MEDICINAL CHEMISTRY THEME

MONASH INSTITUTE OF PHARMACEUTICAL SCIENCES (MIPS)

Ligands for the Malaria Apical Membrane Antigen 1 as Inhibitors of Parasite Invasion of Host Erythrocytes: Screening, Structure and Dynamics

Doctor of Philosophy (PhD) Candidature

Supervisors: Professor Raymond S. Norton
Professor Peter J. Scammells

Student Name: San Sui Lim

[redacted]

A thesis submitted to the Faculty of Pharmacy and Pharmaceutical Science, Monash University in fulfilment of the requirements for the Doctor of Philosophy candidature.

ERRATA

Page ix, para 2, 8th line: "corresponds to a 5 % hit rate" for "corresponds to 5 % hit rate"

Page ix, para 2, 9th line: "druggable site is present" for "druggable is present"

Page ix, para 3, 3rd line: "3D7 strain was also" for "3D7 strain were also"

Page 2, section 1.2, 10th line: "gametocytes are activated to produce gametes which undergo fertilisation to develop into sporozoites" for "gametocytes fuse and develop into sporozoites"

Page 3, section 1.3, 1st line: "catabolic" for "metabolic"

Page 6, figure 2 caption, 2nd line: "use the moving junction" and "invasion" for "use moving junction" and "invasions", respectively

Page 6, section 1.7, 4th line: "target hepatocyte" for "target the hepatocyte"

Page 6, section 1.7, 5th line: "fewer" for "lesser"

Page 6, section 1.7, 7th line: "small" for "smaller"

Page 6, section 1.7, 8th line: "evidence that AMA1" for "evidence that the AMA1"

Page 6, section 1.8, 2nd line: "spectroscopy" for "septroscopy"

Page 8, para 2, 2nd line: correct reference 67 is "Vulliez-Le Normand, B.; Tonkin, M. L.; Lamarque, M. H.; Langer, S.; Hoos, S.; Roques, M.; Saul, F. A.; Faber, B. W.; Bentley, G. A.; Boulanger, M. J.; Lebrun, M., Structural and functional insights into the malaria parasite moving junction complex. *PLoS Pathog*. **2012,** *8* (6), e1002755."

Page 12, para 2, 8th line: "Despite their weak binding" for "Despite of their weak binding"

Page 26, chapter title: "Binding Sites of Fragments" for "Binding Site of fragment"

ADDENDA

Page 5, section 1.6, para 2, 4th and 5th lines: Replace "RON proteins are derived from rhoptries, and are localised onto the host cells during invasion." <u>with</u> "In *P. falciparum*, RON2, 4 and 5 proteins derived from rhoptries are localised onto the red cells during invasion.[1-3] Orthologs of these RON proteins as well as RON8 protein are found in *T. gondii*, and are translocated in a similar way onto their host cells.[4-6] RON4, 5 and 8 are located entirely inside host cell cytoplasm to interact with host cytoskeleton, and this interaction may serve as an anchor point upon which the parasites can apply traction for invasion.[7-9] RON2 is a transmembrane protein with a C-terminal region exposed on the host cell surface.[2, 7]"

Page 5, section 1.6: Add at the end of para 2: "To date, the precise function of the AMA1 cytoplasmic region remains unclear. Early studies found that this region of AMA1 interacts with aldolase (ALD) in both *P. falciparum*[7] and *T. gondii*[10, 11], thereby providing a link to the actinomyosin motor that generates propulsive forces required for host-cell invasion.[12] However, this model was not supported by a more recent study which showed that disruption of the *Tg*AMA1-*Tg*ALD interactions did not result in the impairment of host-cell invasion.[9, 13] Moreover, contrary to the previously described role, ALD was found not to play essential role in invasion, but rather is primarily important for energy metabolism. The potential role of AMA1 as a signalling

protein has also been investigated.[14] It was shown that phosphorylation of AMA1 cytoplasmic region, particularly at Ser610, by protein kinase A is important for host-cell invasion by *P. falciparum*."

Page 13, section 1.12: Add after "resistance to many existing malaria drugs.": "In the presence of AMA1 ligands, R1 and RON2 peptides, *Plasmodium* parasites were still able to attach to the red cells, followed by a reorientation step that juxtaposed its apical end with the red cell membrane, but failed to invade their host cells.[7, 11] Small-molecule inhibitors of AMA1 would be expected to modulate the invasion process in a similar manner."

Page 15: Add between para 1 and 2: "The detailed binding poses of our fragment hits may be obtained either experimentally or computationally.[15, 16] *In silico* docking is a relatively fast and cost-effective technique, and may be a viable option to obtain structural information in cases where high-resolution structures are available. However, there are several drawbacks with this approach, one of which is the relatively low accuracy of the predictions that may hamper the process of designing inhibitors with desirable properties.[15] Therefore, we elected to obtain experimental structural information about the fragment-AMA1 interactions by NMR spectroscopy and X-ray crystallography."

References

1. Alexander, D. L.; Arastu-Kapur, S.; Dubremetz, J.-F.; Boothroyd, J. C., *Plasmodium falciparum* AMA1 binds a rhoptry neck protein homologous to *Tg*RON4, a component of the moving junction in *Toxoplasma gondii*. *Eukaryot. Cell* **2006,** *5* (7), 1169-1173.
2. Cao, J.; Kaneko, O.; Thongkukiatkul, A.; Tachibana, M.; Otsuki, H.; Gao, Q.; Tsuboi, T.; Torii, M., Rhoptry neck protein RON2 forms a complex with microneme protein AMA1 in *Plasmodium falciparum* merozoites. *Parasitol. Int.* **2009,** *58* (1), 29-35.
3. Proellocks, N. I.; Coppel, R. L.; Waller, K. L., Dissecting the apicomplexan rhoptry neck proteins. *Trends Parasitol.* **2010,** *26* (6), 297-304.
4. Besteiro, S.; Michelin, A.; Poncet, J.; Dubremetz, J.-F.; Lebrun, M., Export of a *Toxoplasma gondii* rhoptry neck protein complex at the host cell membrane to form the moving junction during invasion. *PLoS Pathog.* **2009,** *5* (2), e1000309.
5. Straub, K. W.; Cheng, S. J.; Sohn, C. S.; Bradley, P. J., Novel components of the apicomplexan moving junction reveal conserved and coccidia-restricted elements. *Cell. Microbiol.* **2009,** *11* (4), 590-603.
6. Alexander, D. L.; Mital, J.; Ward, G. E.; Bradley, P.; Boothroyd, J. C., Identification of the moving junction complex of *Toxoplasma gondii*: a collaboration between distinct secretory organelles. *PLoS Pathog.* **2005,** *1* (2), e17.
7. Srinivasan, P.; Beatty, W. L.; Diouf, A.; Herrera, R.; Ambroggio, X.; Moch, J. K.; Tyler, J. S.; Narum, D. L.; Pierce, S. K.; Boothroyd, J. C.; Haynes, J. D.; Miller, L. H., Binding of *Plasmodium* merozoite proteins RON2 and AMA1 triggers commitment to invasion. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (32), 13275-13280.

8. Takemae, H.; Sugi, T.; Kobayashi, K.; Gong, H.; Ishiwa, A.; Recuenco, F. C.; Murakoshi, F.; Iwanaga, T.; Inomata, A.; Horimoto, T.; Akashi, H.; Kato, K., Characterization of the interaction between *Toxoplasma gondii* rhoptry neck protein 4 and host cellular beta-tubulin. *Sci. Rep.* **2013,** *3*.
9. Harvey, K. L.; Yap, A.; Gilson, P. R.; Cowman, A. F.; Crabb, B. S., Insights and controversies into the role of the key apicomplexan invasion ligand, apical membrane antigen 1. *Int. J. Parasitol.* **2014,** *44* (12), 853-857.
10. Sheiner, L.; Santos, J. M.; Klages, N.; Parussini, F.; Jemmely, N.; Friedrich, N.; Ward, G. E.; Soldati-Favre, D., *Toxoplasma gondii* transmembrane microneme proteins and their modular design. *Mol. Microbiol.* **2010,** *77* (4), 912-929.
11. Treeck, M.; Zacherl, S.; Herrmann, S.; Cabrera, A.; Kono, M.; Struck, N. S.; Engelberg, K.; Haase, S.; Frischknecht, F.; Miura, K.; Spielmann, T.; Gilberger, T. W., Functional analysis of the leading malaria vaccine candidate AMA1 reveals an essential role for the cytoplasmic domain in the invasion process. *PLoS Pathog.* **2009,** *5* (3), e1000322.
12. Besteiro, S.; Dubremetz, J.-F.; Lebrun, M., The moving junction of apicomplexan parasites: a key structure for invasion. *Cell. Microbiol.* **2011,** *13* (6), 797-805.
13. Shen, B.; Sibley, L. D., *Toxoplasma* aldolase is required for metabolism but dispensable for host-cell invasion. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111* (9), 3567-3572.
14. Leykauf, K.; Treeck, M.; Gilson, P. R.; Nebl, T.; Braulke, T.; Cowman, A. F.; Gilberger, T. W.; Crabb, B. S., Protein kinase A dependent dhosphorylation of apical membrane antigen 1 plays an important role in erythrocyte invasion by the malaria parasite. *PLoS Pathog.* **2010,** *6* (6), e1000941.
15. Kumar, A.; Voet, A.; Zhang, K. Y., Fragment based drug design: from experimental to computational approaches. *Curr. Med. Chem.* **2012,** *19* (30), 5128-47.
16. Scott, D. E.; Coyne, A. G.; Hudson, S. A.; Abell, C., Fragment-based approaches in drug discovery and chemical biology. *Biochemistry* **2012,** *51* (25), 4990-5003.

Table of Contents

Statement of Originality

To the best of the candidate's knowledge and belief, this thesis contains no material which has been accepted for the award of any other degree or diploma in any University or other institutions and contains no material previously published or written by another person except where due reference is made.

San Sui Lim

General Declaration

In accordance with Monash University Doctorate Regulation 17.2 Doctor of Philosophy and Research Master's regulations the following declarations are made:

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes 2 original papers published in peer-reviewed journals, 1 submitted manuscript and 1 traditional chapter. The core theme of the thesis is medicinal chemistry. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the candidate, working within the Monash Institute of Pharmaceutical Sciences (MIPS) under the supervision of Professor Raymond S. Norton and Professor Peter J. Scammells.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

In the case of chapter 2, 3, 4 and 5 my contribution to the work involved the following:

Thesis chapter	Publication title	Publication status	Nature and extent of candidate's contribution
2	Development of inhibitors of *Plasmodium falciparum* apical membrane antigen 1 based on fragment screening	Published	70 %
3	A critical evaluation of pyrrolo[2,3-*d*]pyrimidine-4-amines as *Plasmodium falciparum* apical membrane antigen 1 (AMA1) inhibitors	Published	50 %
4	Structure and dynamics of apical membrane antigen 1 from *Plasmodium falciparum* FVO	Submitted	70 %

5	Mapping the fragment binding sites on malaria apical membrane antigen 1 using high resolution NMR spectroscopy	Traditional chapter	80 %

I have not renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Signed: [redacted]

Date: 5th October 2014

Acknowledgements

I am fully indebted to my supervisors, Professor Raymond S. Norton and Professor Peter J. Scammells, not only for providing me the opportunity to work in a great area of research, but also for their understanding, wisdom, patience, enthusiasm, and encouragement and for pushing me farther than I thought I could go.

I would like to also express my sincere gratitude to Drs Christopher A. MacRaild, Indu R. Chandrashekaran, Hiromasa Yagi, Shane M. Devine, Marie Ø Pedersen, Eleanor Leung, and Charles A. Galea for all their patience, guidance and constructive ideas. In particular, Dr Christopher A. MacRaild, Dr Indu R. Chandrashekaran, Dr Hiromasa Yagi and Dr Shane M. Devine's recommendations and suggestions have been invaluable for my research progress and without their help I could not have got this far.

I wish to extend my gratitude to Associate Professor Martin J. Scanlon, Emeritus Professor Robin F. Anders and Dr Sheena McGowan whose support has been invaluable. To Dr Jamie S. Simpson and the members of the Scanlon group, especially Dr Martin L. Williams and Dr Mark Mulcair, thank you for your constant help and advice. A thank you also goes to Dr Nyssa Drinkwater.

I also thank my fellow colleagues Parvathy Ashok, Krishnarjuna Bankala, Keith Khoo, Samuel D. Robinson, Tony Wang and Jeff Chang for their support and help.

Last but not least I wish to avail myself of this opportunity to express a sense of gratitude and love to my friends and my beloved parents for their support, strength, help and for everything.

Abbreviations

ACT	Artemisinin combination therapy
AMA1	Apical membrane antigen 1
CPMG	Carr-Purcell-Meiboom-Gill
CSI	Chemical shift index
CSP	Chemical shift perturbation
DMSO	Dimethyl sulfoxide
FBLD	Fragment based ligand design
FPLC	Fast protein liquid chromatography
GdnHCl	Guanidine hydrochloride
GSH	Reduced glutathione
GSSG	Oxidised glutathione
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
HSQC	Heteronuclear single-quantum coherence
HTS	High-throughput screening
IC_{50}	Half maximal inhibitory concentration
IEX	Ion exchange
LB	Luria broth
LE	Ligand efficiency
MD	Molecular dynamics
MJ	Moving junction
MMP3	Matrix metalloproteinase-3
NMR	Nuclear magnetic resonance
PAN	Plasminogen-apple-nematode
PBS	Phosphate buffered saline
Pf	*Plasmodium falciparum*
PPI	Protein-protein interaction
Pv	*Plasmodium vivax*
RON	Rhoptry neck protein
RU	Response unit
SAR	Structure-activity relationship
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SPR	Surface plasmon resonance

STD	Saturation transfer difference
SUB2	Subtilisin-like protease 2
TFA	Trifluoroacetic acid
Tg	*Toxoplasma gondii*
TROSY	Transverse relaxation optimised spectroscopy
WHO	World Health Organisation

Abstract

Malaria is one of the most widespread infectious diseases, causing approximately 250 million clinical cases and claiming more than 650,000 lives each year. Although current artemisinin combination therapies (ACT) have been highly effective against *Plasmodium* parasites, signs of resistance have already emerged. Hence, there is an urgent need for drugs with new modes of action to combat this threat. Apical membrane antigen 1 (AMA1) interacts with rhoptry neck (RON) protein complex to form part of the moving junction (MJ) complex important for the invasion of human red blood cells by *Plasmodium falciparum* (*Pf*). AMA1 has a conserved hydrophobic cleft that is the site of interactions with the rhoptry neck (RON) protein complex. Peptides identified by phage display, such as R1, as well as monoclonal antibodies that target this site on AMA1, are able to inhibit red blood cell invasion, but usually in a strain-specific manner as numerous polymorphic residues are clustered at one end of the cleft.

My goal is to design small molecule inhibitors of AMA1 that have broad strain specificity and we are pursuing this goal using a fragment-based approach. My project began with cocktail screening of a fragment library against AMA1 using saturation transfer difference (STD). The hits found from the cocktail screen were then evaluated in the STD and Carr-Purcell-Meiboom-Gill (CPMG) R1 competition assays to identify hits that bind to the AMA1 hydrophobic cleft. Thereafter, the binding affinities (K_D) and ligand efficiencies (LE) of the R1-competing hits were determined using surface plasmon resonance (SPR). A total of 57 fragment hits were identified in the screening campaign which corresponds to 5 % hit rate. The high hit rate observed strongly suggest that a druggable is present within the cleft. Subsequently, my work focused on mapping the specific binding sites of the hits using ^{1}H-^{15}N HSQC perturbation studies of *Pf*AMA1 backbone amide resonances. To perform this study, the backbone amide resonances were first assigned using a combination of 3D NMR and specific ^{15}N-Lys labelled HSQC experiments. The HSQC perturbation experiments identified fragments that bind to a conserved region on the AMA1 hydrophobic cleft, and these compounds represent promising starting scaffolds for subsequent chemical elaborations.

The first X-ray crystal structure of FVO *Pf*AMA1 was determined to understand the impact of sequence diversity on AMA1 structure and facilitate the design of small-molecule inhibitors. The crystal structure of AMA1 from the *P. falciparum* 3D7 strain were also reproduced at higher resolution in an attempt to obtain binding poses of fragments bound to the antigenically diverse forms of AMA1 (FVO and 3D7). Currently we are working towards

getting the crystal structures of the fragments bound to different forms of AMA1, which will allow more rational design of fragment analogues. In parallel, chemical modifications of the hits based on the structure-activity relationship (SAR) of the analogues are underway to improve their binding affinities.

Chapter 1:

Introduction

1.1 Introduction

Malaria is one of the most widespread infectious diseases, with approximately 40% of the world's population living at risk of contracting the disease.[1] Young children are particularly vulnerable and bear a disproportionately large burden of malaria morbidity and mortality.[2-3] The main species that cause malaria in humans are *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae* and *P. knowlesi*.[1] These *Plasmodium* parasites collectively cause around 250 million clinical cases per year, resulting in more than 600,000 deaths. Among them, *P. falciparum* is the most deadly species, being responsible for 90% of the total malaria-related deaths, as well as being the most prevalent malaria species worldwide and especially common in sub-Saharan Africa. *P. vivax* is the second most common species, with 70 to 80 million cases occurring annually, mainly in Asia and South America.[1, 4] As a result of its substantial health burden, malaria is also a cause of poverty and a major hindrance to economic development.[5]

1.2 Life Cycle of Malaria Parasite

Transmission of *Plasmodium* species occurs through the female *Anopheles* mosquito (Figure 1).[6] During an *Anopheles* mosquito blood meal, *Plasmodium* sporozoites are inoculated into the human skin.[7] Some of these sporozoites enter the circulation and are carried to the liver.[6, 8] The *Plasmodium* parasites then mature into schizonts, each of which produces several thousand merozoites that are released into the bloodstream when the liver cells rupture. Each merozoite invades the red blood cells and then multiply further for subsequent erythrocyte invasions, thus beginning the erythrocytic stage of malaria infection.

In the circulation, some merozoites develop into male and female gametocytes and are taken up into the female mosquito gut during a blood meal.[9] In the mosquito gut, the male and female gametocytes fuse and develop into sporozoites, which then migrate to the salivary glands of the mosquito. Inoculation of the sporozoites into a new human host by the mosquito starts a new malaria life cycle. With *P. vivax* and *P. ovale* (but not *P. falciparum* or *P. malariae*), some of the liver-stage parasites become dormant hypnozoites for up to several years before it is reactivated to cause relapsing malaria.[9]

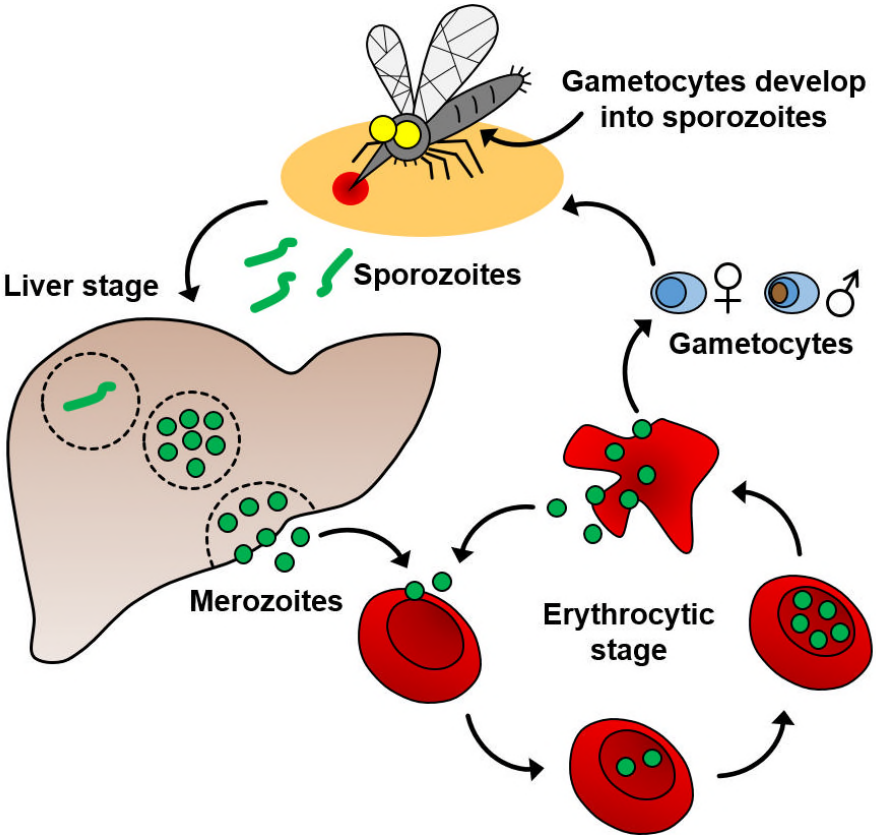


Figure 1. Malaria life cycle. The two hosts in this cycle are human and *Anopheles* mosquito.

1.3 Antimalarial Agents and Drug Resistance

Plasmodium species flourish in erythrocytes by digesting haemogloblin.[9-10] This metabolic reaction generates soluble haem, which is toxic to the parasites. To avoid these toxic effects, the parasite is able to biocrystallise the soluble haem into insoluble haemozin. Chloroquine and perhaps mefloquine appear to interfere with haem sequestration, and despite the widespread resistance to both drugs, haem sequestration remains a viable target for the development of new antimalarials.[11-12]

Pyrimethamine and sulfadoxine combination therapy have long been an important antimalarial combination.[9-10] Pyrimethamine exerts its pharmacological action by inhibiting parasite dihydrofolate reductase, thereby blocking the biosynthesis of purines and pyrimidines. Sulfadoxine inhibits the dihydropteroate synthetase enzyme preventing the utilisation of *para*-aminobenzoic acid in the synthesis of dihydropteroic acid. This antifolate combination was highly effective, cheap and well-tolerated, but unfortunately its clinical role in malaria therapy has declined due to the occurrence of resistance at both enzymes, caused by mutations in *pfdhfr* and *pfdhps* genes.[11, 13]

The World Health Organization (WHO) has recommended the use of artemisinin combination therapies (ACT) as first-line therapy for *P. falciparum* malaria worldwide.[1] The important artemisinin group of antimalarials possesses an endoperoxide moiety that is required for its antimalarial activity.[9-10, 14] Activation of the endoperoxide group occurs in the presence of free Fe^{3+}, which is released during digestion of haemoglobin. This is followed by rearrangement to produce a carbon-centered radical that alkylates and damages macromolecules in the parasites. Although artemisinin combination therapies have been highly effective, signs of resistance have already emerged in Africa and Southeast Asia.[15-17]

1.4 Apical Membrane Antigen 1

Apical membrane antigen 1 (AMA1) is a protein used by malaria parasites to invade their host cells.[18] It is found in all *Plasmodium* species, and orthologues exist in other apicomplexan parasites such as *Toxoplasma gondii*.[19-20] AMA1 is a structurally conserved type I integral membrane protein comprised of an N-terminal ectoplasmic region, a single transmembrane region, and a small C-terminal cytoplasmic region.[21] Based on the connectivities of its eight intramolecular disulfide bonds, the ectodomain was suggested to consist of an N-terminal domain I, a central domain II, and a C-terminal domain III, and this is consistent with the structures subsequently determined by X-ray crystallography.[22-23] AMA1 is synthesised as an 83 kDa precursor protein that is subsequently converted to a 66 kDa form.[24-25] Thereafter, it is translocated to the apical end of the parasite, where it functions in the invasion of red blood cells. Around the point of invasion, the bulk of the AMA1 ectodomain is shed quantitatively as a 48 kDa protein by the membrane-bound subtilisin-like 'sheddase' *Pf*SUB2.[26]

Although recent studies of knockdown and knockout AMA1 mutants of *P. berghei* and *Toxoplasma gondii* showed that AMA1 is dispensable for host cell invasion,[27-28] conditional knockdown of *Pf*AMA1 severely impaired the parasite's ability to invade red cells[29] and a complete gene knockout is not viable in *P. falciparum*.[30] Further studies found AMA1 homologues in *T. gondii* that enable host cell entry when AMA1 was inactivated,[31] but most of these homologues are absent in *P. falciparum*, possibly accounting for these apparently divergent results.[32] Ligands that disrupt the AMA1 function inhibit the *in vitro* growth of *P. falciparum* asexual blood stage.[33-37]

1.5 AMA1 as Vaccine Candidate

Antibodies to AMA1 contribute to the adaptive immune response that partially protects exposed individuals against malaria, and several anti-AMA1 antibodies are capable of disrupting red cell invasion by *P. falciparum in vitro*.[38-39] Immunisation with the AMA1 ectodomain has provided protection against malaria in animal models.[40-41] For these reasons AMA1 has been a leading candidate for inclusion in a vaccine against *P. falciparum*.[42] A Phase 2b clinical trial in Mali that tested a monovalent *Pf*AMA1 (3D7) vaccine formulated in the AS02 adjuvant provided protection against a subset of *P. falciparum* AMA1 genotypes due to the fact that sequence polymorphism exists in AMA1.[43] In an attempt to overcome this problem, two forms of *Pf*AMA1 (3D7 and FVO) were included in the vaccine design.[42, 44] However, such bi-allelic *Pf*AMA1 (3D7 and FVO) vaccine failed to induce protection because of the poor immunogenicity of the aluminum hydroxide formulation.[45-46] In preclinical studies, vaccines containing 4 – 6 alleles of AMA1 elicited antibodies that were more broadly cross reactive.[47-48]

1.6 Moving Tight Junction and Erythrocyte Invasion

Malaria merozoites invade erythrocytes via a multistep process initiated by weak attachment to the host cell surface.[49] Thereafter, the merozoite reorientates itself with its apical end juxtaposed on the erythrocyte membrane. The parasite is then brought closer to the red cells, followed by moving junction (MJ) formation that firmly anchors the merozoite to the host cell. The moving junction moves from the apical to the posterior end of the merozoite as the parasite pulls its way into the erythrocyte. As the merozoite invades the erythrocyte, it also creates a parasitophorous vacuole to protect itself from the host-cell cytoplasm.

The moving junction (MJ) complex is an invasive machinery used by obligate intracellular apicomplexan parasites that include *P. falciparum* and *T. gondii*.[18, 50-52] This complex consists of various components that include AMA1 and RON proteins (Figure 2). AMA1 is stored in the micronemes and is released onto the merozoite surface prior to the invasion process. RON proteins are derived from rhoptries, and are localised onto the host cells during invasion. It has recently been shown that AMA1 uses the rhoptry neck protein, RON2, as a receptor to promote invasion, and this AMA1-RON2 interaction is important for the invasive process of *P. falciparum* and *T. gondii*.[49, 53-54]

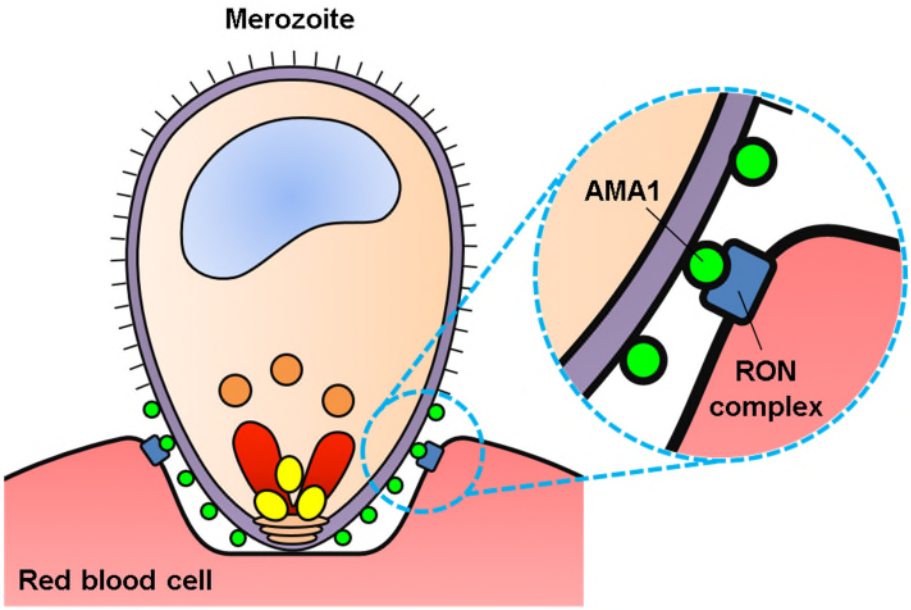


Figure 2. Moving junction consists of AMA1 and RON complex. Apicomplexan parasites use moving junction for host cell invasions.[51]

1.7 AMA1 in Sporozoite Invasion of Hepatocytes

As few as 10 *Plasmodium* parasites are injected initially into the human bloodstream by the mosquito.[55] This number is subsequently expanded by several thousand fold after successful invasion of liver cells, and this is followed by an expansion to several trillion in a mature blood stage infection.[55] Any therapeutic agents that target the hepatocyte invasion by sporozoites are expected to lead to true prevention (causal prophylaxis). As far lesser parasites are present in the initial liver stage, the risk of developing resistance against liver stage therapeutic agents would be smaller compared to those that act on the blood stage.[56]

Anti-AMA1 antibodies inhibit sporozoite invasion, providing evidence that the AMA1 is involved during invasion of hepatocytes.[57] Hence, a therapeutic agent targeting AMA1 not only has the potential to treat or suppress the development of symptomatic malaria (suppressive prophylaxis), but its activity on hepatic invasion (causal prophylaxis) may also provide further protection against the disease.

1.8 Structure of AMA1

Early structural studies were conducted using individual domains of AMA1 with nuclear magnetic resonance (NMR) sepctroscopy.[58-59] More recently the structures of larger AMA1 constructs were resolved by X-ray crystallography (Figure 3).[23, 60] Structural studies in *P. falciparum*[60] and *P. vivax*[23] revealed that both domain I and II possess a PAN (plasminogen-

apple-nematode) motif consisting of a five-stranded β-sheet that packs against a single α-helix. Several loops are extended from their PAN domain cores and the acquisition of these loops during AMA1 evolution may serve to facilitate parasite evasion of the protective human antibody responses.[60] Structural analyses of *Pf*AMA1 and *Pv*AMA1 indicate that they are highly homologous to each other (Figure 3A and B),[23, 60] but show significant levels of divergence from *T. gondii* AMA1 (Figure 3C).[61]

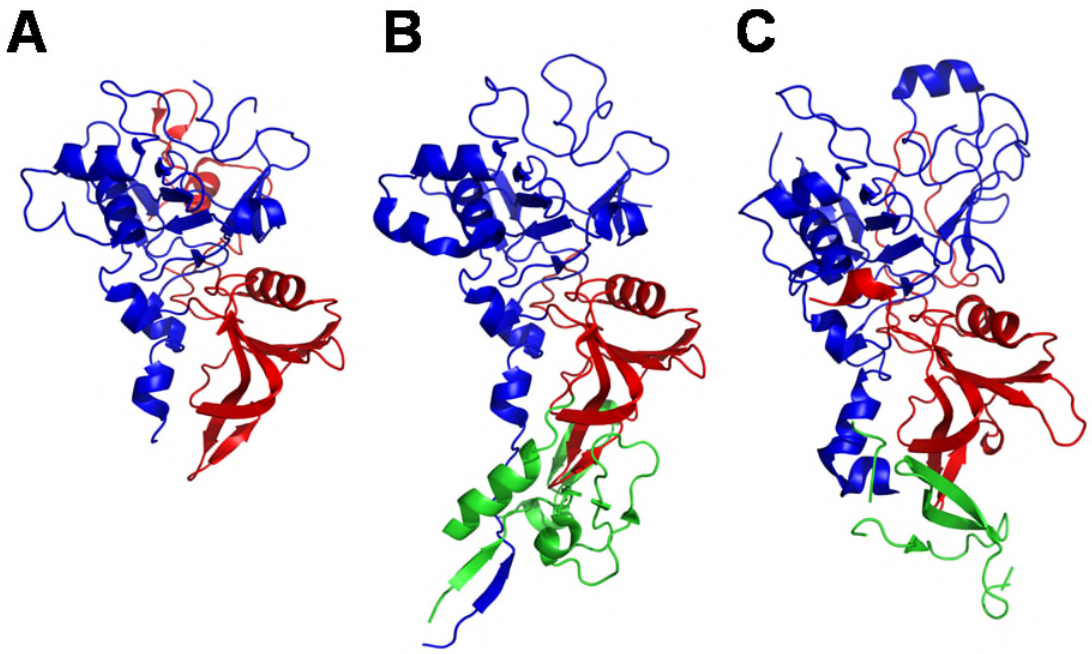


Figure 3. Crystal structures of (A) *Pf*AMA1 DI + II (PDB ID 1Z40)[60] (B) *Pv*AMA1 DI + II + III (PDB ID 1W81)[23] (C) *Tg*AMA1 DI + II + III (PDB ID 2X2Z)[61]. AMA1 DI + II + III are coloured blue, red and green, respectively.

AMA1 has a hydrophobic cleft that is the site of interaction with the RON complex (Figure 4).[60] It is surrounded by six loops of domain I (loops Ia – If) and an extended loop from domain II (DII loop). The AMA1 hydrophobic cleft can be viewed as having three separate elements: (1) At one end of the cleft there is a cluster of domain I (DI) polymorphic residues, probably as a result of diversifying selection to avoid the binding of naturally occurring human antibodies.[60] In particular, polymorphic residues in loop Id have been shown to be the most important region for mediating escape from vaccine induced and naturally acquired anti-AMA1 antibodies.[62-65] (2) The center is marked by an absolutely conserved Tyr251 across all apicomplexan AMA1 sequences, even those of the more distantly related *T. gondii*.[19, 60-61] Mutation of this central tyrosine in the cleft abrogates the binding to RON complex, highlighting the importance of this structural feature in formation

of the MJ complex.[49, 66] (3) The other end of the cleft is formed as a result of interactions between domain I and the domain II (DII) loop. This end of the cleft is highly conserved, with the AMA1 DII loop playing a crucial role in RON2 binding.

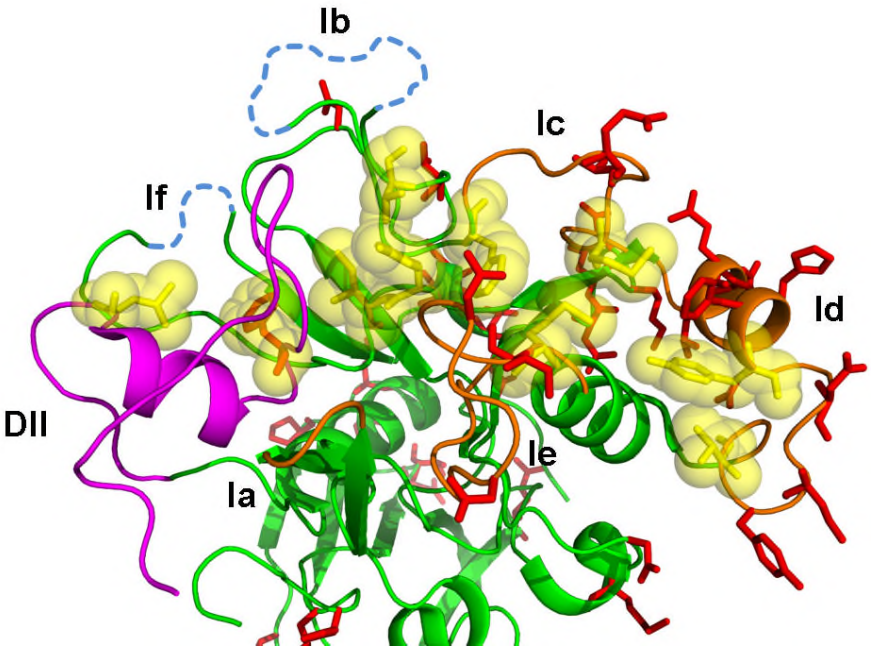


Figure 4. AMA1 hydrophobic cleft (PDB ID 1Z40). The hydrophobic residues that line the cleft are shown as yellow spheres. The cleft is surrounded by the Ia – f (orange) and DII (purple) loops. Disordered loops (Ib and If) are shown as dotted blue lines. Polymorphic residues are defined by red sticks.

