incubated for 5 min at 37 °C. The pellet was washed (1ml PBSa, 1 wash for early/mature schizonts, 3 washes for rings/trophozoites) and snap frozen in liquid N₂.
Soluble protein lysate was extracted by the dilution of pellets two-fold in 0.01% Triton-X in PBS. The mixture was incubated on ice for 20 min, and then the soluble fraction obtained by microcentrifugation.

Large quantities of pure merozoites pellets were collected from egressing culture (Blackman, 1995). Mature schizonts were isolated using Percoll (Section 9.2.1) and washed with RPMI-1640. Schizonts were then diluted in 5% horse serum in RPMI-1640 supplemented with 0.292 g/L L-glutamine and kept at 37°C until merozoites were observed using Giemsa-stained thin blood smears (Section 9.1). The schizonts were removed from the culture by multiple rounds of centrifugation (1200 rpm for 3 min). Further residual schizonts and haemozoin-associated erythrocyte debris was filtered out by passing the merozoite supernatant through a magnet (SuperMACS). The magnet attracts free and parasite-containing haemozoin crystals, thus removing them from the merozoite supernatant. The merozoites were then pelleted by fast centrifugation (17,000 x g) and frozen in liquid nitrogen. Soluble protein was extracted by diluting merozoite pellets 10-fold in PBS, 0.01% Triton X. The mixture was incubated on ice for 20 min, and then the soluble fraction obtained by microcentrifugation. Merozoite lysates were labelled with 1 μM FPP-N3 for 1 hr. The protein concentration of parasite lysates was measured using a BCA Protein Assay Kit. Click-chemistry reactions were then performed on these samples (Section 9.4.2).

### 9.4.2 Tandem-tagging by CuAAC-chemistry

Protein lysate labelled with FPP-N3 was diluted two-fold to 2 mg/ml protein in 1 % Triton X-100 and 0.1 % SDS in PBS (282 μl). The CuAAC-reagents were mixed together in the following order: 3 μl of 10 mM biotin-alkyne; 6 μl of 50 mM CuSO4; 6 μl of 50 mM Tris (2-carboxyethyl)phosphine hydrochloride (TCEP, Aldrich); and 3 μl of 10 mM tris[1-(benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA, Sigma). 18 μl of the CuAAC-mix was added to each sample to give final concentrations of 100 μM biotin-alkyne, 1 mM CuSO4, 1 mM TCEP, 100 μM TBTA. The CuAAC-reaction was incubated at RT for 1h with gentle mixing before quenching with 3.3 μl 5 mM ethylenediaminetetraacetic acid (EDTA).

### 9.4.3 Affinity purification of labelled proteins

Excess probes (and CuAAC-reagents if applicable) were removed by MeOH/CHCl3 precipitation (2 vol MeOH (Fisher), 0.5 vol CHCl3 (Fisher), 1 vol ddH2O (Pierce LC-MS grade), centrifuged at 17,000 x g 2 min, washed with 1ml MeOH and dried in fume hood). Protein was resolubilised in by vortexing for 15 min at RT in 2% SDS (Sigma), 10 mM DL-Dithiothreitol (DTT, Sigma) in PBS. Diluted protein (600 μl, 1 mg/ml, 0.2 % SDS, 1 mM DTT in PBS) was incubated with 30 μl combined NeutrAvidin-agarose slurry (10 μl 4%-crosslinked NeutrAvidin-agarose 50% slurry (Thermo), 20 μl blank-agarose
50% slurry, (Pierce)) for 2 hr at RT with gentle shaking. The bead-mix was centrifuged (3000 x g, 2 min) to collect the supernatant. The beads were then washed sequentially with 3 x 1% SDS in PBS, 2 x 4M Urea in 50mM Ammonium bicarbonate (AMBIC, bioultra 99.5% Sigma in dH2O) and 2 x AMBIC only.

Proteins were then treated to reduce and alkylate free thiols. For the initial label-free SH chemical proteomics with FPP-N\textsubscript{3}/FPP-Biotin, beads were treated with DTT (10 mM in AMBIC) for 1h at 37°C and then washed with 50 mM AMBIC. Beads were then treated with 10 mM iodoacetamide (IAA) (NEM, Sigma) in AMBIC for 30 min at RT before being washed again with AMBIC. For all subsequent experiments TCEP (10 mM in AMBIC) was used in place of DTT and 10 mM NEM (Sigma) in AMBIC was used in place of IAA.

9.4.4 Trypsin digestion of proteins and peptide sample clean-up by stage tipping

To digest proteins 0.12 μg Trypsin was added to the beads in 50 μl 50 mM AMBIC (final concentration 2.4 ng/μl, Trypsin gold Promega UK Ltd, Cat. # V5280) and left to digest proteins shaking overnight at 37°C. The beads were then washed with 70 μl AMBIC and 70 μl 1.5 % Trifluoroacetic acid (TFA, Thermo) in dH\textsubscript{2}O, and the supernatants combined. The combined peptide mixture was filtered and desalted on homemade ‘stage tips’ containing C18 Empore membrane (Empore Octadecyl C18 47mm Extraction Disks 2215 Supleco 66883-U). Tips were washed by adding 150 μl MeOH and then 150 μl H\textsubscript{2}O and centrifuging (2000 x g, 2 min). The peptide sample (140 μl) was loaded (2000 x g, 1 min) and the water wash repeated. Peptides were eluted with 60 μl 40 % Acetonitrile (ACN, 99.8% anhydrous, Sigma) in ddH\textsubscript{2}O (2000 x g, 1 min). The peptide solution was then evaporated to dryness using a speed vacuum (Savant Integrated SpeedVac system ISS 100). Samples for label-free quantification were now ready for solubilisation for injection into the LC-MS/MS (Section 9.4.6). Samples for isobaric-label tagging and multiplexing continued to the TMT-labelling stage. For all steps of proteomics protocols 1.5ml Eppendorf\textsuperscript{®} LoBind microcentrifuge tubes (Sigma) were used to minimise any protein losses.

9.4.5 TMTsix/tenplex\textsuperscript{TM} labelling and high pH fractionation

Peptide samples were dissolved in 25 μl of 50 mM Triethylammonium bicarbonate buffer (TEAB, Sigma). 0.2 mg of TMT label reagents (ThermoFisher TMTsixplex\textsuperscript{TM} product number 90308, lot number RH239931 or TMTtenplex\textsuperscript{TM} product number 90309, lot number RH239932) were dissolved in 65 μl anhydrous ACN. 20 μl of the appropriate TMT label was added to each replicate and samples incubated for 1h at RT. To check labelling efficiency, 1 μl of TMT labelled sample was added to 49 μl 5 % ACN, 0.1 % TFA and analysed by LC-MS/MS as described below (Section 9.4.6). All TMT-
six/tenplex labelled samples were combined into one sample for each replicate. Combined TMT samples were separated into 8 fractions using a high pH reversed-phase peptide fractionation kit (Pierce 84868). Fractionated samples were then evaporated to dryness by vacuum centrifugation (Centrivap Concentrator, Labconco).

### 9.4.6 Proteomics method

The peptide mixture residues were resuspended in 15 μl 0.1% TFA, 10 μl were used per injection unless otherwise stated. Peptides were chromatographically resolved on an Ultimate 3000 nanoRSLC HPLC (Thermo Scientific). A 2 mm x 0.3 mm Acclaim Pepmap C18 trap column (Thermo Scientific) was used at a flow rate of 15 μl/min prior to the trap being switched to elute at 0.25 μl/min through a 50 cm x 75 μm EasySpray C18 column. In a 90-min run the following gradients were used: 9% - 25% solution B over 37 min, then 25% - 40% B over 18 min followed by a short gradient to 100% B and back down to 9% B, and finally a 20 min equilibration in 9% B (A = 2% ACN, 0.1% formic acid; B = 80% ACN, 0.1% formic acid).

HPLC eluent was introduced into an Orbitrap Fusion Lumos (Thermo Scientific). The Orbitrap was operated in “Data Dependent Acquisition” mode with a survey scan at a resolution of 120,000 from m/z 400-1400. This was followed by MS/MS using 38 % high energy collision dissociation (HCD). Dynamic exclusion was used with a time window of 30 s.

### 9.4.7 Proteomics analysis

The raw data files were analysed using MaxQuant version 1.6.2.1 (Tyanova et al., 2016a). Quantification was done at MS1 level by LFQ and iBAQ, or at MS2 level using TMT-six/tenplex labels. All other MaxQuant settings were kept the same as default. The MaxQuant search engine Andromeda (Cox et al., 2011) was used with sequence databases *Homo sapiens* (Uniprot 13/01/2013) and *Plasmodium falciparum* 3D7 (PlasmoDB 15/12/2016). A decoy database of reversed sequences was used to filter false positives at a peptide false detection rate of 1 %.

### 9.4.8 Interpretation of data

The data files generated by MaxQuant were further analysed using Perseus version 1.5.6.0 (Tyanova et al., 2016b). The list of identified proteins was first filtered to remove potential contaminants, proteins only identified by one site and proteins identified from the reverse (decoy) peptides (Section 1.6.2.2). The protein intensity values were then transformed by function log₂(x). The log₂ intensities were normalised by subtracting the median intensities of each replicate followed by the median intensities of each protein, unless otherwise stated. Proteins identified from less than 2 peptides (4 peptides for life
cycle proteomics study) were filtered out at this stage. Student t-tests were performed to identify proteins with statistically significant changes between conditions using stringency parameter $s_0 = 0.5$ and FDR = 0.01.

9.5 Human SHs

9.5.1 Monitoring of parasite replication in the presence of HsSH inhibitors

The human SH inhibitors were purchased commercially (Table 9.1). Standard 72-hreplication assays were set up in triplicate to observe the effect of the HsSH inhibitors on parasite replication (Section 9.3.4). Parasites were fixed after 72 hrs, DNA stained and the parasitaemia measured by flow cytometry (Section 9.3.5). The data were plotted using Prism and fitted to a dose response curve (inhibitor vs response, variable slope) to calculate the EC$_{50}$ of parasite growth inhibition.

<table>
<thead>
<tr>
<th>Human SH</th>
<th>Inhibitor</th>
<th>Catalogue number</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAFAH1B3</td>
<td>P11</td>
<td>CAY17507</td>
<td>Cambridge bioscience</td>
</tr>
<tr>
<td>FASN</td>
<td>TBV-3166</td>
<td>SML1694</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td>PNPLA6</td>
<td>TOCP</td>
<td>S1885</td>
<td>Supelco</td>
</tr>
<tr>
<td>LYPLA1</td>
<td>ML-348</td>
<td>SML1901</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td>LYPLA2</td>
<td>ML-349</td>
<td>SML1918</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td>AChE</td>
<td>PyBr</td>
<td>P9797</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td>NCEH1</td>
<td>JW480</td>
<td>SML0792</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td>APEH</td>
<td>AA74-1</td>
<td>SML0358</td>
<td>Sigma-Aldrich</td>
</tr>
</tbody>
</table>

LYPLA1 and APEH inhibitors (ML-348 and AA74-1) were also tested in an extended replication assay to monitor the effect of the inhibitors on parasite growth. Replication assays were set up in triplicate with DMSO, two control compounds or inhibitor: ML-348 (5 μM); AA71-1 (0.625 μM); C2 (1 μM); or CQ (2 μM). Parasites were cultured under standard conditions (Section 9.1). Samples were fixed (Section 9.3.4) after 24, 41, 48 or 72 hrs. Fixed cells were stained with 1 μg/ml Hoechst and 2 μM 132A RNA dye (kind gift of Prof Young-Tae Change, University of Singapore, Cervantes et al., 2009) in PBSa for 30 min at 37 °C. Samples were analysed by flow cytometry as described (Section 9.3.5), additionally detecting the RNA 132A signal which was excited by the 488 nm blue laser and detected using a the 610/20 nm filter.

Using the DNA content, iRBC and uRBC populations were gated and quantified (Section 9.3.5). Flow cytometry plots were generated in FlowJo. The DNA and RNA content was used to follow parasites throughout the life cycle and therefore distinguish which cycle they belong to: cycle 1 (rings through to schizonts 0-48 hrs) or cycle 2 (newly invaded rings at 48 hrs, grown to trophozoite stage by 72 hrs). The parasitaemia arising from each population was calculated as a percentage of total RBCS (uRBC...
and iRBC). The mean RNA and DNA content for each population was quantified and normalised by dividing the median RNA and DNA content of uRBCs. The parasitaemia, DNA content and RNA content of populations treated with the different inhibitors were plotted for each time point in Prism as the mean of three replicates with error bars showing the standard error.