A 38-mer synthetic *Pf*RON2 peptide corresponding to the critical binding region on *Pf*AMA1 hydrophobic cleft inhibits merozoite invasion at nanomolar concentration.[67] The co-crystal structure of *Pf*AMA1 and the *Pf*RON2 peptide revealed substantial conformational change in the DII loop of *Pf*AMA1 (Figure 5). In the unbound state, the DII loop is intimately associated with the hydrophobic cleft (Figure 5A), but is readily displaced for effective binding of *Pf*RON2 to *Pf*AMA1 (Figures 5B). The *Pf*RON2 peptide has a helical structure at its N-terminal and extended through an ordered coil to a disulfide-closed β-hairpin loop, forming a U shaped conformation (Figure 5B). *Pf*RON2 Arg-2041, located at the tip of the β-hairpin, is an important residue for AMA1 binding. Its guanidyl group fits snugly into a deep pocket of *Pf*AMA1, forming a complex network of seven hydrogen bonds. Similar binding paradigm of AMA1-RON2 was also observed in *T. gondii*.[68]

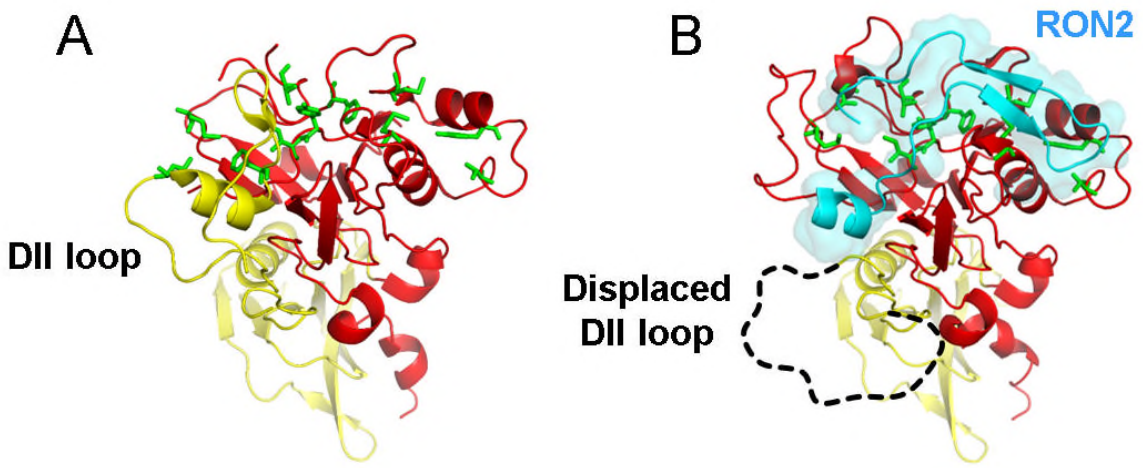


Figure 5. Co-crystal structure of *Pf*AMA1-*Pf*RON2 complex that links apicomplexan parasites and their target host cells.[68] (A) *Pf*AMA1 with ordered DII loop noted (PDB ID 1Z40), (B) Binding of RON2 peptide (cyan cartoon with transparent surface view) to *Pf*AMA1 displaced the DII loop (disordered state) (PDB ID 3ZWZ), *Pf*AMA1 DI + II are coloured red and yellow, respectively. The hydrophobic cleft is shown as green sticks. The disordered DII loop is indicated with black dotted line.

1.9 Antibodies Targeting the Hydrophobic Cleft

As a consequence of extensive polymorphisms, antibodies targeting *Pf*AMA1 often exhibit considerable strain specificity.[69-70] For example, a monoclonal shark antibody IgNAR is capable of inhibiting invasion of 3D7 and W2mef strains of *Plasmodium falciparum*, but not HB3. This limited cross-strain inhibitory effect is a result of binding to the end of the hydrophobic cleft surrounded by polymorphic amino acids (Figure 6).[70] The binding of monoclonal antibody 1F9 to the *Pf*AMA1 hydrophobic cleft was disrupted when the polymorphic residues 197, 200, 201, 204, and 225 were mutated.[71] In contrast, the monoclonal antibody 4G2 binds to the other end of the cleft, which lacks polymorphic residues, inhibits merozoite invasion in a strain-independent manner.[72] All of these antibodies exert their antimalarial activity by binding to the hydrophobic cleft, thus emphasising its importance as an inhibitory ligand-binding site.

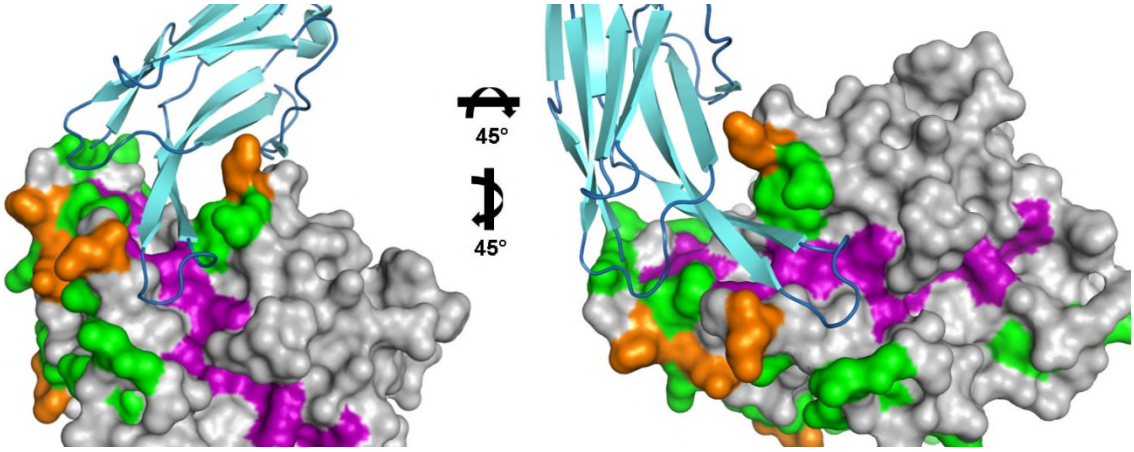


Figure 6. Co-crystal structure of IgNAR-*Pf*AMA1 complex (PDB ID 2Z8W).[70] IgNAR and *Pf*AMA1 are shown in cartoon and surface views, respectively. The hydrophobic cleft is coloured purple. Polymorphic residues are coloured green. Residues are defined as highly polymorphic (orange) if the third most abundant amino acid had a frequency of at least 5% in the non-redundant set of *Pf*AMA1 sequences.[60]

1.10 R1 Peptide

Screening random peptide libraries displayed on the surface of phage has yielded the *Pf*AMA1-binding peptide R1, which inhibits merozoite invasion.[41, 73-75] R1 binds to AMA1 with a K_D of approximately 100 nM and is capable of disrupting *Pf*AMA1-*Pf*RON2 interactions[52, 54 73] As with 1F9 and IgNAR, the potential of R1 as a therapeutic agent is limited by its strain-specific inhibitory activity to a subset of *P. falciparum*. The solution structure determined using NMR revealed that R1 contains two ordered turn-like conformations. The first region, involving residues 5 to 10 is hydrophobic, whereas the second region, encompassing residues 13 to 17, is hydrophilic.[75]

Early evidence that R1 binds across the cleft came from observations that R1 interacts with *Pf*AMA1 competitively with 1F9 and 4G2 antibodies, whose epitopes lie at opposite ends of the hydrophobic cleft.[73] These observations were confirmed using two-dimensional NMR experiments with methionine as probes.[52] Perturbation of signals in ^{1}H-^{13}C HSQC spectra demonstrated that the binding of R1 to *Pf*AMA1 spans the full length of the hydrophobic cleft, from Met273 at one end to the cluster of methionine residues at the other. Surprisingly, the co-crystal structure of R1-*Pf*AMA1 revealed that two molecules of peptides are bound to the AMA1 hydrophobic cleft (Figure 7A). Overall, a very similar binding paradigm was found between R1 and *Pf*RON2 peptide. One of the peptide (R1-major) lies deeply in the binding cleft while the other one lies (R1-minor) above R1-major, making

lesser interactions with *Pf*AMA1 (Figure 7A).[67] R1-major residue Arg15 makes most contacts within the same pocket of *Pf*AMA1 as *Pf*RON2 Arg2041. Analyses of the interactions showed that polymorphisms at positions 175 and 225 limit R1-major inhibitory activity against different forms of *Pf*AMA1. Further studies using NMR in solution reveal that only one R1 peptide binds to AMA1, suggesting that the presence of R1-minor may be resulted from crystallisation artefacts.[76]

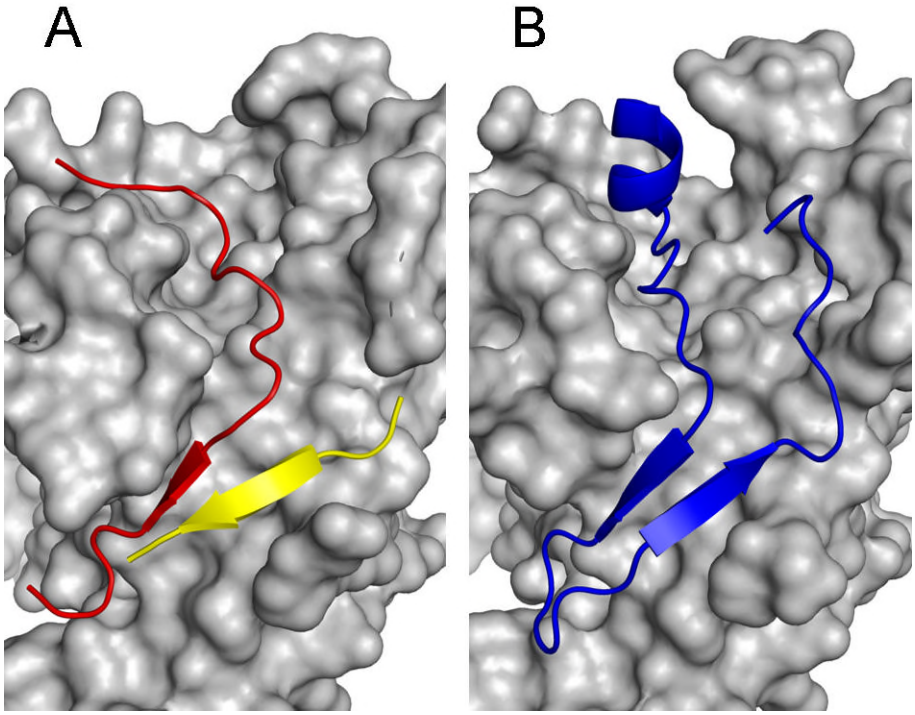


Figure 7. Structures of 3D7 *Pf*AMA1 (grey) with (A) R1-major and R1-minor (PDB ID 3SRJ) and (B) *Pf*RON2 peptides (PDB ID 3ZWZ).[77] R1-major, R1-minor and *Pf*RON2 peptide are coloured red, yellow and blue, respectively.

1.11 Fragment-Based Inhibitor Design

Target-based drug design is currently the most common approach adopted by pharmaceutical companies to discover new drugs.[78] Hit compounds are identified initially in this process and then their activities are improved by chemical modifications guided by structure–activity relationship (SAR) analysis to produce potent molecules with drug-like properties suitable for clinical evaluation.

High-throughput screening (HTS) is the predominant method within large pharmaceutical companies for hit identification.[78] In this method, thousands to millions of relatively complex compounds are screened with the aim of finding potent hits with low μM K_D values.[79-81] Despite continuous improvements in HTS methods, designing an automated system for

screening such large number of compounds remains a considerable challenge with this approach.[78, 80] Many false positive hits can arise in HTS screens, which requires further effort and financial investments to triage and validate.

An alternative method for hit identification is fragment-based lead design (FBLD), which has recently gained momentum in both large pharmaceutical companies and academic institutions as a means for discovering new drug candidates.[78-79, 81-84] Fragment-based lead design is based on screening low molecular weight fragments against the target of interest (Figure 8). The number of fragments typically screened is in the range of hundreds to a few thousand to find suitable hits as starting points for development.[85-86]

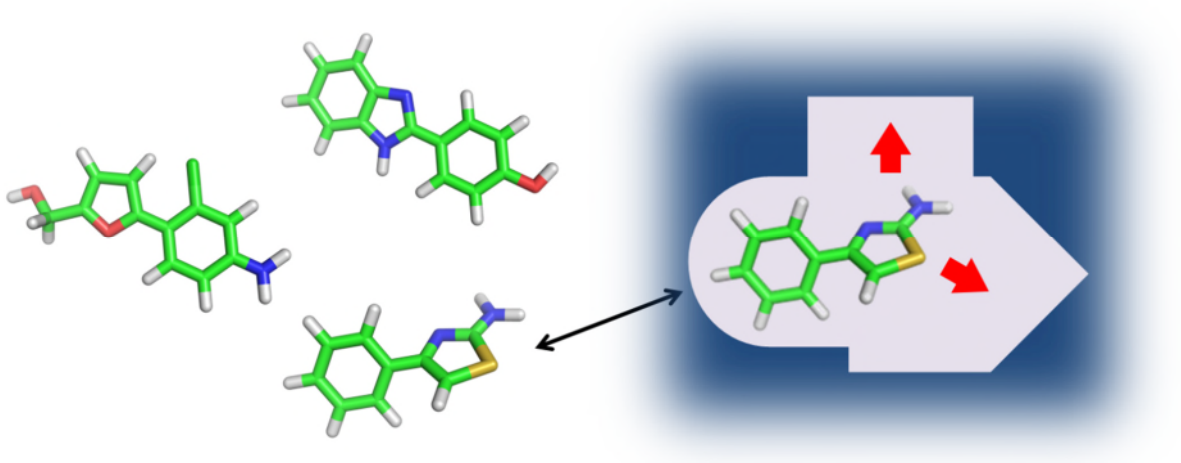


Figure 8. Fragment based ligand design.

As a result of their low molecular weight, fragment hits generally have low binding affinity to their receptors with K_D values in the high µM to mM range. Despite of their weak binding activities, fragments generally have higher ligand efficiency (LE, see equation below) compared to HTS hits as a high proportion of the atoms in a fragment hit are directly involved in the protein-binding interaction. The affinity of these hits can be subsequently improved by either fragment linking or the incorporation of extra chemical functionality.[79] During the process of fragment elaborations, the LE of hits, which is a useful parameter in the lead selection, can be closely monitored and optimised.[82, 87]

$$LE = -(RT \cdot \ln K_D) / N_{HA}$$

Where R = gas constant, T = absolute temperature, and N_{HA} = number of non-hydrogen atoms.

FBLD has been used successfully in generating promising leads for difficult targets in drug companies and academic institutions.[81, 88-90] After initial failures with the HTS approach, Abbott Laboratories performed a fragment screen against matrix metalloproteinase-3 (MMP3) and discovered that acetohydroxamate could bind to the protein simultaneously with biaryl compounds.[79] This was followed by fragment linking guided by three-dimensional structure of a ternary complex that ultimately produced a highly potent compound with nanomolar affinity. Similar to the MMP3 scenario, initial efforts at conventional HTS against Bcl-xl failed to yield productive leads.[79] However, a fragment-based screen again revealed weak-binding fragments that were then elaborated into potent lead compounds.

1.12 AMA1 as a Therapeutic Target

AMA1 represents an attractive target for developing antimalarial drugs.[24] It is unique to *Plasmodium* species and other apicomplexan parasites such as *Toxoplasma gondii*. No homologues exist in the human host, making it possible to design inhibitors that selectively target the malaria parasites. Importantly, AMA1 is expressed on the parasite surface in the human bloodstream, thus avoiding the challenges of delivering a potential drug across the numerous membranes that protect intra-cellular parasite targets. Moreover, this renders AMA1 inhibitors immune to parasite-mediated drug efflux mechanisms that are important sources of resistance to many existing malaria drugs.[91] In addition to its role in merozoite invasion, AMA1 has also been implicated in the invasion of hepatocytes.[57] Therefore, a therapeutic agent targeting AMA1 not only has the potential to treat or suppress the development of symptomatic malaria, but its activity on hepatic invasion may also prevent the *Plasmodium* parasites from establishing initial liver-stage infections.

1.13 Small-Molecule Inhibitors of AMA1

Srinivasan *et al*.[92] recently described small-molecule inhibitors of the AMA1–RON2 complex. The initial hits were identified from an AlphaScreen assay of a ~21,733 member library. Twenty molecules were able to disrupt the AMA1-RON2 interactions in the screen and three of them (NCGC00015280, NCGC00181034 and NCGC00014044) blocked merozoite invasion *in vitro* with IC_{50} values in the range 21 – 29 μM (Figure 10). Analogues of pyrrolo[2,3-*d*]pyrimidine-4-amines scaffold (NCGC00015280) were investigated further, and compounds with improved inhibitory activities were found (IC_{50} = 9.8 and 6 μM for NCGC00262654 and NCGC00262650, respectively). These pyrrolo[2,3-*d*]pyrimidine-4-

amines compounds were able to block red cell invasion by merozoites in a strain-transcending manner.

However, these compounds exhibit weak direct binding activities to AMA1, as determined by SPR, and is inconsistent with low µM inhibitory activities in the invasion assays. Besides, the compounds suffer from unattractive qualities, such poor solubility in aqueous solution.[92] Therefore, further investigations are necessary before using these compounds as lead candidates of AMA1 inhibitors.

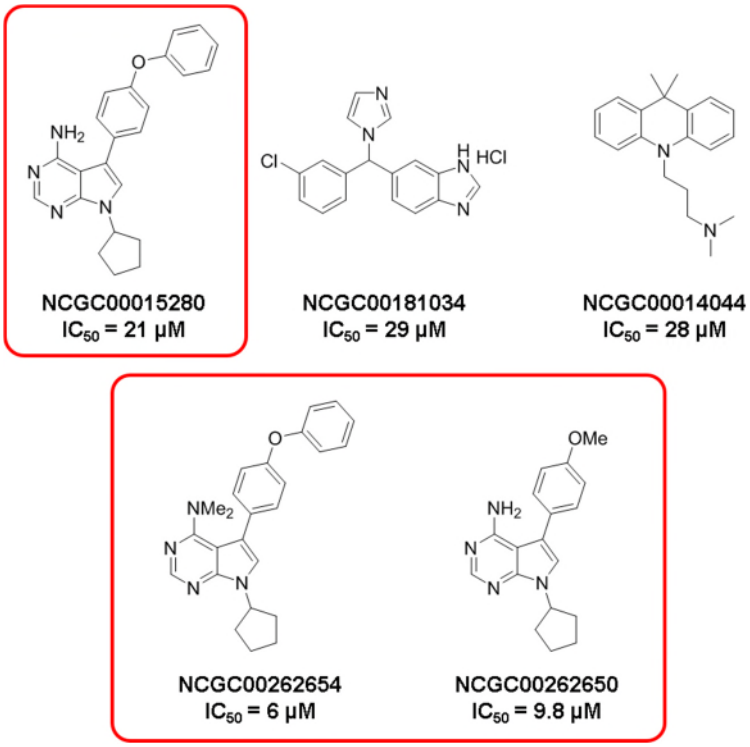


Figure 9. Putative AMA1 inhibitors identified by Srinivasan *et al.*[26] Compound NCGC00015280, NCGC00181034 and NCGC00014044 were initial hits identified from the AlphaScreen assay. NCGC00262654 and NCGC00262650 are analogues of NCGC00015280 and are shown in red brackets.

1.14 Project Scope and Aims

Our aim is to design small molecule inhibitors of AMA1 using fragment based lead design (FBLD). Compared with existing peptides or antibodies targeting AMA1, small molecule inhibitors are better drug candidates for treating malaria due to the lower production costs and the requirement for oral formulation.[93] In the fragment-based approach, the process of growing hit to lead would allow us to sample chemical functionalities that are complementary to a conserved binding pocket, important for the discovery of AMA1 inhibitors with broad strain specificity.

My project began with screening a 1140-compound library in cocktails of 6 fragments using saturation transfer difference (STD) ^{1}H NMR spectroscopy to identify AMA1 binders (Figure 10). The STD hits identified were followed up with STD and Carr-Purcell-Meiboom-Gill (CPMG) R1 competition binding assay to identify fragments that bind to the critical AMA1 hydrophobic cleft. Subsequently, the binding affinities (K_D) and ligand efficiencies (LE) of the fragment hits were characterised using surface plasmon resonance (SPR). Analogues of promising fragment scaffolds were tested with SPR to elucidate their structure-activity relationships (SARs) for further compound development. These results are described in detail in Chapter 2.

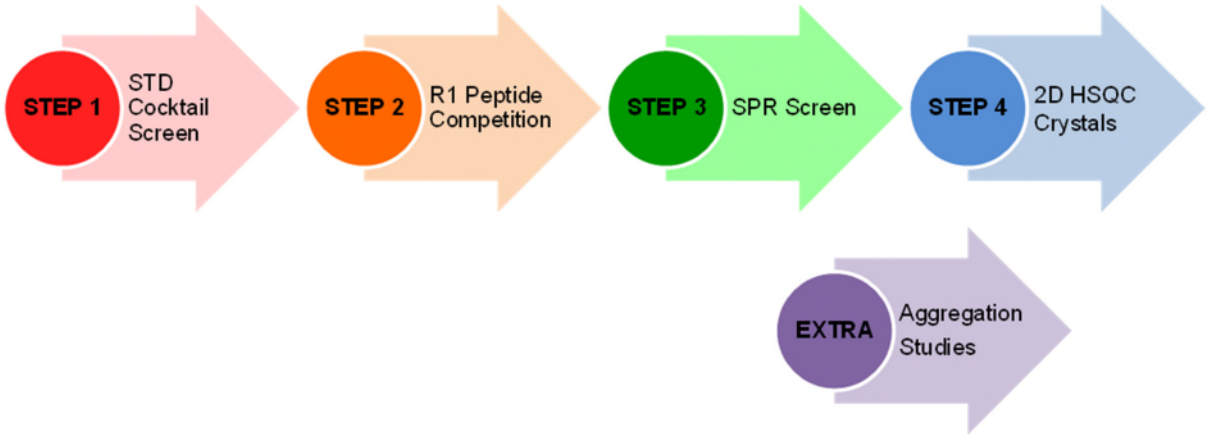


Figure 10. Scope of the thesis. The project involved the use of various biophysical techniques that include NMR, SPR and X-ray crystallography to identify efficient AMA1 binders.

After successful completion of the fragment screen, the project progressed to optimising the fragments into molecules with the desired affinity and properties. Whilst generating analogues for our fragment hits, other group[92] discovered small-molecule inhibitors of AMA1 with low micromolar affinities. This finding represents an opportunity to generate potent AMA1 inhibitors through merging or linking of our fragment hits with the reported pyrrolo[2,3-*d*]pyrimidine-4-amines compounds. However, the compounds are poorly soluble in aqueous solution as acknowledged by Srinivasan *et al.*[92] and the observed SPR sensorgrams were not consistent with low micromolar binding activities. To assess the suitability of the pyrrolo[2,3-*d*]pyrimidine-4-amines scaffolds for further development as AMA1 inhibitors, the solution behaviour of the compounds and their interactions with AMA1 were evaluated with a range of biophysical and cell-based assays (Figure 11). These findings are presented in Chapter 3.

To obtain binding poses of fragments bound to the antigenically diverse AMA1 strains, the crystal structure of *Pf*AMA1 from the 3D7 strain were reproduced at higher resolution (Figure 10). We also solved the first crystal structure of *Pf*AMA1 from the FVO strain. The FVO *Pf*AMA1 structure was compared to those of AMA1 from *P. falciparum* 3D7 and *P. vivax*. A combination of normalised B-factor analysis and molecular dynamic simulations has been used to investigate the flexibility of the domain I and DII loops and how this correlates with their roles in determining the strain specificity of human antibody responses and inhibitory peptides. Collectively, these results provide valuable insights that should contribute to the design of strain-transcending agents targeting *Plasmodium falciparum* AMA1. Currently we are working towards getting the crystal structures of the fragments bound to AMA1, which will allow more rational design of fragment analogues. The results are presented in Chapter 4.

Subsequent work focused on mapping the specific binding sites of the fragments on AMA1 using ^{1}H-^{15}N HSQC perturbation experiments (Figure 10). To perform this study, it was necessary to deuterate as well as label the AMA1 protein with ^{15}N and ^{13}C atoms. The backbone amide peaks on the HSQC spectrum were first assigned using triple resonance NMR experiments (HNCA, HN(CO)CA, HN(CA)CB and HN(COCA)CB) with transverse relaxation optimised spectroscopy (TROSY) effects. Specific ^{15}N-Lys labelled HSQC spectrum was also acquired to assist with the backbone assignments. Studies of the fragment chemical shift perturbation (CSP) on the HSQC spectrum allowed us to identify fragments that bind to only the conserved residues around the hydrophobic cleft for crystallisation trials. These results are described in Chapter 5.

Chapter 2:

Fragment Screening against AMA1

Published paper

Lim, S. S., Debono, C. O., MacRaild, C. A., Chandrashekaran, I. R., Dolezal, O., Anders, R. F., Simpson, J. S., Scanlon, M. J., Devine, S. M., Scammells, P. J., and Norton, R. S. (2013) Development of inhibitors of *Plasmodium falciparum* apical membrane antigen 1 based on fragment screening, *Aust. J. Chem. 66* (12), 1530-1536.

2.1 Declaration for Thesis Chapter 2

Declaration by candidate

In the case of Chapter 2, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Designed and performed experiments, data analysis, manuscript preparation	70

The following co-authors contributed to the work. If co-authors are students at Monash University, the extent of their contribution in percentage terms must be stated:

Name	Nature of contribution	Extent of contribution (%) for student co-authors only
Cael O. Debono	Compound synthesis, data analysis	20
Christopher A. MacRaild	Design experiments, manuscript preparation	
Indu R. Chandrashekaran	Manuscript preparation	
Olan Dolezal	Design experiments, manuscript preparation	
Robin F. Anders	Intellectual input	
Jamie S. Simpson	Intellectual input	
Martin J. Scanlon	Intellectual input, manuscript preparation	
Shane M. Devine	Design experiments, data analysis, manuscript preparation	
Peter J. Scammells	Intellectual input, manuscript preparation	
Raymond S. Norton	Intellectual input, manuscript preparation	

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the candidate's and co-authors' contributions to this work.

Candidate's Signature	[redacted]	**Date 5/10/2014**
Main Supervisor's Signature	[redacted]	**Date 7/10/2014**

RESEARCH FRONT

CSIRO PUBLISHING

Aust. J. Chem.
http://dx.doi.org/10.1071/CH13266

Full Paper

Development of Inhibitors of *Plasmodium falciparum* Apical Membrane Antigen 1 Based on Fragment Screening

San Sui Lim,[A] *Cael O. Debono,*[A] *Christopher A. MacRaild,*[A] *Indu R. Chandrashekaran,*[A] *Olan Dolezal,*[B] *Robin F. Anders,*[C] *Jamie S. Simpson,*[A] *Martin J. Scanlon,*[A,D] *Shane M. Devine,*[A] *Peter J. Scammells,*[A] *and Raymond S. Norton*[A,E]

[A]Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Vic. 3052, Australia.
[B]CSIRO Materials Science and Engineering, Parkville, Vic. 3052, Australia.
[C]Department of Biochemistry, La Trobe University, Bundoora, Vic. 3086, Australia.
[D]Centre of Excellence for Coherent X-Ray Science, Monash University, Parkville, Vic. 3052, Australia.
[E]Corresponding author. [redacted]

Apical membrane antigen 1 (AMA1) is an essential component of the moving junction complex used by *Plasmodium falciparum* to invade human red blood cells. AMA1 has a conserved hydrophobic cleft that is the site of key interactions with the rhoptry neck protein complex. Our goal is to develop small molecule inhibitors of AMA1 with broad strain specificity, which we are pursuing using a fragment-based approach. In our screening campaign, we identified fragments that bind to the hydrophobic cleft with a hit rate of 5 %. The high hit rate observed strongly suggests that a druggable pocket is present within the cleft.

Manuscript received: 22 May 2013.
Manuscript accepted: 29 July 2013.
Published online: 26 August 2013.

Introduction

Malaria is caused by parasites of the genus *Plasmodium* and is one of the most widespread infectious tropical diseases.[1] There are around 250 million clinical cases of malaria each year, resulting in almost one million deaths.[1,2] Amongst the malaria parasites, *P. falciparum* is the most deadly species, being responsible for 90 % of the total malaria-related deaths.[1]

Apical membrane antigen 1 (AMA1) is an essential component of the moving junction (MJ) used by *Plasmodium* merozoites to invade human red blood cells.[3] AMA1 has a conserved hydrophobic cleft that is the site of key interactions with the rhoptry neck (RON) protein complex that forms part of the moving junction (Fig. 1).[4–8] Peptides identified by phage display, such as R1, as well as monoclonal antibodies that target this site on AMA1, are able to inhibit red blood cell invasion, but usually in a strain-specific manner as numerous polymorphic residues are clustered at one end of the cleft (Fig. 2).[5,6,9,10]

AMA1 represents an attractive target for developing antimalarial drugs.[11] It is unique to *Plasmodium* species and other apicomplexan parasites such as *Toxoplasma gondii*. No homologues exist in the human host, facilitating the design of inhibitors that selectively target the malaria parasite. Importantly, AMA1 is expressed on the parasite surface in the human bloodstream, thus avoiding the challenges of delivering a drug across the numerous membranes that protect intra-cellular parasite targets. Moreover, this renders AMA1 inhibitors immune to parasite-mediated drug efflux mechanisms that are important sources of resistance to many existing malaria drugs.[12] Anti-AMA1 antibodies also inhibit sporozoite invasion, providing evidence that the protein is involved in the invasion of hepatocytes.[13] Therefore, a therapeutic agent targeting AMA1 has the potential not only to treat or suppress the development of symptomatic malaria, but also to prevent *Plasmodium* parasites from establishing initial liver-stage infections.

Our aim is to develop small molecule inhibitors of AMA1 using fragment-based ligand design (FBLD). Compared with existing peptides or antibodies targeting AMA1, small molecule inhibitors are preferred candidates for treating malaria because of their lower production costs and likely oral bioavailability.[14] In the fragment-based approach, the process of growing, linking, or merging hits to leads would allow us to install chemical functionalities that are complementary to a conserved binding pocket, important for the discovery of AMA1 inhibitors that have broad strain specificity. More importantly, FBLD has been used successfully in generating promising leads for difficult targets such as those involving protein–protein interactions.[15–19]

Results and Discussion

A diverse fragment library[20] was screened against 3D7 *Pf*AMA1$_{[104–442]}$ corresponding to domains I and II of the AMA1 ectodomain; these fragments were screened in cocktails of six compounds using saturation transfer difference (STD) NMR experiments (Fig. 3a–c).[21,22] Of the 1140 compounds screened, 208 fragments showed positive STD signals in the

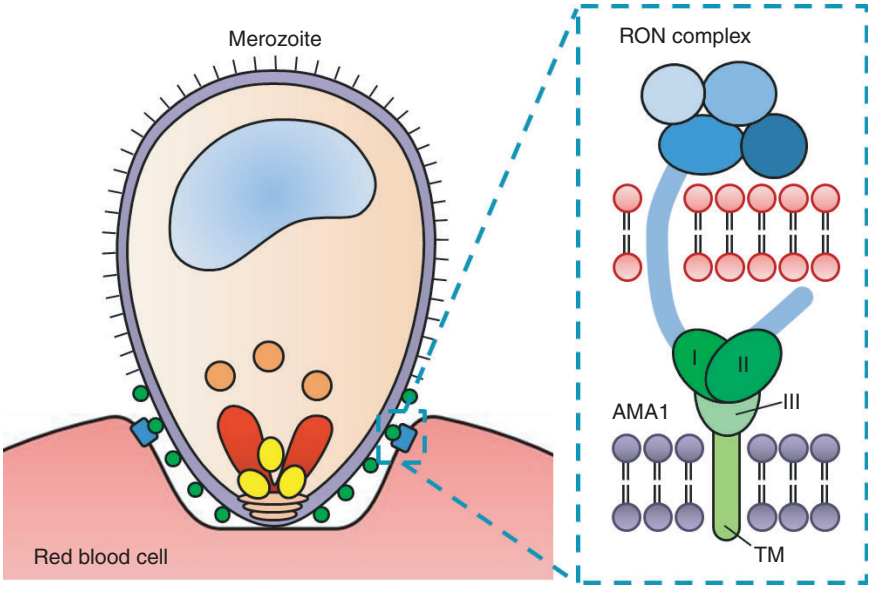


Fig. 1. Graphical representation showing apical membrane antigen 1 (AMA1) (green) interacting with the rhoptry neck (RON) protein complex (blue) to form the moving junction important for merozoite invasion of red blood cells. During invasion, the RON complex is inserted onto the red cell membrane (red) and AMA1 is located on the merozoite membrane surface (purple). AMA1 consists of three extracellular domains (I, II, and III), a transmembrane domain (TM), and a short cytoplasmic tail, with domains I and II directly involved in the binding of RON complex.

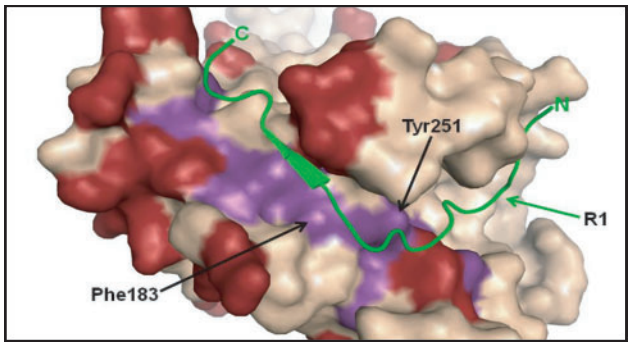


Fig. 2. Co-crystal complex of R1 and 3D7 *Pf*AMA1 DI + II. R1 (green) inhibits apical membrane antigen 1 (AMA1) (light brown) through its binding in the hydrophobic cleft (purple, PDB ID 3SRJ).[8] The R1 inhibitory effect is strain specific as the peptide interacts with several polymorphic residues (red) on AMA1. Residue numbers apply to AMA1. (Phe: phenylalanine, Tyr: tyrosine.)

presence of AMA1, corresponding to a primary hit rate of 18 %. Although STD is a rapid and sensitive method for identifying weak binders, it does not provide any information about the binding sites of these fragment hits.[22] To address this, we made use of R1, an inhibitory peptide that binds to the hydrophobic cleft of AMA1 with a binding affinity (K_D) of 100 nM.[9,23] The hits identified from the initial cocktail screen were followed up as individual compounds using R1 competition experiments to identify hits that bind to the AMA1 hydrophobic cleft.

These competition experiments were conducted using both STD experiments and transverse relaxation rate (T_2) measurements using Carr–Purcell–Meiboom–Gill (CPMG) experiments[24] in both the presence and absence of the competing R1 peptide. Higher concentrations of fragments and 3D7 *Pf*AMA1 (500 and 10 μM, respectively, compared with 300 and 5 μM, respectively, in the initial cocktail screens) were employed in an effort to increase the STD signal intensity and make any competitive effect easier to detect. Upon addition of R1 to these samples, the peptide is expected to compete with and displace any fragments bound to the hydrophobic cleft. This was observed as decreased saturation transfer from the protein to fragments and hence a reduction in STD signal intensities in the R1 STD competition experiments (Fig. 3d). In the case of CPMG, the smaller fraction of bound fragment results in reduced fragment transverse relaxation rates, and thus increased fragment signal after the CPMG period of 600 ms (Fig. 3e).

Among the 208 primary hits evaluated in the R1 competition assays, 65 showed evidence of competition by STD, whilst 98 appeared to compete with the peptide by CPMG. In total, 57 fragments were found to compete with R1 in both assays, corresponding to a 5 % hit rate overall (Fig. 4). We note that 90 compounds that were positive hits in the initial cocktail screen showed no binding activity on AMA1 when examined as individual fragments in the STD experiments. One possible explanation for this observation is that the fragments in the cocktail might have interacted with each other and AMA1 to form higher-order complexes.[25] Despite this intriguing possibility, these compounds were treated as false positives and were not tested further in the R1 competition experiments.

The R1-competing hits were not all structurally distinct, with some fragments having similar chemical structures that could be clustered into different series such as 2-aminothiazoles and 2-aryl furans, among others. Several physicochemical parameters of the hits have been examined to assess whether any properties are important for the compounds to bind to AMA1 (Table 1). From this analysis, there is an obvious difference in log *P* values between R1-competing hits and the library compounds, with the hits appearing to be more hydrophobic on average than the rest of the library compounds. This concurs with the general expectation that non-polar interactions will promote binding in a hydrophobic site. The hits also have slightly higher molecular weights, with the increase in size being attributable to increasing numbers of rings rather than additional rotatable bonds. This trend is consistent with

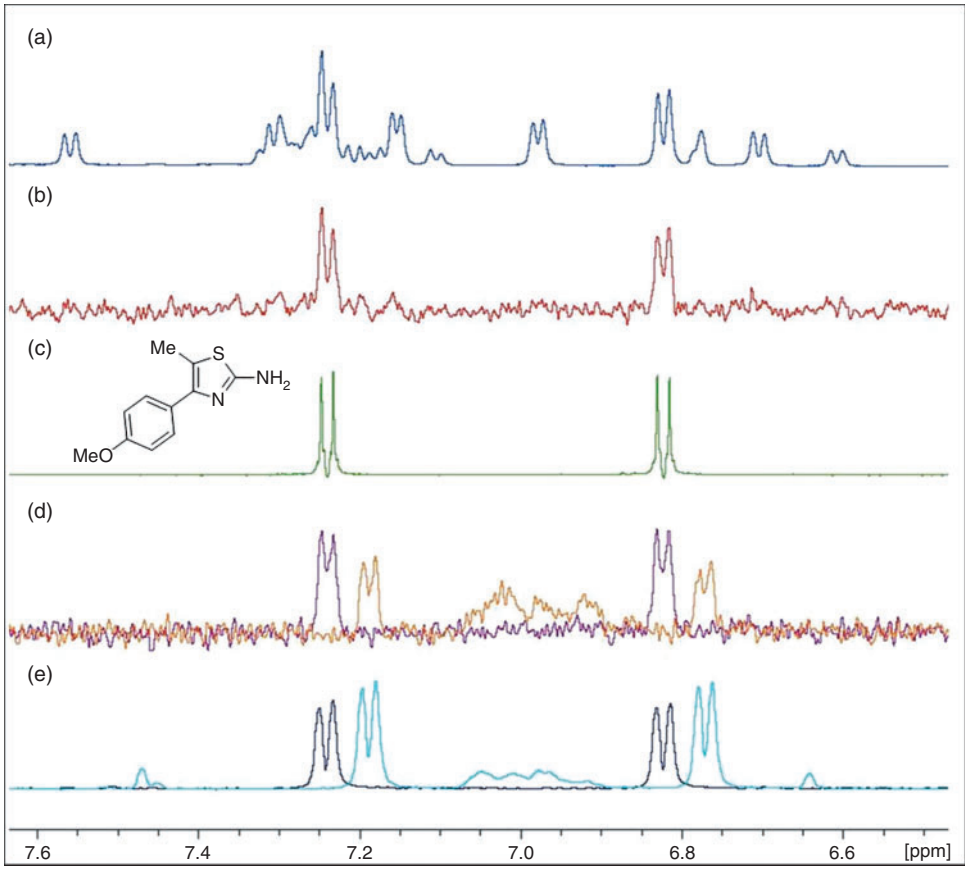


Fig. 3. Saturation transfer difference (STD) results from a cocktail of six fragments at 10°C, 20 mM phosphate buffer, pH 7.4. (a) ^{1}H NMR spectrum for cocktail of six fragments. (b) STD spectrum showing that one fragment out of the cocktail was identified as a hit, with STD signal intensity of 1.3 % (STD signal intensity is the percentage of proton signal reduction relative to the off-resonance spectrum). (c) Reference spectrum for the fragment hit. (d) STD competition NMR experiments. R1 peptide was used to identify fragments that bind to the hydrophobic cleft of 3D7 *Pf*AMA1 DI + II. STD spectra were acquired in the absence (purple) and presence (orange) of R1. (e) CPMG spectra for the fragment hit in the presence of AMA1 and in the absence (grey) and presence (cyan) of R1. In both (d) and (e), spectra in the presence of R1 are offset by −0.07 ppm for clarity. Proton signals between 6.9 and 7.1 ppm are resonances of the R1 peptide.

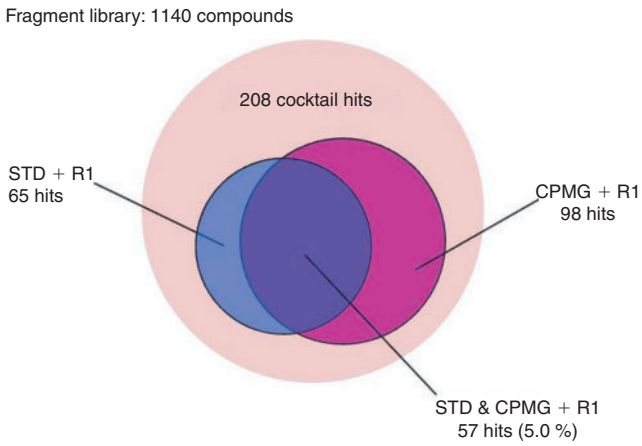


Fig. 4. Venn diagram showing the numbers of hits identified in the cocktail screen and R1 competition experiments. There were 57 hits that showed competition with the peptide in both saturation transfer difference (STD) and Carr–Purcell–Meiboom–Gill (CPMG) experiments.

Table 1. Comparison of physicochemical properties of R1-competing hits and library compounds (average values are shown for each property)

Log *P*: partition coefficient; TPSA: topological polar surface area

Properties	Hits	Fragment library
Molecular weight	221.7	210.6
Log *P*	2.16	1.39
TPSA	47.64	47.50
H bond acceptor	2.23	2.47
H bond donor	1.14	0.99
Rotatable bonds	1.98	1.97
Ring count	2.12	1.82

other fragment screening results reported for protein–protein interactions, which generally have shallow pockets on their interacting surfaces and require more extended scaffolds to make sufficient contacts with the binding sites.[26,27] There is also a slight difference in the numbers of hydrogen bond acceptors and donors (Table 1).

All R1-competing hits were characterised further using surface plasmon resonance (SPR). These compounds were screened against immobilized 3D7 *Pf*AMA1 at 50, 100, and 200 μM to estimate their binding affinities. Fragment concentrations of more than 200 μM were not tested since, at higher concentrations, many of the selected fragments displayed so-called 'SPR-undesirable-behaviours', similar to those

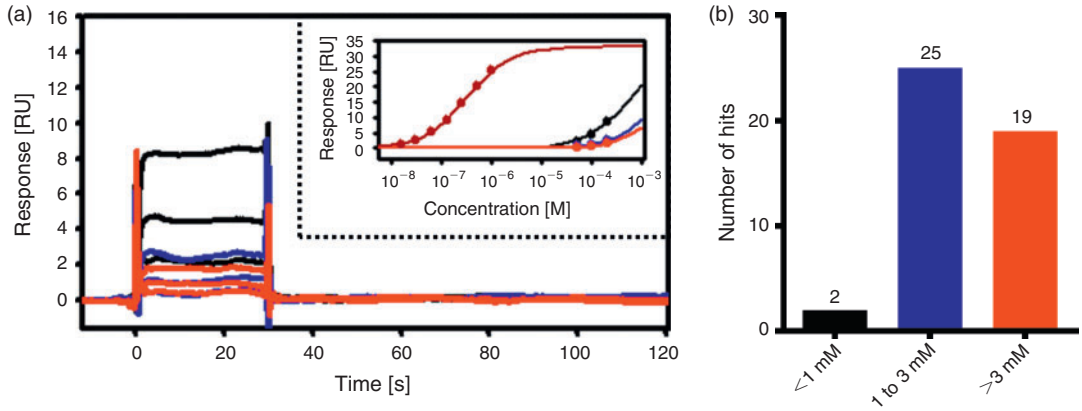


Fig. 5. (a) Surface plasmon resonance (SPR) sensorgram for R1-competing hits tested at three different concentrations (50, 100 and 200 μM). Shown are overlayed sensorgrams for fragments MIPS0000865 (black sensorgrams), MIPS0000713 (blue lines) and MIPS0000873 (red lines); the inset shows binding responses at equilibrium ($t = 20$–25 s, plotted against injected concentration) fitted to a Langmuir adsorption isotherm. R_{max} value (maximal binding response), derived by fitting the R1 peptide responses to the same isotherm (brown dots), was applied to non-saturating responses obtained with fragment hits. Using this approach, the estimated binding affinities (K_D values) for these three fragments were 0.6 mM (MIPS0000865), 2 mM (MIPS0000713) and 3 mM (MIPS0000873) (b) Distribution of the binding affinities (K_D) of R1-competing hits. A total of 46 compounds showed binding activity by SPR.

Table 2. Ligand efficiency (LE) of R1 competing hits

LE [kcal mol^{-1} heavy atom^{-1}]	Number of hits	Percentage [%]
LE <0.2	12	26.1
0.2 ≤ LE <0.3	31	67.4
LE ≥ 0.3	3	6.5

described by Giannetti et al.[28] Thus, for example, several fragments appeared to undergo concentration-dependent aggregation, which, at high (>200 μM) concentrations, resulted in non-stoichiometric binding. Of 57 NMR hits evaluated, 46 compounds showed binding in the SPR experiments. The 11 NMR hits not identified by SPR might have very weak interactions with the target protein that are beyond the detection limit of our SPR experiments. For all fragments tested, binding to AMA1 failed to reach saturation over the concentration range tested, consistent with the relatively weak affinity expected for primary fragment hits. Not unexpectedly, SPR sensorgrams (Fig. 5a) also revealed that the interaction between fragments and AMA1 displayed fast dissociation kinetics that could not be fitted to a kinetic binding model (the data collection rate in Biacore T200 is not sufficient to determine dissociation rate constants, k_d, that are greater than 0.5–1 s^{-1}). In order to rank the selected fragment hits, their binding responses were normalized using a scheme similar to that described by Giannetti et al.[28] Thus, the normalized maximal binding response (R_{max}), predicted from fitting the control compound (R1 peptide) sensorgrams, was applied to response curves obtained with fragments at non-saturating concentrations. To determine K_D values, the Langmuir adsorption isotherm was then fitted to the normalized binding data. Based on SPR experiments, all except two hits showed binding affinities weaker than 1 mM, with the strongest hits having K_D values of 600 μM (Fig. 5b). The weak binding activities observed are a consequence of the size of the fragments and are consistent with other fragment-based drug discovery projects reported in the literature.[15,29] Most of the fragments (67.4 %) bind AMA1 with ligand efficiencies (LE) of between 0.2 and 0.3 kcal mol^{-1} heavy atom^{-1} (Table 2). Three fragments that bind AMA1 have ligand efficiencies

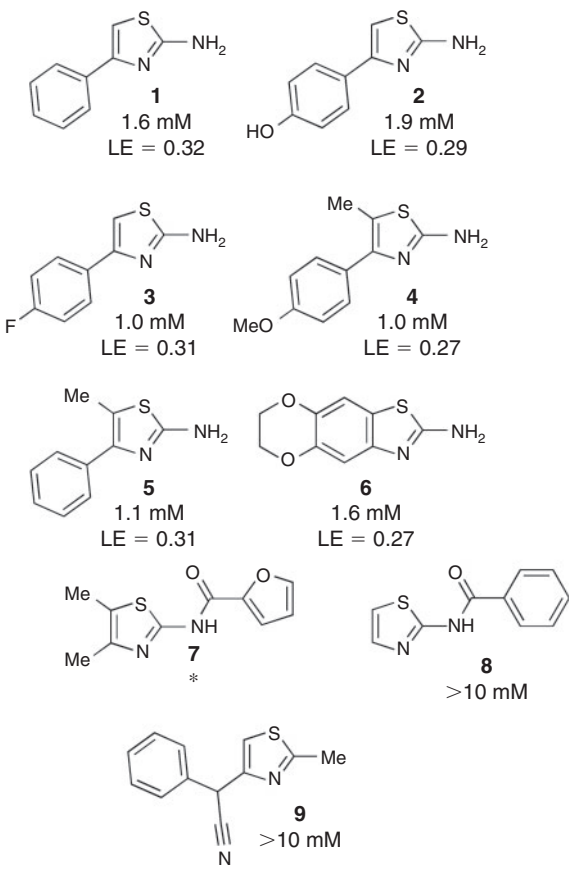


Fig. 6. Thiazole series identified from the R1 competition assay, showing binding affinities (K_D) determined by surface plasmon resonance (SPR) and ligand efficiencies (LE). *The binding affinity of **7** could not be determined as it gave an SPR response of zero at the three test concentrations.

of ≥0.3 kcal mol^{-1} heavy atom^{-1}, all of which belong to the 2-aminothiazole series. Indeed, from the total of 57 hits identified, nine have the common feature of a thiazole core, representing 15 % of the hits. These closely related R1-competing hits are shown in Fig. 6, along with their binding affinities as

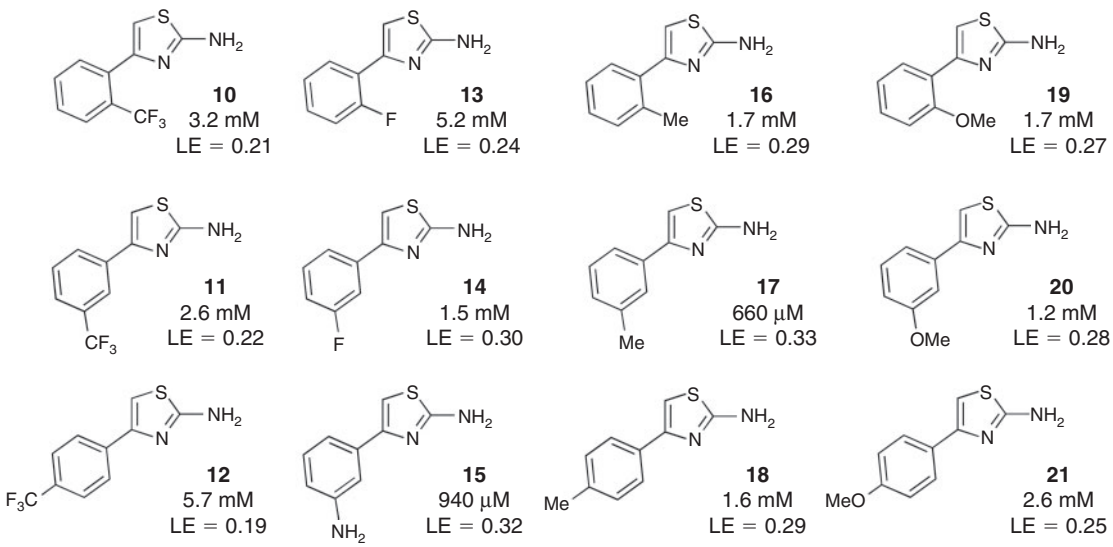


Fig. 7. Synthesised 2-aminothiazole series with binding affinities (K_D) determined by surface plasmon resonance (SPR) and ligand efficiencies (LE) indicated for each analogue.

determined by SPR. From this small set of molecules, some trends were evident. The thiazoles with the 2-amino moiety (**1**–**6**) have binding affinities ranging from 1.0 to 1.9 mM, whereas the 2-amidothiazoles **7** and **8** are virtually non-binders by SPR, with **7** showing no response. The 2-methylthiazole **9** has greatly reduced activity (>10 mM) compared with the 2-aminothiazoles. This 2-aminothiazole binding scaffold should thus serve as a good starting point for chemical elaboration.

Ligand efficiency has gained wide acceptance as a basis for initially ranking and tracking the progress of fragment elaboration.[30] The goal is to maintain the LE throughout the optimization process to give a compound with an LE > 0.3 kcal mol^{-1} heavy atom^{-1}. However, as the LE can decrease during fragment optimization, starting with a high LE hit will make it easier to grow the fragment into an inhibitor with drug-like properties.[30] We have begun to develop preliminary SAR around this scaffold by synthesizing a small library of 4-aryl substituted 2-aminothiazoles (see Supplementary Material), as shown in Fig. 7. Strongly electron-withdrawing substituents on the 4-aryl group, such as the trifluoromethyl group, are not well tolerated, with the *m*-substituent faring slightly better, with a binding affinity of 2.6 mM, over the *o*- and *p*-substituents. The *p*-fluoro (**3**) (1.0 mM) and *m*-fluoro (**14**) (1.5 mM) compounds were substantially better binders than the corresponding *o*-fluoro (**13**) compound (5.2 mM). In the case of the methoxy- and methyl-substituted analogues, *meta*-substituents were found to support the greatest activity, with binding affinities of 660 μM for **17** and 1.2 mM for **20**. The electron-donating affects of the *m*-amino group (**15**) also gave sub-millimolar affinity of 940 μM. Generally speaking, these molecules are weak binders of AMA1 and we are currently synthesizing analogues of this promising scaffold to enable evolution into higher affinity ligands.

Although various inhibitory peptides and antibodies that bind to the hydrophobic cleft of AMA1 have been described, the question of whether a druggable pocket for small molecule inhibitors exists has to date remained unclear.[5,6,8,9,31] Fragments are low-complexity molecules that allow the efficient sampling of large parts of chemical space. This allows the assessment of the protein druggability based on the fragment screening hit rate.[32] The high hit rate observed in our screening strongly suggests that at least one hot spot capable of binding small molecule ligands with high affinity is present within the AMA1 hydrophobic cleft, and our current efforts are directed towards exploiting this by enhancing the binding affinities of promising fragment hits. Efforts are also underway to identify the binding sites of candidate fragments within the hydrophobic cleft by means of high-resolution NMR spectroscopy and X-ray crystallography.

Conclusions

Our fragment screen has identified several *Pf*AMA1-binding scaffolds that bind to the AMA1 hydrophobic cleft. We observed a high hit rate of 5 %, which supports the existence of a druggable pocket within the hydrophobic cleft that is amenable to the design of small molecule inhibitors. The R1-competing hits discovered in the screening process are expected to serve as building blocks for the development of AMA1 inhibitors with broad strain specificity.

Experimental

High-Cell-Density 3D7 Pf*AMA1 DI+II Expression*

3D7 *Pf*AMA1$_{[104–442]}$ (corresponding to domains I + II) was expressed from a pPROEX HTb expression vector (Novagen) in *Escherichia coli* BL21 (DE3) using a high-cell-density methodology.[33] A single colony of the freshly transformed cells was inoculated into L-Broth containing 100 μg mL^{-1} ampicillin. The culture was grown overnight at 37°C with constant shaking at 225 rpm. After ~18 h, cells were harvested by centrifugation and resuspended in two volumes of optimized minimal medium.[33] The mixture was incubated with shaking at 37°C for 1 h before being induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and allowed to grow for 3 h before harvesting. The cell pellets were frozen at −20°C until further use.

Solubilization and Ni^{2+} *Affinity Chromatography*

Frozen cell pellets were thawed at room temperature for 30 min and resuspended with a small volume (2 mL) of cold phosphate

buffered saline (PBS). The mixture was then solubilized with buffer containing 7 M guanidine (Gdn)-HCl and 10 mM Tris (tris(hydroxymethyl)aminomethane) at pH 8.0. The total solubilization volume of mixture was 10-fold that of the weight of the cell paste, giving a final Gdn-HCl concentration of 6 M. The pH was adjusted to 8.0 using 1 M NaOH after 10 min, and incubated for another hour at room temperature with gentle shaking and intermittent 30 s sonication (10 times over 1 h). The mixture was then centrifuged at 20000*g* for 30 min. The supernatant was collected and sterile-filtered, and then 2 % (v/v) Ni^{2+}-charged chelating Sepharose (GE Healthcare) was added. After 2 h incubation, the supernatant was isolated from the resin. A further 2 % fresh Ni^{2+}-charged chelating Sepharose was added and left to stir overnight at room temperature. The mixture was transferred into a column housing and the flow through was collected. The resin was washed with 10 column volumes of guanidine buffers (6 M Gdn-HCl, 100 mM Na_2HPO_4, 10 mM Tris) of decreasing pH, starting at pH 8.0, then pH 6.3 and pH 5.9. Proteins bound to the Ni^{2+}-charged resin were eluted in 5×2 mL fractions of 6 M Gdn-HCl buffer, pH 4.5. Protein content was assessed by UV absorbance at 280 nm using a calculated extinction coefficient of 58000 M^{-1} cm^{-1}.

Refolding

The denatured protein solution was first diluted with 6 M Gdn-HCl buffer, pH 4.5, and then with 1/9 volume of 2 M Tris-HCl buffer, pH 8.0, to give a protein concentration of less than 2 mg mL^{-1}. Under constant stirring at 4°C, this solution was diluted 1 : 50 into the freshly made refold buffer (0.5 M urea, 100 mM NaCl, 20 mM Tris, pH 8.0) that had been filtered and purged with nitrogen gas for a minimum of 1 h. Reduced glutathione (GSH) was added to give a concentration of 2 mM and the mixture was stirred for 3–4 min. Oxidized glutathione (GSSG) was then added at a final concentration of 0.5 mM and the mixture stirred for another 1–2 min. The refold mixture was then incubated overnight at room temperature.

Ion-Exchange Chromatography

The refold mixture was passed through a sterile 0.2 μm filter, diluted 5-fold with 20 mM Tris, pH 8.0, and pumped through a 5 mL HiTRAP QFF (GE Healthcare) column. Refolded 3D7 *Pf*AMA1 was eluted using a linear gradient of 0 to 300 mM NaCl in 20 mM Tris, pH 8.0, over 40 column volumes. UV absorbance at 280 nm was measured to identify the desired product. Fractions containing 3D7 *Pf*AMA1 were combined and analyzed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE). Protein concentration was determined using UV absorbance at 280 nm. The product was then dialyzed against 4×100 volumes of 20 mM ammonium bicarbonate solution at 4°C over 2 days before it was lyophilized.

STD NMR Screen

The initial STD screen was conducted on an 1140-fragment library in 190 cocktails, each containing six fragments dissolved in d_6-DMSO. Samples were prepared by dissolving 3D7 *Pf*AMA1$_{[104–442]}$ and cocktails of six fragments in 20 mM phosphate buffer at pH 7.4 with 10 % D_2O and 1 % d_6-DMSO. The final concentrations of AMA1 and each fragment were 5 and 300 μM, respectively. All STD experiments were performed on a Bruker Avance III 600 MHz spectrometer at 10°C with 128 scans.[21] Saturation was achieved with a 5 s train of 50 ms Gaussian pulses at 45 dB attenuation. The irradiation frequency of the saturation pulse train was changed after every scan (on- and off-resonance frequencies were −480 and −20000 Hz, respectively). The fragment screen was automated using IconNMR with temperature check (precision ± 0.1 K) and 3 min equilibration time between samples.

STD and CPMG R1 Competition Experiments

Samples were prepared for STD and CPMG R1 competition experiments by dissolving 3D7 *Pf*AMA1$_{[104–442]}$ and individual fragments in 20 mM phosphate buffer at pH 7.4 containing 10 % D_2O and 1 % d_6-DMSO. The final concentrations of AMA1 and each fragment were 10 and 500 μM, respectively. All NMR experiments were performed on a Bruker Avance III 600 MHz spectrometer at 10°C. In both STD and CPMG experiments, reference spectra for R1-free samples were first acquired before adding R1 from a 5 mM stock solution to give a final peptide concentration of 100 μM. In the STD experiments, the parameters used to acquire the R1-containing and R1-free spectra were the same as those from the initial cocktail STD screen. For CPMG experiments, spin-locks of 0 and 0.6 s were applied to both samples with or without R1 peptide. A total of 32 scans were acquired for each sample. Small variations in signal intensities between different experiments were normalized manually using the spectra acquired at 0 s spin-lock.

SPR Screen

A Biacore T200 biosensor was employed to estimate binding affinities of R1-competing hits. 3D7 *Pf*AMA1$_{104–442]}$ was immobilized on the CM5 sensor chip using an amine-coupling methodology similar to that described by Harris et al.[31] Approximately 10000 RU of protein was coupled in a single flow cell (1000 RU = 1 ng of protein per mm^2). A reference flow cell on the same chip was prepared by subjecting it to the identical amine coupling procedure with no AMA1 protein being injected. All SPR binding experiments were conducted at 25°C with HBS-EP+ (50 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES), 150 mM NaCl, 3.4 mM ethylenediaminetetraacetic acid (EDTA), 0.05 % Tween 20, 5 % DMSO, pH 7.4) as the instrument running buffer at a constant flow rate of 60 μL min^{-1}. Immobilized AMA1 was equilibrated with 30 blank buffer injections of 30 s contact time before the start of the binding experiments. Solvent correction curves[34] were obtained from a series of injections of running buffer containing 4.55 to 5.95 % (v/v) DMSO. R1 peptide was injected over immobilized AMA1 at concentrations of 0, 15.6, 31.3, 62.5, 125, 250, 500, and 1000 nM with 2 min contact time. Fragment samples and elaborated molecules were screened at 50, 100, and 200 μM with 30 s contact times. All binding data were processed and analyzed using Biacore T200 evaluation software version 1.0. The binding affinities of fragment hits were estimated based on the response (RU) from three different concentrations with a fixed fragment R_{max} (maximal binding capacity of AMA1 protein surface). The R_{max} value for the R1 peptide was determined experimentally by fitting the dose–response curves to a 1 : 1 steady-state affinity model. R_{max} values for each fragment were adjusted according to the following normalization formula:

$$R_{max[fragment]} = R_{max[R1\ peptide]} \times MW_{[fragment]}/MW_{[R1\ peptide]}$$
$$(MW = \text{molecular weight}).$$

Supplementary Material

The experimental techniques required to synthesize compounds **10**–**21**, the ^{1}H NMR, ^{13}C NMR (and where appropriate, ^{19}F NMR) spectra, and HRMS spectra associated with these compounds are available on the Journal's website.

Acknowledgements

The authors thank Martin L. Williams and Mark D. Mulcair for their assistance in this project. This work was supported in part by an Australian National Health and Medical Research Council project grant (1025150). R.S.N. acknowledges fellowship support from the NHMRC.

References

[1] World Health Organization (WHO), *World Malaria Report* **2012**.

[2] N. J. White, A. M. Dondorp, A. Faiz, S. Mishra, T. T. Hien, *Lancet* **2012**, *380*, 559. doi:10.1016/S0140-6736(12)61321-X

[3] T. Triglia, J. Healer, S. R. Caruana, A. N. Hodder, R. F. Anders, B. S. Crabb, A. F. Cowman, *Mol. Microbiol.* **2000**, *38*, 706. doi:10.1046/J.1365-2958.2000.02175.X

[4] T. Bai, M. Becker, A. Gupta, P. Strike, V. J. Murphy, R. F. Anders, A. H. Batchelor, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12736. doi:10.1073/PNAS.0501808102

[5] A. M. Coley, A. Gupta, V. J. Murphy, T. Bai, H. Kim, R. F. Anders, M. Foley, A. H. Batchelor, *PLoS Pathog.* **2007**, *3*, e138. doi:10.1371/JOURNAL.PPAT.0030138

[6] K. A. Henderson, V. A. Streltsov, A. M. Coley, O. Dolezal, P. J. Hudson, A. H. Batchelor, A. Gupta, T. Bai, V. J. Murphy, R. F. Anders, M. Foley, S. D. Nuttall, *Structure* **2007**, *15*, 1452. doi:10.1016/J.STR.2007.09.011

[7] M. L. Tonkin, M. Roques, M. H. Lamarque, M. Pugnière, D. Douguet, J. Crawford, M. Lebrun, M. J. Boulanger, *Science* **2011**, *333*, 463. doi:10.1126/SCIENCE.1204988

[8] B. Vulliez-Le Normand, M. L. Tonkin, M. H. Lamarque, S. Langer, S. Hoos, M. Roques, F. A. Saul, B. W. Faber, G. A. Bentley, M. J. Boulanger, M. Lebrun, *PLoS Pathog.* **2012**, *8*, e1002755. doi:10.1371/JOURNAL.PPAT.1002755

[9] K. S. Harris, J. L. Casey, A. M. Coley, R. Masciantonio, J. K. Sabo, D. W. Keizer, E. F. Lee, A. McMahon, R. S. Norton, R. F. Anders, M. Foley, *Infect. Immun.* **2005**, *73*, 6981. doi:10.1128/IAI.73.10.6981-6989.2005

[10] E. F. Lee, S. Yao, J. K. Sabo, W. D. Fairlie, R. A. Stevenson, K. S. Harris, R. F. Anders, M. Foley, R. S. Norton, *Biopolymers* **2011**, *95*, 354. doi:10.1002/BIP.21582

[11] C. A. MacRaild, R. F. Anders, M. Foley, R. S. Norton, *Curr. Top. Med. Chem.* **2011**, *11*, 2039. doi:10.2174/156802611796575885

[12] C. P. Sanchez, A. Dave, W. D. Stein, M. Lanzer, *Int. J. Parasitol.* **2010**, *40*, 1109. doi:10.1016/J.IJPARA.2010.04.001

[13] O. Silvie, J. F. Franetich, S. Charrin, M. S. Mueller, A. Siau, M. Bodescot, E. Rubinstein, L. Hannoun, Y. Charoenvit, C. H. Kocken, A. W. Thomas, G. J. van Gemert, R. W. Sauerwein, M. J. Blackman, R. F. Anders, G. Pluschke, D. Mazier, *J. Biol. Chem.* **2004**, *279*, 9490. doi:10.1074/JBC.M311331200

[14] S. Kar, S. Kar, *Nat. Rev. Drug Discov.* **2010**, *9*, 511. doi:10.1038/NRD3207

[15] M. Congreve, G. Chessari, D. Tisi, A. J. Woodhead, *J. Med. Chem.* **2008**, *51*, 3661. doi:10.1021/JM8000373

[16] D. A. Erlanson, R. S. McDowell, T. O'Brien, *J. Med. Chem.* **2004**, *47*, 3463. doi:10.1021/JM040031V

[17] D. C. Rees, M. Congreve, C. W. Murray, R. Carr, *Nat. Rev. Drug Discov.* **2004**, *3*, 660. doi:10.1038/NRD1467

[18] D. E. Scott, A. G. Coyne, S. A. Hudson, C. Abell, *Biochemistry* **2012**, *51*, 4990. doi:10.1021/BI3005126

[19] J. A. Wells, C. L. McClendon, *Nature* **2007**, *450*, 1001. doi:10.1038/NATURE06526

[20] C. J. Morton, B. Doak, J. S. Simpson, M. J. Scanlon, *Aust. J. Chem.* **2013**, in press.

[21] M. Mayer, B. Meyer, *Angew. Chem. Int. Ed.* **1999**, *38*, 1784. doi:10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q

[22] B. Meyer, T. Peters, *Angew. Chem. Int. Ed.* **2003**, *42*, 864. doi:10.1002/ANIE.200390233

[23] D. Richard, C. A. MacRaild, D. T. Riglar, J. Chan, M. Foley, J. Baum, S. A. Ralph, R. S. Norton, A. F. Cowman, *J. Biol. Chem.* **2010**, *285*, 14815. doi:10.1074/JBC.M109.080770

[24] P. J. Hajduk, E. T. Olejniczak, S. W. Fesik, *J. Am. Chem. Soc.* **1997**, *119*, 12257. doi:10.1021/JA9715962

[25] B. J. Davis, D. A. Erlanson, *Bioorg. Med. Chem. Lett.* **2013**, *23*, 2844. doi:10.1016/J.BMCL.2013.03.028

[26] I. J. Chen, R. Hubbard, *J. Comput. Aided Mol. Des.* **2009**, *23*, 603. doi:10.1007/S10822-009-9280-5

[27] S. Surade, T. L. Blundell, *Chem. Biol.* **2012**, *19*, 42. doi:10.1016/J.CHEMBIOL.2011.12.013

[28] A. M. Giannetti, B. D. Koch, M. F. Browner, *J. Med. Chem.* **2008**, *51*, 574. doi:10.1021/JM700952V

[29] P. Hajduk, J. Greer, *Nat. Rev. Drug Discov.* **2007**, *6*, 211. doi:10.1038/NRD2220

[30] S. Schultes, C. de Graaf, E. E. J. Haaksma, I. J. P. de Esch, R. Leurs, O. Krämer, *Drug Discov. Today. Technol.* **2010**, *7*, e157. doi:10.1016/J.DDTEC.2010.11.003

[31] K. S. Harris, J. L. Casey, A. M. Coley, J. A. Karas, J. K. Sabo, Y. Y. Tan, O. Dolezal, R. S. Norton, A. B. Hughes, D. Scanlon, M. Foley, *J. Biol. Chem.* **2009**, *284*, 9361. doi:10.1074/JBC.M808762200

[32] P. J. Hajduk, J. R. Huth, S. W. Fesik, *J. Med. Chem.* **2005**, *48*, 2518. doi:10.1021/JM049131R

[33] A. Sivashanmugam, V. Murray, C. Cui, Y. Zhang, J. Wang, Q. Li, *Protein Sci.* **2009**, *18*, 936. doi:10.1002/PRO.102

[34] G. A. Papalia, S. Leavitt, M. A. Bynum, P. S. Katsamba, R. Wilton, H. Qiu, M. Steukers, S. Wang, L. Bindu, S. Phogat, A. M. Giannetti, T. E. Ryan, V. A. Pudlak, K. Matusiewicz, K. M. Michelson, A. Nowakowski, A. Pham-Baginski, J. Brooks, B. C. Tieman, B. D. Bruce, M. Vaughn, M. Baksh, Y. H. Cho, M. D. Wit, A. Smets, J. Vandersmissen, L. Michiels, D. G. Myszka, *Anal. Biochem.* **2006**, *359*, 94. doi:10.1016/J.AB.2006.08.021

10.1071/CH13266_AC

Australian Journal of Chemistry 2013, 66(12), 1530-1536

SUPPLEMENTARY MATERIAL

Development of inhibitors of *Plasmodium falciparum* apical membrane antigen 1 based on fragment screening

San Sui Lim[A], *Cael O. Debono*[A], *Christopher A. MacRaild*[A], *Indu R. Chandrashekaran*[A], *Olan Dolezal*[B], *Robin F. Anders*[C], *Jamie S. Simpson*[A], *Martin J. Scanlon*[A,D], *Shane M. Devine*[A], *Peter J. Scammells*[A], *Raymond S. Norton*[A,E]

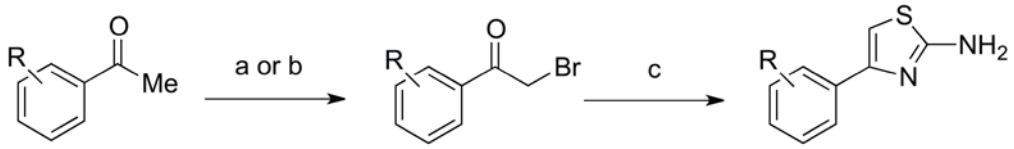
[A] Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville 3052, Australia.

[B] CSIRO Materials Science and Engineering, Parkville 3052, Australia.

[C] Department of Biochemistry, La Trobe University, Bundoora 3086, Australia.

[D] Centre of Excellence for Coherent X-ray Science, Monash University, Parkville 3052, Australia

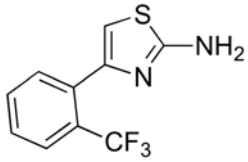
[E] Corresponding author. Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, 381 Royal Parade, Parkville 3052, Australia. [redacted]



Synthesis of the 2-aminothiazoles. (a) For **10**, **11**, **12**, **13**, **14**, **16**, **17**, **18**: Br_2, MeCN, 100 ^oC, 2 h; (b) For **15**, **19**, **20**, **21**: Br_2, $CHCl_3$, 50 ^oC, 2 h; (c) thiourea, MeCN, 100 ^oC, 2 h.

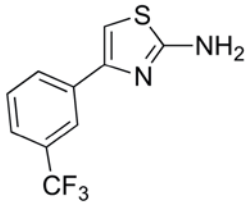
General methods: Melting points were determined on an Electrothermal melting point apparatus and are uncorrected. All NMR spectra were recorded on a Bruker Avance III 400-MHz Ultrashield Plus spectrometer and ^1H, ^{19}F and ^{13}C NMR spectra were recorded at 400, 376 and 101 MHz, respectively. Thin-layer chromatography was conducted on 0.2 mm plates using Merck silica gel 60 F254. Column chromatography was achieved using Merck silica gel 60 (article size 0.063–0.200μm, 70–230 mesh). High resolution mass spectra were obtained on a Waters 2795 Alliance Separations Module. LCMS were routinely run to verify reaction outcome using an Agilent 6100 Series Single Quad couple to an Agilent 1200 Series HPLC. All compounds were of > 95% purity.

4-(2-(Trifluoromethyl)phenyl)thiazol-2-amine (10)



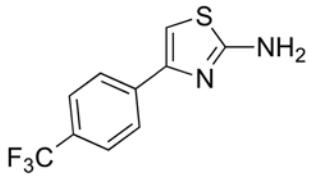
2'-(Trifluoromethyl)acetophenone (200 mg, 1.06 mmol, 159 µL) was added to MeCN (3 mL). Br_2 (93 mg, 1.17 mmol, 30 µL) was added dropwise to the mixture at room temperature. The mixture was then heated for 2 h at 100 °C. At this time, thiourea (89 mg, 1.17 mmol) was added to the cooled reaction mixture, which was then heated for a further 2 h at 100 °C. The reaction was cooled on ice and the precipitate that formed was filtered. This solid was dissolved in EtOAc (20 mL) and washed twice with Na_2CO_3 (1M) (20 mL). The organic fraction was dried over $MgSO_4$ and the filtrate reduced *in vacuo* to yield a white compound. The compound was purified using column chromatography using a step gradient from Hexane: EtOAc (1:1) to EtOAc 100 % to yield a crystalline white solid (48 mg, 20 %). m.p: 116-118 ^oC. Analytical HPLC indicates greater than 97 % purity at 254 nm; ^1H NMR (400 MHz; DMSO): δ 7.76 (d, *J* = 7.9 Hz, 1H, ArH), 7.70-7.64 (m, 1H, ArH), 7.63-7.58 (m, 1H, ArH), 7.55 (dd, *J* = 11.2, 3.9 Hz, 1H, ArH), 7.02 (br s, 2H, NH_2), 6.58 (s, 1H, ArH); ^{19}F NMR (376 MHz; DMSO) δ -56.3 (s, CF_3); ^{13}C NMR (101 MHz, DMSO): δ 167.7 (C), 147.5 (C), 135.2 (q, *J* = 3.0 Hz, C), 132.1 (CH), 131.8 (CH), 128.1 (CH), 126.6 (q, *J* = 32.0 Hz, *C*CF_3), 126.2 (q, *J* = 5.5 Hz, CH), 124.2 (q, *J* = 273.6 Hz, C*C*F_3), 104.9 (q, *J* = 3.0 Hz, CH). *m/z* (ESI-HRMS) Calc $[M+H]^+$ = 245.0355; Observed $[M+H]^+$ = 245.0353.