9.6 Genetic validation of SHs

9.6.1 Molecular biology techniques for the generation of plasmids for genetic modification.

A conditional allelic replacement strategy was designed and applied to investigate four SHs in \textit{P. falciparum}. Molecular biology techniques were used to create conditional allelic replacement plasmids, as well as control plasmids, for each GOI (Fig. 9.6). The plasmid backbone contained a 3’UTR from \textit{P. berghei} dihydrofolate reductase thymidylate synthase and a \textit{hdhfr} cassette. The plasmid was targeted to the GOI by the HR. A synthetic, recodonised version of the remaining 3’ sequence of the gene (recodonised region 1, RR1) was cloned downstream of the HR. RR1 was followed by a triple HA tag (3xHA) and linked to \textit{npt} by a T2A sequence, this was all flanked by two LoxPint sites. A second RR followed (RR2) containing either an Ala (pT2A-GOI-cMut) or Ser (pT2A-GOI-cWT) catalytic codon, this was linked to \textit{gfp} reporter by a second T2A sequence.

9.6.1.1 Preparation of the plasmid vector backbone

All transfection constructs were based on the pT2A_Ddi1-comp construct (kind gift of Sophie Ridewood which was in turn based on the pT2A_FIKK10_cKO construct, kind gift of Dr Moritz Treeck, The Francis Crick Institute). The pT2A_Ddi1-comp construct was digested with SalI and BglII restriction enzymes (New England Biosciences, as per manufacturer’s instructions) (Fig. 9.6). The digestion products were separated by agarose electrophoresis and the linear vector purified using QIAquick Gel Extraction Kit (Qiagen) as per manufacturer’s instructions.

9.6.1.2 PCR amplification of genomic HRs and synthetic DNA fragments

Genomic DNA was extracted from parasite pellets (Section 9.6.1.2) using a DNeasy Blood and Tissue Kit (Qiagen), as per manufacturer’s instructions and stored at 4°C. Homology regions (HR) were amplified from each GOI from \textit{P. falciparum} B11 gDNA (Perrin et al., 2018), using primers that created overlapping regions with the vector and synthetic DNA fragments on each end (Fig. 9.6A, primers 1 and 2, Table 9.2-9.4).
Synthetic DNA fragments containing the rest of the construct design including the RR1, T2A and \textit{npt} gene flanked by \textit{loxP}ints and the second RR and T2A were made by GENEWIZ (Fig. 9.6A). The synthetic DNA was amplified in two pieces using primers that left overlaps with the vector and HR fragments on each end (Fig. 9.6, primers 3 and 4, primers 5 and 6, Table 9.2-9.4).

Phusion High-fidelity DNA polymerase was used in PCR reactions (New England Biolabs) as per manufacturer’s instructions. Oligonucleotide primers were used at 200nM (Table 9.2-9.4). PCR products were purified using QIAGen QIAquick PCR Purification Kit as per manufacturer’s instructions. Oligonucleotide primers were obtained from Sigma (Table 9.2-9.4). PCRs were carried out using a thermocycler and amplicons were analysed by examining 1% agarose gel electrophoresis stained with SYBR Safe (ThermoFisherScientific). DNA fragments were purified using QIAGen QIAquick PCR Purification Kit. DNA concentration was measured using a NanoDrop 2000 spectrophotometer (ThermoFischerScientific) via the absorbance at 260 nm.

9.6.1.3 Site directed mutagenesis

Site directed mutagenesis was performed on the synthetic DNA fragment to change the Ala-encoding codon back to a catalytic Ser-encoding codon for the conditional WT constructs (Fig. 9.6B). Oligonucleotide primers were designed to template the new desired sequence, changing one or two bases to give a codon change from the mutant Ala in the synthetic DNA back to wild type Ser catalytic residue codon (Fig. 9.6B). The primers were used to amplify the synthetic DNA amplicon containing RR2 in two sections, 3’ (primers 5 and 7) and 5’ (primers 8 and 6, Table 9.2-9.4) to the catalytic residue (Fig. 9.6B). The two pieces were combined by overlapping PCR amplification (primers 5 and 6, Table 9.2-9.4) to reconstitute the synthetic DNA amplicon now containing the Ala-Ser mutation (Fig. 9.6B).

9.6.1.4 In-Fusion

The linearised pT2A vector, HR and synthetic DNA were then combined by In-Fusion reactions to give the final pT2A\_GOI-cMUT and pT2A\_GOI-cWT constructs (Fig. 9.6C). Molecular cloning steps were designed such that adjacent DNA fragments had 15 nucleotide overlaps with each other and the vector. Multi-fragment ligation was performed using the In-Fusion HD Cloning Kit (ClonTech) as per the manufacturer’s instructions (Fig. 9.6).

9.6.1.5 Transformation of \textit{E. coli} and preparation of plasmid DNA

In order to amplify final integration plasmids, XL10-Gold Ultracompetant cells (Ailgent) were transformed with plasmid, as per manufacturer’s instructions, and grown in lysogeny broth at 37 °C with shaking (225 rpm). Plasmid DNA was purified from bacterial cultures using QIAprep Midi Kits (Qiagen). Colony PCRs were performed on bacterial clones using KAPA2G Fast HotStart ReadyMix.
(Kapa Biosystems) as per manufacturer’s instructions (Fig. 9.6 primers 10 and 4). The final constructs were checked via capillary sequencing by Source Bioscience.

### Table 9.2 PCRs for the construction of transfection plasmids

<table>
<thead>
<tr>
<th>Gene</th>
<th>PCR ID</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Product / bp</th>
<th>Polymerase</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ID</td>
<td>Figure label</td>
<td>ID</td>
<td>Figure label</td>
</tr>
<tr>
<td>Pst1 / PF3D7_1476700</td>
<td>HR amplification</td>
<td>218</td>
<td>1</td>
<td>225</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Synthetic DNA amplification 1/2</td>
<td>223</td>
<td>3</td>
<td>256</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Synthetic DNA amplification 2/2</td>
<td>257</td>
<td>5</td>
<td>236</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Mutation PCR amplification 2/2 SER a</td>
<td>257</td>
<td>5</td>
<td>240</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Mutation PCR amplification 2/2 SER b</td>
<td>239</td>
<td>8</td>
<td>235</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Overlapping PCR 2/2 SER</td>
<td>257</td>
<td>5</td>
<td>235</td>
<td>6</td>
</tr>
<tr>
<td>Colony PCR</td>
<td>119</td>
<td>10</td>
<td>256</td>
<td>4</td>
<td>1488</td>
</tr>
<tr>
<td>Pst2 / PF3D7_1476800</td>
<td>HR amplification</td>
<td>220</td>
<td>1</td>
<td>229</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Synthetic DNA amplification 1/2</td>
<td>232</td>
<td>3</td>
<td>256</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Synthetic DNA amplification 2/2</td>
<td>257</td>
<td>5</td>
<td>236</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Mutation PCR amplification 2/2 SER a</td>
<td>257</td>
<td>5</td>
<td>242</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Mutation PCR amplification 2/2 SER b</td>
<td>241</td>
<td>8</td>
<td>236</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Overlapping PCR 2/2 SER</td>
<td>257</td>
<td>5</td>
<td>236</td>
<td>6</td>
</tr>
<tr>
<td>Colony PCR</td>
<td>119</td>
<td>10</td>
<td>256</td>
<td>4</td>
<td>1545</td>
</tr>
<tr>
<td>abH112 / PF3D7_1120400</td>
<td>HR amplification</td>
<td>284</td>
<td>1</td>
<td>269</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Synthetic DNA amplification 1/2</td>
<td>271</td>
<td>3</td>
<td>256</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Synthetic DNA amplification 2/2</td>
<td>257</td>
<td>5</td>
<td>270</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Mutation PCR amplification 2/2 SER a</td>
<td>257</td>
<td>5</td>
<td>278</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Mutation PCR amplification 2/2 SER b</td>
<td>277</td>
<td>8</td>
<td>270</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Overlapping PCR 2/2 SER</td>
<td>257</td>
<td>5</td>
<td>270</td>
<td>6</td>
</tr>
<tr>
<td>Colony PCR</td>
<td>119</td>
<td>10</td>
<td>256</td>
<td>4</td>
<td>1363</td>
</tr>
<tr>
<td>abH114 / PF3D7_1143000</td>
<td>HR amplification</td>
<td>283</td>
<td>1</td>
<td>274</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Synthetic DNA amplification 1/2</td>
<td>275</td>
<td>3</td>
<td>256</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Synthetic DNA amplification 2/2</td>
<td>257</td>
<td>5</td>
<td>276</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Mutation PCR amplification 2/2 SER a</td>
<td>257</td>
<td>5</td>
<td>281</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Mutation PCR amplification 2/2 SER b</td>
<td>280</td>
<td>8</td>
<td>276</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Overlapping PCR 2/2 SER</td>
<td>257</td>
<td>5</td>
<td>276</td>
<td>6</td>
</tr>
<tr>
<td>Colony PCR</td>
<td>119</td>
<td>10</td>
<td>256</td>
<td>4</td>
<td>1581</td>
</tr>
</tbody>
</table>

### Table 9.3 PCRs for the interrogation of parasite lines

<table>
<thead>
<tr>
<th>Gene</th>
<th>PCR ID</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Product / bp</th>
<th>Polymerase</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ID</td>
<td>Figure label</td>
<td>ID</td>
<td>Figure label</td>
</tr>
<tr>
<td>Pst1 / PF3D7_1476700</td>
<td>Endogenous gene PCR</td>
<td>307</td>
<td>11</td>
<td>309</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>5' integration PCR</td>
<td>307</td>
<td>11</td>
<td>291</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>3' integration PCR</td>
<td>119</td>
<td>10</td>
<td>209</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Excision PCR</td>
<td>289</td>
<td>15</td>
<td>198</td>
<td>16</td>
</tr>
<tr>
<td>Pst2 / PF3D7_1476800</td>
<td>Endogenous gene PCR</td>
<td>311</td>
<td>11</td>
<td>313</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>5' integration PCR</td>
<td>311</td>
<td>11</td>
<td>294</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>3' integration PCR</td>
<td>119</td>
<td>10</td>
<td>313</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Excision PCR</td>
<td>292</td>
<td>15</td>
<td>198</td>
<td>16</td>
</tr>
<tr>
<td>abH112 / PF3D7_1120400</td>
<td>Endogenous gene PCR</td>
<td>315</td>
<td>11</td>
<td>317</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>3' integration PCR</td>
<td>119</td>
<td>10</td>
<td>296</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Excision PCR</td>
<td>295</td>
<td>15</td>
<td>198</td>
<td>16</td>
</tr>
<tr>
<td>abH114 / PF3D7_1143000</td>
<td>Endogenous gene PCR</td>
<td>319</td>
<td>11</td>
<td>321</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>3' integration PCR</td>
<td>119</td>
<td>10</td>
<td>299</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Excision PCR</td>
<td>257</td>
<td>15</td>
<td>198</td>
<td>16</td>
</tr>
</tbody>
</table>