4-(3-(Trifluoromethyl)phenyl)thiazol-2-amine (11)



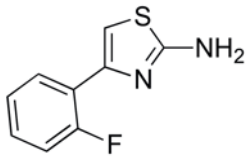
Same procedure as for **10**, from 3'-(trifluoromethyl)acetophenone gave a golden yellow solid (59 mg, 23 %). m.p: 78-80 ^{o}C. Analytical HPLC indicates greater than 95 % purity at 254 nm; ^{1}H NMR (400 MHz; MeOD): δ 8.05 (s, 1H, ArH), 7.95 (d, *J* = 7.1 Hz, 1H, ArH), 7.54-7.48 (m, 2H, ArH), 6.94 (s, 1H, ArH); ^{19}F NMR (376 MHz; MeOD) δ -64.1 (s, CF_3); ^{13}C NMR (101 MHz, MeOD): δ 171.4 (d, *J* = 6.8 Hz, C), 149.6 (d, *J* = 13.1 Hz, C), 136.8 (d, *J* = 2.2 Hz, C), 131.9 (q, *J* = 32.1 Hz, *C*CF_3), 130.3 (CH), 130.2 (d, *J* = 1.0 Hz, CH), 125.8 (q, *J* = 271.1 Hz, C*C*F_3), 124.9 (q, *J* = 3.9 Hz, CH), 123.5 (q, *J* = 3.9 Hz, CH), 104.5 (CH). *m/z* (ESI-HRMS) Calc $[M+H]^+$ = 245.0355; Observed $[M+H]^+$ = 245.0355.

4-(4-(Trifluoromethyl)phenyl)thiazol-2-amine (12)



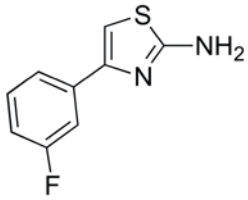
Same procedure as for **10**, from 4'-(trifluoromethyl)acetophenone gave a light orange solid (39 mg, 15 %). m.p: 158-160 ^{o}C. Analytical HPLC indicates greater than 97 % purity at 254 nm; ^{1}H NMR (400 MHz; DMSO): 8.00 (d, *J* = 8.0 Hz, 2H, ArH), 7.72 (d, *J* = 8.1 Hz, 2H, ArH), 7.25 (s, 1H, ArH), 7.21 (br s, 2H, NH_2); ^{19}F NMR (376 MHz; DMSO) δ -60.8 (s, CF_3); ^{13}C NMR (101 MHz, DMSO): δ 168.5 (C), 148.1 (C), 138.4 (C), 127.2 (q, *J* = 31.7 Hz, *C*CF_3), 126.0 (2 × CH), 125.5 (q, *J* = 3.7 Hz, 2 × CH), 124.4 (q, *J* = 272.0 Hz, C*C*F_3), 104.4 (CH). *m/z* (ESI-HRMS) Calc $[M+H]^+$ = 245.0355; Observed $[M+H]^+$ = 245.0353.

4-(2-Fluorophenyl)thiazol-2-amine (13)



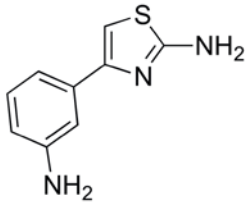
Same procedure as for **10**, from 2'-fluoroacetophenone gave a pink solid (31 mg, 11 %). m.p: 76-80 ^{o}C. Analytical HPLC indicates greater than 97 % purity at 254 nm; R_f: 0.3 (DCM: MeOH: NH_4 OH; 96: 2: 2); ^{1}H NMR (400 MHz; $CDCl_3$): δ 8.01 (td, *J* = 7.8, 1.9 Hz, 1H, ArH), 7.25 (dtd, *J* = 7.1, 5.4, 1.9 Hz, 1H, ArH), 7.18 (td, *J* = 7.5, 1.4 Hz, 1H, ArH), 7.10 (ddd, *J* = 11.8, 8.0, 1.3 Hz, 1H, ArH), 7.02 (d, *J* = 2.2 Hz, 1H, ArH), 5.17 (br s, 2H, NH_2); ^{19}F NMR (376 MHz; $CDCl_3$) δ -114.2 (s, F); ^{13}C NMR (101 MHz; $CDCl_3$): 166.6 (C), 160.3 (d, *J* = 250.0 Hz, CF), 144.8 (d, *J* = 3.5 Hz, C), 129.8 (d, *J* = 3.3 Hz, CH), 128.9 (d, *J* = 8.6 Hz, CH), 124.4 (d, *J* = 3.5 Hz, CH), 122.4 (d, *J* = 11.3 Hz, C), 116.0 (d, *J* = 22.6 Hz, CH), 108.0 (d, *J* = 15.0 Hz, CH). *m/z* (ESI-HRMS) Calc $[M+H]^+$ = 195.0387; Observed $[M+H]^+$ = 195.0390.

4-(3-Fluorophenyl)thiazol-2-amine (14)



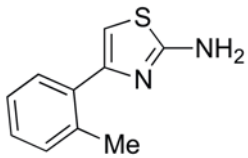
Same procedure as for **10**, from 3'-fluoroacetophenone gave a yellow powder (44 mg, 16 %) m.p: 75-78 ^{o}C. Analytical HPLC indicates greater than 97 % purity at 254 nm; R_f: 0.2, (DCM; 100); ^{1}H NMR (400 MHz; DMSO): δ 7.66-7.61 (m, 1H, ArH), 7.57 (ddd, *J* = 10.9, 2.6, 1.5 Hz, 1H, ArH), 7.39 (td, *J* = 8.0, 6.2 Hz, 1H, ArH), 7.14 (s, 1H, ArH), 7.13-7.02 (m, 3H, ArH, NH_2); ^{19}F NMR (376 MHz; DMSO) δ -113.4 (s, F); ^{13}C NMR (101 MHz, DMSO): δ 168.3 (C), 162.5 (d, *J* = 242.1 Hz, CF), 148.5 (d, *J* = 2.8 Hz, C), 137.3 (d, *J* = 8.2 Hz, C), 130.4 (d, *J* = 8.5 Hz, CH), 121.5 (d, *J* = 2.6 Hz, CH), 113.8 (d, *J* = 21.2 Hz, CH), 112.0 (d, *J* = 22.8 Hz, CH), 103.0 (CH). *m/z* (ESI-HRMS) Calc $[M+H]^+$ = 195.0387; Observed $[M+H]^+$ = 195.0389.

4-(3-Aminophenyl)thiazol-2-amine (15)



3'-Aminoacetophenone (500 mg, 3.70 mmol) was dissolved in $CHCl_3$ (3 mL). Br_2 (698 mg, 3.88 mmol, 225 µL) was added dropwise over 10 min whilst stirring at 0 ^{o}C. The mixture was then heated at 50 ^{o}C for 2 h, at which time the reaction mixture was cooled and quenched using Na_2CO_3 (1M). The biphasic mixture was added to $CHCl_3$ (20 mL) and washed twice with Na_2CO_3 (1M) (20 mL). The organic fraction was dried over $MgSO_4$ and the filtrate reduced *in vacuo* to yield a cream-coloured powder. The crude powder was dissolved in MeCN (4 mL), thiourea (187 mg, 2.45 mmol) was added and the mixture heated at 100 ^{o}C for 2 h. The reaction was cooled on ice and the precipitate that formed was filtered. This solid was dissolved in EtOAc (20 mL) and washed twice with Na_2CO_3 (1M) (20 mL). The organic fraction was dried over $MgSO_4$ and the filtrate reduced *in vacuo* to yield a white powder. The compound was purified using column chromatography using a step gradient from Hexane: EtOAc (1:1) to EtOAc 100 % to yield a light brown solid (36 mg, 5 %). m.p: 160-163 ^{o}C; Analytical HPLC indicates greater than 97 % purity at 254 nm; ^{1}H NMR (400 MHz; DMSO): δ 7.03-6.90 (m, 5H, 3 × ArH, NH_2), 6.77 (s, 1H, ArH), 6.45 (ddd, J = 7.7, 2.3, 1.2 Hz, 1H, ArH), 5.04 (br s, 2H, NH_2); ^{13}C NMR (101 MHz, DMSO): δ 167.8 (C), 150.7 (C), 148.6 (C), 135.5 (C), 128.8 (CH), 113.5 (CH), 113.0 (CH), 111.5 (CH), 100.5 (CH). *m/z* (ESI-HRMS) Calc $[M+H]^+$ = 192.0590; Observed $[M+H]^+$ = 192.0589.

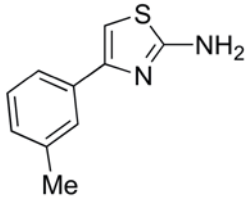
4-(o-Tolyl)thiazol-2-amine (16)



Same procedure as for **10**, from 2'-methylacetophenone gave a pink solid (84.5 mg, 30 %). m.p: 73-75 ^{o}C. Analytical HPLC indicates greater than 97 % purity at 254 nm; ^{1}H NMR (400 MHz; $CDCl_3$): δ 7.52 (ddd, J = 5.1, 3.7, 1.9 Hz, 1H, ArH), 7.39-7.11 (m, 3H, 3 × ArH), 6.45 (s, 1H, ArH), 5.28 (br s, 2H, NH_2), 2.44 (s, 3H, CH_3); ^{13}C NMR (101 MHz, $CDCl_3$): δ 166.7 (C), 151.4 (C), 136.2 (C), 135.1 (C),

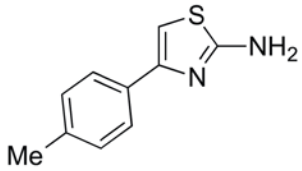
130.8 (CH), 129.7 (CH), 127.9 (CH), 125.9 (CH), 105.8 (CH), 21.1 (CH_3). *m/z* (ESI-HRMS) Calc $[M+H]^+$ = 191.0637; Observed $[M+H]^+$ = 191.0632.

4-(*m*-Tolyl)thiazol-2-amine (17)



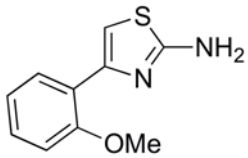
Same procedure as for **10**, from 3'-methylacetophenone gave a bright yellow solid (63.3 mg, 22 %). m.p: 56-58 ^oC. Analytical HPLC indicates greater than 97 % purity at 254 nm; ^1H NMR (400 MHz; DMSO): δ 7.62 (s, 1H, ArH), 7.57 (d, *J* = 7.8 Hz, 1H, ArH), 7.23 (t, *J* = 7.6 Hz, 1H, ArH), 7.06 (dd, *J* = 8.1, 0.6 Hz, 1H, ArH), 7.02 (br s, 2H, NH_2), 6.96 (s, 1H, ArH), 2.32 (s, 3H, CH_3); ^{13}C NMR (101 MHz, DMSO): δ 168.1 (C), 149.9 (C), 137.4 (C), 134.8 (C), 128.3 (CH), 127.8 (CH), 126.2 (CH), 122.7 (CH), 101.3 (CH), 21.1 (CH_3). *m/z* (ESI-HRMS) Calc $[M+H]^+$ = 191.0637; Observed $[M+H]^+$ = 191.0637.

4-(*p*-Tolyl)thiazol-2-amine (18)



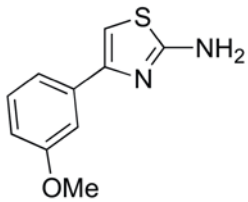
Same procedure as for **10**, from 4'-methylacetophenone gave a light yellow powder (40 mg, 14 %). m.p: 100-102 ^oC. Analytical HPLC indicates greater than 97 % purity at 254 nm; ^1H NMR (400 MHz, $CDCl_3$): δ 7.67 (d, *J* = 8.2 Hz, 2H, 2 × ArH), 7.18 (d, *J* = 7.9 Hz, 2H**,** 2 × ArH), 6.67 (s, 1H, ArH), 5.03 (br s, 2H, NH_2), 2.36 (s, 3H, CH_3); ^{13}C NMR (101 MHz, $CDCl_3$): 167.2 (C), 151.5 (C), 137.7 (C), 132.1 (C), 129.4 (2 × CH), 126.0 (2 × CH), 102.2 (CH), 21.4 (CH_3). *m/z* (ESI-HRMS) Calc $[M+H]^+$ = 191.0637; Observed $[M+H]^+$ = 191.0639.

4-(2-Methoxyphenyl)thiazol-2-amine (19)



Same procedure as for **15**, from 2'-methoxyacetophenone gave a white powder (262 mg, 38%). m.p: 83-85^oC. Analytical HPLC indicates greater than 98 % purity at 254 nm; ^1H NMR (400 MHz; DMSO): δ 8.02 (dd, *J* = 7.7, 1.8 Hz, 1H, ArH), 7.23 (ddd, *J* = 8.3, 7.3, 1.8 Hz, 1H, ArH), 7.11 (s, 1H, ArH), 7.05 (dd, *J* = 8.3, 0.9 Hz, 1H, ArH), 7.01-6.90 (m, 3H, ArH, NH_2), 3.87 (s, 3H, OCH_3); ^{13}C NMR (101 MHz, DMSO): δ 166.3 (C), 156.5 (C), 145.8 (C), 129.3 (CH), 128.0 (CH), 123.1 (C), 120.3 (CH), 111.4 (CH), 105.7 (CH), 55.3 (OCH_3). *m/z* (ESI-HRMS) Calc $[M+H]^+$ = 207.0587 ; Observed $[M+H]^+$ = 207.0593.

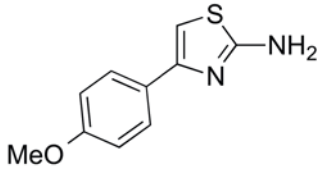
4-(3-Methoxyphenyl)thiazol-2-amine (20)



Same procedure as for **15**, from 3'-methoxyacetophenone gave a yellow solid (95 mg, 14 %). m.p: 98-100 ^oC. Analytical HPLC indicates greater than 97 % purity at 254 nm; ^1H NMR (400 MHz; DMSO): δ 7.38-7.35 (m, 2H, ArH), 7.26 (t, *J* = 7.9 Hz, 1H, ArH), 7.05 (br s, 2H, NH_2), 7.02 (s, 1H, ArH), 6.82 (ddd, *J* = 8.1, 2.6, 1.0 Hz, 1H, ArH), 3.77 (s, 3H, OCH_3); ^{13}C NMR (101 MHz, DMSO): δ 168.0 (C), 159.4 (C), 149.7 (C), 136.3 (C), 129.5 (CH), 117.9 (CH), 113.0 (CH),

110.8 (CH), 101.9 (CH), 55.0 (OCH_3). *m/z* (ESI-HRMS) Calc $[M+H]^+$ = 207.0587; Observed $[M+H]^+$ = 207.0585

4-(4-Methoxyphenyl)thiazol-2-amine (21)

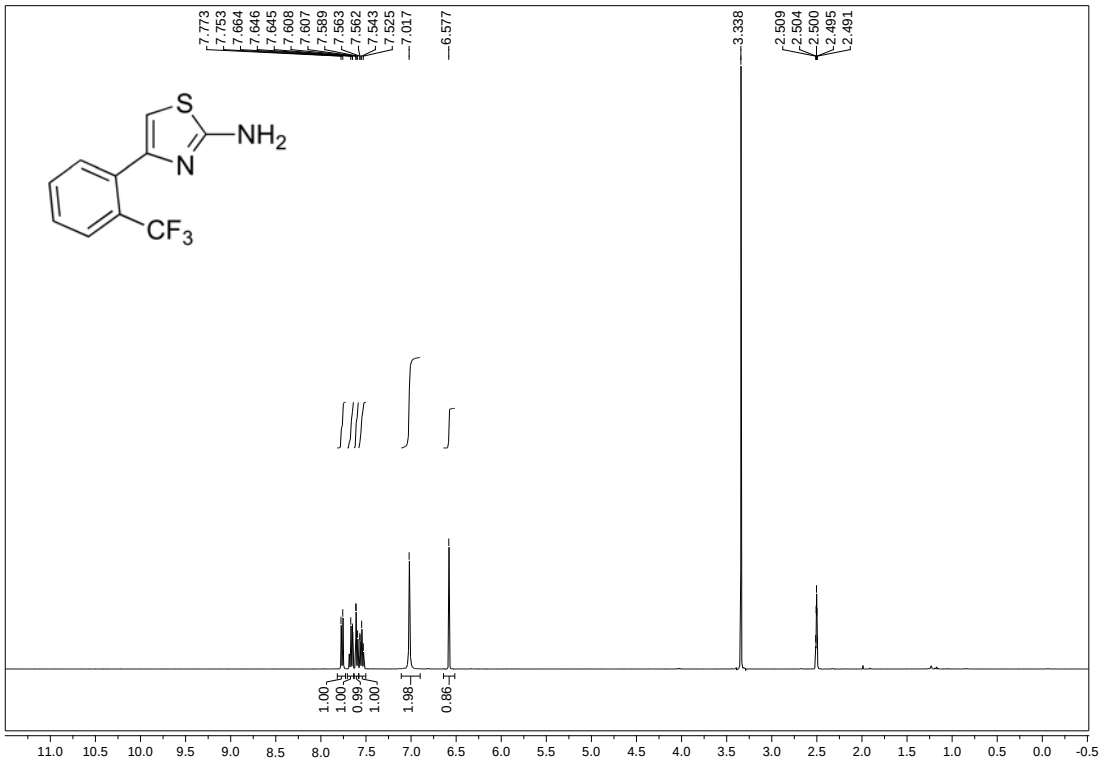


Same procedure as for **15**, from 4'-methoxyacetophenone gave a beige solid (96 mg, 14 %). m.p: 185-187°C**.** Analytical HPLC indicates greater than 98 % purity at 254 nm; ^{1}H NMR (400 MHz; DMSO): δ 7.72 (d, *J* = 8.9 Hz, 2H, 2 × ArH), 6.99 (br s, 2H, NH_2), 6.92 (d, *J* = 8.9 Hz, 2H, 2 × ArH), 6.82 (s, 1H, ArH), 3.76 (s, 3H, OCH_3); ^{13}C NMR (101 MHz, DMSO): δ 168.1 (C), 158.5 (C), 149.7 (C), 127.9 (C), 126.8 (2 × CH), 113.8 (2 × CH), 99.3 (CH), 55.1 (OCH_3). *m/z* (ESI-HRMS) Calc $[M+H]^+$ = 207.0587 ; Observed $[M+H]^+$ = 207.0593.

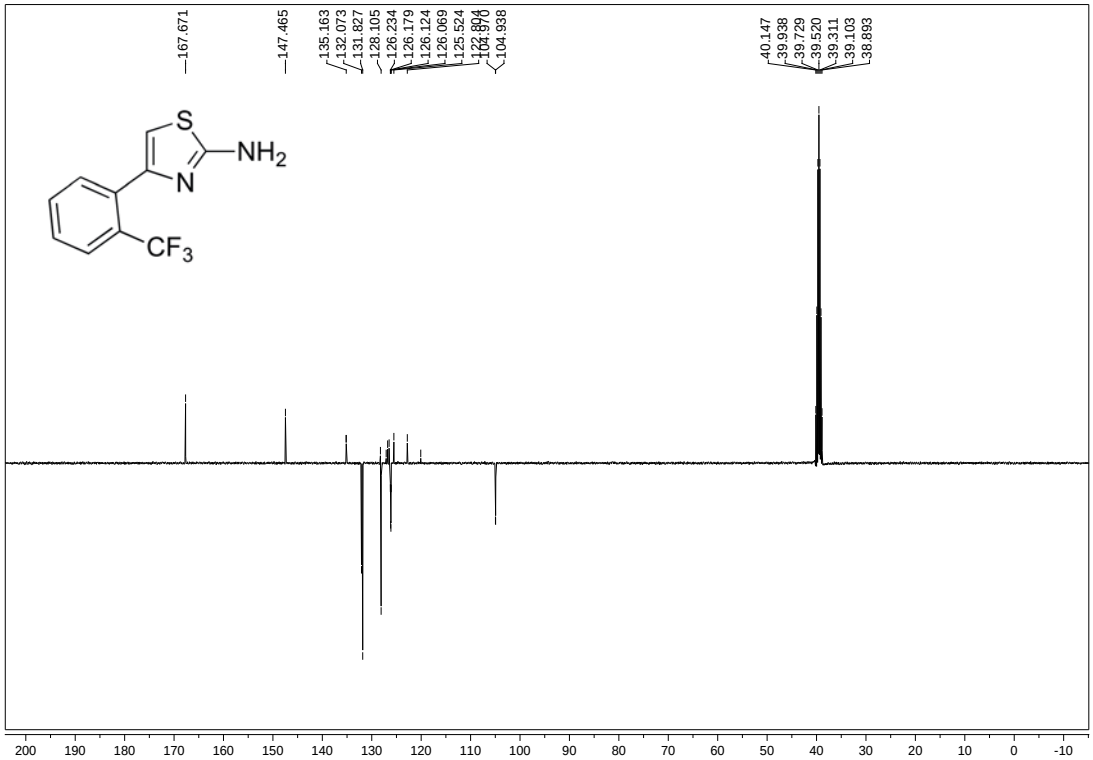
NMR Spectra

4-(2-(Trifluoromethyl)phenyl)thiazol-2-amine (10)

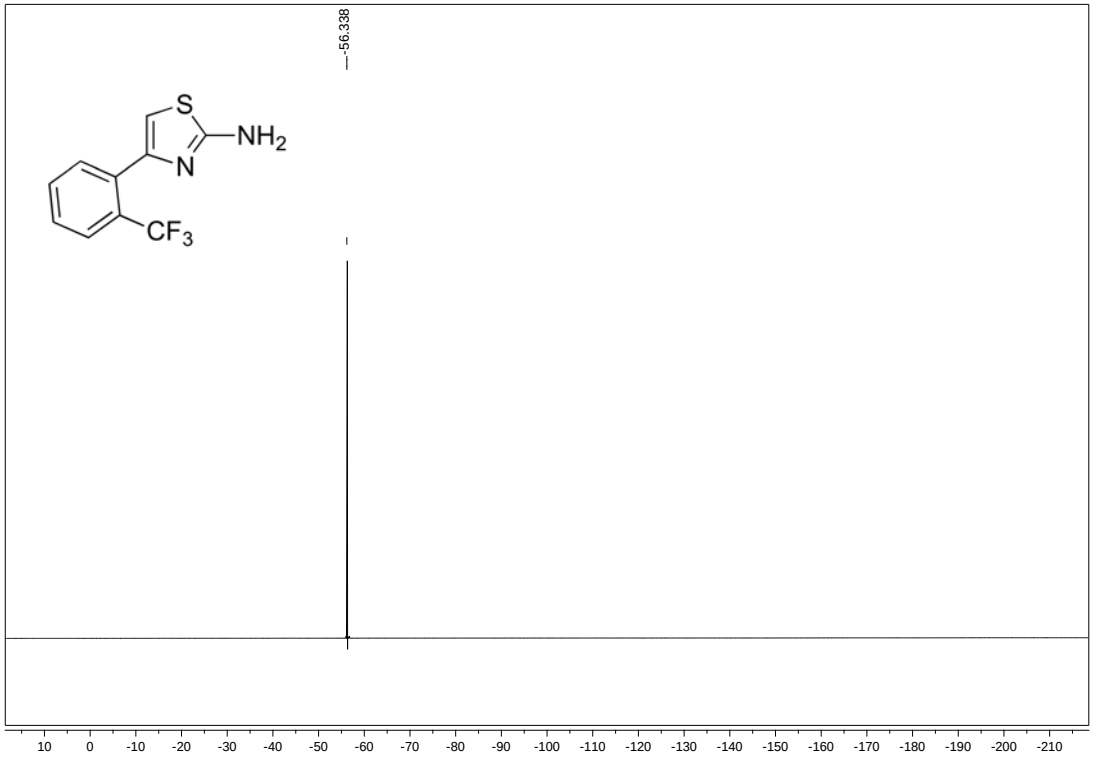
^{1}H NMR Spectrum (400 MHz, DMSO)



^{13}C DEPTQ NMR Spectrum (101 MHz, DMSO)

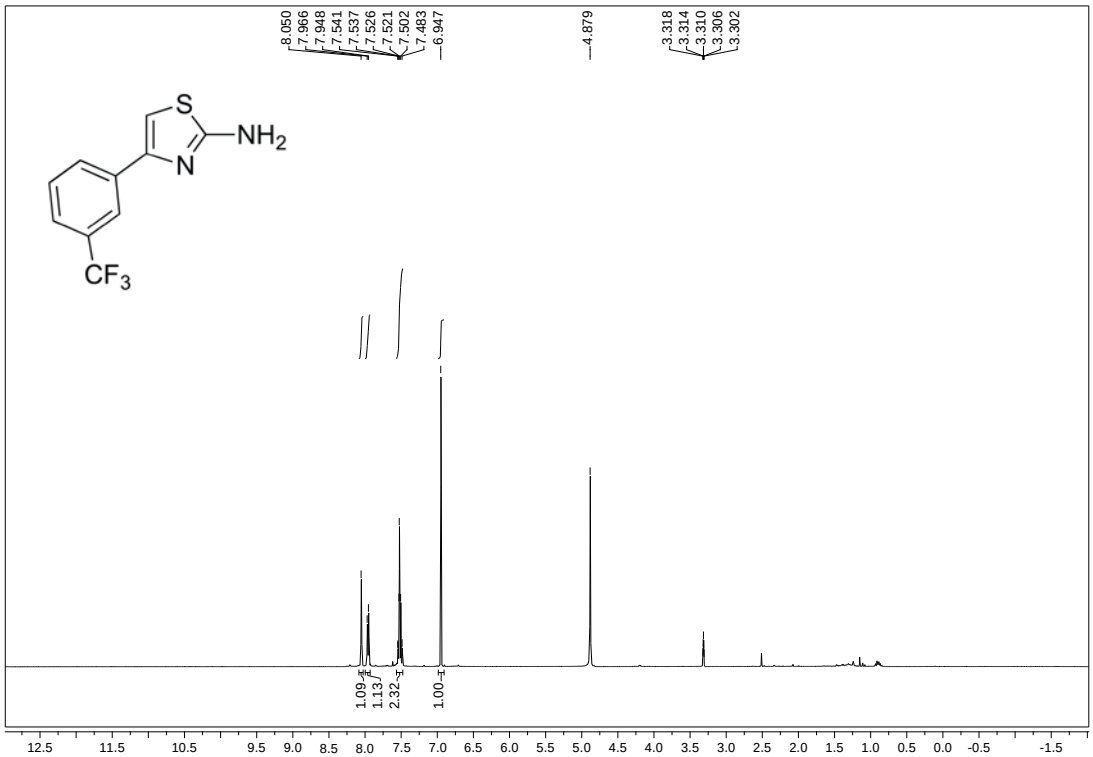


^{19}F NMR Spectrum (376 MHz, DMSO)

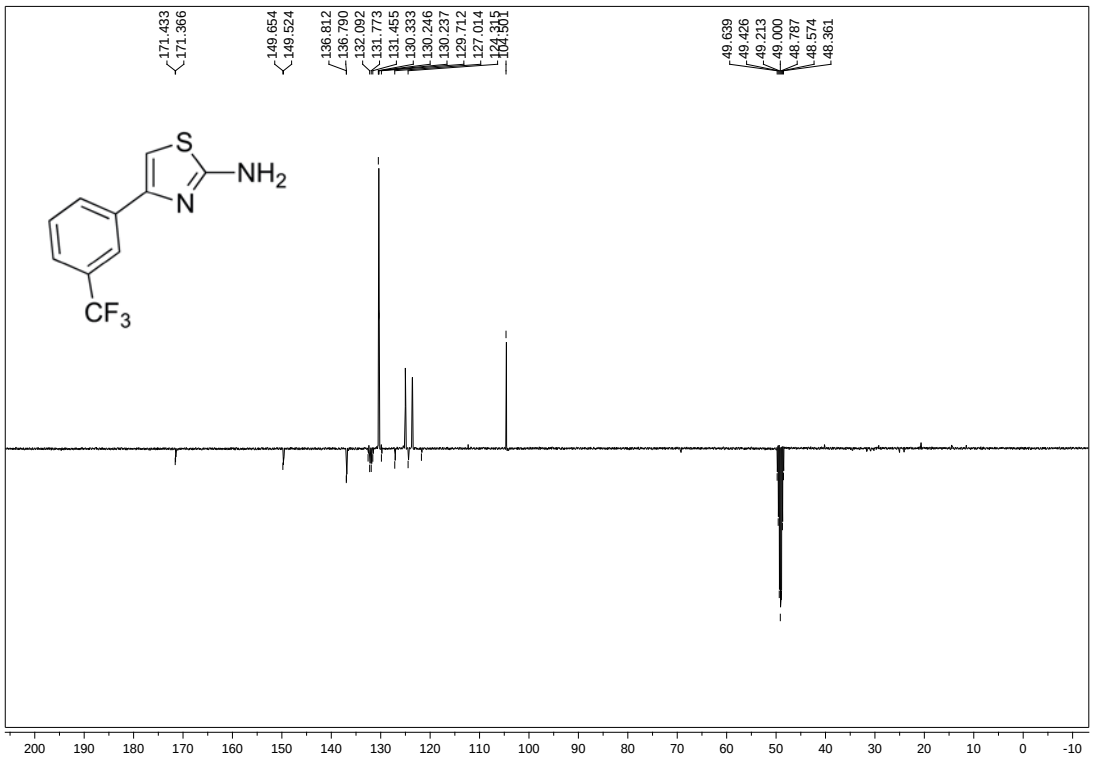


4-(3-(Trifluoromethyl)phenyl)thiazol-2-amine (11)

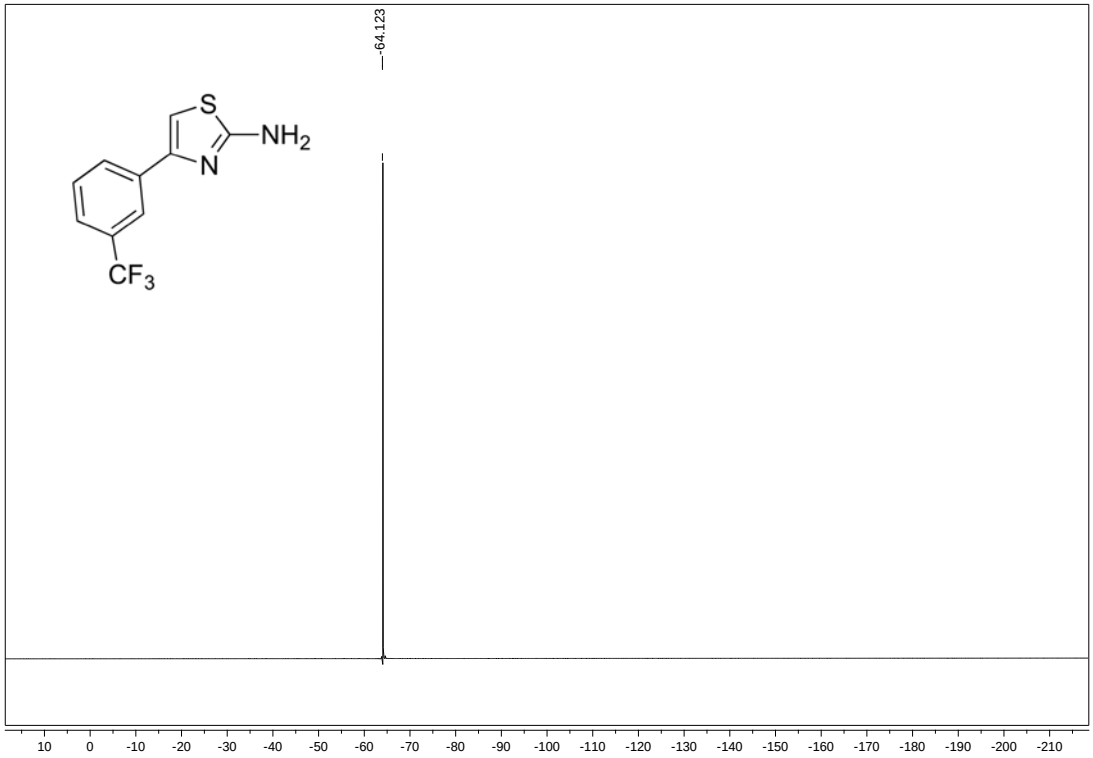
^{1}H NMR Spectrum (400 MHz, MeOD)



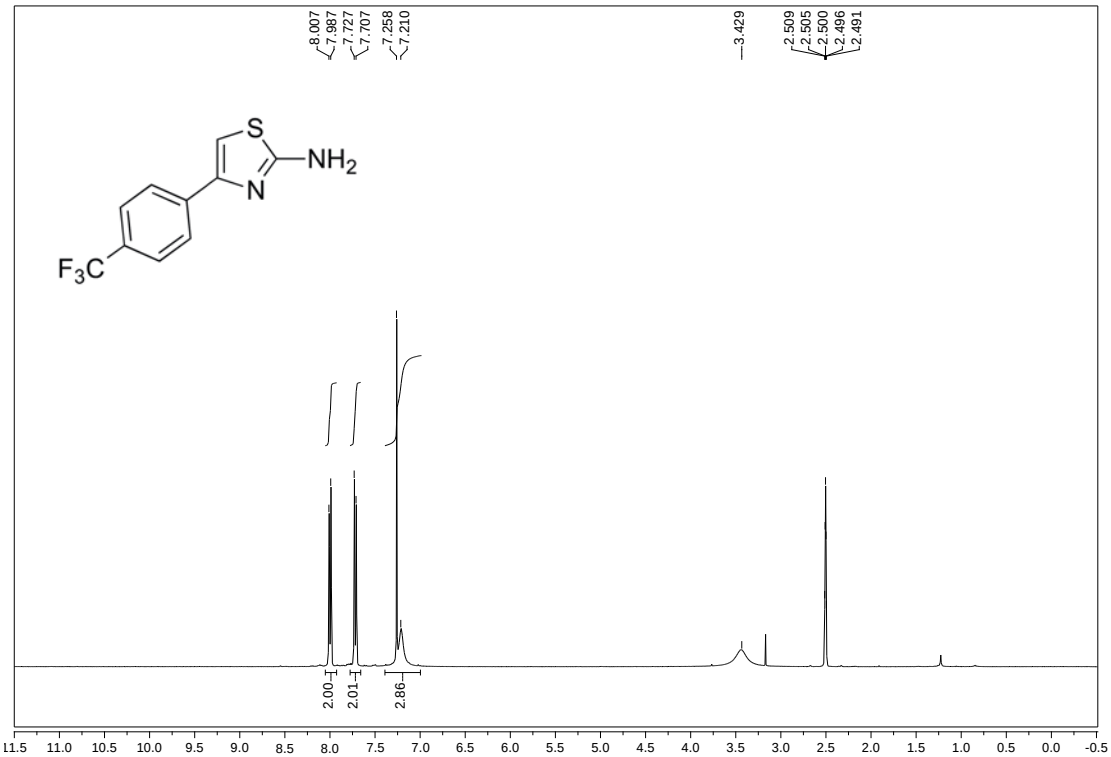
^{13}C DEPTQ NMR Spectrum (101 MHz, MeOD)