*The product size before/after rapamycin-mediated excision is indicated
<table>
<thead>
<tr>
<th>Primer ID</th>
<th>Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>119</td>
<td>ACACGTATGACCATGATTACGC</td>
</tr>
<tr>
<td>198</td>
<td>CCTCTCCACTGACAGAAAATTTGTGCC</td>
</tr>
<tr>
<td>218</td>
<td>ACTTAGAGGGAGATCTTAATATATAATAACCTATTGAGAA</td>
</tr>
<tr>
<td>220</td>
<td>ACTTAGAGGGAGATCTCAAGAAATATATTATATTGAGAA</td>
</tr>
<tr>
<td>223</td>
<td>ATAGCGATATATAGGTAAATAAATATACATAAATACGTAGCATACAT</td>
</tr>
<tr>
<td>225</td>
<td>CTATAATAGTTATCTCTTTATACATTTGGTTC</td>
</tr>
<tr>
<td>229</td>
<td>CTGCTCCTCTTATCAATGTGTAGCTTCTCTGGTTC</td>
</tr>
<tr>
<td>232</td>
<td>TTAGATAGAGACGTGTAATAAATATACATAAATACGTAGCATACAT</td>
</tr>
<tr>
<td>235</td>
<td>TTCTTCTCCGTGACACTCA</td>
</tr>
<tr>
<td>236</td>
<td>TTCTTCTCCGTGACACTTC</td>
</tr>
<tr>
<td>239</td>
<td>TTGTTATAGTAGTGAGCTAATAATACCTTAGGCC</td>
</tr>
<tr>
<td>240</td>
<td>ATACTATAACCAAACTGGAATATTTGGTA</td>
</tr>
<tr>
<td>241</td>
<td>GCTTTTATGATGAGTGCCAAATATATGTAGCCG</td>
</tr>
<tr>
<td>242</td>
<td>ACCGATCATAAAAAACGGCAGAAGATATTGGGTA</td>
</tr>
<tr>
<td>256</td>
<td>CGAATGCTCCCTCCACCA</td>
</tr>
<tr>
<td>257</td>
<td>TGGAGAGGCTATCCGTGCTATGAC</td>
</tr>
<tr>
<td>269</td>
<td>CTAAATCATAACCACTATAACTGATGAA</td>
</tr>
<tr>
<td>270</td>
<td>CTCTCTCTCTCTCTCTCGACCTCTGGTTC</td>
</tr>
<tr>
<td>271</td>
<td>GTGTTATGATGAGTAAATAAATACAGCTAGTG</td>
</tr>
<tr>
<td>274</td>
<td>ATGCATATATTCTGCTGATGCTTGAATTCC</td>
</tr>
<tr>
<td>275</td>
<td>CCAGGAAATATCGGAGATTGAGAAATTGGTA</td>
</tr>
<tr>
<td>276</td>
<td>AAATAAATCCTCCTCTCTCTCGACCTAGGG</td>
</tr>
<tr>
<td>277</td>
<td>TGGCATTTCTGTGATCTGCAACCAGTGGTAA</td>
</tr>
<tr>
<td>278</td>
<td>CCAGAGAAGTGCTCATAACAAATATATCTC</td>
</tr>
<tr>
<td>280</td>
<td>GATCTTCTTATAGAGAAAGAGACGCTATGC</td>
</tr>
<tr>
<td>281</td>
<td>CCTCTCTAAAGGAGACTCATAAGCAGAC</td>
</tr>
<tr>
<td>283</td>
<td>AGACTTGGAGGAGAGATCTCTCCCCATATAAATCCTGCAAGAAGAGCTGC</td>
</tr>
<tr>
<td>284</td>
<td>AGACTTGGAGGAGAGATCTGGAAGGCTAGGGTTAGGGAGCATGGGTGAAGAAGCTGC</td>
</tr>
<tr>
<td>289</td>
<td>AGTACATGGGTAGCTCTCTCACATAAAGAT</td>
</tr>
<tr>
<td>291</td>
<td>GCTACATATGCCAATGATATATCTCC</td>
</tr>
<tr>
<td>292</td>
<td>TACATGGGTAGCTCTCTCACATAAAGG</td>
</tr>
<tr>
<td>294</td>
<td>CTTAGATATGCTACGCTGACCTCTTCTTCTAGCTGACTC</td>
</tr>
<tr>
<td>295</td>
<td>GAACTACAGATATTTGTTTATATGACAC</td>
</tr>
<tr>
<td>296</td>
<td>AGCATTTCACCTCCCAGAATCTGGAAG</td>
</tr>
<tr>
<td>299</td>
<td>GGAATCTCTTAAATAGTGGTGCTAGCGTGAAGG</td>
</tr>
<tr>
<td>307</td>
<td>CTACTATTAAATATAAATAGTGAAACACA</td>
</tr>
<tr>
<td>309</td>
<td>CATATATATTATATCATAAGCCTTTG</td>
</tr>
<tr>
<td>311</td>
<td>TTACATGTGTATAGTTGGAA</td>
</tr>
<tr>
<td>313</td>
<td>AAAGGATCTCAGTGTAGACAAA</td>
</tr>
<tr>
<td>315</td>
<td>CATCTCTGTGTATGACCATATAGGTAG</td>
</tr>
<tr>
<td>317</td>
<td>ATCTCTATAATCTGATAGCAGTAAACATC</td>
</tr>
<tr>
<td>319</td>
<td>CATATATACCTTTATACAGGTTGTGTGTGT</td>
</tr>
<tr>
<td>321</td>
<td>CATACATATAAAACTCTCTTTATATAGGC</td>
</tr>
</tbody>
</table>
Figure 9.6: The generation of pT2A-GOI-cMut/WT constructs for transfection. A) The GOI homology region (HR) was amplified using PCRs of genomic DNA. The LoxPint sites, recodonised regions (RR) and npt DNA was ordered from GeneWiz and amplified in two parts from the synthetic plasmid. The template construct was digested by BgII and SalI to give a linear vector. The four DNA fragments were combined via an In-Fusion reaction to give the new construct for allelic replacement strategy (C) B) Overlapping PCR was used to perform site directed mutagenesis on the RR2 containing synthetic DNA fragment to change the Ala codon to a Ser codon. C) The common architecture of all conditional allelic replacement plasmids for transfection. The plasmid backbone contains a 3'UTR from P. berghei dihyrofolate reductase thymidylate synthase, and a hdtf cassette. The plasmid is targeted to the GOI by the HR. The rest of the gene follows recodonised (RR1), tagged with triple HA tag (3xHA) and linked to npt by a T2A sequence, this is all flanked by two LoxPint sites. A second RR follows (RR2) containing either an Ala (pT2A-GOI-cMut) or Ser (pT2A-GOI-cWT) catalytic codon, this is linked to gfp reporter by a second T2A sequence.
9.6.2 Transfection, cloning and testing of *P. falciparum* conditional allelic replacement lines

9.6.2.1 Transfection of *P. falciparum*

Plasmid DNA (10 μg) to be transfected was precipitated in EtOH and re-suspended in 10 μl sterile TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). The DNA was combined with 100 μl Amaza Primary cell solution P3 (P3 Primary cell 4-D Nucleofector X kit, Lonza) and used to resuspend 15-20 μl of Percoll-purified *P. falciparum* B11 schizonts. The schizont solution was electroporated using Amaza 4-D electroporator (Lonza) at FP158 pre-set condition. Immediately, the schizonts were transferred to 2 ml of RBCs suspended in complete media at 20% haematocrit and incubated for 30 min at 37°C with shaking (100 rpm) to allow egress and invasion to take place. A further 8 ml of complete media was then added, and the cultures were incubated at 37 °C under standard conditions. After approximately 18 hrs, the complete media was replaced and 2.5 nM WR99210 (WR, Jacobus Pharmaceuticals) was added to select for parasites containing the pT2A-cMUT/WT plasmid that expressed the *hdhfr* selection marker.

The transfected parasites were cultured under standard conditions in the presence WR, and parasitaemia was monitored twice weekly using light microscopy (Section ). When WR resistant parasites were observed at 1% parasitaemia (approximately 2 weeks), Geneticin selective Antibiotic G418 Sulphate (G418, Gibco) was added at 450 μl/ml to select for chromosomal integration of the episome (selection linked integration, (SLI); Section 1.7.2.2; Birnbaum et al., 2017). When 1% parasitaemia of WR and G418 resistant parasites was observed (2-4 weeks) samples were taken for storage by cryopreservation (Section 9.6.2.2) and confirmation of integration by PCR (Section 9.6.1.2). Integrated parasite lines were cultured as standard with the addition of 2.5 nM WR and 225 μg/ml G418 (media+WR+G418).

9.6.2.2 Cryopreservation and thawing of parasite lines

To cryopreserve parasite lines, concentrated samples of ring-stage parasite culture at 1-5% parasitaemia were re-suspended 1:1 in freezing solution (28% (v/v) glycerol, 3% (w/v) sorbitol, 65% (w/v) NaCl in dH₂O), and 1ml of the mixture was transferred to cryovials (Nunc) and stored in liquid N₂. To thaw cryopreserved parasite lines, cryovials were warmed at 37 °C before the contents was washed twice with 5 ml thawing solution (3.4 % (w/v) NaCl). Parasites were then transferred to complete media+WR+G418 and cultured as described.

9.6.2.3 Cloning of *P. falciparum* lines

WR- and G418-resistant transfected parasite lines were cloned using a plaque assay (Thomas et al., 2016). Mixed stage parasite culture at 0.75 % haematocrit was 5-fold serial diluted in complete
media+WR+G418 across a 96-well flat-bottomed plate (Corning Costar). At least 3 plates were set up for each mixed line and were incubated under standard conditions. After 12 days, wells that contained a single plaque in the RBC layer, (corresponding to a single clonal parasite population) were identified by inverted light microscopy. These wells (at least 5 for each mixed line) were given fresh media+WR+G418 with 1% haematocrit and transferred to 96-well round-bottomed plates. Plates were fed with fresh media+WR+G418 every 4-5 days. When parasites were detectable by light microscopy (Section 9.1) they were tested for integration by direct blood PCR (Chapter 3.1.2). Established integrated clonal lines were transferred to flasks and grown until at least 1% parasitaemia was observed. Genomic DNA was extracted to further support clonal identity (Section 9.6.1.2), and samples were taken for cryopreservation (at least 2 clonal lines per transfection; Section 9.6.2.2).

Diagnostic PCRs on potential clonal parasite lines were performed in the first instance using Phusion Blood Direct PCR Master Mix (Thermoscientific). Follow-up diagnostic PCRs to check for integration/excision were performed using Phusion High-fidelity DNA polymerase (New England Biolabs) on purified genomic DNA (Section 9.6.1.2).

### 9.6.2.4 RAP treatment to induce DiCre mediated DNA excision

Synchronous, ring-stage (0 hpi, cycle 1), clonal lines from the transfection of *P. falciparum* B11 parasites were treated with 20 nM RAP (Sigma-Aldrich) or an equal volume of DMSO for 12-18 h at 37 °C. The media was then replaced with fresh complete media + WR and cultured as standard. Genomic DNA was extracted for diagnostic PCRs to confirm excision (Section 9.6.1.2). Parasite samples were also taken for protein analysis by western blot (Section 9.2.1). Small quantities of merozoites from Pstal-cMUT lines for use in PCRs and western blots were obtained simply by removing schizonts from egressing culture by multiple rounds of centrifugation (1200 rpm for 3 min) and then pelleting merozoites from the resulting supernatant (17,000 x g).

The effect of the RAP-induced conditional allelic replacement was assessed using a standard replication assay over three or four cycles (Section 9.3.4). Immediately after rapamycin was removed at approximately 18 hpi, parasites were diluted to 0.1% parasitaemia and 0.5% haematocrit in complete media and plated in triplicate in 96-well round-bottom plates. Samples were taken for fixation at trophozoite stage in each cycle at approximately 30 hpi (Section 9.3.4). Parasitaemia at each time point was measured by flow cytometry (Section 9.3.5) and the data were plotted in Prism.
9.6.2.5 Western blot analysis

Soluble parasite protein (Section 9.2.1) was obtained and run on SDS-PAGE gels (Section 9.2.3). After fluorescent imaging (Section 9.2.3), gels were transferred onto nitrocellulose membrane (ThermoFisherScientific) over 5-15 hrs using Appleton Wood wet blotter apparatus in transfer buffer (25 mM Tris-HCL, 190 mM glycine, 20 % methanol). Blots were blocked in 5 % Bovine Serum Albumin (BSA, Sigma) in PBST (PBSa with 1 % Tween-20) for 2 h at RT. The primary antibody (Ab) against the HA tag (Anti-HA 3F10 monoclonal Rat Ab, Roche) was used at 1/1000 in 2% BSA. The blot was incubated with the Anti-HA Ab for 1 h with agitation before it was washed 3 times for 5 min in PBST. A secondary Ab linked to horseradish peroxidase (HRP) was then used. Anti-rat-HRP (Goat, polyclonal, anti-rat-HRP, Bio-Rad) was diluted 1/10000 in 2 % BSA for 1 h with gentle agitation followed by washing with PBST, 3 times for 5 min. HRP signal was activated by adding Immobilon Western Chemiluminescent HRP Substrate (Millipore) and visualised using a ChemiDoc Imager (Bio-Rad).
10.1 Optimisation of lysis conditions for ABPP

Supplementary figure 10.1: Lysis of parasite pellets with 0.1% Triton X allowed more parasite proteins to be labelled than any other method. The clearest and most efficient labelling is seen with parasite pellets from saponin lysed parasites. Schizonts were purified from synchronous culture, some were treated with 0.15% saponin and washed before freezing at -80°C, some were frozen directly. The saponin fraction was kept. Three different methods were used in combinations to extract proteins from defrosted parasite pellets: 5 cycles of freeze-thaw in liquid N\textsubscript{2}, treatment with 1% NP40 for 20 min at 4°C, and treatment with 0.1% Triton X for 20 min at 4°C. In each case the soluble protein fraction was obtained by centrifugation. The protein concentration of soluble lysates was tested using a Nanodrop and all samples were diluted to 5mg/ml. The lysates and saponin fraction were labelled with probe (1 \, \mu M/500 \, nM W-sCy5-VS or 100 \, nM/ 50 \, nM FPP-TMR) for 30 min. Labelling reactions were quenched by addition of loading buffer (LB) and immediate boiling. 12.5 % SDS-PAGE gels were run and a fluorescence scanner was used to visualise TMR (red channel Ex/Em 552/572nm) and sCy5 (green channel Ex/Em 647/663nm) labelling.