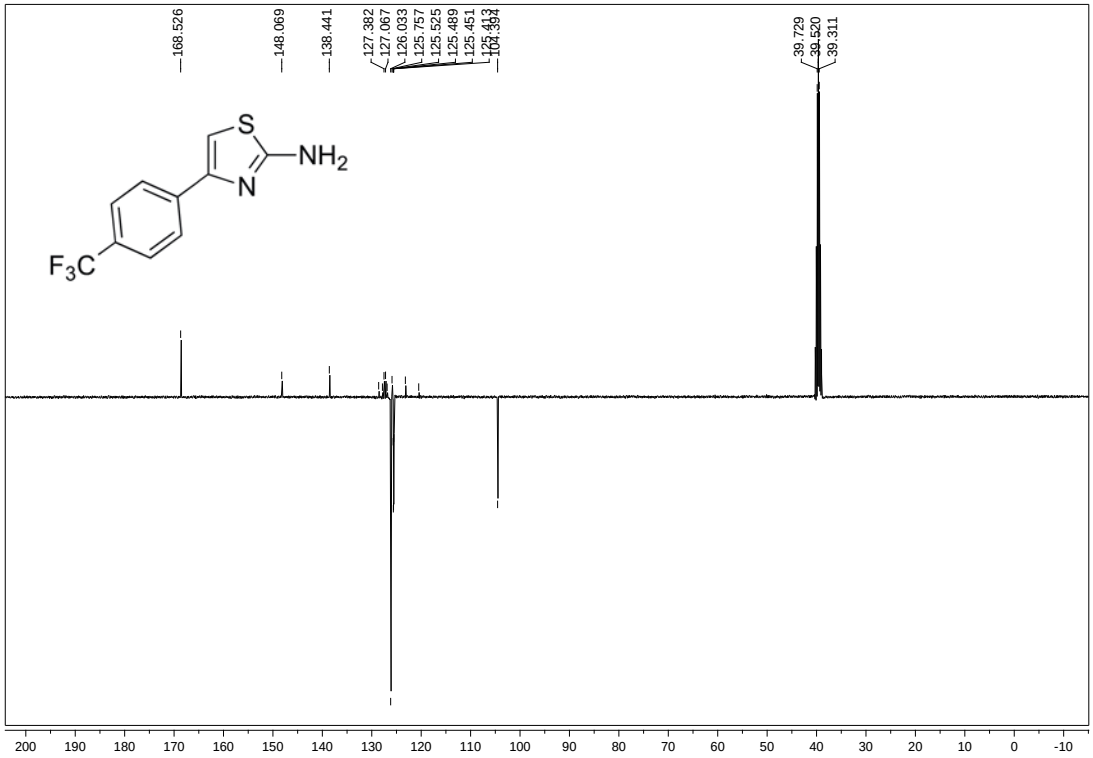
^{19}F NMR Spectrum (376 MHz, MeOD)



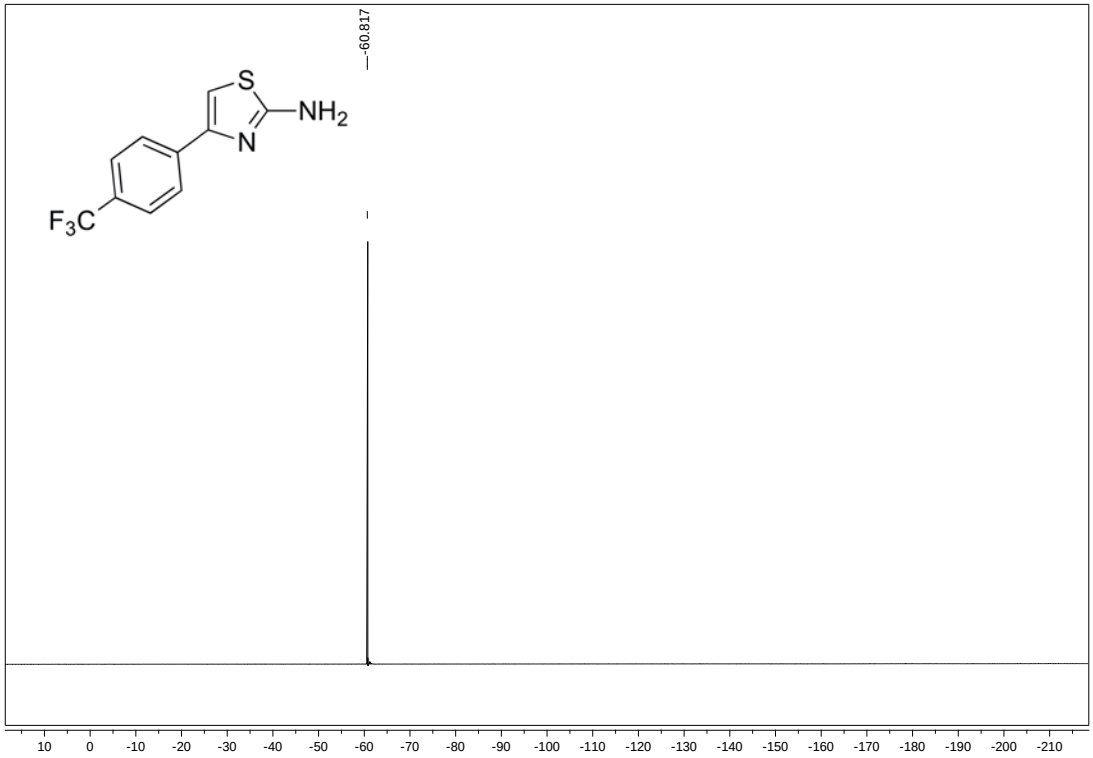
4-(4-(Trifluoromethyl)phenyl)thiazol-2-amine (12)
^{1}H NMR Spectrum (400 MHz, DMSO)



^{13}C DEPTQ NMR Spectrum (101 MHz, DMSO)

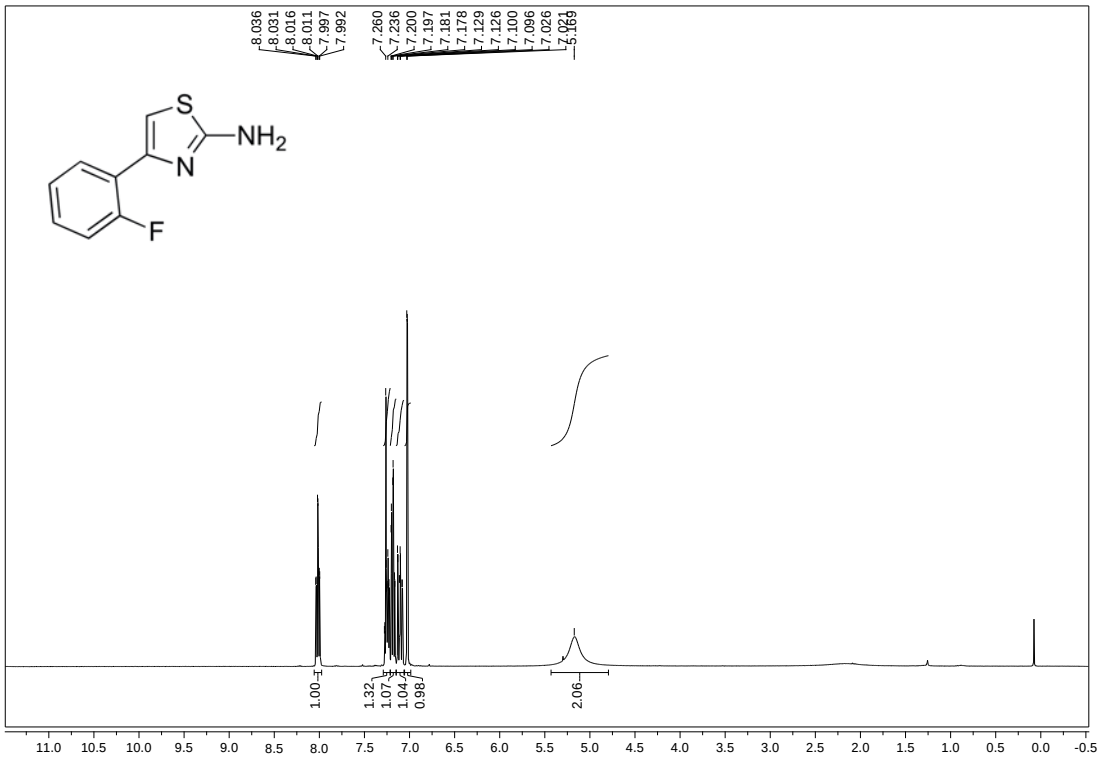


^{19}F NMR Spectrum (376 MHz, DMSO)

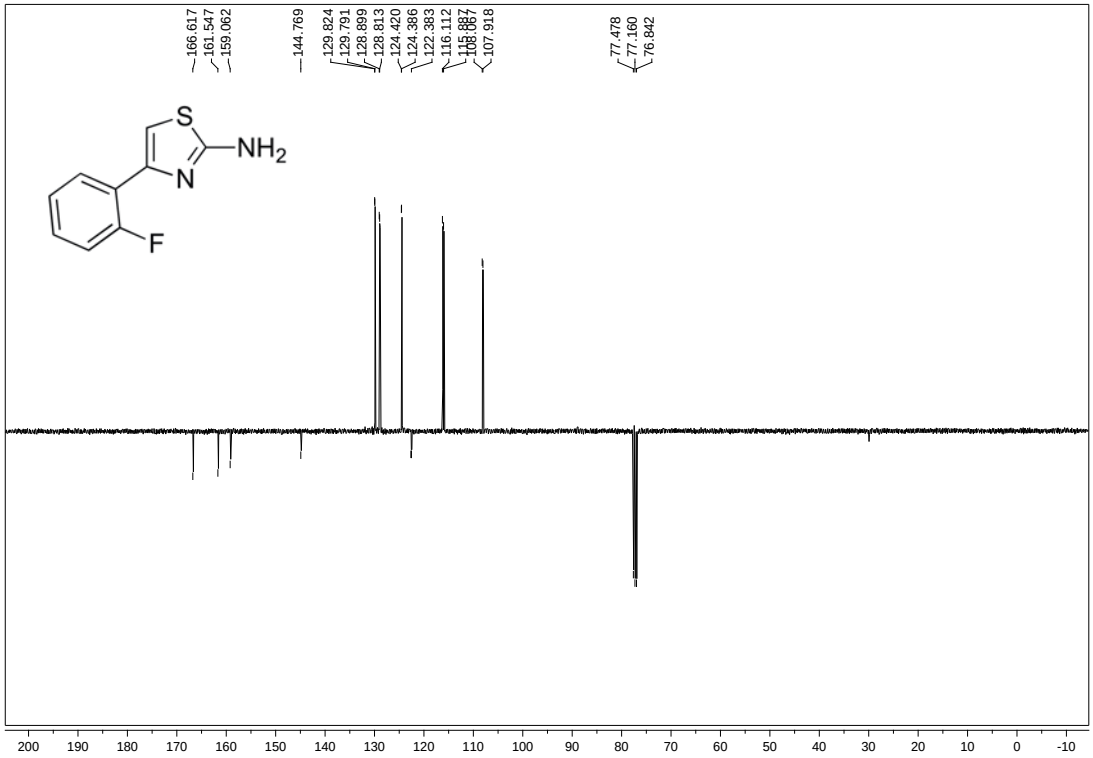


4-(2-Fluorophenyl)thiazol-2-amine (13)

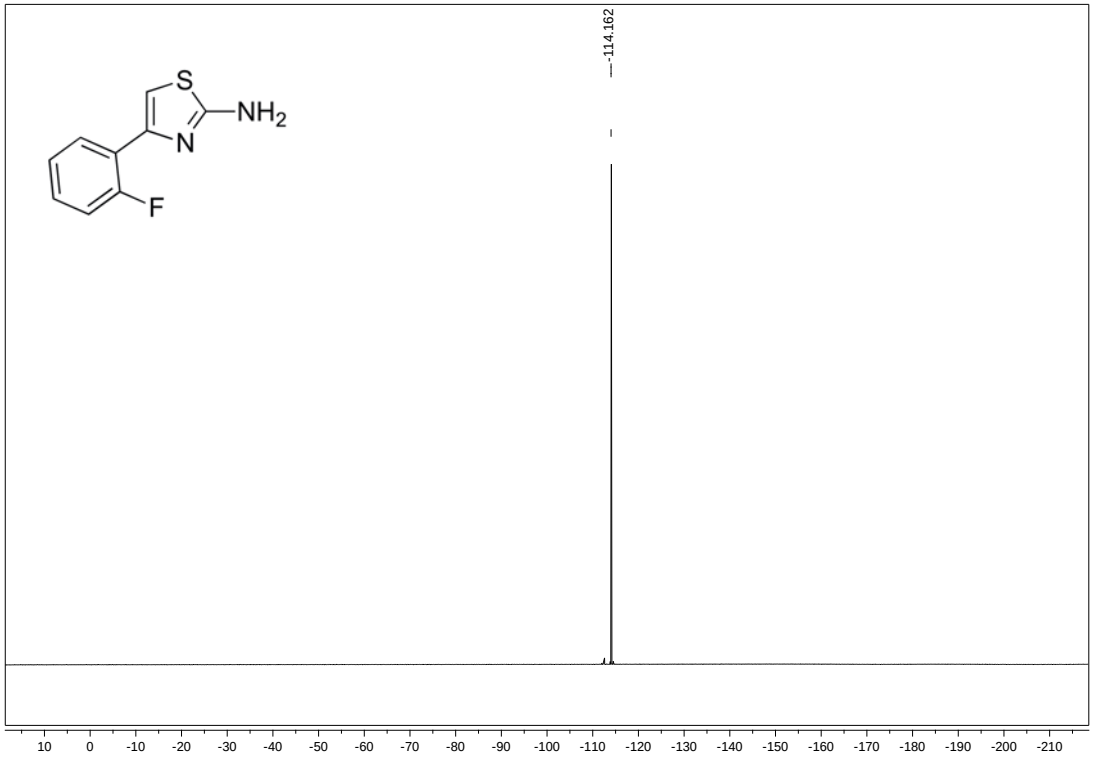
^{1}H NMR Spectrum (400 MHz, $CDCl_3$)



^{13}C DEPTQ NMR Spectrum (101 MHz, $CDCl_3$)

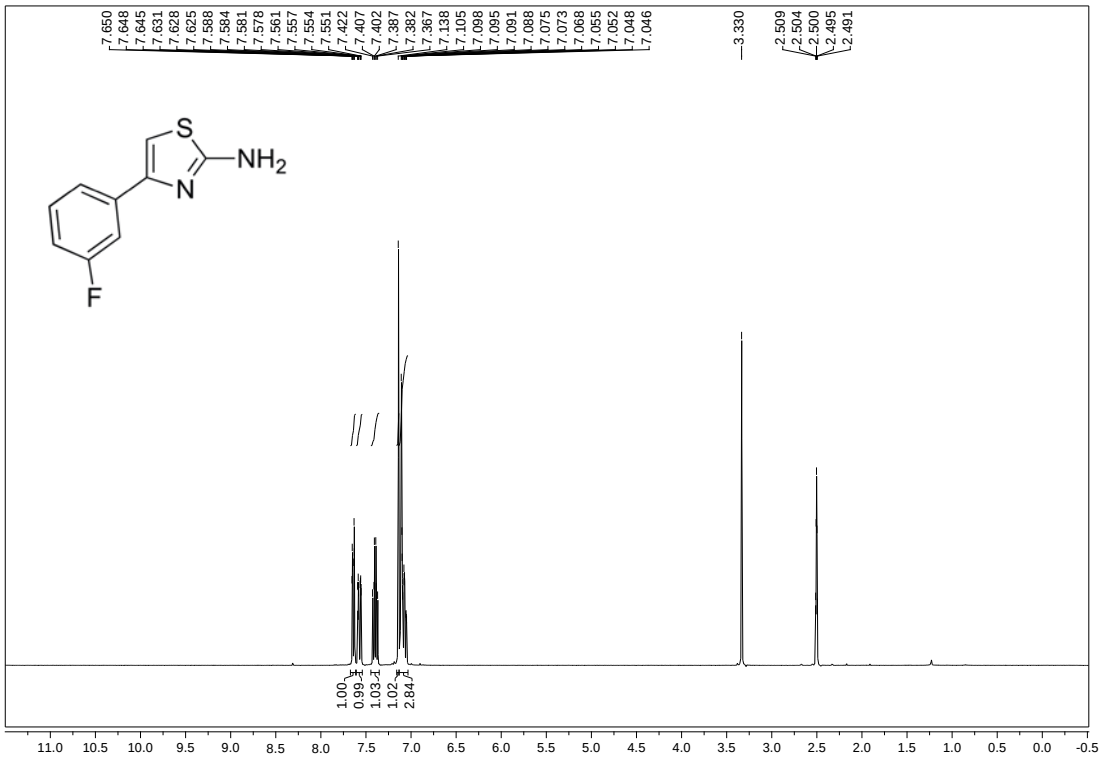


^{19}F NMR Spectrum (376 MHz, $CDCl_3$)

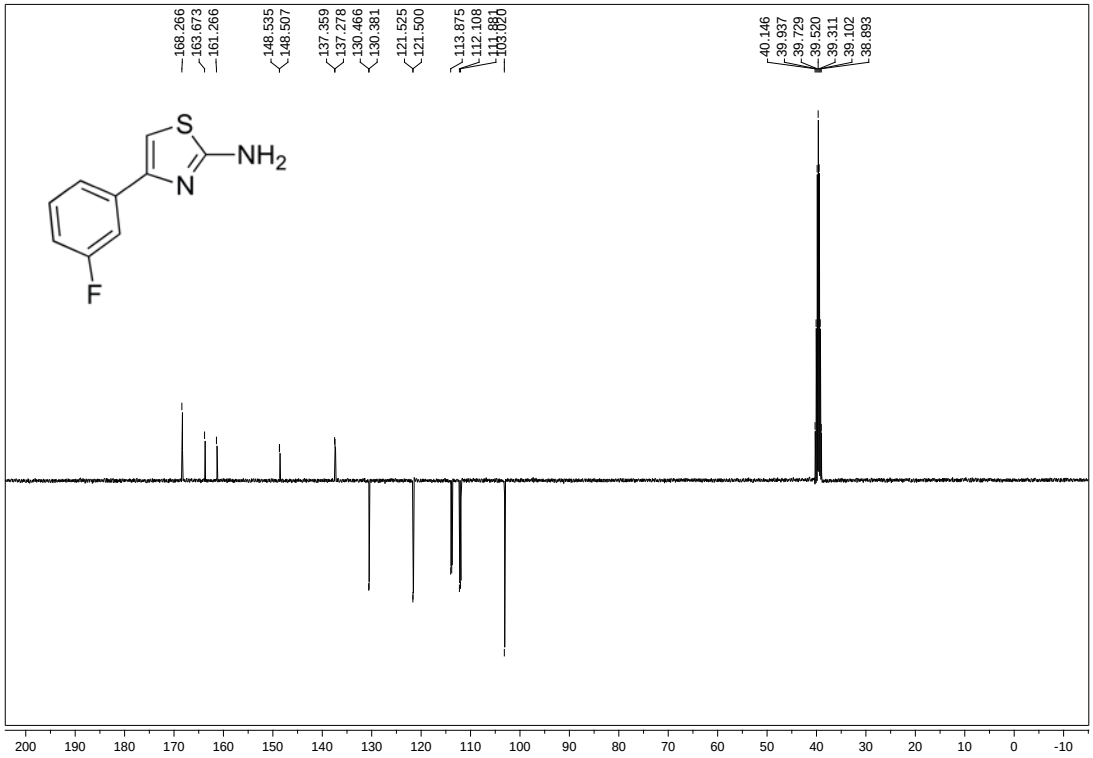


4-(3-Fluorophenyl)thiazol-2-amine (14)

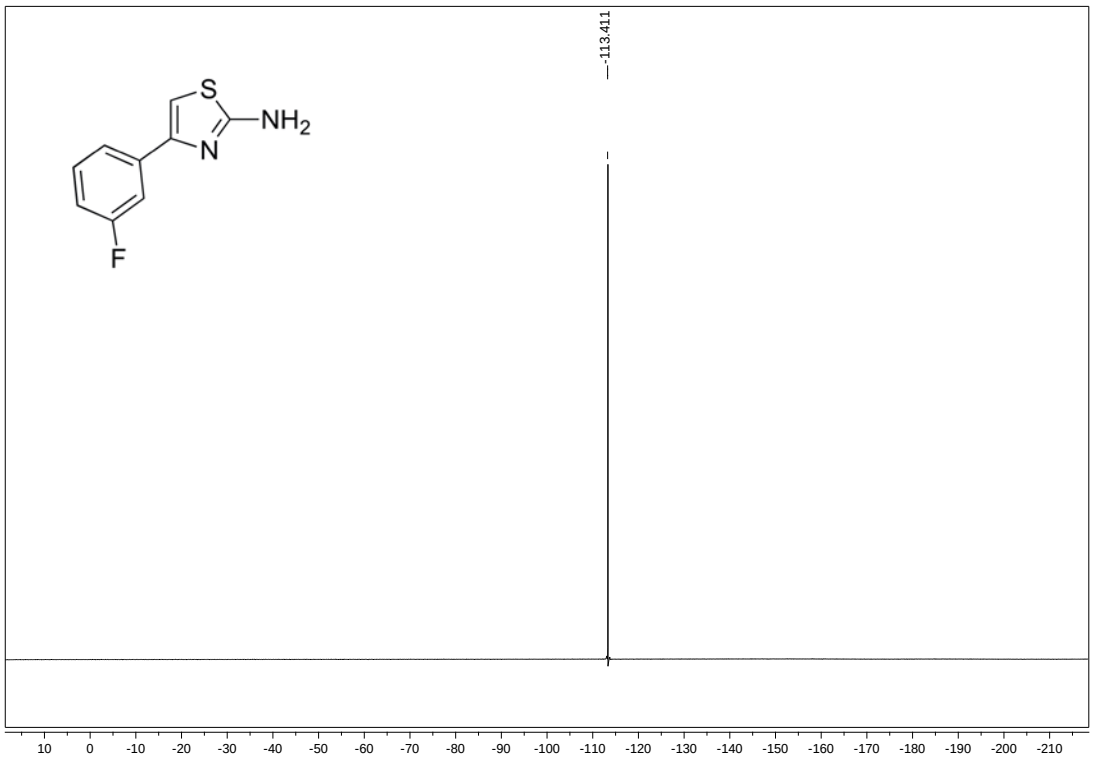
^{1}H NMR Spectrum (400 MHz, DMSO)



^{13}C DEPTQ NMR Spectrum (101 MHz, DMSO)

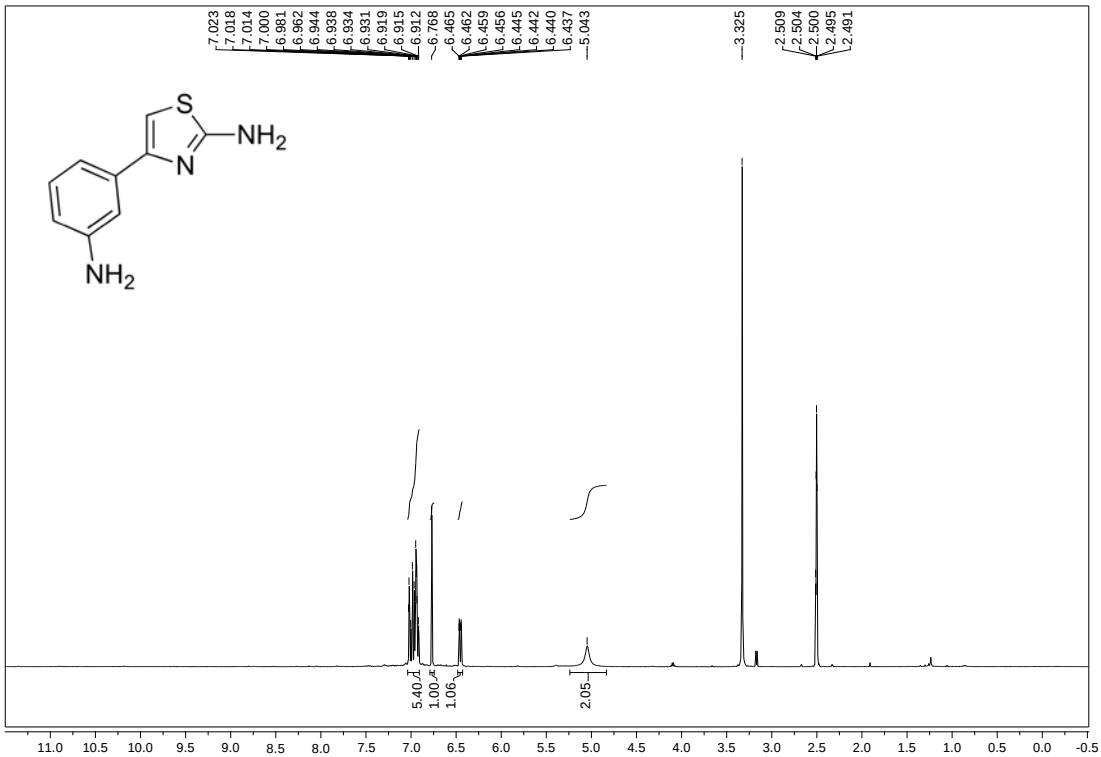


^{19}F NMR Spectrum (376 MHz, DMSO)

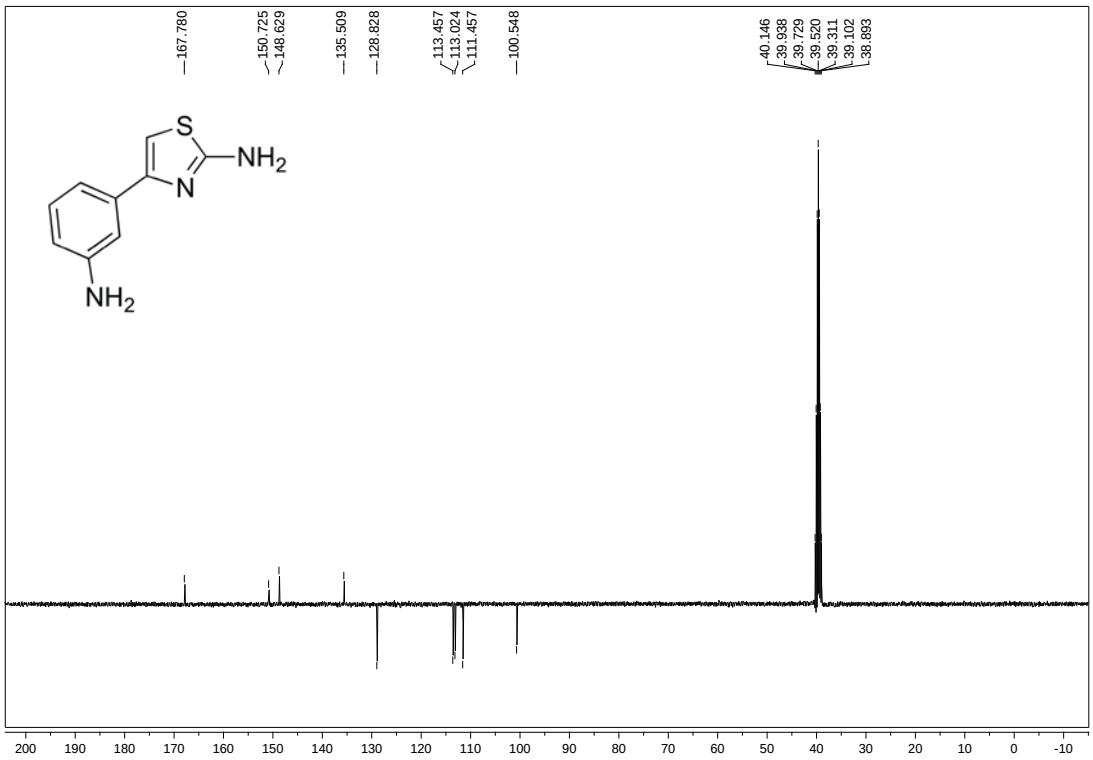


4-(3-Aminophenyl)thiazol-2-amine (15)

^{1}H NMR Spectrum (400 MHz, DMSO)

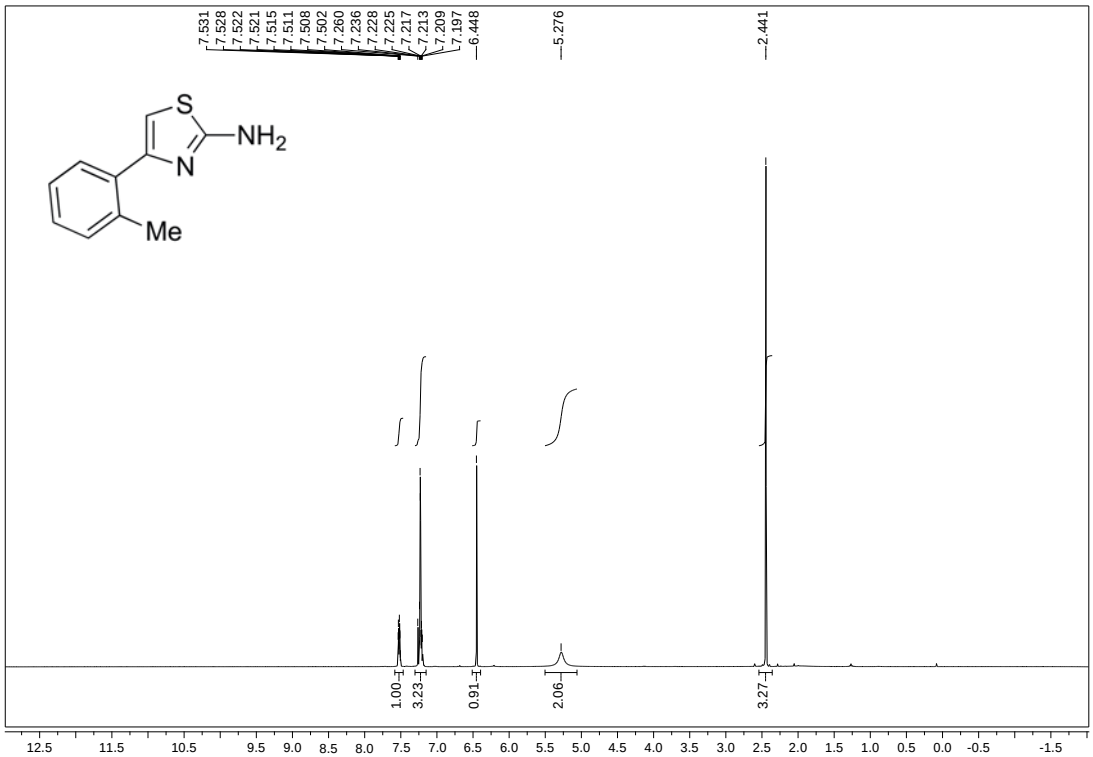


^{13}C DEPTQ NMR Spectrum (101 MHz, DMSO)

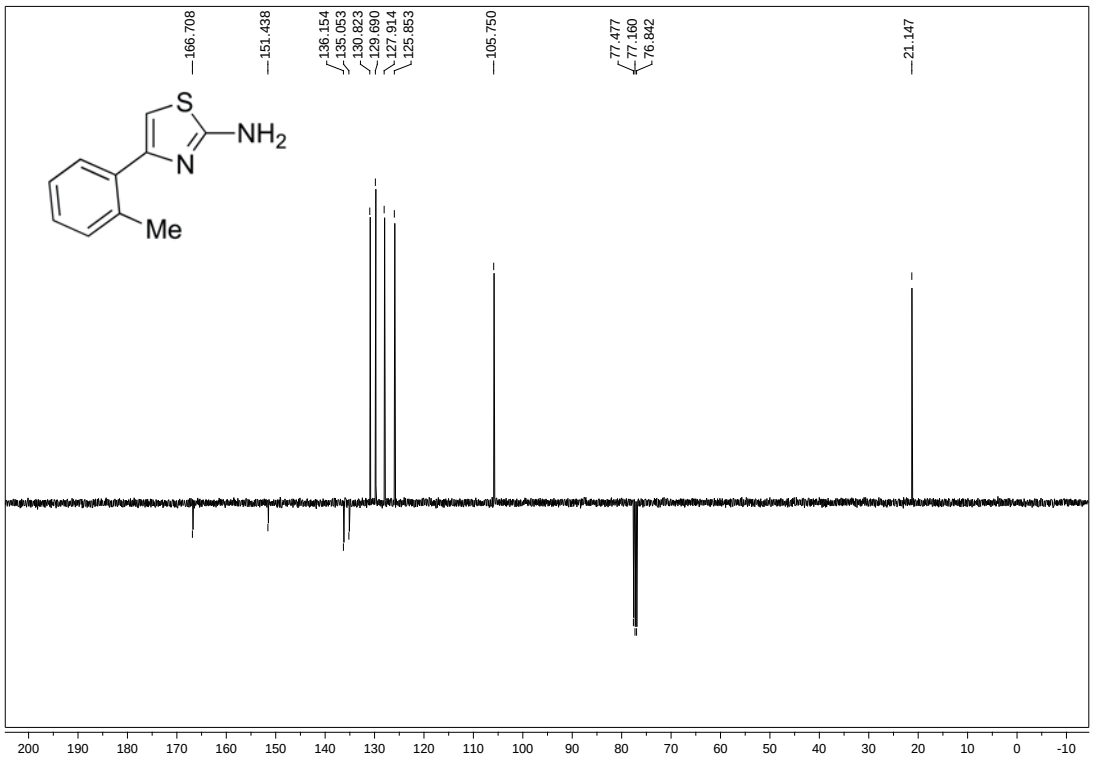


4-(o-Tolyl)thiazol-2-amine (16)

^{1}H NMR Spectrum (400 MHz, $CDCl_3$)

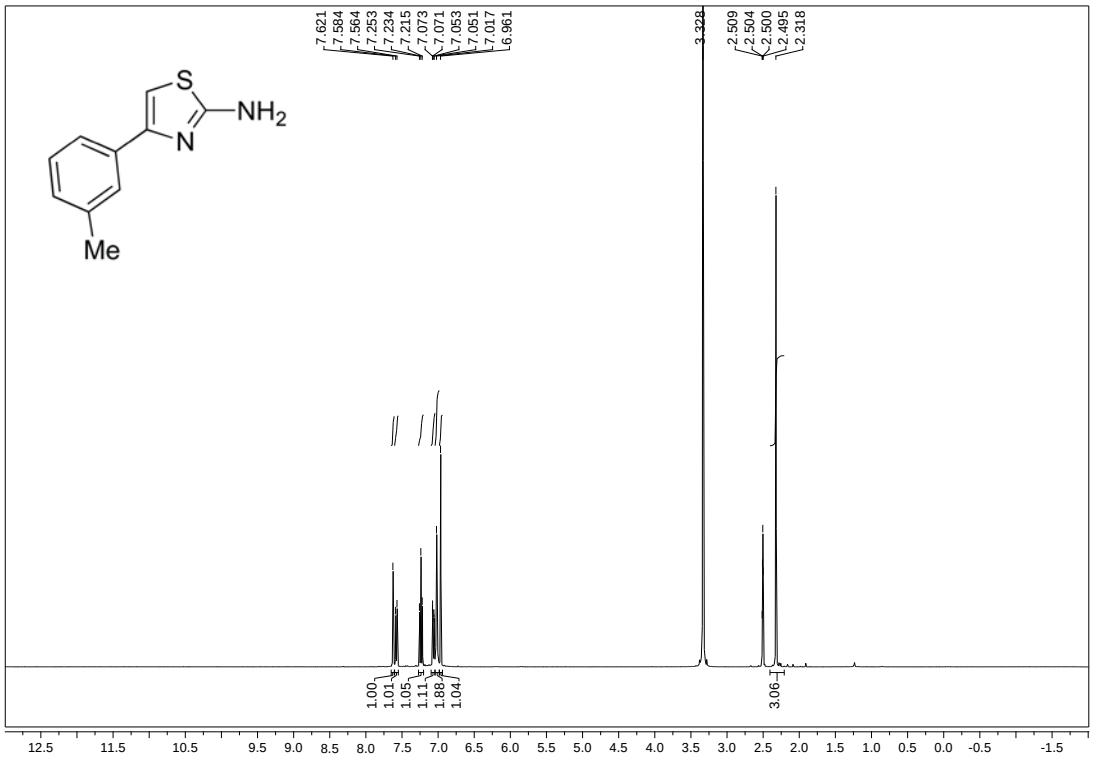


^{13}C DEPTQ NMR Spectrum (101 MHz, $CDCl_3$)

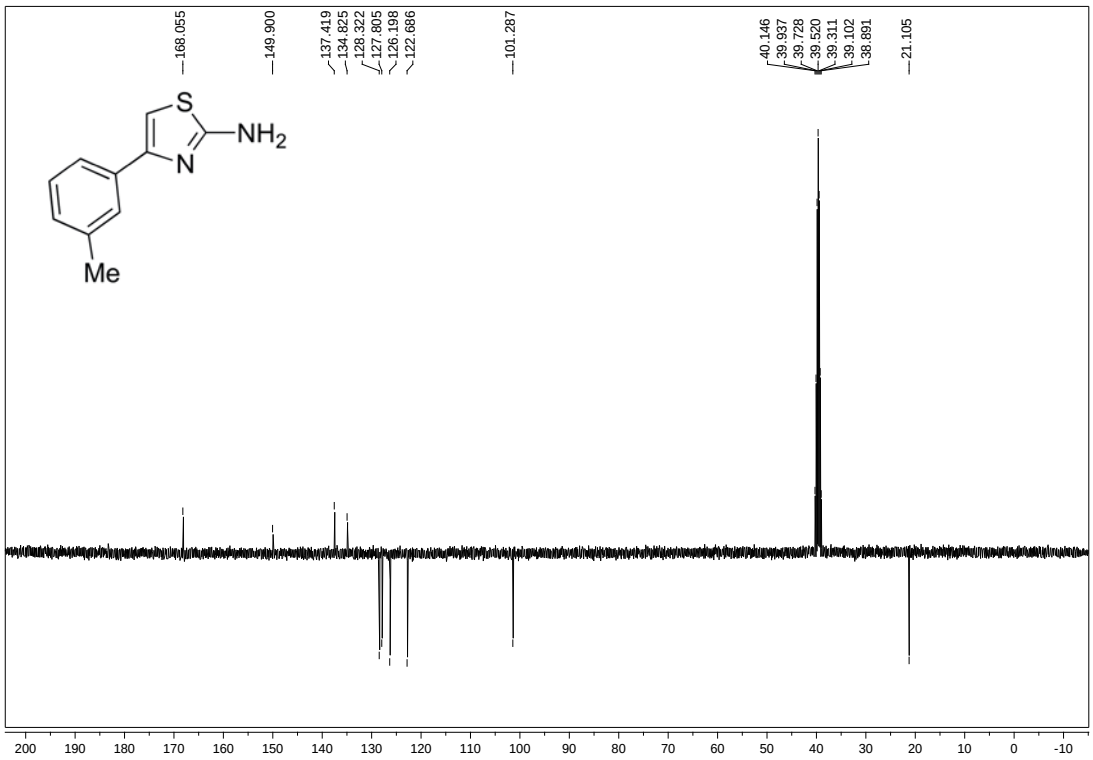


4-(*m*-Tolyl)thiazol-2-amine (17)

^{1}H NMR Spectrum (400 MHz, DMSO)

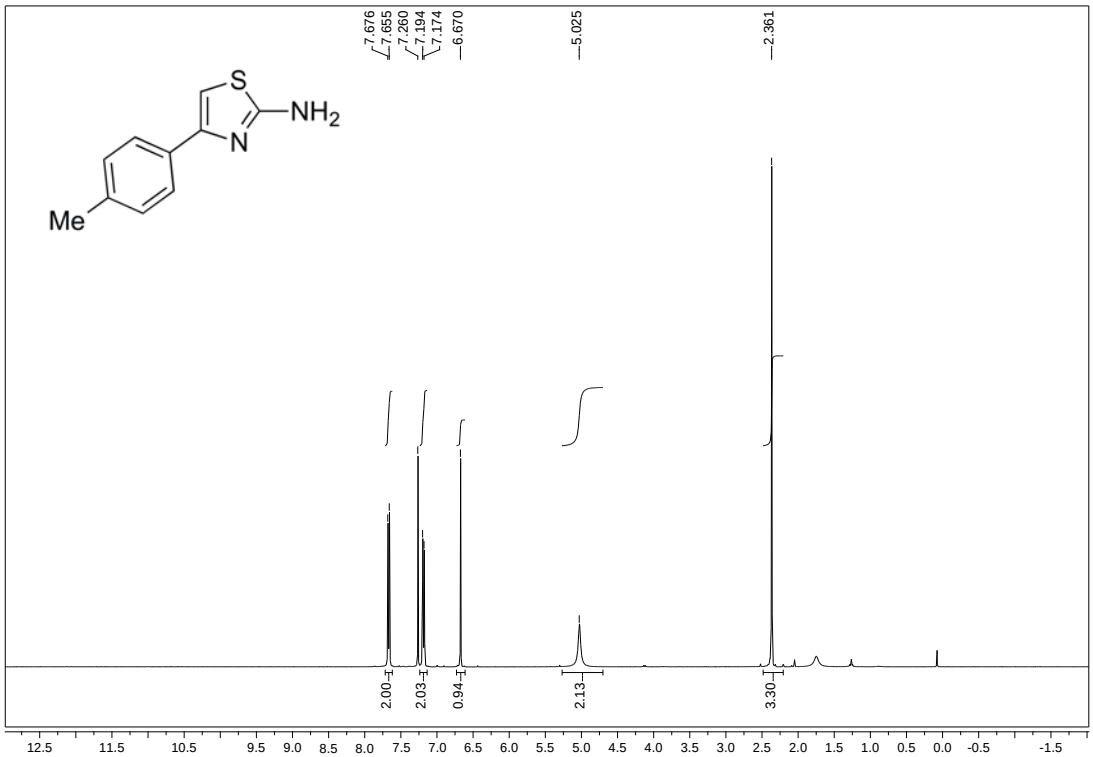


^{13}C DEPTQ NMR Spectrum (101 MHz, DMSO)

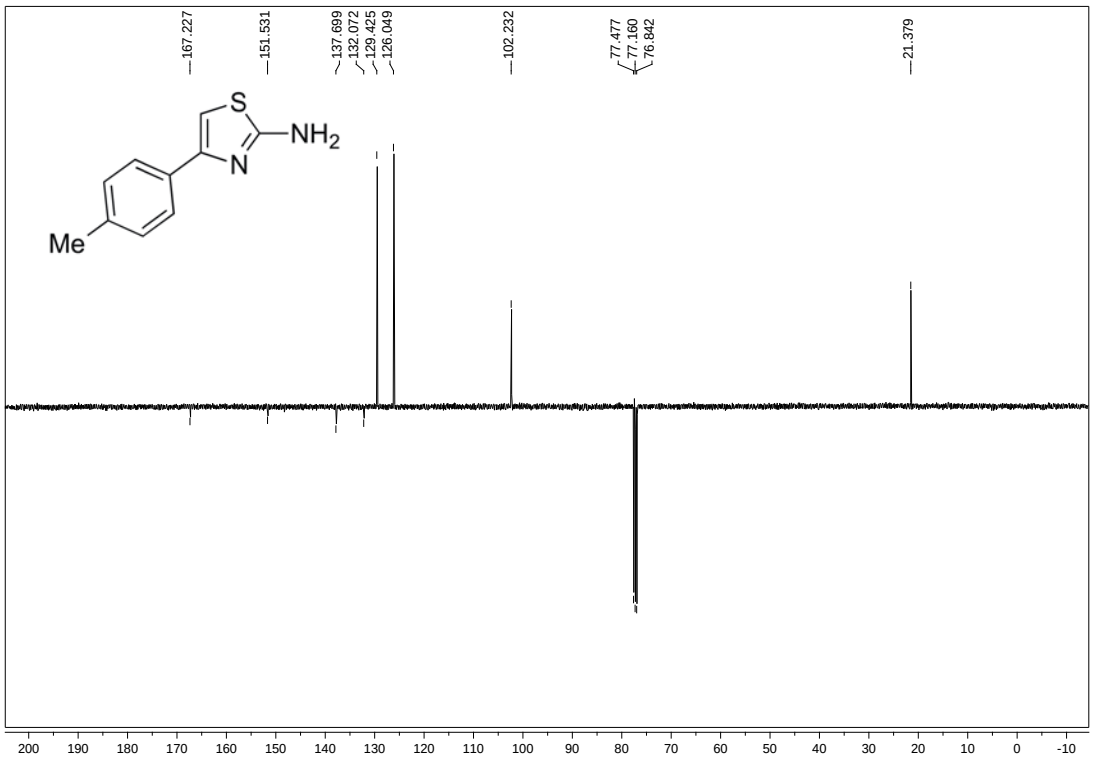


4-(*p*-Tolyl)thiazol-2-amine (18)

^{1}H NMR Spectrum (400 MHz, $CDCl_3$)

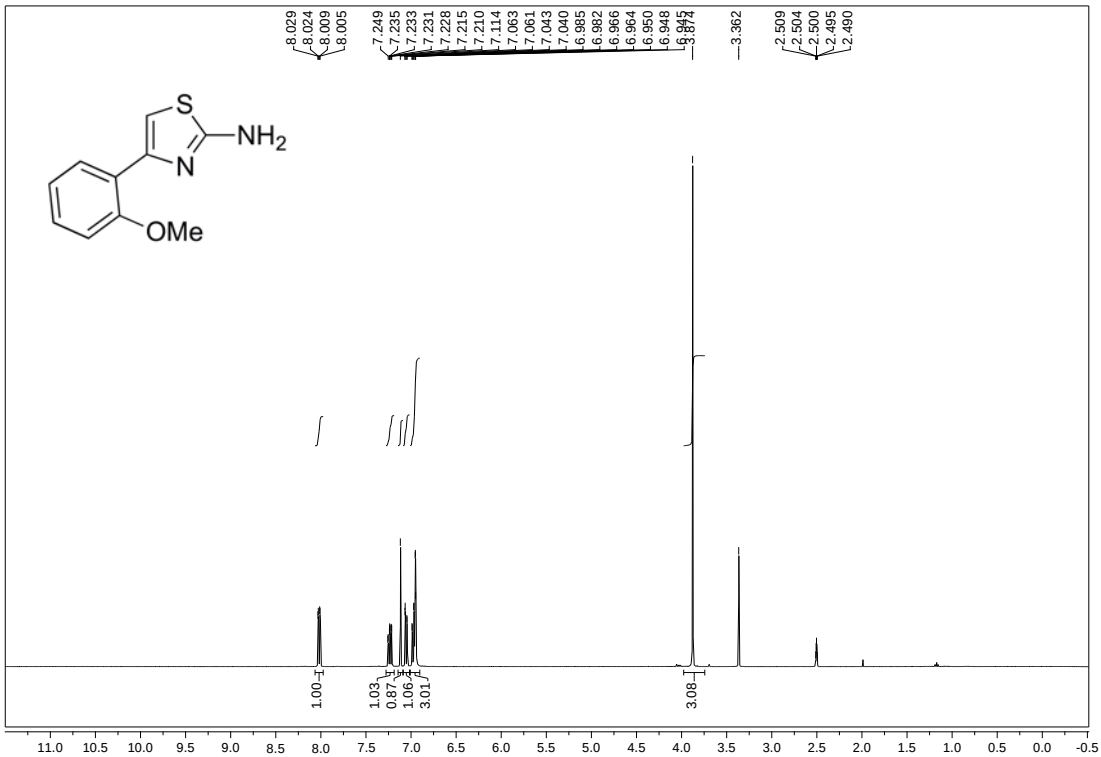


^{13}C DEPTQ NMR Spectrum (101 MHz, $CDCl_3$)

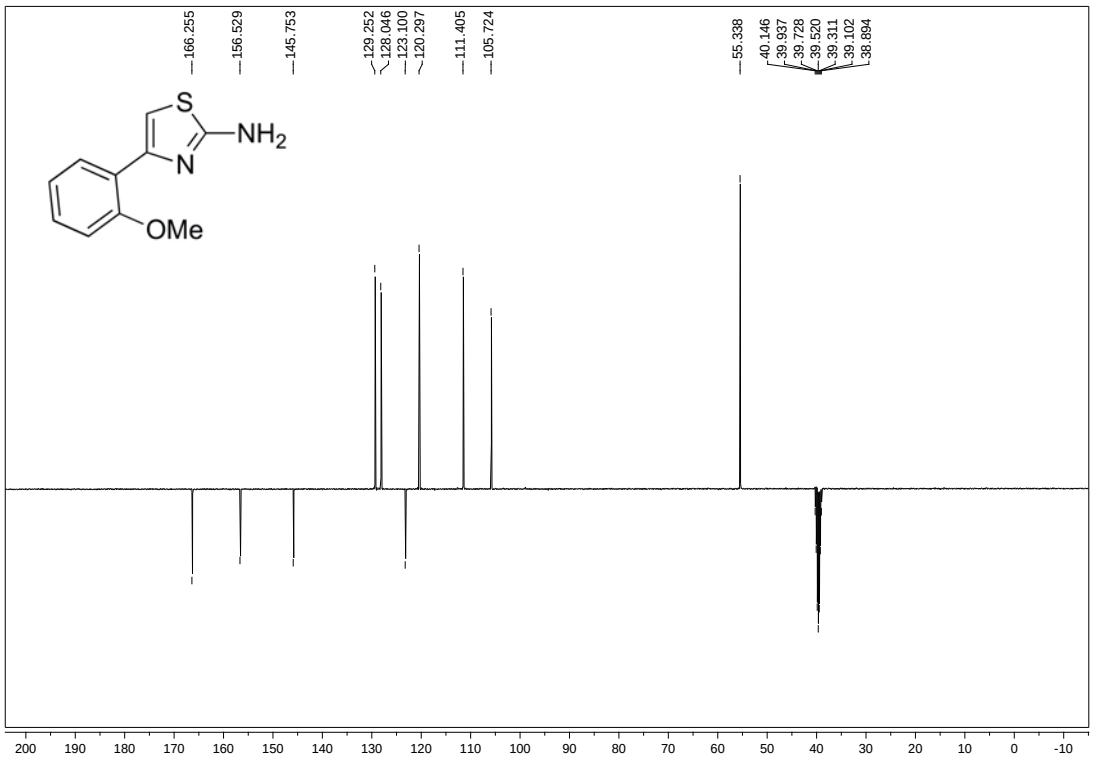


4-(2-Methoxyphenyl)thiazol-2-amine (19)

^{1}H NMR Spectrum (400 MHz, DMSO)

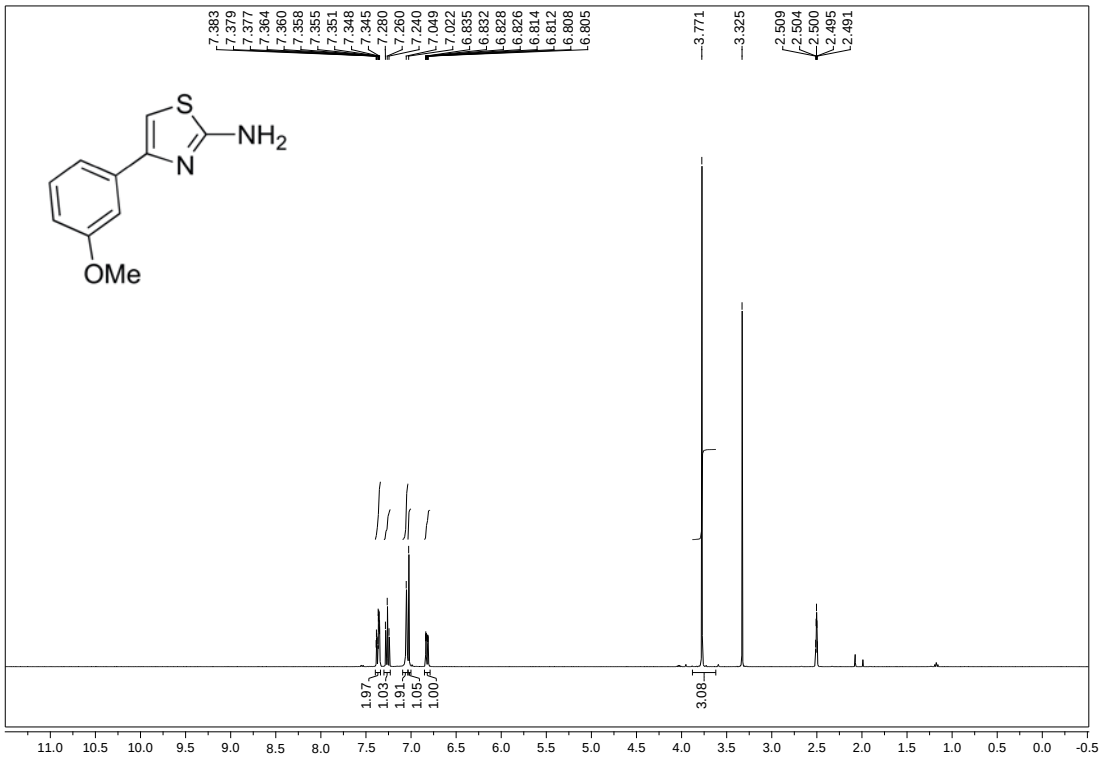


^{13}C DEPTQ NMR Spectrum (101 MHz, DMSO)

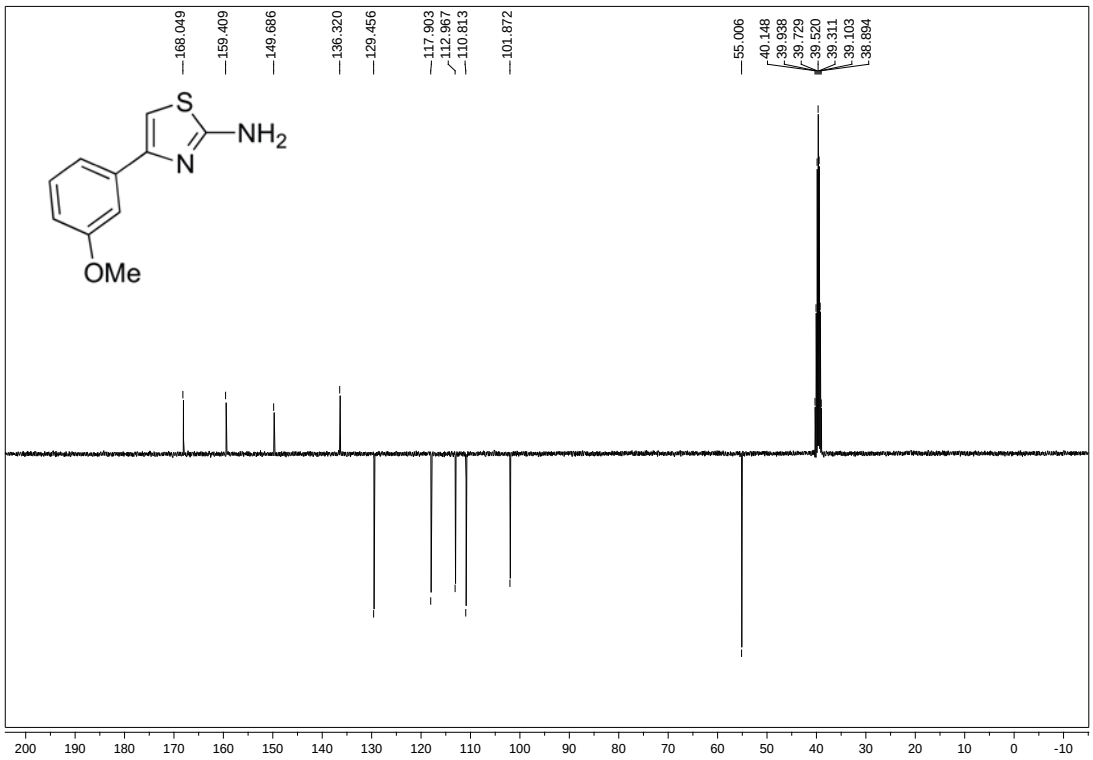


4-(3-Methoxyphenyl)thiazol-2-amine (20)

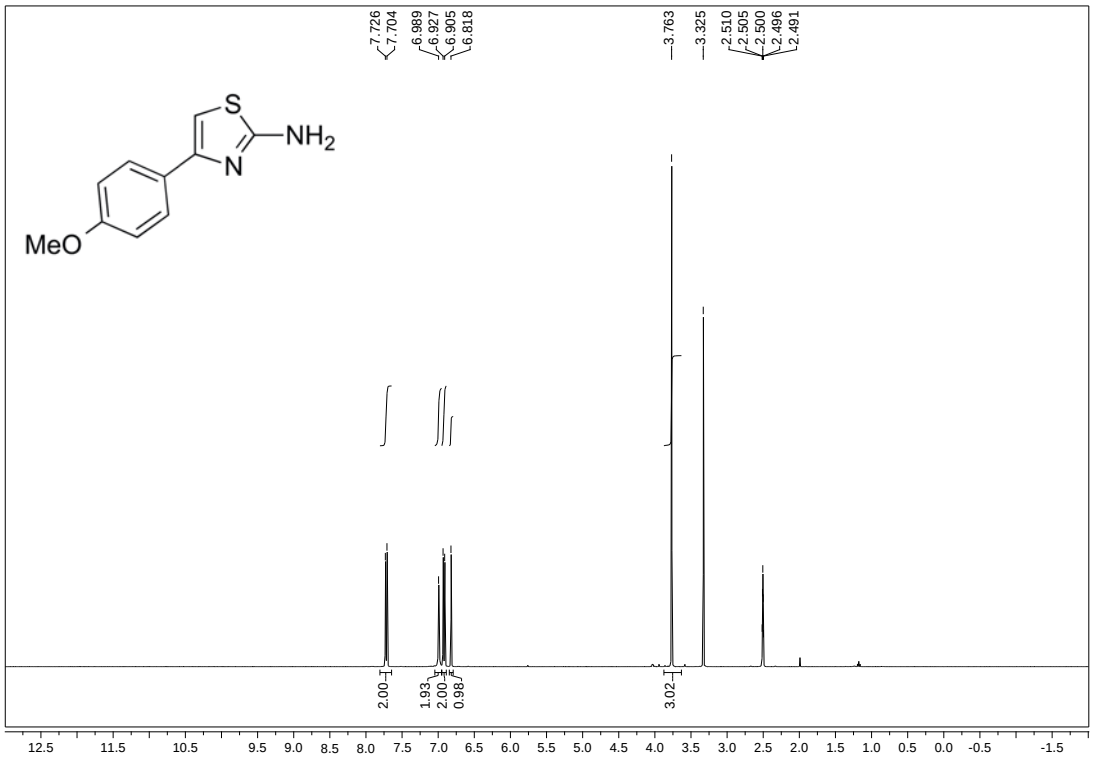
^{1}H NMR Spectrum (400 MHz, DMSO)



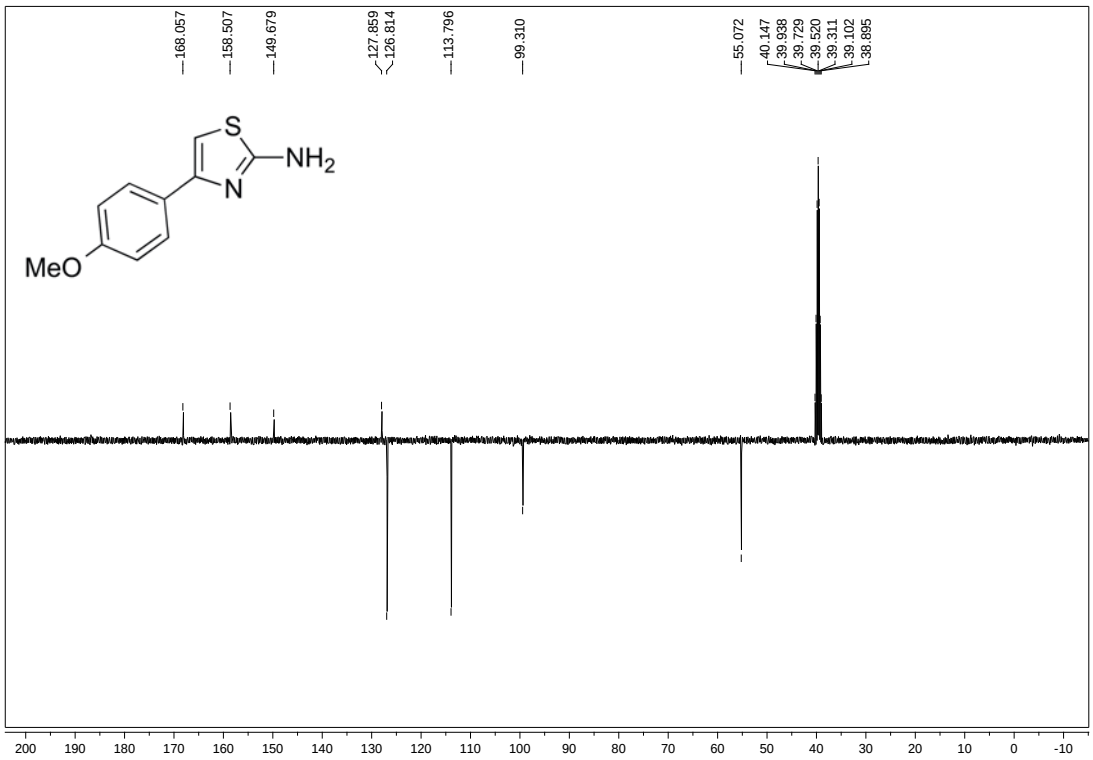
^{13}C DEPTQ NMR Spectrum (101 MHz, DMSO)



4-(4-Methoxyphenyl)thiazol-2-amine (21)
^{1}H NMR Spectrum (400 MHz, DMSO)

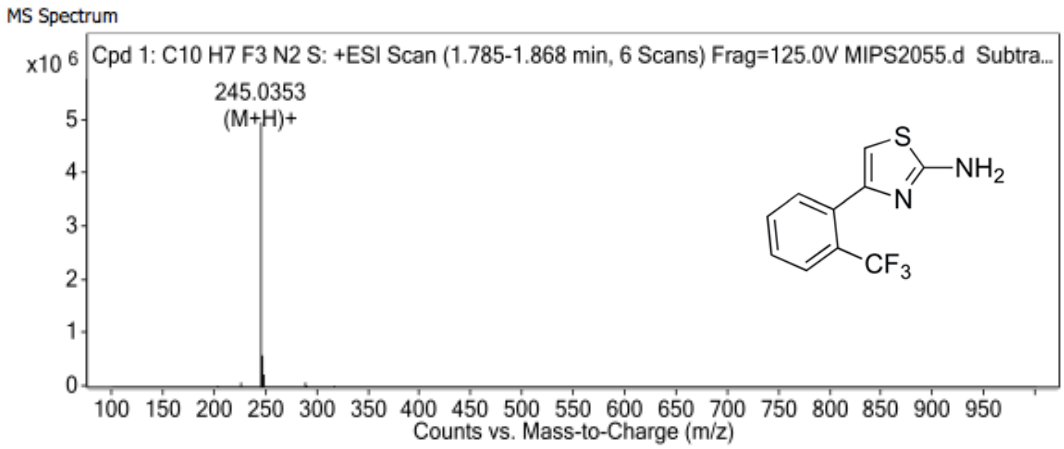


^{13}C DEPTQ NMR Spectrum (101 MHz, DMSO)

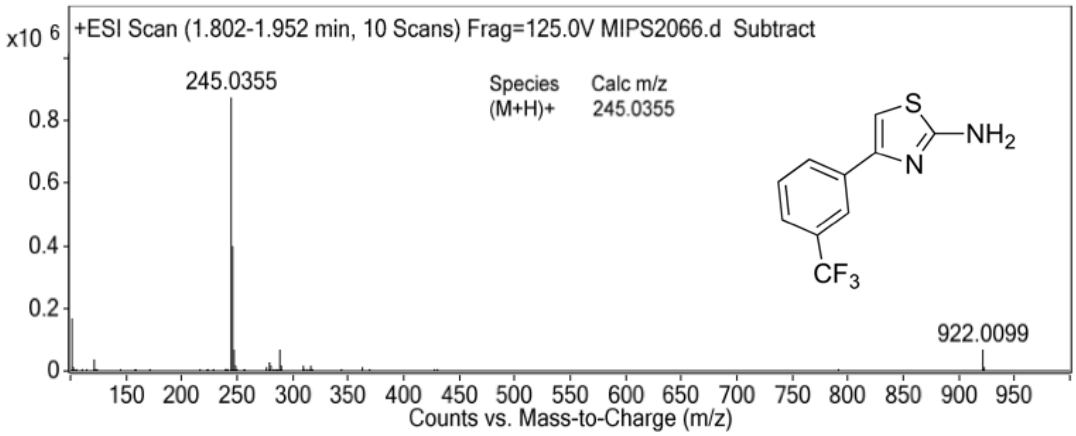


High-Resolution Mass Spectra

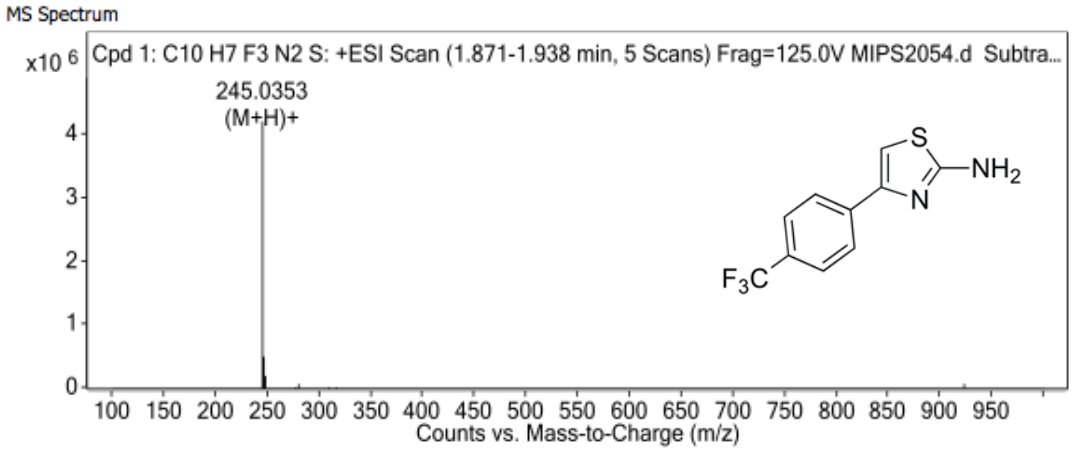
4-(2-(Trifluoromethyl)phenyl)thiazol-2-amine (10)



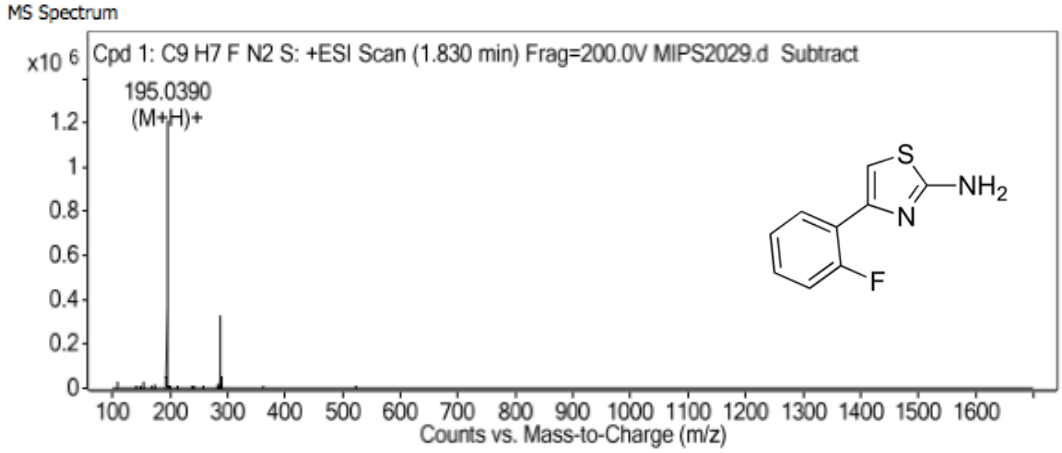
4-(3-(Trifluoromethyl)phenyl)thiazol-2-amine (11)



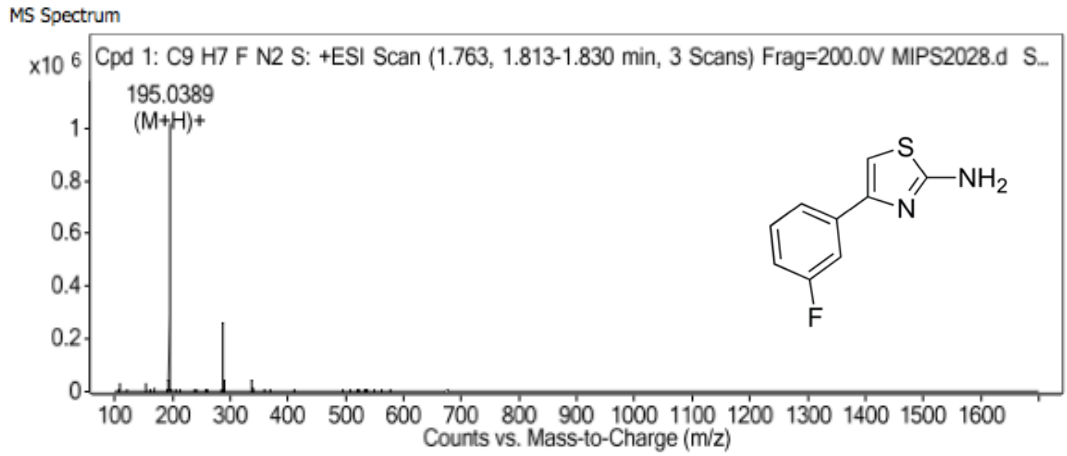
4-(4-(Trifluoromethyl)phenyl)thiazol-2-amine (12)



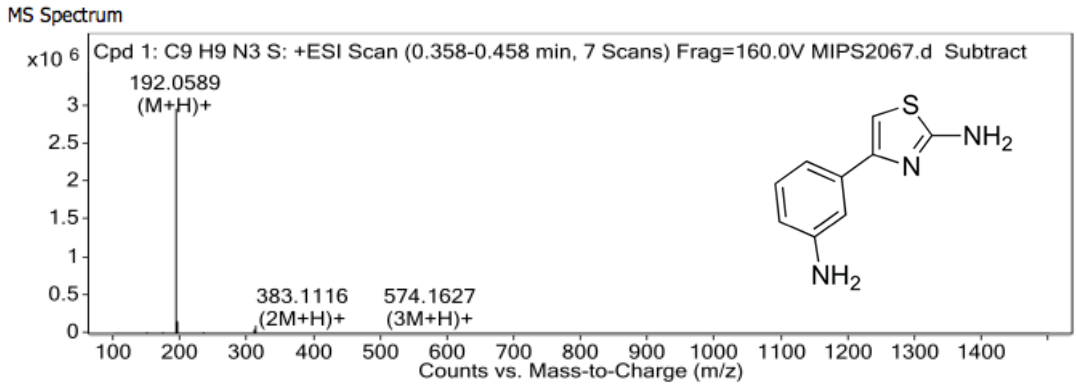
4-(2-Fluorophenyl)thiazol-2-amine (13)



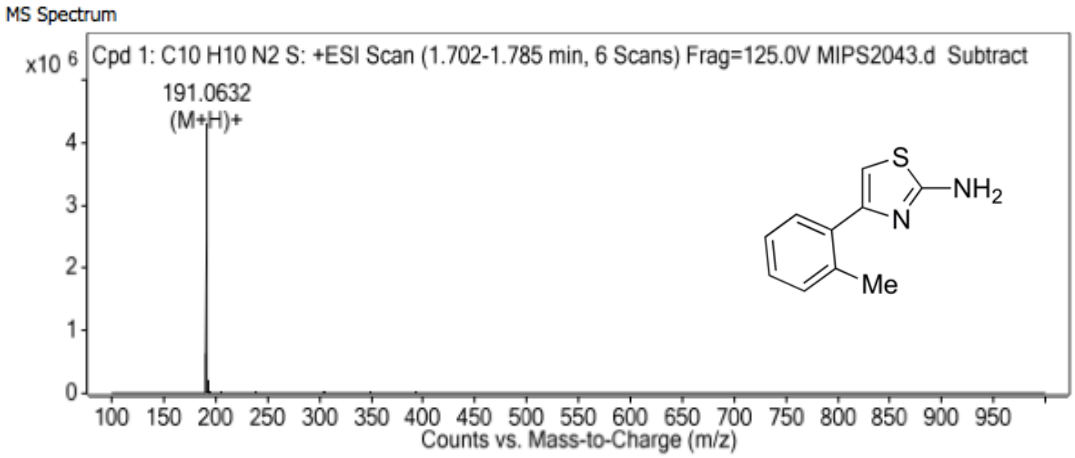
4-(3-Fluorophenyl)thiazol-2-amine (14)



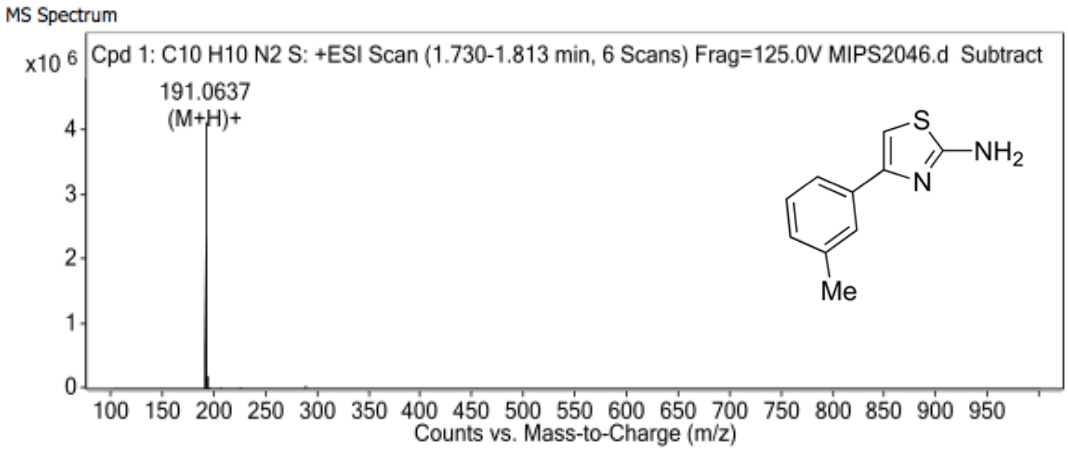
4-(3-Aminophenyl)thiazol-2-amine (15)