10.2 Malaria Box ABPP screening data

10.2.1 Malaria Box screening gels
Schizont screen

Plate A: W-sCy5-VS 0.5μM
Plate B: W-sCy5-VS 0.5µM
Plate B: FP-TMR 0.05μM
Plate C: W-sCy5-VS 0.5μM

12.5 % gel
Plate C: FP-TMR 0.05μM

12.5 % gel
Plate D: FP-TMR 0.05μM

MW/ kDa

190
150
120
100
90
80
70
60
50
40
30
20
10

sat + A2 A3 A4 A5 A6 A7 A8 A9 A10 A11 - sat - B2 B3 B4 B5 B6 B7 B8 B9 B10 B11 +

sat + C2 C3 C4 C5 C6 C7 C8 C9 C10 C11 - sat - D2 D3 D4 D5 D6 D7 D8 D9 D10 D11 +

sat + E2 E3 E4 E5 E6 E7 E8 E9 E10 E11 - sat - F2 F3 F4 F5 F6 F7 F8 F9 F10 F11 +

sat + G2 G3 G4 G5 G6 G7 G8 G9 G10 G11 - sat - H2 H3 H4 H5 H6 H7 H8 H9 H10 H11 +

12.5 % gel
Plate E: W-sCy5-VS 0.5μM
Plate E: FP-TMR 0.05µM

MW/ kDa

sat  +  A2  A3  A4  A5  A6  A7  A8  A9  A10  A11  -

sat  -  B2  B3  B4  B5  B6  B7  B8  B9  B10  B11  +

MW/ kDa

sat  +  C2  C3  C4  C5  C6  C7  C8  C9  C10  C11  -

sat  -  D2  D3  D4  D5  D6  D7  D8  D9  D10  D11  +

MW/ kDa

sat  +  E2  E3  E4  E5  E6  E7  E8  E9  E10  E11  -

sat  -  F2  F3  F4  F5  F6  F7  F8  F9  F10  F11  +

MW/ kDa

sat  +  G2  G3  G4  G5  G6  G7  G8  G9  G10  G11  -

-  H2  H3  H4  H5  H6  H7  H8  H9  H10  H11  +

12.5 % gel

189
Trophozoite screen

Plate A: W-sCy5-VS 0.5μM
Plate B: FP-TMR 0.05μM

12.5% gel
Plate D: W-sCy5-VS 0.5μM
Plate E: FP-TMR 0.05μM

12.5% gel
Ring screen

Plate A: W-sCy5-VS 0.5μM
Plate BCDE: W-sCy5-VS 0.5μM

MW/ kDa

- A2 A3 A4 A5 A6 A7 A8 A9 A10 A11 + sat + B2 B3 B4 B5 B6 B7 B8 B9 B10 B11 - sat

MW/ kDa

- C2 C3 C4 C5 C6 C7 C8 C9 C10 C11 + sat + D2 D3 D4 D5 D6 D7 D8 D9 D10 D11 - sat

MW/ kDa

- E2 E3 E4 E5 E6 E7 E8 E9 E10 E11 + sat + F2 F3 F4 F5 F6 F7 F8 F9 F10 F11 - sat

MW/ kDa

- G2 G3 G4 G5 G6 G7 G8 G9 G10 G11 + sat + H2 H3 H4 H5 H6 H7 H8 H9 H10 H11 - sat
10.2.2 Data for unconfirmed Malaria Box screen hits

Supplementary figure 10.3.1: Additional unconfirmed hits from the Malaria Box screen: MMV665798 (M798), MMV019313 (M313), MMV020549 (M549)  
(A) Trophozoite lysate was pre-incubated with 50 μM of a Malaria Box compound or DMSO. Lysates were then labelled with 500 nM W-sCy5-VS (M798) or 50 nM FP-TMR (M549 and M313) or for 10 min.  
(B) ImageLab was used to extract the densitometry profiles of gel lanes from (B).  
(C) Potential hits were freshly purchased and retested with trophozoite lysates as described in (B) using different concentrations of Malaria Box compounds. No dose dependent inhibition was observed.  
(A-C) Saturation of labelling controls were done using 1 μM W-sCy5-VS or 100nM FP-TMR for 30 min. Labelling reactions were quenched by addition of LB and immediate boiling. 12.5% SDS-PAGE gels were run and a fluorescence scanner was used to visualise TMR (red channel Ex/Em 552/572nm) and sCy5 (green channel Ex/Em 647/663 nm) labelling. The bands affected by the hits compounds are labelled with a black arrow.
10.2.3 Inhibition of labelling by Malaria Box compounds was not seen in intact parasites.

<table>
<thead>
<tr>
<th></th>
<th>M563</th>
<th>M438</th>
<th>M914</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intact</td>
<td>Lysate</td>
<td>Intact</td>
</tr>
<tr>
<td>FPP-TMR</td>
<td>500nM</td>
<td>50nM</td>
<td>500nM</td>
</tr>
<tr>
<td>Inhib/</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>50 μM</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Supplementary figure 10.3.2: Intact labelling with FPP-TMR in the presence of Malaria Box compounds. The inhibition of labelling of FPP-TMR by Malaria Box compounds was compared in schizonts using lysates and intact parasites. For intact labelling, schizonts or trophozoites were diluted 1/10 in media. Trophozoites (M563) and schizonts (M438 and M914) were pre-treated with 50 μM inhibitor for 30 min at RT. 500nM or 50nM FPP-TMR was then added and incubated for 30 min at RT with gentle shaking. Parasites were saponin lysed and the soluble protein extracted. Loading buffer (LB) was added and samples heated at 95°C. 12.5 % SDS-PAGE gels were run and a fluorescence scanner was used to visualise TMR (red channel Ex/Em 552/572nm) and sCy5 (green channel Ex/Em 647/663nm) labelling.
10.2.4 Analogues of the M563 and M438 did not inhibit SHs

Supplementary figure 10.3.3: Testing of M563 and M438 malaria box analogues.  
(A) Structures of the malaria box compounds- tetracyclic benzothiazepines (M563, MMV007127 (M127) and MMV645672 (M672)) and fused ring compounds (M438, MM085583 (M583) and MMV011436 (M436)). The arrow on M563 points to a possible place of nucleophillic attack. (B) Trophozoite or schizont lysates were pre-treated with DMSO or the malaria box compounds (50 μM) for 30 mins before 10 min labelling with FP-TMR (50nM) at RT. Labelling reactions were quenched by addition of LB and immediate boiling. 12.5% SDS-PAGE gels were run and a fluorescent scanner was used to visualise TMR (red channel Ex/Em 552/572nm).

10.3 ABPP of SHs during the *P. falciparum* blood-stage asexual life cycle tables.
Table 10.4 ABPP of SHs during the P. falciparum blood-stage asexual life cycle tables
Ring

Trophozoite
Foldenrichment
(log2)
2.59
3.20
2.18
2.70
2.33
2.47

Early Schizont
Foldenrichme
nt (log2)
2.77
3.61
1.97
3.03
2.07
2.66

Mature Schizont
Foldenrichment
(log2)
2.68
3.09
1.83
2.27
2.10
2.51

Merozoite

1.31
0.85
0.22

Foldenrichment
(log2)
1.24
0.90
0.29

Neutral cholesterol ester hydrolase 1
Acetylcholinesterase
Neuropathy target esterase

0.28
0.07
0.14

Foldenrichment
(log2)
0.35
-0.14
-0.46

81.7
18.0
17.7

1.28
1.53
1.42

0.80
4.27
2.66

3.65
1.54
1.58

2.23
4.51
3.83

2.97
0.53
0.69

1.97
1.67
1.30

2.85
0.41
0.68

1.58
1.28
1.27

5.41
1.21
1.11

3.75
3.07
2.04

Acylamino-acid-releasing enzyme
Acyl-protein thioesterase 1
Acyl-protein thioesterase 2

FASN

273.2

0.95

1.74

2.23

3.42

1.30

1.13

0.75

0.72

0.82

1.21

Gene ID

Protein
name

Human
Human
Human

NCEH1
ACHE
PNPLA6

Mol.
weight
[kDa]
49.1
65.6
146.2

Human
Human
Human

APEH
LYPLA1
LYPLA2

Human

-Log (t-test
p value)

-Log (t-test
p value)

-Log (t-test
p value)

-Log (t-test
p value)

-Log (t-test
p value)

T: Description

Human

PAFAH1B3

25.7

0.16

-0.33

0.45

-0.53

0.04

-0.06

0.23

-0.15

3.25

2.60

PF3D7_0209100
PF3D7_0218600

PNPLA1

78.3
283.6

0.43
0.01

0.44
0.01

2.16
3.47

2.00
3.57

3.25
3.52

3.68
3.65

3.30
4.51

3.09
3.17

3.50
2.97

4.20
2.41

Fatty acid synthase
Platelet-activating factor
acetylhydrolase IB subunit gamma
Phospholipase A2
Patatin-like phospholipase, putative

PF3D7_0301300
PF3D7_0321500
PF3D7_0403800

EH1
S9A
S9C

50.5
125.9
83.4

0.91
0.42
0.03

2.03
-0.47
-0.04

2.30
0.44
0.56

3.67
0.20
0.63

3.54
2.74
2.94

5.11
1.92
1.67

2.88
1.34
1.87

4.20
0.82
1.99

2.61
1.36
3.53

4.38
0.94
4.69

α/β-hydrolase
Putative acylaminoacyl-peptidase
α/β-hydrolase

PF3D7_0629300
PF3D7_0709700
PF3D7_0731800

LCAT
PARE
S33G

99.2
42.4
78.4

0.16
1.01
0.60

0.39
1.16
0.53

0.80
3.92
2.08

0.31
3.78
3.61

3.29
3.93
2.31

3.86
5.30
4.60

3.71
4.12
2.48

4.26
4.54
4.89

2.83
3.68
1.40

2.81
4.31
2.16

PC-sterol acyltransferase, putative
Esterase
α/β-hydrolase

PF3D7_0805000
PF3D7_0818600
PF3D7_1001600

S9B
PBLP
XL2

28.5
34.9
88.6

0.88
0.19
0.68

1.10
0.16
0.55

2.17
2.00
2.87

3.44
1.59
2.20

2.86
4.36
3.02

5.32
3.88
3.72

2.80
3.95
4.06

4.83
3.65
3.17

3.72
3.90
4.41

3.96
3.86
4.28

α/β-hydrolase
BEM46-like protein
Exported lipase 2

30.6
41.8
217.1

0.12
0.55
0.11

-0.10
0.43
-0.27

0.48
2.34
0.37

0.66
1.92
0.65

1.59
4.00
2.12

1.85
4.21
3.05

1.65
2.52
1.31

1.02
2.44
2.13

2.01
4.31
1.61

2.36
3.96
3.22

Proteasome beta type-5
Patatin-like phospholipase, putative
Serine esterase, putative

PF3D7_1011400
PF3D7_1038900
PF3D7_1116100
PF3D7_1117300
PF3D7_1120400
PF3D7_1129300

abH112

15.3
44.7
221.7

0.03
0.34
0.48

0.05
0.77
0.61

0.35
2.00
0.22

-0.20
2.85
0.36

1.04
3.30
2.38

-0.79
4.79
3.71

2.39
2.52
2.87

-1.77
3.28
3.53

0.66
1.04
0.73

-0.60
1.36
1.08

unknown transmembrane protein
Alpha/beta hydrolase, putative
Alpha/beta hydrolase, putative

PF3D7_1134500
PF3D7_1143000
PF3D7_1252600

abH114

210.5
44.8
52.9

0.04
0.44
0.83

-0.05
0.82
0.58

1.53
3.37
1.89

1.99
1.25
2.95

3.46
3.66
2.65

4.53
3.81
4.20

2.61
3.00
2.85

4.18
2.20
3.60

1.03
5.24
2.98

1.02
3.55
3.58

Alpha/beta hydrolase, putative
Alpha/beta hydrolase, putative
Putative esterase

115.9
238.2
43.7

0.08
0.01
2.49

0.14
0.00
2.18

0.94
0.67
2.80

1.08
0.67
4.82

2.01
2.62
2.74

1.86
3.96
6.52

1.23
2.40
1.54

0.71
2.18
4.81

4.00
1.69
2.65

5.17
0.98
5.67

Alpha/beta hydrolase, putative
Patatin-like phospholipase, putative
Lysophospholipase, putative

PF3D7_1328500
PF3D7_1358000
PF3D7_1401500

MLPL

PF3D7_1427100
PF3D7_1449400
PF3D7_1458300

MUS81

155.7
150.1
183.5

0.58
0.00
0.14

0.50
1.49
0.27

2.70
0.21
0.56

3.28
0.35
0.55

1.69
0.11
3.66

1.74
-0.63
4.05

2.46
0.15
3.54

2.65
-0.32
2.30

2.41
2.02
0.97

2.71
5.52
1.08

Lipase
Endonuclease
Alpha/beta hydrolase, putative

PF3D7_1476700
PF3D7_1476800

Psta1
Psta2

41.3
42.7

0.28
0.09

0.46
0.13

1.70
2.84

1.84
1.93

2.26
4.24

1.90
4.12

1.90
2.37

1.88
3.57

3.02
2.39

3.60
2.80

lysophospholipase, putative
lysophospholipase, putative




10.4 Human SH inhibitor values.

Table 10.4: Significance values corresponding to Figure 6.2.3: Unpaired, one-way ANOVA using Dunnett’s multiple comparisons test.

<table>
<thead>
<tr>
<th>Parasitaemia</th>
<th>Test</th>
<th>Mean Diff.</th>
<th>95.00% CI of diff.</th>
<th>Summary</th>
<th>Adjusted P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 hpi</td>
<td>DMSO vs. CQ</td>
<td>-0.07442</td>
<td>-0.7175 to 0.5687</td>
<td>ns</td>
<td>0.9899</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. C2</td>
<td>-0.4617</td>
<td>-1.105 to 0.1814</td>
<td>ns</td>
<td>0.1787</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. ML-348</td>
<td>0.2346</td>
<td>-0.4085 to 0.8777</td>
<td>ns</td>
<td>0.6718</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. AA74-1</td>
<td>0.003233</td>
<td>-0.7158 to 0.7222</td>
<td>ns</td>
<td>0.9999</td>
</tr>
<tr>
<td>41 hpi</td>
<td>DMSO vs. CQ</td>
<td>2.178</td>
<td>1.083 to 3.273</td>
<td>***</td>
<td>0.0008</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. C2</td>
<td>1.771</td>
<td>0.6762 to 2.866</td>
<td>**</td>
<td>0.0034</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. ML-348</td>
<td>2.21</td>
<td>1.115 to 3.305</td>
<td>***</td>
<td>0.0007</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. AA74-1</td>
<td>2.363</td>
<td>1.138 to 3.587</td>
<td>**</td>
<td>0.001</td>
</tr>
<tr>
<td>48 hpi</td>
<td>DMSO vs. CQ</td>
<td>7.725</td>
<td>5.844 to 9.606</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. C2</td>
<td>7.617</td>
<td>5.736 to 9.497</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. ML-348</td>
<td>6.828</td>
<td>4.947 to 8.709</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. AA74-1</td>
<td>7.476</td>
<td>5.373 to 9.579</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td>72 hpi</td>
<td>DMSO vs. CQ</td>
<td>7.854</td>
<td>7.353 to 8.354</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. C2</td>
<td>7.981</td>
<td>7.48 to 8.482</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. ML-348</td>
<td>6.953</td>
<td>6.453 to 7.454</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. AA74-1</td>
<td>7.403</td>
<td>6.844 to 7.963</td>
<td>****</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DNA content</th>
<th>Test</th>
<th>Mean Diff.</th>
<th>95.00% CI of diff.</th>
<th>Summary</th>
<th>Adjusted P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 hpi</td>
<td>DMSO vs. CQ</td>
<td>1.727</td>
<td>0.8177 to 2.636</td>
<td>**</td>
<td>0.0011</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. C2</td>
<td>0.6102</td>
<td>-0.299 to 1.519</td>
<td>ns</td>
<td>0.2186</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. ML-348</td>
<td>0.7717</td>
<td>-0.1375 to 1.681</td>
<td>ns</td>
<td>0.0997</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. AA74-1</td>
<td>1.842</td>
<td>0.8257 to 2.859</td>
<td>**</td>
<td>0.0016</td>
</tr>
<tr>
<td>41 hpi</td>
<td>DMSO vs. CQ</td>
<td>86.1</td>
<td>83.65 to 88.55</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. C2</td>
<td>81.59</td>
<td>79.14 to 84.04</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. ML-348</td>
<td>83.54</td>
<td>81.08 to 85.99</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. AA74-1</td>
<td>85.61</td>
<td>82.86 to 88.35</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td>48 hpi</td>
<td>DMSO vs. CQ</td>
<td>86.56</td>
<td>79.14 to 93.99</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. C2</td>
<td>76.69</td>
<td>69.27 to 84.12</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. ML-348</td>
<td>78.35</td>
<td>70.93 to 85.78</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. AA74-1</td>
<td>85.17</td>
<td>76.87 to 93.47</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td>72 hpi</td>
<td>DMSO vs. CQ</td>
<td>86.852</td>
<td>3.129 to 10.57</td>
<td>**</td>
<td>0.0014</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. C2</td>
<td>-22.41</td>
<td>-26.14 to -18.69</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. ML-348</td>
<td>-17.44</td>
<td>-21.17 to -13.72</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. AA74-1</td>
<td>5.635</td>
<td>1.473 to 9.797</td>
<td>*</td>
<td>0.0102</td>
</tr>
<tr>
<td>RNA content</td>
<td>Test</td>
<td>Mean Diff.</td>
<td>95.00% CI of diff.</td>
<td>Summary</td>
<td>Adjusted P Value</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
<td>------------</td>
<td>--------------------</td>
<td>---------</td>
<td>------------------</td>
</tr>
<tr>
<td>24 hpi</td>
<td>DMSO vs. CQ</td>
<td>1.432</td>
<td>1.152 to 1.713</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. C2</td>
<td>0.3835</td>
<td>0.1034 to 0.6637</td>
<td>**</td>
<td>0.0096</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. ML-348</td>
<td>0.6972</td>
<td>0.4171 to 0.9773</td>
<td>***</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. AA74-1</td>
<td>0.9782</td>
<td>0.665 to 1.291</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td>41 hpi</td>
<td>DMSO vs. CQ</td>
<td>23.72</td>
<td>22.45 to 24.98</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. C2</td>
<td>20.42</td>
<td>19.15 to 21.69</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. ML-348</td>
<td>20.1</td>
<td>18.83 to 21.37</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. AA74-1</td>
<td>22.74</td>
<td>21.33 to 24.16</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td>48 hpi</td>
<td>DMSO vs. CQ</td>
<td>12.73</td>
<td>11.71 to 13.76</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. C2</td>
<td>9.344</td>
<td>8.317 to 10.37</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. ML-348</td>
<td>8.667</td>
<td>7.639 to 9.694</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. AA74-1</td>
<td>12.01</td>
<td>10.87 to 13.16</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td>72 hpi</td>
<td>DMSO vs. CQ</td>
<td>3.985</td>
<td>3.028 to 4.943</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. C2</td>
<td>0.2049</td>
<td>-0.7527 to 1.163</td>
<td>ns</td>
<td>0.9179</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. ML-348</td>
<td>-4.121</td>
<td>-5.078 to -3.163</td>
<td>****</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>DMSO vs. AA74-1</td>
<td>3.287</td>
<td>2.216 to 4.358</td>
<td>****</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
10.5 Genetic interrogation

10.5.1 Constructs for transfection
10.5.2 Integration PCRs for GOI-cWT lines.

Figure 10.6.2: Integration of conditional mutants. Endogenous and 3'/5' integration PCRs were done on genomic DNA extracted from mixed and clonal populations of Psta1-cWT (2G6, 2G7), Psta2-cMUT (4B6, 4B10, 4B11), abH112-cMUT (6C8, 6D11) and abH114-cMUT (8B6, 8C5) as well as the parent B11 line as a control.

10.6 Papers published in the course of this thesis
RESEARCH ARTICLE

Novel broad-spectrum activity-based probes to profile malarial cysteine proteases

Michele S. Y. Tan1, Dara Davison1, Mateo I. Sanchez2, Bethany M. Anderson3, Stephen Howell1, Ambrosius Snijders1, Laura E. Edgington-Mitchell3,4,5, Edgar Deu1*

1 The Francis Crick Institute, London, United Kingdom, 2 Department of Genetics, Stanford School of Medicine, Stanford, California, United States of America, 3 Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville Victoria, Australia, 4 Drug Discovery Biology, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Victoria, Australia, 5 Department of Maxillofacial Surgery, College of Dentistry, New York University, New York, New York, United States of America

These authors contributed equally to this work.

Abstract

Clan CA cysteine proteases, also known as papain-like proteases, play important roles throughout the malaria parasite life cycle and are therefore potential drug targets to treat this disease and prevent its transmission. In order to study the biological function of these proteases and to chemically validate some of them as viable drug targets, highly specific inhibitors need to be developed. This is especially challenging given the large number of clan CA proteases present in Plasmodium species (ten in Plasmodium falciparum), and the difficulty of designing selective inhibitors that do not cross-react with other members of the same family. Additionally, any efforts to develop antimalarial drugs targeting these proteases will also have to take into account potential off-target effects against the 11 human cysteine cathepsins. Activity-based protein profiling has been a very useful tool to determine the specificity of inhibitors against all members of an enzyme family. However, current clan CA proteases broad-spectrum activity-based probes either target endopeptidases or dipeptidyl aminopeptidases, but not both subfamilies efficiently. In this study, we present a new series of dipeptidyl vinyl sulfone probes containing a free N-terminal tryptophan and a fluorophore at the P1 position that are able to label both subfamilies efficiently, both in Plasmodium falciparum and in mammalian cells, thus making them better broad-spectrum activity-based probes. We also show that some of these probes are cell permeable and can therefore be used to determine the specificity of inhibitors in living cells. Interestingly, we show that the choice of fluorophore greatly influences the specificity of the probes as well as their cell permeability.

Introduction

Malaria remains one of the most devastating infectious diseases worldwide killing close to half a million people and affecting over 200 million every year[1]. Malaria incidence has significantly decreased over the last 15 years mainly due to the distribution of insecticide-
In *Plasmodium* species, this family of proteases has been shown to be involved in essential biological process throughout the parasite life cycle[14]. This family is composed of 10 members in *P. falciparum* and includes 3 dipeptidyl aminopeptidases (DPAPs) and 7 endopeptidases: 4 falcipains (FPs) and 3 SERAs (serine repeat antigens). In order to better understand the biological functions of these proteases and to chemically evaluate their potential as antimalarial targets, highly specific inhibitors are required. However, given the large number of clan CA proteases present in *P. falciparum*, off-target effects within this family is one of the main hurdles to developing specific compounds. In addition, eleven cysteine cathepsins are expressed in humans. Therefore, highly specific inhibitors that do not target host proteases are needed to study the biological function of malarial proteases within the host context, either in animal models of malaria or during liver stages.