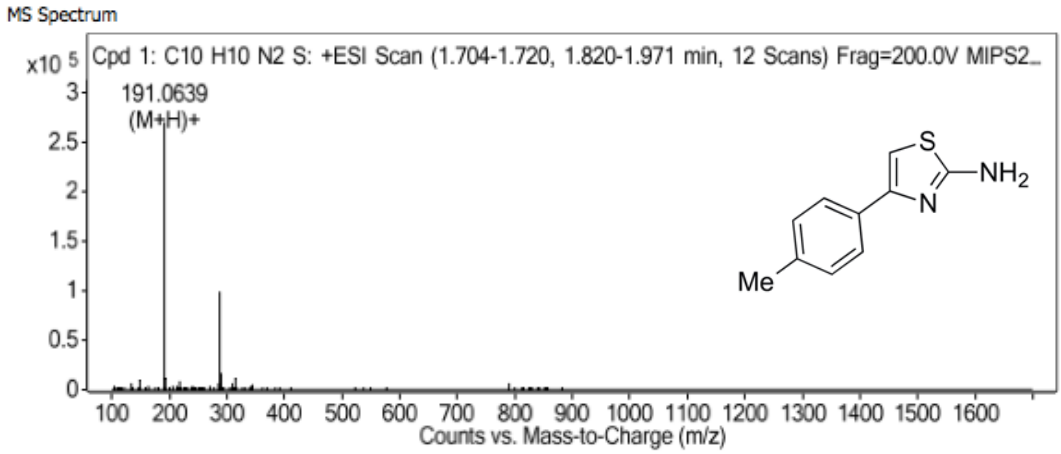
4-(o-Tolyl)thiazol-2-amine (16)



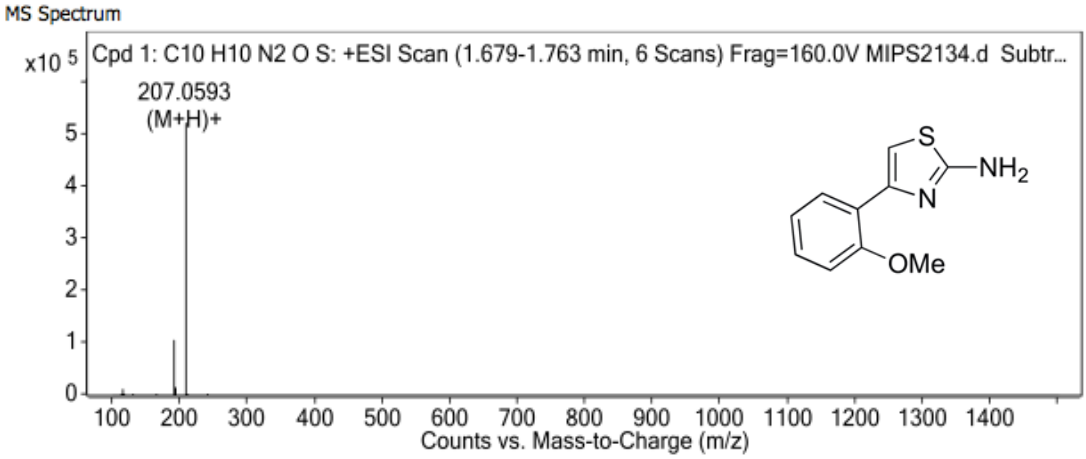
4-(*m*-Tolyl)thiazol-2-amine (17)



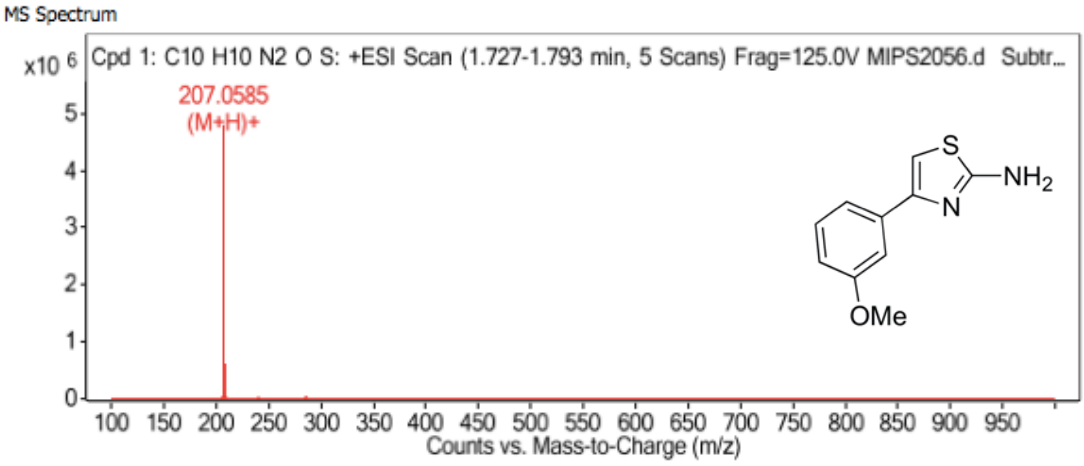
4-(*p*-Tolyl)thiazol-2-amine (18)



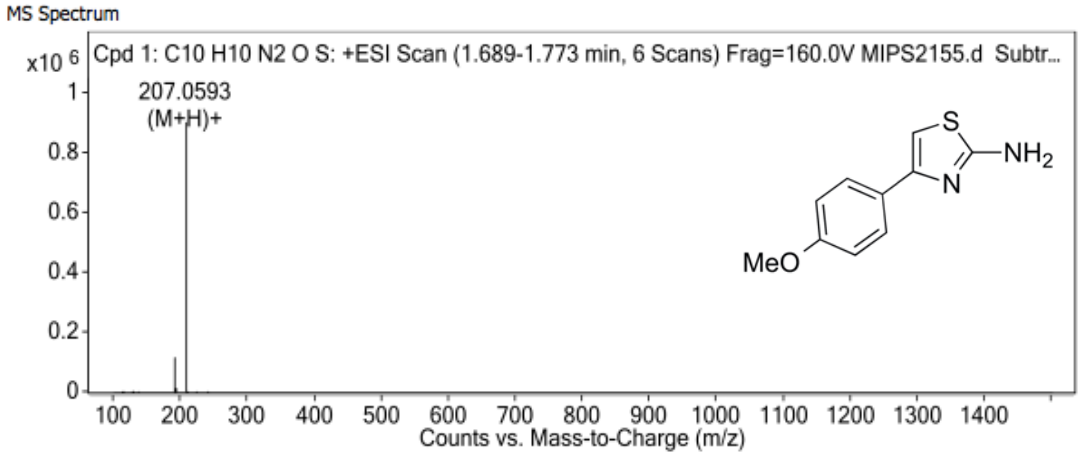
4-(2-Methoxyphenyl)thiazol-2-amine (19



4-(3-Methoxyphenyl)thiazol-2-amine (20)



4-(4-Methoxyphenyl)thiazol-2-amine (21)



Chapter 3:

Aggregation Study of Putative Inhibitors of AMA1

Published paper

Devine, S. M.; Lim, S. S.; Chandrashekaran, I. R.; MacRaild, C. A.; Drew, D. R.; Debono, C. O.; Lam, R.; Anders, R. F.; Beeson, J. G.; Scanlon, M. J.; Scammells, P. J.; Norton, R. S., A critical evaluation of pyrrolo[2,3-d]pyrimidine-4-amines as *Plasmodium falciparum* apical membrane antigen 1 (AMA1) inhibitors. *MedChemComm* **2014,** *5* (10), 1500-1506. (co-first author paper)

3.1 Declaration for Thesis Chapter 3

Declaration by candidate

In the case of Chapter 3, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Designed and performed experiments, data analysis, manuscript preparation	50

The following co-authors contributed to the work. If co-authors are students at Monash University, the extent of their contribution in percentage terms must be stated:

Name	Nature of contribution	Extent of contribution (%) for student co-authors only
Shane M. Devine	Compound synthesis, data analysis, manuscript preparation	
Indu R. Chandrashekaran	Performed experiments	
Christopher A. MacRaild	Intellectual input	
Damien R. Drew	Design experiments, data analysis, manuscript preparation	
Cael O. Debono	Performed experiments	2.5
Raymond Lam	Performed experiments	2.5
Robin F. Anders	Intellectual input	
James G. Beeson	Intellectual input, manuscript preparation	
Martin J. Scanlon	Intellectual input, manuscript preparation	
Peter J. Scammells	Intellectual input, manuscript preparation	
Raymond S. Norton	Intellectual input, manuscript preparation	

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the candidate's and co-authors' contributions to this work.

Candidate's Signature	[redacted]	**Date 5/10/2014**
Main Supervisor's Signature	[redacted]	**Date 7/10/2014**

MedChemComm

CONCISE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Med. Chem. Commun.*, 2014, **5**, 1500

A critical evaluation of pyrrolo[2,3-*d*]pyrimidine-4-amines as *Plasmodium falciparum* apical membrane antigen 1 (AMA1) inhibitors†

Shane M. Devine,‡*[a] San Sui Lim,‡[a] Indu R. Chandrashekaran,[a] Christopher A. MacRaild,[a] Damien R. Drew,[b] Cael O. Debono,[a] Raymond Lam,[a] Robin F. Anders,[c] James G. Beeson,[b] Martin J. Scanlon,[ad] Peter J. Scammells[a] and Raymond S. Norton[a]

Received 28th February 2014
Accepted 23rd July 2014

DOI: 10.1039/c4md00090k

www.rsc.org/medchemcomm

We have determined that a previously reported class of pyrrolo[2,3-*d*]pyrimidine-4-amines exhibit low binding to apical membrane antigen 1 (AMA1) and suffer from unattractive qualities, such as aggregation. We attempted to remove these traits by generating molecules with improved solubility, but this did not translate into enhanced binding affinity or inhibition of parasite growth in erythrocytes. These results indicate that anti-malarial activity is not primarily due to inhibition of AMA1 function, but mediated by an alternate or additional mechanism of action.

Introduction

Malaria is a major health concern for many of the world's most vulnerable and impoverished societies, affecting peoples throughout sub-Saharan Africa and Southeast Asia, in particular. This infectious disease continues to threaten a large number of people, with more than 40% of the global population at risk of infection and the cause of 2% of human mortalities worldwide.[1] Malaria is caused by intracellular parasites of the genus *Plasmodium*, which are transmitted to humans when a female *Anopheles* mosquito takes a blood meal. All of the symptoms of malaria arise as a result of the parasite's asexual reproductive cycle, that occurs within erythrocytes of the human host.[2] For this reason existing treatments target this blood-stage infection.

Treatment strategies for malaria have changed markedly over recent decades, as the parasite has developed resistance to previously effective drugs. The current frontline approach is that of the artemisinin-combination therapies (ACTs).[1] However, recent evidence shows that resistance to artemisinin is emerging in Southeast Asia, particularly along the Thailand–Myanmar border.[3] Although there are other potential therapeutics in the pipeline, these are not sufficiently advanced to be therapeutically relevant at this stage. Moreover, the current portfolio of candidates in development, although improving, is still lacking in diversity.[4] Therefore, there is a clear and present need for new targets for drugs to combat malaria.[5]

Mature apical membrane antigen 1 (AMA1), a 66 kDa type 1 integral membrane protein with a short well-conserved cytoplasmic region, forms a complex with parasite rhoptry neck (RON) proteins as part of a moving junction that forms between the invading parasite and the host cell.[6,7] Crystal structures of the ectodomain of AMA1 from *Plasmodium falciparum* and related species reveal three domains, including two closely-packed PAN domains.[8,9] From these PAN domains, seven flexible loops extend to surround a large hydrophobic cleft consisting of 12 well-defined and conserved residues.[8] This cleft is the site of interaction between AMA1 and RON2,[6,10] and numerous inhibitory peptides and antibodies have also been shown to target this site.[11–13] Although recent studies have used genetic knockouts to challenge previous evidence that the AMA1–RON2 interaction is essential to host cell invasion by *P. falciparum* and related parasites in cell culture,[14,15] it remains clear that diverse inhibitors of this interaction do inhibit invasion.[6,11,12,16–19] Moreover, the strong conservation of the AMA1–RON2 interaction in the Apicomplexa phylum,[7,20] even in the face of strong selective pressure from host immune systems,[21,22] implies an important functional role *in vivo*.

On this basis, we and others have proposed AMA1 as a potential drug target against malaria.[13,23,24] Several additional factors contribute to its attractiveness in this regard. Firstly, there are no human homologues of AMA1 or the RON proteins. Secondly, inhibitors of the AMA1–RON2 interaction will have their site of action in the bloodstream, thereby avoiding the

[a]*Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, 381 Royal Parade, Parkville, Victoria 3052, Australia.* [redacted]

[b]*Centre for Biomedical Research, Burnet Institute, Melbourne, Victoria 3004, Australia*

[c]*Department of Biochemistry, La Trobe University, Melbourne, Victoria 3086, Australia*

[d]*Centre of Excellence for Coherent X-Ray Science, Monash University, 381 Royal Parade, Parkville, Victoria 3052, Australia*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c4md00090k

‡ These authors contributed equally to this paper.

difficulties associated with targeting the intracellular stages of the malaria parasite. This point is particularly relevant in light of the role of drug transporters in mediating resistance to known anti-malarials.[25]

Recently, Srinivasan *et al.* reported the first example of small-molecule inhibitors of the AMA1–RON2 complex, identified *via* an AlphaScreen assay of a ~21 000 member library, utilising a truncated RON2 peptide.[26] This screen identified seven molecules, three of which blocked merozoite invasion *in vitro* with IC_{50} values in the range 21–29 μM (Fig. 1).

Re-synthesis of NCGC00015280 and chemical elaboration of this scaffold identified two related molecules NCGC00262650 and NCGC00262654 that showed enhanced inhibition with reported IC_{50} values of 9.8 μM and 6 μM, respectively, compared to the re-synthesised NCGC00015280's activity of 30 μM (Fig. 2).

Srinivasan *et al.* present an array of cell-based assays to support their proposed mode of action for these compounds. They also attempted to demonstrate a direct interaction with AMA1 by surface plasmon resonance (SPR).[26] However, these experiments were compromised by the poor solubility of the compounds, and the shape and concentration dependence of the observed sensorgrams are consistent with low-affinity super-stoichiometric interaction with the SPR biosensor surface, rather than the high-affinity stoichiometric interaction expected of a specific inhibitor. Two of the identified compounds have calculated partition co-efficients (clog *P*) that fall outside the traditional Lipinski 'rule of 5' upper limit of clog $P = 5$.[27] Another useful metric of inhibitor quality is the lipophilic ligand efficiency (LLE_{AT}), as described by Astex, which incorporates the number of heavy atoms as an indicator of lipophilicity.[28] By this metric as well, these molecules are far from the attractive value of ≥0.3 kcal mol^{-1} per heavy atom (Fig. 2). To assess the suitability of this class of compounds for further development as AMA1 inhibitors, we synthesised a panel of analogues and studied their solution behaviour and interactions with AMA1.

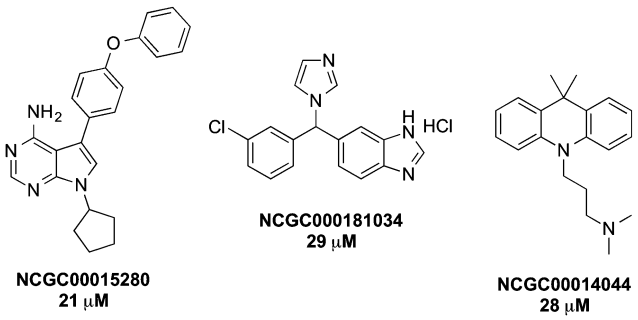


Fig. 1 Putative AMA1 inhibitors identified by Srinivasan *et al.*[26]

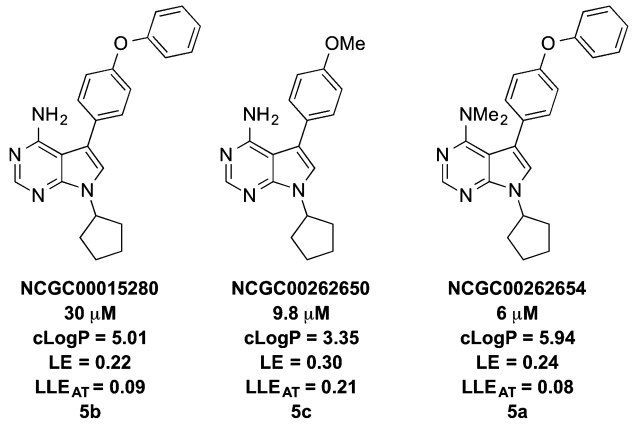


Fig. 2 Elaboration of NCGC00015280 by Srinivasan *et al.*[26]

Results

We endeavoured to evaluate the pyrrolo[2,3-*d*]pyrimidine scaffold by using our methods for monitoring AMA1 ligand binding[29] and to ameliorate the unfavourable traits, such as the high clog *P* and low LLE_{AT} values, whilst maintaining activity. Although compound lipophilicity is frequently an important factor for inhibitor potency, it can also contribute to off-target effects, leading to unwanted toxicity.[30] Small molecules that exhibit poor solubilities in aqueous buffers are often identified as promiscuous hits in a wide range of assays, complicating the interpretation of binding and/or activity data.[30–33] Understanding the physicochemical properties of a compound in an aqueous environment is therefore essential in evaluating chemical entities as lead candidates. To assess these issues for the class of compounds reported by Srinivasan *et al.*,[26] **5a–5c** were synthesised (see ESI† for details). They were then assessed at a range of concentrations (5, 10, 20, 40 and 80 μM) using 1D ^{1}H NMR spectroscopy, following the methods outlined by LaPlante *et al.*[31,32] This allowed direct observations of the aggregation propensities of these molecules in aqueous buffers.

The ^{1}H NMR spectra of **5a** and **5b** (Fig. 3A and B) show clear evidence of extensive aggregation. Both compounds gave very weak NMR signals that did not increase with concentration. As there were no visible precipitates in these samples, these observations suggest that the compounds were self-associating to form colloidal aggregates over the entire concentration range tested.[31] These large aggregates tumble more slowly than non-aggregating compounds and therefore exhibit faster NMR relaxation and signals that are broadened beyond detection.

To confirm this interpretation of these data, surfactant (Tween 20) was added into these samples. The NMR signals were consistent with those expected for these molecules, albeit with significant residual line broadening (top panel, Fig. 3A and B). These observations may be explained by the fact that surfactant is capable of dissociating large assemblies into smaller entities with better relaxation properties for NMR detection.[32] In contrast to **5a** and **5b**, **5c** gave sharp proton signals with peak intensities that increased with concentration over the range tested (Fig. 3C).

However, the concentration dependence of the peak intensities is not linear (Fig. 3D), while chemical shifts for a number of resonances showed a weak concentration dependence (Fig. 3E), indicating that compound **5c** also aggregates in aqueous solution, albeit to a much lesser extent than do **5a** and **5b**. The signals of compound **5c** were shifted and broadened in

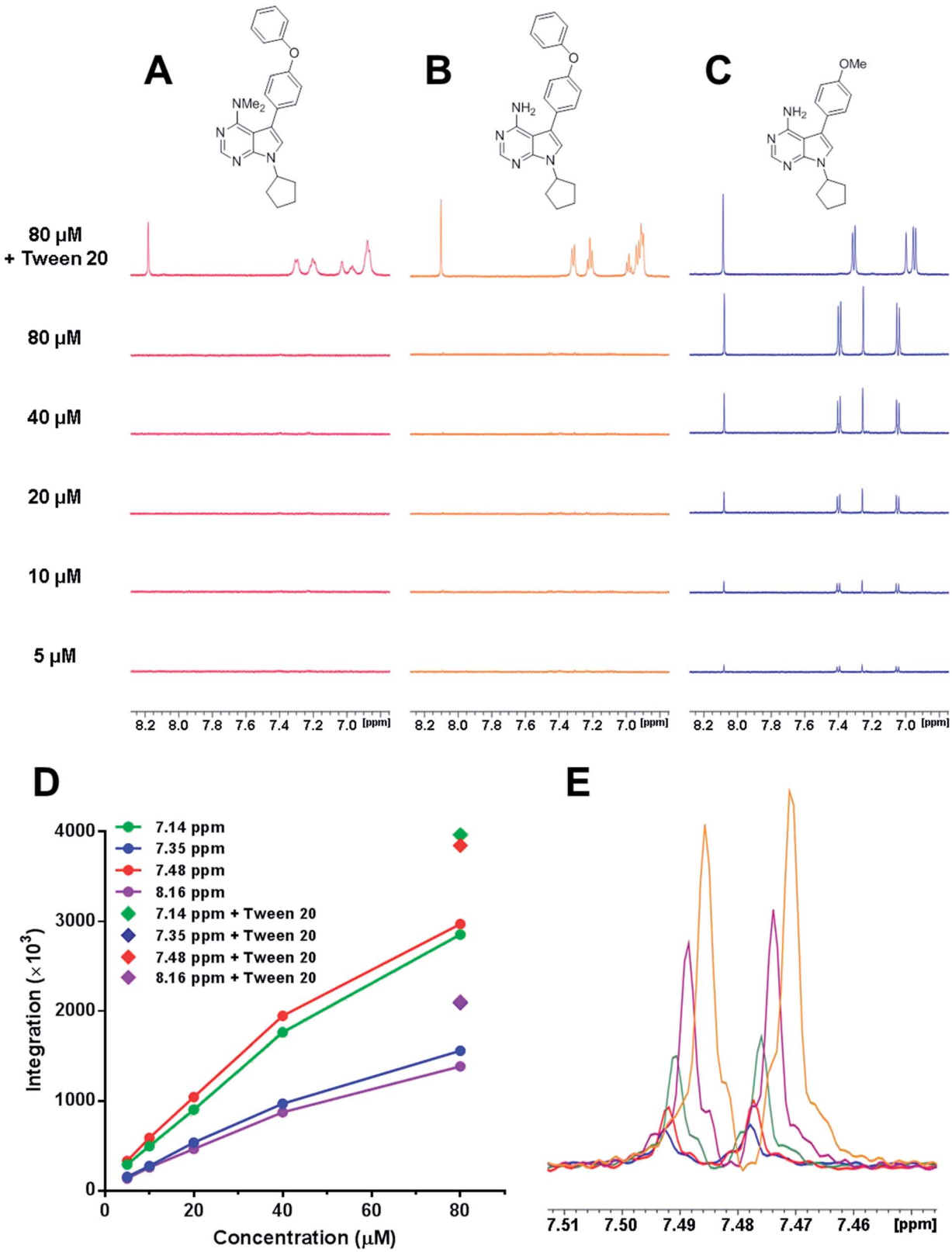


Fig. 3 1D ^{1}H NMR spectra of (A) **5a**, (B) **5b** and (C) **5c** at different concentrations and in the presence of 0.05% Tween 20. (D) Peak intensities of **5c** proton signals (7.14, 7.35, 7.48 and 8.16 ppm) at increasing compound concentrations and with the addition of surfactant. (E) Concentration dependence of the chemical shifts of **5c** at 5, 10, 20, 40 and 80 μM, respectively.

the presence of Tween 20, presumably as a result of the interactions between the compound and surfactant. Furthermore, peak intensities increased in the presence of the surfactant (Fig. 3D), again indicating the presence of large aggregates that are disrupted by the addition of Tween 20 into the sample.

We then assessed the interaction of these compounds with AMA1 using techniques developed in our own search for inhibitors of the AMA1–RON2 interaction.[29] The aggregation behaviour of **5a** and **5b**, described above, precluded the use of ligand-detected NMR strategies, so only **5c** was evaluated in the Carr–Purcell–Meiboom–Gill (CPMG) binding assay (Fig. 4). This method utilises the CPMG spin-lock filter to eliminate signals of the rapidly relaxing AMA1 protein and bound ligands, with the signals of free compounds less affected.[34] CPMG spectra for detergent-free samples containing 80 μM **5c** in the presence and absence of 10 μM 3D7 *Pf*AMA1 were first acquired with 0 and 200 ms spin-relaxation filters. Thereafter, R1 and RON2 peptides were added into the samples and the same set of CPMG spectra was acquired to evaluate the binding activities of **5c** at the AMA1 hydrophobic cleft. The signal intensities and chemical shifts of a number of peaks for **5c** were affected by the presence of AMA1 (Fig. 4A). These effects are exemplified using the proton resonance of **5c** at around 7.35 ppm in Fig. 4B. The signals were broadened and shifted slightly upfield in the presence of AMA1 and, when a 200 ms spin-lock filter was applied, the signal in the presence of AMA1 relaxed significantly faster than the signal in the absence of protein (R_2 relaxation rates of 8.2 and 1.4 s^{-1}, respectively). These results are indicative of **5c** binding to AMA1.

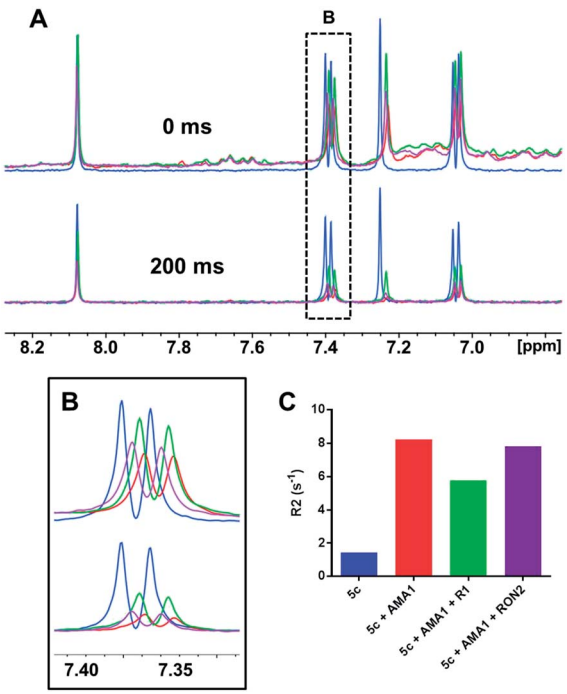
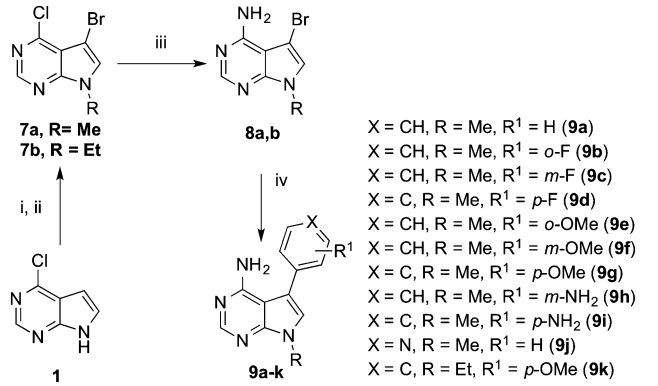


Fig. 4 CPMG binding assay. (A and B) All CPMG spectra were acquired with both 0 (top spectra) and 200 ms (bottom spectra) spin-relaxation filters in the absence of detergent. Blue and red spectra were results for samples containing **5c** in the presence or absence of 3D7 *Pf*AMA1, respectively. Green and purple spectra correspond to samples containing **5c** and 3D7 *Pf*AMA1 with the additions of R1 and RON2 peptides, respectively. (C) Transverse relaxation rate (R2) for different samples used in the binding studies.

When R1 and RON2 peptides were added to the samples containing both AMA1 and **5c**, the compound peaks shifted slightly downfield towards the chemical shift of the free compound. Also, partial restoration of the free compound line shape and relaxation rate was observed in both samples with the peptides added. These observations are consistent with competition between **5c** and R1 and RON2 peptides for binding sites on AMA1. The presence of the competing peptides reduced the fraction of bound **5c**, which in turn decreased the average transverse relaxation rate of the compound, and thus increased **5c** signals after the CPMG period (Fig. 4B and C). Similar results were observed when the same set of experiments was conducted in the presence of Tween 20 (Fig. S1†). This result confirms that **5c** in its monomeric form is able to bind AMA1.

Finally, we interrogated these compounds, including their synthetic precursors, by SPR, to estimate the binding affinities to AMA1 using methods described previously.[29] Not unexpectedly, the precursor molecule fragments **1–3**, showed minimal interaction with AMA1 at concentrations up to 200 μM. Once the 4-amino or 4-dimethylamino group was introduced, we saw evidence of super-stoichiometric binding, in the form of responses that exceeded the maximal response expected for these compounds and that failed to saturate, even at the highest concentrations studied. These issues were particularly acute in the case of **5a** and **5b**, and precluded any estimate of the affinity of these molecules with AMA1. For **5c**, we observed unambiguous over-binding only at higher concentrations. However, the response observed at lower concentrations is inconsistent with an affinity for AMA1 tighter than ~1 mM, while the IC_{50} for this compound reported by Srinivasan *et al.* is 9.8 μM.[26] This discrepancy suggests that this series of compounds exert their effects on host cell invasion by *P. falciparum* merozoites by some mechanism other than direct inhibition of AMA1, or by an additional mechanism of action.

To address the shortcomings of the three molecules, as outlined above, a series of pyrrolo[2,3-*d*]pyrimidine-4-amines was synthesised in an attempt to improve solubility whilst maintaining or improving affinity for AMA1 (Scheme 1). This was achieved by alkylation, bromination, amination and Suzuki coupling. We envisaged that replacement of the 7-cyclopentyl with a methyl group should aid in solubility, and we also explored substitutions on the 5-aryl group while retaining the 4-amino group. This scaffold featuring the *N*-methyl substitution is the core of a known protein kinase R (PKR)-like endoplasmic reticulum kinase (PERK) inhibitor that has been selected as a preclinical candidate for tumour inhibition.[35,36] The compounds produced had clog *P* values of 0.93–2.35, which, in contrast to the clog *P* value of 3.35 for **5c**, provided soluble compounds for study. Aggregation propensities of **9a–k** were evaluated using 1D ^{1}H NMR spectroscopy (Fig. S2†) as described earlier for **5a–c**. The 7-methyl and 7-ethyl series exhibited substantially reduced aggregation behaviour, when compared to the 7-cyclopentyl series (**5a–c**), with or without the addition of Tween 20, demonstrating our success in eliminating some of the unwanted behaviour of **5a–c**.



Scheme 1 Reagents and conditions: (i) NaH, DMF, 0 °C, 30 min for **6a**, MeI and for **6b**, EtI, 60 °C, 4 h; (ii) NBS, DCM, 25 °C; (iii) NH_4OH, *i*-PrOH, 100 °C, 40 h; (iv) R^1-$B(OH)_2$, $PdCl_2(PPh_3)_2$, THF/1 M Na_2CO_3, 100 °C, 2 h.

All pyrrolo[2,3-*d*]pyrimidines (**5a–5c** and **9a–9k**) were analysed for binding to AMA1 using SPR, utilising similar conditions to those employed by Srinivasan *et al.*[26] In general, the sensorgrams showed weak binding affinities, with K_D values ≥1 mM. There were no visible precipitates in the samples and all compounds were soluble in the buffer conditions employed. From this small library of compounds, however, no meaningful structure–activity relationships (SAR) could be inferred and no increased affinity for AMA1 was observed, suggesting that these compounds are not potent inhibitors of AMA1 interactions. Finally, **9a–9k** were tested to ascertain their ability to inhibit parasite growth in erythrocytes of *P. falciparum* (Fig. S3†).[37] Compounds **5c** and **9a–k** showed substantial inhibitory activity in this assay, although this activity did not correlate with their affinity for AMA1 by SPR (Fig. S3†) . **5a** and **5b** demonstrated IC_{50} values > 1 mM, compared to the reported values of 6 and 30 μM, respectively.[26] The IC_{50} value for **5c** was 63 μM, in comparison to the reported value of 9.8 μM and the best analogue, **9c**, had an IC_{50} value < 31 μM. This represents still further evidence in support of an alternative mechanism of anti-malarial action by these compounds.

Discussion

Our results suggest that these compounds in their current form are not suitable for development as AMA1 inhibitors given their apparent low binding affinity to AMA1, sub-optimal potency, relative insolubility and tendency to form aggregates. Our findings suggest that these compounds may exhibit an alternative or additional mechanism of anti-malarial action. Compound **5b** was identified initially as a Src-family kinase inhibitor[38] and molecules containing the pyrrolo[2,3-*d*]pyrimidine scaffold have been shown to be involved in a number of therapeutically relevant areas such as Huntington's disease[39] and acute myeloid leukaemia,[40] where they are believed to act *via* a Src-family kinase mechanism. Srinivasan *et al.* demonstrated that a structurally different Src-inhibitor did not block invasion.[26] However, this does not preclude the possibility of the pyrrolo[2,3-*d*]pyrimidine inhibitors acting as kinase inhibitors in this context. A structurally similar framework, incorporating an extra nitrogen at the 6-position, namely the pyrazolo[3,4-*d*]-pyrimidin-4-amine scaffold with comparable N1- and 3-position modifications, has been implicated in a related apicomplexan species, *Toxoplasma gondii*, as a calcium-dependent protein kinase 1 (CDPK1) inhibitor.[41,42] CDPK is known to control microneme secretion and consequently block invasion in *T. gondii*, as well as *Plasmodium falciparum*.[43] Therefore, it is conceivable that these pyrrolo[2,3-*d*]pyrimidines could be acting *via* a CDPK mechanism, or in addition to AMA1-mediated inhibition. Colloidal aggregation of compounds is well known to contribute to promiscuity in high-throughput screens and in a range of other assays.[30–33,44] The observed aggregation properties of **5a** and **5b**, may be responsible for their inhibitory effects,[26] rather than being specific inhibitors of the AMA1–RON2 interaction.[44] Compound **5c** seems to be genuinely binding AMA1, but with insufficient affinity (using recombinant AMA1) to explain its reported activity. It remains possible that the binding affinity to native AMA1 is different from that measured using recombinant AMA1, although we note that these two preparations afford essentially identical results for genuine AMA1 inhibitors such as the peptide R1. Our attempts to produce more soluble analogues (**9a–k**) with increased activity did not generate any meaningful SAR.

Conclusion

Candidate pyrrolo[2,3-*d*]pyrimidine-4-amines (**5a–c**) appear to be moderately potent inhibitors of erythrocyte invasion by *P. falciparum* merozoites that prevent the formation of both the moving junction and the AMA1–RON complex.[26] These effects may not be mediated primarily by direct stoichiometric interaction with AMA1; inhibitory activity could include off-target mechanisms, which may be related to their tendency to form colloidal aggregates in aqueous solution, and/or their activity as kinase inhibitors. Molecules with lower clog *P* values were synthesised (**9a–k**), but failed to achieve greater inhibitory activity. Furthermore, this series of compounds carries significant physicochemical liabilities that are likely to impede their further development.

Acknowledgements

The authors would like to thank Brad Sleebs for fruitful discussions on this project, Samuel Hughes for preliminary work on the synthesis, and Jason Dang for obtaining HRMS data. This work was supported in part by an Australian National Health and Medical Research Council (NHMRC) project grant (1025150) and program grant (637406). R.S.N. and J.G.B acknowledge fellowship support from the NHMRC. The Burnet Institute is supported by the NHMRC Independent Research Institutes Infrastructure Support Scheme, and Operational Infrastructure Support from the State Government of Victoria, Australia.