Activity-based probes (ABPs) are small molecules that use the enzymatic mechanism of an enzyme to covalently modify its active site[15]. A tag embedded within the structure of the probe, usually a biotin or fluorophore, allows for visualization of the labelled enzymes in a gel-based format. Biotinylated probes can also be used to pull-down the labelled proteins for MS identification. Broad-spectrum ABPs are designed to covalently modify all members of an enzyme family, thus allowing profiling of the activity of each member in a biological sample. This method known as activity-based protein profiling (ABPP) has been broadly applied to determine the specificity of inhibitors given that binding to the active site will prevent probe labelling. Although several ABPs have been developed for clan CA proteases, these probes generally either target endopeptidases or DPAPs, but not both subfamilies. For example, DCG04, which was developed based on the structure of the general covalent cysteine protease inhibitor E64[16], labels most mammalian cysteine cathepsins and the parasite FPs very efficiently, but is very inefficient at labelling *Plasmodium* DPAPs or CatC[17,18]. On the other hand, the DPAP selective probe, FY01, does not label all endopeptidases efficiently[18,19]. Therefore, to profile the specificity of inhibitors against all clan CA proteases two probes are usually required. In this study, we report a new series of broad-spectrum ABPs that is able to label both of these subfamilies.

*Plasmodium* clan CA proteases play important roles throughout the parasite life cycle. Malaria is transmitted through the bite of *Anopheles* mosquitoes. Parasites initially establish an asymptomatic infection in the liver where they multiply into thousands of merozoites (extracellular form that is able to infect erythrocytes). Release of parasites from infected hepatocytes into the blood stream initiates the erythrocytic cycle, which consists of red blood cell (RBC) invasion, parasite growth (ring to trophozoite stage transition), nuclear division and parasite replication (schizogony), and egress from the infected RBC (iRBC). This exponential asexual replication is responsible for the symptoms and pathology of malaria. A fraction of parasites develop into male and female gametocytes, which, after being ingested during a blood meal,
will mature into gametes and sexually reproduce in the mosquito midgut. Parasites will then multiply and travel to the salivary glands from where they will be transmitted to the next human host.

Three subfamilies of clan CA proteases are conserved in *Plasmodium* species. In *P. falciparum* the SERA subfamily is composed of 9 members, each containing a papain-like domain, but only 3 have a catalytic cysteine (SERA6-8)[20]. The others have a serine instead and are predicted to be inactive. These proteases are secreted into the parasitophorous vacuole (PV) aszymogens. The PV is a membrane bound vacuole within which parasites develop and replicate isolated from the host cytosol. SERAs are proteolytically activated by a subtilisin-like protease that is released into the PV 15 min before egress, and therefore are predicted to be active for a very short period of time[21,22]. SERA6 has been shown to be essential for parasite egress from iRBCs, and this function is dependent on its catalytic cysteine[22]. Finally, the homologue of SERA8 in *P. berghei* has been shown to be essential for parasite egress in insect stages[23].

FPs are endopeptidases active at acidic pH and are expressed during the erythrocytic cycle[24–26]. FP2a, FP2b, and FP3 localize to the digestive vacuole and are at the top of the proteolytic pathway responsible for haemoglobin degradation[14]. This pathway provides a source of amino acids for protein synthesis and liberates space within the iRBC to allow parasite growth. Because FP2a, FP2b, and FP3 perform redundant functions, their individual knockout (KO) does not result in parasite death[27]. However, to the best of our knowledge, a triple KO has not yet been attempted in *P. falciparum*. The biological role of the fourth falcipain, FP1, is not well understood. Studies with small molecules suggested a role in RBC invasion[17], but KO of FP1 in *P. falciparum* showed no clear phenotype in asexual stages[28]. However, KO of the FP1 homologue in the *P. berghei* murine model of malaria impairs invasion of mature RBCs albeit not of reticulocytes[29].

Finally, three DPAPs are conserved in *Plasmodium* species. DPAPs are homologues of CatC and cleave dipeptides off the N-terminus of substrates. Compared to other clan CA proteases, DPAPs have an additional domain, known as the exclusion domain, that interacts with the free N-terminal amine of substrates via its N-terminal Asp side chain. This exclusion domain also prevents binding beyond the P2 position, thus preventing endopeptidase activity. DPAP1 is expressed at all stages of parasite development and has been proposed to play a role during the latest stages of the haemoglobin degradation pathway[30,31]. DPAP2 is only expressed in sexual stages, localizes to apical secretory organelles in gametocytes, and has been shown to play an important role in gamete egress from iRBCs[32,33]. Finally, DPAP3 localizes to novel apical organelles in merozoites, it is secreted immediately before parasite egress, and its activity is required for efficient RBC invasion[34].

In this study we present a new series of fluorescent broad-spectrum ABPs able to label clan CA endopeptidases and DPAPs. These dipeptidic vinyl sulfone probes are able to efficiently target the FPs and DPAPs in *P. falciparum*, as well as most mammalian clan CA proteases, both in lysates and in intact cells. These new probes are therefore perfect tools to determine the specificity of inhibitors not only against parasite targets but also against potential off-targets in the host.

**Materials and methods**

**Probe synthesis**

The synthesis and characterization of W-hPG-VS, D-W-hPG-VS, SAK1, and SAK2 have been previously described[34,35]. To synthesize the new probes presented here, the trifluoroacetic (TFA) salt of W-hPG-VS (10 mg, 0.018 mmol) and 0.09 mmol of the azide derivates of the tag...
(Lumiprobe) were dissolved in an eppendorf tube in 0.9 mL of 1:1 water:DMSO. CuSO₄ (0.018 mmol, 40 μL of a solution 500 mM) and sodium L-ascorbate (14.2 mg, 0.072 mmol) were added, and the reaction was followed by HPLC-MS until total consumption of the azide derivative. The crude reaction was directly purified by preparative reverse-phase chromatography. The combined fractions were concentrated and freeze-dried, yielding a solid that was identified as the TFA salt of the desired product. The LCMS traces of the high-resolution mass for each probe are shown in S1 Fig.

Parasite culture and lysates preparation

Anonymized human blood to culture malaria parasites was purchased from the United Kingdom National Health System Blood and Transplant Special Health Authority. No ethical approval is required for its use. *Plasmodium falciparum* 3D7 parasites were grown at 2% hematocrit in RPMI 1640 (Gibco) media supplemented with Albumax (Invitrogen) as previously described[35]. Parasite cultures were synchronised by purifying mature schizonts by sedimentation using a Percoll gradient, and incubating them under shaking conditions for 2–3 h with fresh RBCs and media to allow parasite egress and RBC invasion. Unruptured schizonts were removed from the culture using a Percoll gradient, and any remaining schizonts in our ring-stage culture were removed by sorbitol treatment. Mature schizonts were obtained by culturing purified schizonts for 3 h in the presence of 1 μM of the cGMP-dependent protein kinase inhibitor Compound 2[21].

Trophozoite stage parasite pellets were collected 24 h after synchronization, and schizonts were purified using a Percoll gradient 44 h after synchronization. For both, the RBC and PV membranes were lysed using saponin, and the parasite pellets washed multiple times with PBS. Merozoites were purified as previously described[35]. Briefly, purified schizonts were allowed to egress in media under shaking conditions. Free merozoites were separated from unruptured schizonts by centrifugation, and the merozoite-containing supernatant passed through a magnet (SuperMACS) to remove any leftover schizonts and hemozoin-containing residual bodies of ruptured iRBCs. All parasite pellets were snap frozen in liquid nitrogen and stored at -80°C. To obtain lysates, parasite pellets were incubated on ice for 1 h in two volumes of 1% NP40 in PBS followed by a 5 min centrifugation at 17,000 x g to obtain the soluble fraction of the lysate.

Parasite labelling conditions

Parasite lysates were diluted 10-fold in acetate buffer (50 mM sodium acetate, 100 mM NaCl, 5 mM MgCl₂, and 5 mM DTT, pH 5.5) and treated with different concentrations of probes for 1 h at RT. For competition assays, lysates were pre-treated for 30 min with inhibitor prior to probe labelling. Labelling was stopped by adding 4X loading buffer and boiling the samples for 5 min. Labelling of cysteine proteases in live parasites was performed by diluting purified mature schizonts 10-fold in RPMI containing 1 μM of Compound 2, and treating the samples for 1 h with increasing concentrations of probes at 37°C. The reaction was stopped by added 4X loading buffer and boiling the samples for 10 min. All samples were loaded on a 12% SDS-PAGE gel, and the in-gel fluorescence measured on a PharosFX fluorescence scanner (Biorad) either using a 532 nm laser for probes containing a Cy3, sCy3 or TAMRA (605 nm emission filter), and a 635 nm laser for those with a Cy5 or sCy5 fluorophore (695 nm emission filter).

RAW macrophage culture, lysates preparation, and labelling

RAW264.7 cells (immortalized murine macrophages) were cultured in DMEM containing 10% fetal bovine serum and 1% antibiotic/antimycotic. Cells were passaged by scraping with a
rubber policeman. Cells were plated in 12-well plates (300,000 cells/per well) and allowed to adhere overnight. The indicated probes were added at a final concentration of 1 μM for 3 h. Cells were lysed in 30 μl citrate buffer (50 mM citrate, 0.5% CHAPS, 0.1% Triton X-100, and 4 mM DTT, pH 5.5). Lysates were centrifuged, and cleared supernatants were solubilized with 7.5 μl 5x sample buffer (200 mM Tris-Cl, 8% SDS, 0.04% bromophenol blue, 5% β-mercaptoethanol, and 40% glycerol, pH 6.8). Samples (20 μl) were then resolved on a 15% SDS-PAGE gel, and labeling was detected by scanning for Cy5 or Cy3 fluorescence on a Typhoon 5 flatbed laser scanner (GE Healthcare). For lysate labeling, cells were first lysed as above in citrate buffer and 20 μl lysate was incubated with 1 μM probe for 20 min at 37°C. Sample buffer was added to stop the reaction, and samples were analyzed as above. All samples were run as duplicates.

**Chemical proteomics**

Schizont pellets were resuspended in one volume of 0.1% Triton X-100 in PBS buffer, incubated at 4°C for 20 min, and the soluble fraction of the lysate collected after centrifugation at 17,000 x g for 5 min. Protein concentration was determined using a NanoDrop 2000 spectrometer (Thermo Scientific). Lysates were diluted in acetate buffer to a protein concentration of 2 mg/mL, and six 100 μl samples were treated for 10 min either with DMSO or 1 μM W-BF-VS. Protein was then precipitated by adding 200 μL of methanol, 50 μL of chloroform and 100 μL of water, and the pellet collected by centrifugation (2 min at 17,000 x g). Protein was first resolubilized in 60 μL of 2% SDS and 10 mM DTT in PBS, and diluted 10-fold in PBS to achieve a concentration of 1 mg/mL. A mixture of 10 μL of neutravidin-agarose beads and 20 μL of blank agarose beads (both from Pierce) was added to the samples and incubated at RT for 2 h under shaking conditions. Beads were then collected by centrifugation (2 min at 3000 x g) and washed three times with 1% SDS in PBS, two times with 4 M urea in PBS, and two times with 50 mM ammonium bicarbonate (AMBIC). Beads were then treated with 10 mM tris(2-carboxyethyl)phosphine hydrochloride in 50 mM AMBIC for 30 min at RT, washed, resuspended in 30 μL of 50 mM AMBIC, and treated with 0.12 μg of trypsin gold (Promega) overnight at 37°C under shaking conditions. After collecting the supernatant, peptides were further eluted with 70 μL of 1.5% TFA.

The combined peptide mixture was filtered and desalted on homemade ‘stage tips’ containing C18 Empore membrane (Empore Octadecl C18 47mm Extraction Disks 2215 Supelco 66883-U). Tips were washed by centrifugation (2 min at 2000 x g) with 150 μL of MeOH and 150 μL of water. The peptide sample (140 μl) was loaded, the tips washed with 150 μL of water, and eluted with 60 μl of 40% acetonitrile in water. The peptide solution was then evaporated to dryness using a speed vacuum (SpeedVac system ISS 100, Savant Integrated). Peptide samples were dissolved in 25 μL of 50 mM triethylammonium bicarbonate buffer (Sigma), and 20 μl of the appropriate TMT label (ThermoFisher TMTsixplex product number 90308, lot number RH239931) freshly dissolved in anhydrous acetonitrile was added to each of the six samples and incubated for 1 h at RT. Combined TMT samples were then fractioned into 8 samples using a high pH reversed-phase peptide fractionation kit (Pierce 84868). Fractionated samples were then evaporated to dryness using a speed vacuum.

Peptides were chromatographically resolved on an Ultimate 3000 nanoRSLC HPLC (Thermo Scientific). Each peptide fraction was acidified to a final concentration of 0.1% TFA, and 1–10 μl loaded onto a 2 x 0.3 mm Acclaim Pepmap C18 trap column (Thermo Scientific) at 15 μl/min of 0.1% TFA prior to the trap being switched to elute at 0.25 μl/min through a 50 cm x 75 μm EasySpray C18 column. In a 90 min run, the following gradients of solution A (2% acetonitrile, 0.1% formic acid) and B (80% acetonitrile, 0.1% formic acid) were used: 9–25%
solution B over 37 min, 25–40% solution B over 18 min, 100% solution B over 15 min, and equilibration back to 2% solution B for 20 min. HPLC eluant was introduced into an Orbitrap Fusion Lumos (Thermo Scientific). The Orbitrap was operated in “Data Dependent Acquisition” mode with a survey scan at a resolution of 120,000 from m/z 400–1400. This was followed by “TopS” precursor ion selection MS/MS using 38% high energy collision dissociation. Dynamic exclusion was used with a time window of 30 s.

Proteomics data analysis

The raw data files were analysed using MaxQuant (version 1.6.2.1.). Quantification was done at MS2 level using 6-plex TMT labels, no label free quantification was performed. Default MaxQuant settings were used with the Homo sapiens (Uniprot 13/01/2013) and Plasmodium falciparum 3D7 (PlasmoDB 15/12/2016) sequence databases. A decoy database of reversed sequences was used to filter false positives at a peptide false detection rate of 1%.

The data files generated by MaxQuant were further analyzed using Perseus (version 1.5.6.0.). The list of proteins identified by MaxQuant was filtered to remove proteins that were only identified by a modification site, proteins identified from the reverse decoy peptide database, and proteins annotated as potential contaminants. Proteins that were identified from 2 peptides or less were also excluded from the analysis. After log2(x) transformation of the TMT reporter ion intensities, the data was normalised by subtracting the median intensity from each channel to correct for differences in loading. Two sample t-tests with multiple hypothesis testing correction (S0 = 1, permutation based FDR = 0.05 using default Perseus settings) were carried out between DMSO and probe treated samples in order to identify statistically significant differences. The results were visualised on a “volcano plot” (S2 Fig).

Results and discussion

Probe design

Our initial intention was to develop highly selective DPAP3 probes based on the observation that dipeptidic vinyl sulfone inhibitors with a free N-terminal Trp are DPAP3 specific[34,36]. Because the S1 pocket is generally solvent exposed in clan CA proteases, we decided to introduce an alkyne group at the P1 position and use copper-catalyzed click chemistry to conjugate different azido-tagged fluorophores. We therefore synthesize the precursor inhibitor Trp-hPG-VS (Fig 1A) and conjugated different fluorophores (Cy5, sulfoCy5, Cy3, or sulfoCy3) to the homoprolylglycine (hPG) side chain (Fig 1B). We named these probes W-Cy5-VS, W-sCy5-VS, W-Cy3-VS, and W-sCy3-VS, respectively. We then determined the specificity of these probes in parasite lysates and compared them to ABPs previously used to label malarial cysteine proteases, namely FY01 and DCG04 (Fig 1C).

Probe specificity in parasite lysates

We first tested our probes in merozoite lysates since DPAP3 is most abundant at this stage. Lysates were treated for 1 h with a probe concentration ranging from 1 nM to 1 μM. Samples were then run on a SDS-PAGE gel, and the labelled proteins visualized using a fluorescence scanner. At the highest probe concentration lysates were also pre-treated with the DPAP3-selective covalent inhibitor SAK1[18,36] to confirm whether the probes labelled DPAP3 (Fig 1A). In merozoite lysates under acidic conditions FY01 has been shown to label three different forms of DPAP3 running at 120, 95, and 42 kDa[18]. As shown in Fig 2A, all probes are able to label the three forms efficiently with the exception of W-Cy3-VS. W-sCy5-VS and W-sCy3-VS are able to label DPAP3 at 1 nM, and the latter is DPAP3-selective when used between 1 and 40 nM.
100 nM. However, all probes start labelling other targets between 0.1 and 1 μM. In particular, W-sCy5-VS labels a large number of proteins, making it a potential broad-spectrum ABP.

To further determine the specificity of these probes, we tested them at 100 nM in parasite lysates collected at different stages of development (merozoites, trophozoites and schizonts). To determine whether these probes target other known clan CA proteases, lysates were pre-treated with covalent inhibitors of DPAP1 (SAK2), DPAP3 (SAK1), or the FPs (E64)[18]. We also included the W-hPG-VS, compound from which the probes were derived, in this competition assay, as well as the diastereomer control D-W-hPG-VS (Fig 1A) that does not inhibit DPAP3 or any other clan CA protease[36]. Despite being DPAP3-selective in merozoite lysates (Fig 2A and 2B), most probes also target DPAP1 and the FPs in trophozoite and schizont lysates (Fig 2C and 2D), probably because these other proteases are much more abundant at these stages than in merozoites. W-sCy5-VS is able to very robustly label all DPAPs and FPs, as well as the pro-form of DPAP1 running at 38 kDa (Fig 2C and 2D). Although W-Cy5-VS also has broad specificity, it is a much weaker probe than W-sCy5-VS. Surprisingly, the probes containing the Cy3 and sCy3 fluorophores show very poor labelling of cysteine proteases under these conditions. These results illustrate how the choice of fluorophore can have a major influence on the potency and specificity of probes.

As reported previously, the W-hPG-VS inhibitor is a specific inhibitor of DPAP3[36], but conjugating a fluorophore to the P1 position via cupper-catalyzed click chemistry results in a very significant loss of specificity, especially with sCy5. This was surprising as the P1 position

---

**Fig 1. Structures of inhibitors and ABPs.** (A) Structures of inhibitors used in this study. All the new ABPs described in this paper were synthesized by conjugating different azido-tags (B. fluorophore or D, a biotin/TAMRA bifunctional tag) to the alkyne group of W-hPG-VS. (C) Structures of previously published ABPs used in this study for comparison purposes.

https://doi.org/10.1371/journal.pone.0227341.g001
Fig 2. Labelling of cysteine proteases in parasite lysates. (A) Merozoite lysates diluted 1:10 in acetate buffer were treated for 1 h with 1–1000 nM of the indicated ABPs. For the highest ABP concentration, samples were also pre-treated for 30 min with 1 μM of the DPAP3 inhibitor SAK1, which results in the loss of labelling of the three isoforms of DPAP3 running at 120, 95, and 42 kDa. (B-D) Lysates collected at merozoite (B), trophozoite (C), or schizont (D) stages were diluted in acetate buffer (pH 5.5), pre-treated for 30 min with DMSO or 10 μM of different known covalent inhibitors of DPAP1 (SAK2), DPAP3 (SAK1 or W-hPG-VS), the FPs (E64), or the negative control compound D-W-hPG-VS. This was followed by 1 h labelling with the different ABPs at 0.1 μM except for DCG04 that was used at 1 μM concentration. (A-D) The fluorescent bands corresponding to DPAP1, DPAP3, FP1, and FP2/3 are indicated by blue, red, light green, and dark green arrowheads, respectively. Two additional biological replicates of these experiments are shown in S3 Fig.

https://doi.org/10.1371/journal.pone.0227341.g002
is generally solvent exposed in clan CA proteases. However, substrate specificity studies on DPAPs have shown that long and hydrophobic non-natural amino acids at the P1 position greatly increase substrate turnover, possibly allowing the P1 residue to interact with an adjacent hydrophobic pocket.[35]. This might also be the case for the FPs and might explain why our probes are no longer specific for DPAP3 when compared to W-hPG-VS. Another potential explanation is that the fluorophore might allow the probe to make non-specific hydrophobic interactions, increasing the time available for the catalytic cysteine to attack the vinyl sulfone warhead. All labelling experiments in parasite lysates were performed in triplicate, the two replicates not shown in Fig 2 are shown in S3 Fig.

Probe specificity in live parasites
We then tested the specificity of these probes in live parasites by treating very mature schizonts for 1 h with different concentrations of probes. Purified schizonts were arrested 15 min before egress by treating them with an inhibitor of the cGMP-dependent protein kinase[21]. We used these labelling conditions because DPAPs and FPs are active and abundant at this stage, and because the RBC and PV membranes are still intact, thus allowing us to determine the cell-permeability of our probes. For all probes, clear labelling of cysteine proteases only occurs at 1 μM (Fig 3). It is also quite clear that the sulfate groups in W-sCy3-VS and W-sCy5-VS prevent the probes from crossing membranes and reaching their targets. Under these conditions, DCG04 only labels the FPs, and FY01 the DPAPs. However, W-Cy5-VS is able to label all of these proteases, thus making it a very useful broad-spectrum ABP that can be used in live parasites. Note that only the p120 form of DPAP3 is labelled in live parasites. This is consistent with our previous results showing that DPAP3 processing to the p95 and p42 form is an artefact of parasite lysis, and that the only form found in live parasites is the p120 form.[34]. Intact labelling experiments were performed in triplicate, the two replicates not shown in Fig 3 are shown in S4 Fig.

Confirmation of probe selectivity through chemical proteomics
To better determine the specificity of these probes in parasite lysates, we synthesized a probe with a bifunctional tag containing a TAMRA fluorophore and a desthiobiotin moiety for pull-down and MS ID purposes (W-BF-VS, Fig 1D). Since W-sCy5-VS is the probe with broader...
specificity in parasite lysates, we compared its labelling profile to that of W-BF-VS in schizont lysates at 0.5 μM. We also tried a competition labelling experiment where both probes were added to the lysate simultaneously. Most proteins labelled by both probes run at the same MW (Fig 4A). However, as the MW decreases, the TAMRA-labelled proteins run at a slightly higher MW, which is consistent with W-BF-VS being a much larger probe than W-sCy5-VS (Fig 1). This is very evident when comparing the labelling of the DPAP1 doublet. Quantification of these labelling patterns by densitometry clearly show very similar profiles for both probes (Fig 4C). In addition, for most of the bands we observed a clear decrease in fluorescence intensity in both fluorescence channels when the probes were competed against each other. Thus, both these probes seem to have very similar specificities. Labelling experiments comparing W-sCy5-VS and W-BF-VS were performed in triplicate. Those not shown in Fig 4 are shown in S5 Fig.

We then used the W-BF-VS probe to affinity-purify its targets and identify them by MS. Briefly, schizont lysates were treated with DMSO or W-BF-VS under acidic conditions, the modified proteins were then pulled-down with streptavidin agarose beads and after multiple stringent washes, trypsin was added to the beads. This was performed in triplicate, and the tryptic peptides from each sample labelled with a different tandem mass tag (TMT). TMT-labelled peptides were pooled and fractionated into 8 samples by high pH reversed-phase fractionation. Peptides in each fraction were detected and quantified by LCMS/MS/MS analysis. Table 1 shows the 10 most enriched proteins (Table 1 and S1 Fig) when comparing probe vs DMSO treatment. Among these, we could identify FP2, FP3 and both DPAPs, but not FP1. Note that FP2a and FP2b only differ in a single amino acid, therefore, the peptides identified for these proteins could not be assigned to either one of them. Interestingly, among the most enriched proteins we also identified several enzymes involved in redox catalysis such as thioredoxin or a disulphide isomerase, which are known to contain highly reactive catalytic cysteines. These might react with the vinyl sulfone group in our probes. Therefore, some of the unidentified bands labelled by W-sCy5-VS or W-BF-VS might correspond to proteins containing highly nucleophilic cysteines. Overall, this chemical proteomic experiment confirms that the main targets of our probes are indeed clan CA proteases.

Fig 4. Comparison of labelling profiles of W-sCy5-VS and W-BF-VS. (A) Schizont lysates were treated for 1 h either with 0.5 μM of W-sCy5-VS, W-BF-VS, or a mixture of both probes, each at 0.5 μM (Mix). After running the samples in a SDS-PAGE gel, the gel was scanned either in the Cy5 and Cy3 channels. The composite image shows very similar labelling profiles for both probes and a clear co-migration of the labelled bands in the Mix sample. (B) Coomassie staining of the gel shown in A showing equal protein loading. (C) Quantification of the labelling profiles for each probe by densitometry. Fluorescent intensity vs. migration distance (Rf) is shown. The position of FP2/3 and DPAP1 are indicated in A and C. Two additional biological replicates of this experiment are shown in S5 Fig.
Probe specificity in mammalian cells

Finally, to determine whether our probes, especially W-sCy5-VS and W-Cy5-VS, could also be used to profile mammalian clan CA proteases, we tested them in RAW macrophages, both in lysates and intact cells (Fig 5). The labelling profiles of these probes were compared to that of Table 1. Top ten enriched proteins using the W-BF-VS probe.

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Protein Name</th>
<th>Fold-Enrichment</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF3D7_1116700</td>
<td>DPAP1</td>
<td>8.9</td>
<td>Cysteine Protease</td>
</tr>
<tr>
<td>PF3D7_0827900</td>
<td>PDI8</td>
<td>6.4</td>
<td>Protein disulphide isomerase</td>
</tr>
<tr>
<td>PF3D7_1115400</td>
<td>FP3</td>
<td>4.5</td>
<td>Cysteine Protease</td>
</tr>
<tr>
<td>PF3D7_1115300/700</td>
<td>FP2a/b</td>
<td>3.9</td>
<td>Cysteine Protease</td>
</tr>
<tr>
<td>PF3D7_1205600</td>
<td>Tetrapact peptide repeat protein</td>
<td>3.8</td>
<td>Putative protein, unknown function</td>
</tr>
<tr>
<td>PF3D7_1302100</td>
<td>G27/25</td>
<td>3.3</td>
<td>Gamete antigen 27/25</td>
</tr>
<tr>
<td>PF3D7_1352500</td>
<td>Thioredoxin-related protein</td>
<td>3.3</td>
<td>Putative thioredoxin</td>
</tr>
<tr>
<td>PF3D7_1108600</td>
<td>ERC</td>
<td>3.1</td>
<td>ER-resident thioredoxin</td>
</tr>
<tr>
<td>PF3D7_0404700</td>
<td>DPAP3</td>
<td>3.0</td>
<td>Cysteine Protease</td>
</tr>
<tr>
<td>PF3D7_0622800</td>
<td>Leu-tRNA ligase</td>
<td>2.8</td>
<td>Putative protein</td>
</tr>
</tbody>
</table>

Clan CA proteases are indicated in bold, redox enzymes containing a catalytic cysteine in italic.

https://doi.org/10.1371/journal.pone.0227341.t001

Fig 5. Labelling of cysteine cathepsins in mammalian cells.

(A) Lysates from RAW macrophages were treated with 1 μM of the indicated ABPs for 30 min. (B) Live RAW cells were treated with 1 μM of the indicated ABPs for 3 h. Samples were run on a SDS-PAGE gel, and in-gel fluorescence measured using a fluorescence scanner. The identity of the different cysteine cathepsins are indicated with different coloured arrowheads.

https://doi.org/10.1371/journal.pone.0227341.g005
FY01 and BMV109-sCy5, a non-peptidic quenched ABP that has been shown to efficiently label most cysteine cathepsin endopeptidases in mammalian cells[37]. As shown in Fig 5A, both FY01 and our probes label a diffuse band running below the CatL band that corresponds to CatC[19]. This band is more clearly observable in the W-Cy5-VS and FY01 treated samples because these two probes do not label CatL. W-sCy3-VS and W-sCy5-VS are able to label all the cysteine cathepsins labelled by BMV109 in addition to CatC. In intact macrophages, W-Cy5-VS is able to label most clan CA endopeptidases but not as efficiently as BMV109. However, contrary to BMV109, W-Cy5-VS is also able to label CatC in intact cells. Here, we also observed that probes having sCy3 and sCy5 label much less efficiently that those containing Cy3 or Cy5, indicating that negatively charged fluorophores impair cell permeability. Similar results were obtained when these probes were tested in mouse lung and spleen tissues, or in DC1940 mouse dendritic cells and SSC-9 human oral cancer cells (S6 Fig), thus showing that W-Cy5-VS and W-sCy5-VS are able to label most clan CA cysteine proteases in a variety of mammalian samples.

**Conclusion**

Overall, we have shown that W-sCy5-VS is a broad-spectrum clan CA proteases ABP able to efficiently label endopeptidases and DPAPs both in *P. falciparum* and mammalian cell lysates, thus making it a better probe to profile this enzyme family than currently available ones. Unfortunately, the negatively charge nature of the sCy5 fluorophore significantly reduces its cell permeability. However, replacement of the sCy5 with Cy5 allows the probe to cross membranes, thus making W-Cy5-VS an ideal tool to profile the activity of these proteases in living cells.

**Supporting information**

- **S1 Fig. Probes purity.** LCMS traces (left) of the different ABPs synthesized for this study and accurate mass spectra (right) detected at the maximum peak of absorbance. (PDF)

- **S2 Fig. Enrichment of proteins labelled with the W-BF-VS affinity probe.** Volcano plot showing enrichment of clan CA proteases and enzymes containing reactive cysteine when schizont lysates were treated with W-BF-VS versus DMSO. (TIFF)

- **S3 Fig. Labelling of cysteine proteases in parasite lysates, additional biological replicates.** (A) Merozoite lysates diluted 1:10 in acetate buffer were treated for 1 h with 1–1000 nM of the indicated ABPs. For the highest ABP concentration, samples were also pre-treated for 30 min with 10 μM of SAK1, W-hPG-VS D-W-hPG-VS, SAK2 or E64. (B–C) Lysates collected at trophozoite (B) or schizont (C) stages were diluted in acetate buffer (pH 5.5), pre-treated for 30 min with DMSO or 10 μM of different known covalent inhibitors of DPAP1 (SAK2), DPAP3 (SAK1 or W-hPG-VS), the FPs (E64), or the negative control compound D-W-hPG-VS. This was followed by 1 h labelling with the different ABPs at 0.1 μM except for DCG04 that was used at 1 μM concentration. (A–C) The fluorescent bands corresponding to DPAP1, DPAP3, FP1, and FP2/3 are indicated by blue, red, light green, and dark green arrowheads, respectively. Each page represent a different biological replicate. (PDF)

- **S4 Fig. Labelling of cysteine protease in live parasites, additional biological replicates.** Very mature schizonts were diluted ten-fold in RPMI and treated with different concentrations of probes for 1 h. Samples were run on a SDS-PAGE gel, and the labelled proteins detected using

---

*PLOS ONE | https://doi.org/10.1371/journal.pone.0227341 January 10, 2020*
at fluorescence scanner. Bands corresponding to DPAP1, DPAP3, FP1, and FP2/3 are indicated by blue, red, light green, and dark green arrowheads, respectively. Top and bottom panels represent different biological replicates.

(TIF)

**S5 Fig. Comparison of labelling profiles of W-sCy5-VS and W-BF-VS, additional biological replicates.** (A) Schizont lysates were treated for 1 h either with 0.5 μM of W-sCy5-VS, W-BF-VS, or a mixture of both probes, each at 0.5 μM (Mix). (B) Schizont lysates were pre-treated for 30 min with either 10 μM of E64 or 0.5 μM of one of the probes followed by 1h labelling with the other probe or a combination of both, each at 0.5 μM (last two lanes). After running the samples in a SDS-PAGE gel, the gel was scanned either in the Cy5 and Cy3 channels. The composite image (merge channels images) shows very similar labelling profiles for both probes and a clear co-migration of the labelled bands in the Mix sample. Note that E64 is able to outcompete probe labelling for most of the bands. While pre-treatment with W-sCy5-VS results in a decrease in the labelling intensity of the bands labelled by W-BF-VS, pre-treatment with W-BF-VS does not decrease the labelling of band by W-sCy5-VS. This is consistent with W-sCy5-VS being a more potent and reactive probe.

(TIF)

**S6 Fig. Labelling of mammalian cysteine cathepsins, additional experiments.** Lysates from the indicated cell lines or from mouse lung or spleen tissues were treated with 1 μM of the different ABPs for 30 min. The identity of the different cysteine cathepsins are indicated with different coloured arrowheads.

(TIF)

**Acknowledgments**

We would like to thanks Prof. Matthew Bogyo for providing the FY01, DCG04, and BMV109 probes, and MRCT for providing Compound 2.

**Author Contributions**

**Conceptualization:** Dara Davison, Stephen Howell, Ambrosius Snijders, Laura E. Edgington-Mitchell, Edgar Deu.

**Data curation:** Michele S. Y. Tan, Dara Davison, Mateo I. Sanchez, Bethany M. Anderson, Stephen Howell, Ambrosius Snijders, Laura E. Edgington-Mitchell, Edgar Deu.

**Formal analysis:** Michele S. Y. Tan, Dara Davison, Mateo I. Sanchez, Stephen Howell, Laura E. Edgington-Mitchell, Edgar Deu.

**Funding acquisition:** Laura E. Edgington-Mitchell, Edgar Deu.

**Investigation:** Michele S. Y. Tan, Dara Davison, Mateo I. Sanchez, Bethany M. Anderson, Stephen Howell, Edgar Deu.

**Methodology:** Michele S. Y. Tan, Stephen Howell, Edgar Deu.

**Project administration:** Ambrosius Snijders, Laura E. Edgington-Mitchell, Edgar Deu.

**Resources:** Laura E. Edgington-Mitchell.

**Supervision:** Ambrosius Snijders, Laura E. Edgington-Mitchell, Edgar Deu.

**Validation:** Stephen Howell, Laura E. Edgington-Mitchell.
Visualization: Michele S. Y. Tan, Dara Davison, Mateo I. Sanchez, Bethany M. Anderson.

Writing – original draft: Mateo I. Sanchez, Edgar Deu.

Writing – review & editing: Michele S. Y. Tan, Dara Davison, Mateo I. Sanchez, Stephen Howell, Ambrosius Snijders, Laura E. Edgington-Mitchell, Edgar Deu.

References


References


Burda, P.C., Roelli, M.A., Schaffner, M., Khan, S.M., Janse, C.J., Heussler, V.T., 2015. A Plasmodium phospholipase is involved in disruption of the liver stage parasitophorous


Fujino, T., Watanabe, K., Beppu, M., Kikugawa, K., Yasuda, H., 2000. Identification of oxidized protein hydrolase of human erythrocytes as acylpeptide hydrolase1. The nucleotide sequence of human ACPH described in this paper appears in the GenBank databases with accession number AF141383, Homo sapiens oxidized protein hydrolase (OPH) mRNA, complete cds. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology 1478, 102–112. https://doi.org/10.1016/S0167-4838(00)00004-2


242


Identified from the Open Access Malaria Box. PLoS Negl Trop Dis 8, e2610. https://doi.org/10.1371/journal.pntd.0002610


Käll, L., Storey, J.D., MacCoss, M.J., Noble, W.S., 2008a. Posterior Error Probabilities and False Discovery Rates: Two Sides of the Same Coin. J. Proteome Res. 7, 40–44. https://doi.org/10.1021/pr700739d


Tiwari, N.K., Reynolds, P.J., Calderón, A.I., 2016. Preliminary LC-MS Based Screening for Inhibitors of Plasmodium falciparum Thioredoxin Reductase (PfTrxR) among a Set of Antimalarials from the Malaria Box. Molecules 21, 424. https://doi.org/10.3390/molecules21040424


Pathways, and Reprogramming Gene Expression. EBioMedicine 2, 808–824. https://doi.org/10.1016/j.ebiom.2015.06.020