References

1 World Health Organization World Malaria Report, 2013, pp. 1–284.
2 L. H. Miller, D. I. Baruch, K. Marsh and O. K. Doumbo, *Nature*, 2002, **415**, 673–679.
3 A. P. Phyo, S. Nkhoma, K. Stepniewska, E. A. Ashley, S. Nair, R. McGready, C. ler Moo, S. Al-Saai, A. M. Dondorp, K. M. Lwin, P. Singhasivanon, N. P. J. Day, N. J. White, T. J. C. Anderson and F. Nosten, *Lancet*, 2012, **379**, 1960–1966.
4 M. P. Anthony, J. N. Burrows, S. Duparc, J. J. Moehrle and T. N. C. Wells, *Malar. J.*, 2012, **11**, 316.
5 M. A. Biamonte, J. Wanner and K. G. Le Roch, *Bioorg. Med. Chem. Lett.*, 2013, **23**, 2829–2843.
6 D. Richard, C. A. MacRaild, D. T. Riglar, J.-A. Chan, M. Foley, J. Baum, S. A. Ralph, R. S. Norton and A. F. Cowman, *J. Biol. Chem.*, 2010, **285**, 14815–14822.
7 M. Lamarque, S. Besteiro, J. Papoin, M. Roques, B. Vulliez-Le Normand, J. Morlon-Guyot, J.-F. Dubremetz, S. Fauquenoy, S. Tomavo, B. W. Faber, C. H. Kocken, A. W. Thomas, M. J. Boulanger, G. A. Bentley and M. Lebrun, *PLoS Pathog.*, 2011, **7**, e1001276.
8 T. Bai, M. Becker, A. Gupta, P. Strike, V. J. Murphy, R. F. Anders and A. H. Batchelor, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 12736–12741.
9 J. C. Pizarro, B. Vulliez-Le Normand, M.-L. Chesne-Seck, C. R. Collins, C. Withers-Martinez, F. Hackett, M. J. Blackman, B. W. Faber, E. J. Remarque, C. H. M. Kocken, A. W. Thomas and G. A. Bentley, *Science*, 2005, **308**, 408–411.
10 M. L. Tonkin, M. Roques, M. H. Lamarque, M. Pugnière, D. Douguet, J. Crawford, M. Lebrun and M. J. Boulanger, *Science*, 2011, **333**, 463–467.
11 S. Dutta, L. S. Dlugosz, D. R. Drew, X. Ge, D. Ababacar, Y. I. Rovira, J. K. Moch, M. Shi, C. A. Long, M. Foley, J. G. Beeson, R. F. Anders, K. Miura, J. D. Haynes and A. H. Batchelor, *PLoS Pathog.*, 2013, **9**, e1003840.
12 K. S. Harris, J. L. Casey, A. M. Coley, R. Masciantonio, J. K. Sabo, D. W. Keizer, E. F. Lee, A. McMahon, R. F. Anders and M. Foley, *Infect. Immun.*, 2005, **73**, 6981–6989.
13 B. Vulliez-Le Normand, M. L. Tonkin, M. H. Lamarque, S. Langer, S. Hoos, M. Roques, F. A. Saul, B. W. Faber, G. A. Bentley, M. J. Boulanger and M. Lebrun, *PLoS Pathog.*, 2012, **8**, e1002755.
14 D. Y. Bargieri, N. Andenmatten, V. Lagal, S. Thiberge, J. A. Whitelaw, I. Tardieux, M. Meissner and R. Ménard, *Nat. Commun.*, 2013, **4**, 2552.
15 D. Giovannini, S. Späth, C. Lacroix, A. Perazzi, D. Bargieri, V. Lagal, C. Lebugle, A. Combe, S. Thiberge, P. Baldacci, I. Tardieux and R. Ménard, *Cell Host Microbe*, 2011, **10**, 591–602.
16 M. J. Boyle, D. W. Wilson, J. S. Richards, D. T. Riglar, K. K. A. Tetteh, D. J. Conway, S. A. Ralph, J. Baum and J. G. Beeson, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 14378–14383.
17 K. S. Harris, J. L. Casey, A. M. Coley, J. A. Karas, J. K. Sabo, Y. Y. Tan, R. S. Norton, A. B. Hughes, D. Scanlon and M. Foley, *J. Biol. Chem.*, 2009, **284**, 9361–9371.
18 K. A. Henderson, V. A. Streltsov, A. M. Coley, O. Dolezal, P. J. Hudson, A. H. Batchelor, A. Gupta, T. Bai, V. J. Murphy, R. F. Anders, M. Foley and S. D. Nuttall, *Structure*, 2007, **15**, 1452–1466.
19 A. Yap, M. F. Azevedo, P. R. Gilson, G. E. Weiss, M. T. O'Neill, D. W. Wilson, B. S. Crabb and A. F. Cowman, *Cell. Microbiol.*, 2014, **16**, 642–656.
20 N. I. Proellocks, R. L. Coppel and K. L. Waller, *Trends Parasitol.*, 2010, **26**, 297–304.
21 A. Cortés, M. Mellombo, I. Mueller, A. Benet, J. C. Reeder and R. F. Anders, *Infect. Immun.*, 2003, **71**, 1416–1426.
22 S. D. Polley, W. Chokejindachai and D. J. Conway, *Genetics*, 2003, **165**, 555–561.
23 C. A. MacRaild, R. F. Anders, M. Foley and R. S. Norton, *Curr. Top. Med. Chem.*, 2011, **11**, 2039–2047.
24 G. Santos and N. V. Torres, *PLoS One*, 2013, **8**, e59968.
25 C. P. Sanchez, A. Dave, W. D. Stein and M. Lanzer, *Int. J. Parasitol.*, 2010, **40**, 1109–1118.
26 P. Srinivasan, A. Yasgar, D. K. Luci, W. L. Beatty, X. Hu, J. Andersen, D. L. Narum, J. K. Moch, H. Sun, J. D. Haynes, D. J. Maloney, A. Jadhav, A. Simeonov and L. H. Miller, *Nat. Commun.*, 2013, **4**, 2261.
27 C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Delivery Rev.*, 1997, **23**, 3–25.
28 P. N. Mortenson and C. W. Murray, *J. Comput.-Aided Mol. Des.*, 2011, **25**, 663–667.
29 S. S. Lim, C. O. Debono, C. A. MacRaild, I. R. Chandrashekaran, O. Dolezal, R. F. Anders, S. M. Devine, P. J. Scammells and R. S. Norton, *Aust. J. Chem.*, 2013, **66**, 1530.
30 B. J. Davis and D. A. Erlanson, *Bioorg. Med. Chem. Lett.*, 2013, **23**, 2844–2852.
31 S. R. LaPlante, R. Carson, J. Gillard, N. Aubry, R. Coulombe, S. Bordeleau, P. Bonneau, M. Little, J. O'Meara and P. L. Beaulieu, *J. Med. Chem.*, 2013, **56**, 5142–5150.
32 S. R. LaPlante, N. Aubry, G. Bolger, P. Bonneau, R. Carson, R. Coulombe, C. Sturino and P. L. Beaulieu, *J. Med. Chem.*, 2013, **56**, 7073–7083.
33 A. Vom, S. Headey, G. Wang, B. Capuano, E. Yuriev, M. J. Scanlon and J. S. Simpson, *Aust. J. Chem.*, 2013, **66**, 1518.
34 P. J. Hajduk, G. Sheppard, D. G. Nettesheim, E. T. Olejniczak, S. B. Shuker, R. P. Meadows, D. H. Steinman, G. M. Carrera, P. A. Marcotte, J. Severin, K. Walter, H. Smith, E. Gubbins, R. Simmer, T. F. Holzman, D. W. Morgan, S. K. Davidsen, J. B. Summers and S. W. Fesik, *J. Am. Chem. Soc.*, 1997, **119**, 5818–5827.
35 J. M. Axten, J. R. Medina, Y. Feng, A. Shu, S. P. Romeril, S. W. Grant, W. H. H. Li, D. A. Heerding, E. Minthorn, T. Mencken, C. Atkins, Q. Liu, S. Rabindran, R. Kumar, X. Hong, A. Goetz, T. Stanley, J. D. Taylor, S. D. Sigethy, G. H. Tomberlin, A. M. Hassell, K. M. Kahler,

L. M. Shewchuk and R. T. Gampe, *J. Med. Chem.*, 2012, **55**, 7193–7207.

36 J. M. Axten, S. P. Romeril, A. Shu, J. Ralph, J. R. Medina, Y. Feng, W. H. H. Li, S. W. Grant, D. A. Heerding, E. Minthorn, T. Mencken, N. Gaul, A. Goetz, T. Stanley, A. M. Hassell, R. T. Gampe, C. Atkins and R. Kumar, *ACS Med. Chem. Lett.*, 2013, **4**, 964–968.

37 D. R. Drew, A. N. Hodder, D. W. Wilson, M. Foley, I. Mueller, P. M. Siba, A. E. Dent, A. F. Cowman and J. G. Beeson, *PLoS One*, 2012, **7**, e51023.

38 L. D. Arnold, D. J. Calderwood, R. W. Dixon, D. N. Johnston, J. S. Kamens, R. Munschauer, P. Rafferty and S. E. Ratnofsky, *Bioorg. Med. Chem. Lett.*, 2000, **10**, 2167–2170.

39 N. Zhang, B. Li, I. Al-Ramahi, X. Cong, J. M. Held, E. Kim, J. Botas, B. W. Gibson and L. M. Ellerby, *J. Biol. Chem.*, 2012, **287**, 21204–21213.

40 Y. Saito, H. Yuki, M. Kuratani, Y. Hashizume, S. Takagi, T. Honma, A. Tanaka, M. Shirouzu, J. Mikuni, N. Handa, I. Ogahara, A. Sone, Y. Najima, Y. Tomabechi, M. Wakiyama, N. Uchida, M. Tomizawa-Murasawa, A. Kaneko, S. Tanaka, N. Suzuki, H. Kajita, Y. Aoki, O. Ohara, L. D. Shultz, T. Fukami, T. Goto, S. Taniguchi, S. Yokoyama and F. Ishikawa, *Sci. Transl. Med.*, 2013, **5**, 181.

41 R. C. Murphy, K. K. Ojo, E. T. Larson, A. Castellanos-Gonzalez, B. G. K. Perera, K. R. Keyloun, J. E. Kim, J. G. Bhandari, N. R. Muller, C. L. M. J. Verlinde, A. C. White Jr, E. A. Merritt, W. C. Van Voorhis and D. J. Maly, *ACS Med. Chem. Lett.*, 2010, **1**, 331–335.

42 E. T. Larson, K. K. Ojo, R. C. Murphy, S. M. Johnson, Z. Zhang, J. E. Kim, D. J. Leibly, A. M. W. Fox, M. C. Reid, E. J. Dale, B. G. K. Perera, J. Kim, S. N. Hewitt, W. G. J. Hol, C. L. M. J. Verlinde, E. Fan, W. C. Van Voorhis, D. J. Maly and E. A. Merritt, *J. Med. Chem.*, 2012, **55**, 2803–2810.

43 A. Bansal, S. Singh, K. R. More, D. Hans, K. Nangalia, M. Yogavel, A. Sharma and C. E. Chitnis, *J. Biol. Chem.*, 2013, **288**, 1590–1602.

44 J. Seidler, S. L. McGovern, T. N. Doman and B. K. Shoichet, *J. Med. Chem.*, 2003, **46**, 4477–4486.

Supplementary Information

A Critical Evaluation of Pyrrolo[2,3-*d*]pyrimidine-4-amines as *Plasmodium falciparum* Apical Membrane Antigen 1 (AMA1) Inhibitors

Shane M. Devine,[*‡a] San Sui Lim,[‡a] Indu R. Chandrashekaran,[a] Christopher A. MacRaild,[a] Damien R. Drew,[b] Cael O. Debono,[a] Raymond Lam,[a] Robin F. Anders,[c] James G. Beeson,[b] Martin J. Scanlon,[a,d] Peter J. Scammells,[a] Raymond S. Norton[a]

[a] Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, VIC 3052, Australia

[b] Centre for Biomedical Research, Burnet Institute, Melbourne, Victoria 3004, Australia

[c] Department of Biochemistry, La Trobe University, Melbourne, VIC 3086, Australia

[d] Centre of Excellence for Coherent X-Ray Science, Monash University, Parkville, VIC 3052,

* Corresponding author. [redacted]

‡ These authors contributed equally to this paper.

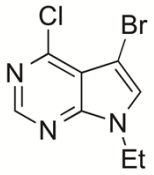
Contents

General Experimental

NMR spectra (^{1}H, ^{19}F, ^{13}C) were recorded on a Bruker Avance Nanobay III 400 MHz Ultrashield Plus spectrometer at 400.13, 376.50 and 100.62 MHz, respectively coupled to a BACS 60 automatic sample changer at 25 °C. Chemical shifts (δ) are recorded in parts per million (ppm) by correction with reference to the chemical shift of the solvent, according to the procedure described by Gottlieb.[45] Coupling constants (*J*) are recorded in Hz, and the significant multiplicities described by singlet (s), doublet (d), triplet (t), quadruplet (q), broad (br), multiplet (m), doublet of doublets (dd), and doublet of triplets (dt). LC-MS were run to verify reaction outcome and purity using an Agilent 6100 series Single Quad coupled to an Agilent 1200 series HPLC. The following buffers were used: buffer A, 0.1% formic acid in H_2O; buffer B, 0.1% formic acid in MeCN. The following gradient was used with a Phenomenex Luna 3 µM C8(2) 15 mm × 4.6 mm column, and a flow rate of 0.5 mL/min and total run time of 12 min; 0–4 min 95% buffer A and 5% buffer B, 4–7 min 0% buffer A and 100% buffer B, 7–12 min 95% buffer A and 5% buffer B. Mass spectra were acquired in positive and negative ion mode with a scan range of 0–1000 *m*/*z* at 5 V. UV detection was carried out at 254 nm. All compounds were of >95% purity. Thin layer chromatography was conducted on 0.2 mm plates using Merck silica gel 60 F_{254}. Column chromatography was achieved using Merck silica gel 60 (particle size 0.063–0.200 µm, 70–230 mesh). Calculated partition co-efficient values (cLogP) were calculated using ChemAxon's Instant JChem program. Instant JChem was used for structure database management, search and prediction (Instant JChem 5.9.4, 2013, ChemAxon; http://www.chemaxon.com). 4-Chloro-7*H*-pyrrolo[2,3-*d*]pyrimidine (**1**) was purchased from Astatech and used to synthesise 5-bromo-4-chloro-7-cyclopentyl-7*H*-pyrrolo[2,3-*d*]pyrimidine (**3**)[26] and 5-bromo-4-chloro-7-methyl-7*H*-pyrrolo[2,3-*d*]pyrimidine (**7a**)[35] via literature procedures.

Synthetic procedures

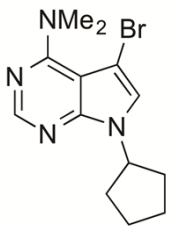
5-Bromo-4-chloro-7-ethyl-7*H*-pyrrolo[2,3-*d*]pyrimidine (7b)



To a solution of 4-chloro-7*H*-pyrrolo[2,3-*d*]pyrimidine (**1**) (1.00 g, 6.51 mmol) in DMF (8 mL) at 0 °C, was added NaH, as a 60% dispersion in mineral oil (389 mg, 9.77 mmol), portionwise in a teflon-capped sealed tube. The mixture was stirred at 0 °C for 30 min, whereby pressure builds up. The mixture was briefly degassed and EtI (1.04 mL, 13.02 mmol) was added and the whole heated at 60 °C for 5 h. The mixture was extracted with EtOAc (100 mL) and washed with H_2O (100 mL). The aqueous fraction was re-extracted with EtOAc (100 mL). The organic fractions were combined and washed with H_2O (5 × 50 mL), then sat. NaCl (10 mL). The organic phase was dried with $MgSO_4$, filtered and the filtrate evaporated under reduced pressure to give a brown gum (1.31 g). The residue was purified by column chromatography (petroleum spirits/EtOAc, 9:1) to give 4-chloro-7-ethyl-7*H*-pyrrolo[2,3-*d*]pyrimidine (**6b**) (994 mg, 84%) as a colourless oil. ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.54 (s, 1H), 7.19 (d, *J* = 3.6 Hz, 1H), 6.49 (d, *J* = 3.6 Hz, 1H), 4.24 (q, *J* = 7.3 Hz, 2H), 1.40 (t, *J* = 7.3 Hz, 3H). ^{13}C NMR (101 MHz, $CDCl_3$): δ 152.0 (C), 150.6 (C), 150.4 (CH), 128.7 (CH), 117.5 (C), 99.3 (CH), 40.0 (CH_2), 15.5 (CH_3).

To a solution of 4-chloro-7-ethyl-7*H*-pyrrolo[2,3-*d*]pyrimidine (**6b**) (510 mg, 2.82 mmol) in DCM (10 mL) was added *N*-bromosuccinimide (551 mg, 3.09 mmol) and the whole was stirred at 25 °C for 16 h. At this time the solution was evaporated under reduced pressure and H_2O (100 mL) added. The precipitate which formed was filtered, washed with H_2O and dried to give 5-bromo-4-chloro-7-ethyl-7*H*-pyrrolo[2,3-*d*]pyrimidine (**7b**) (695 mg, 95%) as a light pink solid. ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.61 (s, 1H), 7.31 (s, 1H), 4.31 (q, *J* = 7.3 Hz, 2H), 1.47 (t, *J* = 7.3 Hz, 3H). ^{13}C NMR (101 MHz, $CDCl_3$): δ 152.2 (C), 151.0 (CH), 150.1 (C), 128.7 (CH), 115.0 (C), 87.5 (C), 40.4 (CH_2), 15.5 (CH_3).

5-Bromo-7-cyclopentyl-7*H*-pyrrolo[2,3-*d*]pyrimidine-4-dimethylamine (4a)

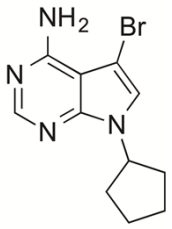


3 (400 mg, 1.33 mmol) was dissolved in *i*-PrOH (1.5 mL) and Me_2NH (as a 33% solution in EtOH) (300 µL) was added and the solution heated at 100 °C in a sealed tube for 16 h. H_2O (50 mL) was added to and the mixture extracted with EtOAc (2 × 20 mL), dried with $MgSO_4$, filtered and the filtrate evaporated under reduced pressure. The residue was purified by column chromatography (petroleum spirits/EtOAc, 1:1) to give **4a** (327 mg, 79%) as a pale yellow solid. ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.34 (s, 1H), 7.11 (s, 1H), 5.21 – 5.14 (m, 1H), 3.27 (s, 6H), 2.23 – 2.19 (m, 2H), 1.88 – 1.74 (m,

6H). ^{13}C NMR (101 MHz, $CDCl_3$): δ 159.5 (C), 150.8 (C), 150.3 (CH), 122.4 (CH), 104.2 (C), 87.2 (C), 55.5 (CH), 42.6 (2 × CH_3), 32.9 (2 × CH_2), 24.2 (2 × CH_2).

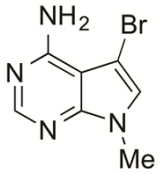
General procedure for formation of 5-bromo-7-substituted-7*H*-pyrrolo[2,3-*d*]pyrimidine-4-amines, (**4b**, **8a** and **8b**): Concentrated NH_4OH (5 mL) and *i*-PrOH (100 μL) were added to the required chloro intermediate (**3**, **7a** or **7b**) (500 mg) and the mixture was heated at 100 °C in a sealed tube for 40 h. At this time, H_2O (50 mL) was added and the reaction was either filtered and washed with H_2O to give **4b** or extracted with EtOAc (2 × 20 mL), dried with $MgSO_4$, filtered and the filtrate evaporated under reduced pressure to give **8a** and **8b**. These were then purified by column chromatography (petroleum spirits/EtOAc, 1:1).

5-Bromo-7-cyclopentyl-7*H*-pyrrolo[2,3-*d*]pyrimidine-4-amine (4b)



3 (500 mg, 1.66 mmol) gave **4b** (315 mg, 67%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.25 (s, 1H), 7.02 (s, 1H), 5.94 (br s, 2H), 5.17 – 5.10 (m, 1H), 2.24 – 2.19 (m, 2H), 1.88 – 1.75 (m, 6H). ^{13}C NMR (101 MHz, $CDCl_3$): δ 156.6 (C), 151.7 (CH), 149.4 (C), 121.2 (CH), 102.0 (C), 86.4 (C), 55.6 (CH), 33.1 (2 × CH_2), 24.2 (2 × CH_2).

5-Bromo-7-methyl-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine (8a)[35]



7a (900 mg, 3.65 mmol) gave **8a** (576 mg, 69%).
^{1}H NMR (400 MHz, $CDCl_3$): δ 8.27 (s, 1H), 6.91 (s, 1H), 5.92 (br s, 2H), 3.75 (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$): 157.0 (C), 152.8 (CH), 149.8 (C), 124.7 (CH), 102.0 (C), 86.1 (C), 31.5 (CH_3).

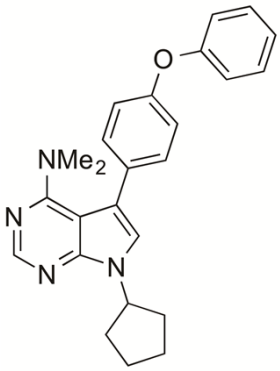
5-Bromo-7-ethyl-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine (8b)

7b (500 mg, 1.91 mmol) gave **8b** (365 mg, 79%).
^{1}H NMR (400 MHz, $CDCl_3$): δ 8.27 (s, 1H), 6.97 (s, 1H), 5.75 (br s, 2H), 4.20 (q, *J* = 7.3 Hz, 2H), 1.43 (t, *J* = 7.3 Hz, 3H). ^{13}C NMR (101 MHz, $CDCl_3$): δ 157.0 (C), 152.6 (CH), 149.3 (C), 123.1 (CH), 102.1 (C), 86.1 (C), 39.8 (CH_2), 15.8 (CH_3).

General procedure for formation of 5-substituted-7-substituted-7*H*-pyrrolo[2,3-*d*]pyrimidine-4-amines, (**5a**–**c** and **9a**–**k**): To a degassed biphasic solution of THF (1.5 mL) and 1M Na_2CO_3 (0.5 mL), was added the required bromo intermediate (**4a**–**b** or **8a**–**b**) (for **4a**, 50 mg, 0.16 mmol, for **4b**, 50 mg, 0.18 mmol, for **8a**, 50 mg, 0.22 mmol and for **8b**, 50 mg, 0.21 mmol), R^1–$B(OH)_2$ (3.0 eq.) and $PdCl_2(PPh_3)_2$ (0.1 eq.) and the mixture heated at 100 °C for 2 h. The reaction mixture was

diluted with 3 mL EtOAc and the organic layer filtered through cotton wool. The filtrate was directly applied to a silica column and eluted with EtOAc/Et_3N (99:1). The purified fractions were evaporated at reduced pressure to give compounds **5a**–**c** or **9a**–**k** typically as off-white amorphous solids.

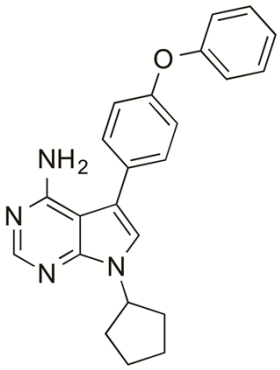
7-cyclopentyl-5-(4-phenoxyphenyl)-7*H*-pyrrolo[2,3-*d*]pyrimidine-4-dimethylamine (5a)[26]



4a (50 mg, 0.16 mmol) gave **5a** (42 mg, 65%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.43 (s, 1H), 7.44 – 7.41 (m, 2H), 7.37 – 7.33 (m, 2H), 7.13 – 7.08 (m, 1H), 7.06 – 7.02 (m, 5H), 5.28 – 5.20 (m, 1H), 2.85 (s, 6H), 2.28 – 2.21 (m, 2H), 1.91 – 1.76 (m, 6H). ^{13}C NMR (101 MHz, $CDCl_3$): δ 160.5 (C), 157.4 (C), 156.0 (C), 151.9 (C), 150.2 (CH), 131.7 (C), 129.9 (4 × CH), 123.4 (CH), 120.2 (CH), 119.1 (2 × CH), 119.0 (2 × CH), 116.7 (C), 102.8 (C), 55.1 (CH), 41.3 (2 × CH_3), 33.0 (2 × CH_2), 24.2 (2 × CH_2).

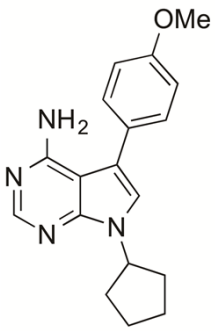
7-cyclopentyl-5-(4-phenoxyphenyl)-7*H*-pyrrolo[2,3-*d*]pyrimidine-4-amine (5b)[26]



4b (50 mg, 0.18 mmol) gave **5b** (36 mg, 54%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.33 (s, 1H), 7.46 – 7.42 (m, 2H), 7.40 – 7.35 (m, 2H), 7.17 – 7.12 (m, 1H), 7.11 – 7.05 (m, 4H), 7.02 (s, 1H), 5.27 – 5.17 (m, 1H), 5.07 (br s, 2H), 2.31 – 2.21 (m, 2H), 1.92 – 1.76 (m, 6H). ^{13}C NMR (101 MHz, $CDCl_3$): δ 157.1 (C), 157.0 (C), 156.9 (C), 151.9 (CH), 150.8 (C), 130.3 (2 × CH), 130.1 (C), 130.0 (2 × CH), 123.8 (CH), 119.9 (CH), 119.3 (2 × CH), 119.2 (2 × CH), 115.8 (C), 101.3 (C), 55.1 (CH), 33.1 (2 × CH_2), 24.3 (2 × CH_2).

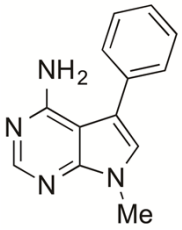
7-cyclopentyl-5-(4-methoxyphenyl)-7*H*-pyrrolo[2,3-*d*]pyrimidine-4-amine (5c)[26]



4b (50 mg, 0.18 mmol) gave **5c** (29 mg, 53%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.29 (s, 1H), 7.39 (d, J = 8.7 Hz, 2H), 7.00 – 6.96 (m, 3H), 5.37 (br s, 2H), 5.22 – 5.17 (m, 1H), 3.85 (s, 3H), 2.25 – 2.20 (m, 2H), 1.90 – 1.75 (m, 6H). ^{13}C NMR (101 MHz, $CDCl_3$): δ 159.1 (C), 156.8 (C), 151.0 (CH), 150.5 (C), 130.1 (2 × CH), 127.2 (C), 119.8 (CH), 116.2 (C), 114.6 (2 × CH), 101.3 (C), 55.5 (CH_3), 55.1 (CH), 33.1 (2 × CH_2), 24.3 (2 × CH_2).

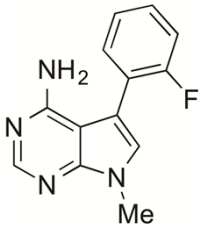
7-Methyl-5-phenyl-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine (9a)



8a (50 mg, 0.22 mmol) gave **9a** (22 mg, 45%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.34 (s, 1H), 7.50−7.43 (m, 4H), 7.38−7.34 (m, 1H), 6.94 (s, 1H), 5.24 (br s, 2H), 3.84 (s, 3H). HR-ESMS calcd. for $C_{13}H_{13}N_4^+$ [M + H] 225.1135, found 225.1135.

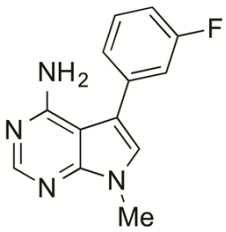
5-(2-Fluorophenyl)-7-methyl-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine (9b)



8a (50 mg, 0.22 mmol) gave **9b** (27 mg, 51%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.36 (s, 1H), 7.43 (td, *J* = 7.6, 1.8 Hz, 1H), 7.35 (tdd, *J* = 7.1, 5.1, 1.9 Hz, 1H), 7.27−7.18 (m, 2H), 7.04 (d, *J* = 1.1 Hz, 1H), 5.04 (br s, 2H), 3.86 (s, 3H). HR-ESMS calcd. for $C_{13}H_{12}FN_4^+$ [M + H] 243.1041, found 243.1040.

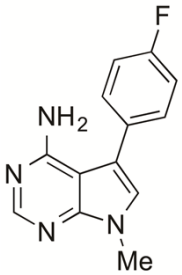
5-(3-Fluorophenyl)-7-methyl-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine (9c)



8a (50 mg, 0.22 mmol) gave **9b** (16 mg, 30%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.36 (s, 1H), 7.42 (td, *J* = 7.9, 6.1 Hz, 1H), 7.28−7.24 (m, 1H), 7.18 (ddd, *J* = 9.7, 2.4, 1.7 Hz, 1H), 7.05 (tdd, *J* = 8.5, 2.6, 1.0 Hz, 1H), 6.96 (s, 1H), 5.16 (br s, 2H), 3.85 (s, 3H). HR-ESMS calcd. for $C_{13}H_{12}FN_4^+$ [M + H] 243.1041, found 243.1040.

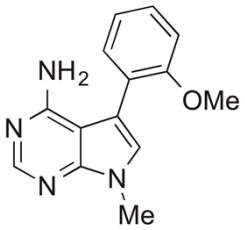
5-(4-Fluorophenyl)-7-methyl-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine (9d)



8a (50 mg, 0.22 mmol) gave **9d** (25 mg, 47%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.34 (s, 1H), 7.45−7.41 (m, 2H), 7.17−7.12 (m, 2H), 6.90 (s, 1H), 5.14 (br s, 2H), 3.83 (s, 3H). HR-ESMS calcd. for $C_{13}H_{12}FN_4^+$ [M + H] 243.1041, found 243.1041.

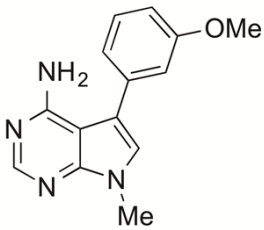
5-(2-Methoxyphenyl)-7-methyl-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine (9e)



8a (50 mg, 0.22 mmol) gave **9e** (28 mg, 50%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.32 (s, 1H), 7.36−7.29 (m, 2H), 7.07−7.00 (m, 2H), 6.96 (s, 1H), 5.09 (br s, 2H), 3.83 (s, 6H). HR-ESMS calcd. for $C_{14}H_{15}N_4O^+$ [M + H] 255.1240, found 255.1240.

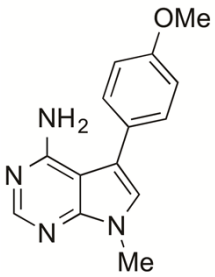
5-(3-Methoxyphenyl)-7-methyl-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine (9f)



8a (50 mg, 0.22 mmol) gave **9f** (5 mg, 9%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.35 (s, 1H), 7.39−7.35 (m, 1H), 7.07−7.04 (m, 1H), 7.02−6.99 (m, 1H), 6.95 (s, 1H), 6.91 (ddd, *J* = 8.3, 2.6, 0.9 Hz, 1H), 5.16 (br s, 2H), 3.86 (s, 3H), 3.85 (s, 3H). HR-ESMS calcd. for $C_{14}H_{15}N_4O^+$ [M + H] 255.1240, found 255.1241.

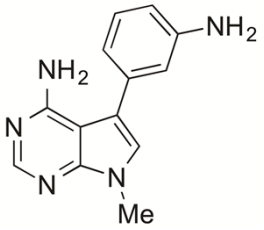
5-(4-Methoxyphenyl)-7-methyl-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine (9g)



8a (50 mg, 0.22 mmol) gave **9g** (27 mg, 48%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.34 (s, 1H), 7.39 (d, *J* = 8.8 Hz, 2H), 6.99 (d, *J* = 8.8 Hz, 2H), 6.88 (s, 1H), 5.05 (br s, 2H), 3.86 (s, 3H), 3.83 (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$): δ 159.1 (C), 157.2 (C), 152.1 (CH), 150.3 (C), 130.2 (2 × CH), 127.2 (C), 128.5 (CH), 115.9 (C), 114.6 (2 × CH), 101.4 (C), 55.5 (CH_3), 31.3 (CH_3). HR-ESMS calcd. for $C_{14}H_{15}N_4O^+$ [M + H] 255.1240, found 255.1240.

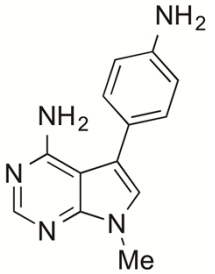
5-(3-Aminophenyl)-7-methyl-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine (9h)



8a (50 mg, 0.22 mmol) gave **9h** (24 mg, 46%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.34 (s, 1H), 7.23 (t, *J* = 7.7 Hz, 1H), 6.92 (s, 1H), 6.85 (ddd, *J* = 7.5, 1.5, 1.0 Hz, 1H), 6.83−6.76 (m, 1H), 6.67 (ddd, *J* = 8.0, 2.4, 1.0 Hz, 1H), 5.20 (s, 2H), 3.83 (s, 3H). HR-ESMS calcd. for $C_{13}H_{14}N_5^+$ [M + H] 240.1244, found 240.1236.

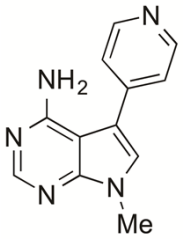
5-(4-Aminophenyl)-7-methyl-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine (9i)



8a (50 mg, 0.22 mmol) gave **9i** (16 mg, 30%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.35 (s, 1H), 7.28 (d, *J* = 8.5 Hz, 2H), 6.87 (s, 1H), 6.79 (d, *J* = 8.5 Hz, 1H), 5.14 (br s, 2H), 3.84 (s, 3H), 3.78 (br s, 2H). HR-ESMS calcd. for $C_{13}H_{14}N_5^+$ [M + H] 240.1244, found 240.1238.

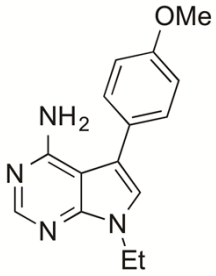
7-Methyl-5-(pyridin-4-yl)-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine (9j)



8a (50 mg, 0.22 mmol) gave **9j** (10 mg, 20%).

^{1}H NMR (400 MHz, DMSO): δ 8.58 (dd, *J* = 4.4, 1.6 Hz, 2H), 8.19 (s, 1H), 7.57 (s, 1H), 7.45 (dd, *J* = 4.4, 1.6 Hz, 2H), 6.30 (br s, 2H), 3.76 (s, 3H). HR-ESMS calcd. for $C_{12}H_{12}N_5^+$ [M + H] 226.1087, found 226.1081.

7-Ethyl-5-(4-methoxyphenyl)- 7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine (9k)



8b (50 mg, 0.21 mmol) gave **9k** (38 mg, 68%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.33 (s, 1H), 7.40 (d, *J* = 8.7 Hz, 2H), 6.99 (d, *J* = 8.7 Hz, 2H), 6.93 (s, 1H), 5.07 (br s, 2H), 4.27 (q, *J* = 7.3 Hz, 2H), 3.86 (s, 3H), 1.49 (t, *J* = 7.3 Hz, 3H). ^{13}C NMR (101 MHz, $CDCl_3$): δ 159.0 (C), 157.3 (C), 152.0 (CH), 150.3 (C), 130.1 (2 × CH), 127.3 (C), 121.8 (CH), 115.9 (C), 114.6 (2 × CH), 101.4 (C), 55.5 (CH_3), 39.4 (CH_2), 15.8 (CH_3). HR-ESMS calcd. for $C_{15}H_{17}N_4O^+$ [M + H] 269.1397, found 269.1397.

3D7 *Pf*AMA1 DI+II Expression

3D7 *Pf*AMA1$_{[104-442]}$ was produced according to the procedures described by Lim *et al*.[29] In brief, the protein was first expressed using the high-cell-density methodology and then purified using Ni^{2+}-affinity chromatography. The denatured protein was refolded in a redox environment containing reduced and oxidised glutathione at 4:1 ratio, and the refolded protein was subsequently purified using anion-exchange chromatography. Pure fractions were combined and buffer-exchanged into 20 mM phosphate buffer, pH 7.4, using an Amicon Ultra-4 centrifugal unit (Millipore) for NMR experiments.

NMR Aggregation Studies of 5a−c

All three compounds (**5a**, **5b** and **5c**) were dissolved at concentrations of 5, 10, 20, 40 and 80 μM in 20 mM phosphate buffer at pH 7.4 containing 10 % ^2H_2O, 1 % $^2H_6$-DMSO and 1 μM 4,4-dimethyl-4-silapentane-1-sulfonic acid. Tween 20 at a final concentration of 0.05 % v/v was added into samples containing 80 μM of the compounds. ^{1}H NMR spectra for all the samples were acquired using Bruker Avance III 600 MHz spectrometer at 25°C with 128 scans. Excitation sculpting was employed to suppress the water proton resonance. The relaxation delay was 5 s. Bruker TopSpin 3.2 software was used to integrate the proton resonances of the compounds.

SPR Binding Experiments

The interaction of **5a−5c** and **9a−9k** with AMA1 was determined using a Biacore T200 biosensor. The methodology was similar to the protocol described by Lim *et al*.,[29] except that running buffer containing phosphate buffered saline (137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4{\cdot}2H_2O$, 2 mM KH_2PO_4), 3 mM EDTA (ethylenediaminetetraacetic acid), 0.05 % v/v Tween 20 and 5 % DMSO, pH 7.4 was used instead. Compounds were tested at 6.25, 12.5, 25, 50 and 100 μM.

CPMG Binding Assays

CPMG experiments were conducted using 20 mM phosphate buffer at pH 7.4 containing 10 % ^2H_2O and 1 % $^2H_6$-DMSO. CPMG spin-locks of 0 and 200 ms with 1 ms delay between the hard 180° pulses were employed. Spectra for samples with 80 µM of **5c** in the presence and absence of 10 µM 3D7 *Pf*AMA1$_{[104-442]}$ were acquired first before adding R1 (VFAEFLPLFSKFGSRMHILK)[12] or RON2 (DITQQAKDIGAGPVASCFTTRMSPPQQICLNSVVNTALS) peptides to a final concentration of 15 µM.

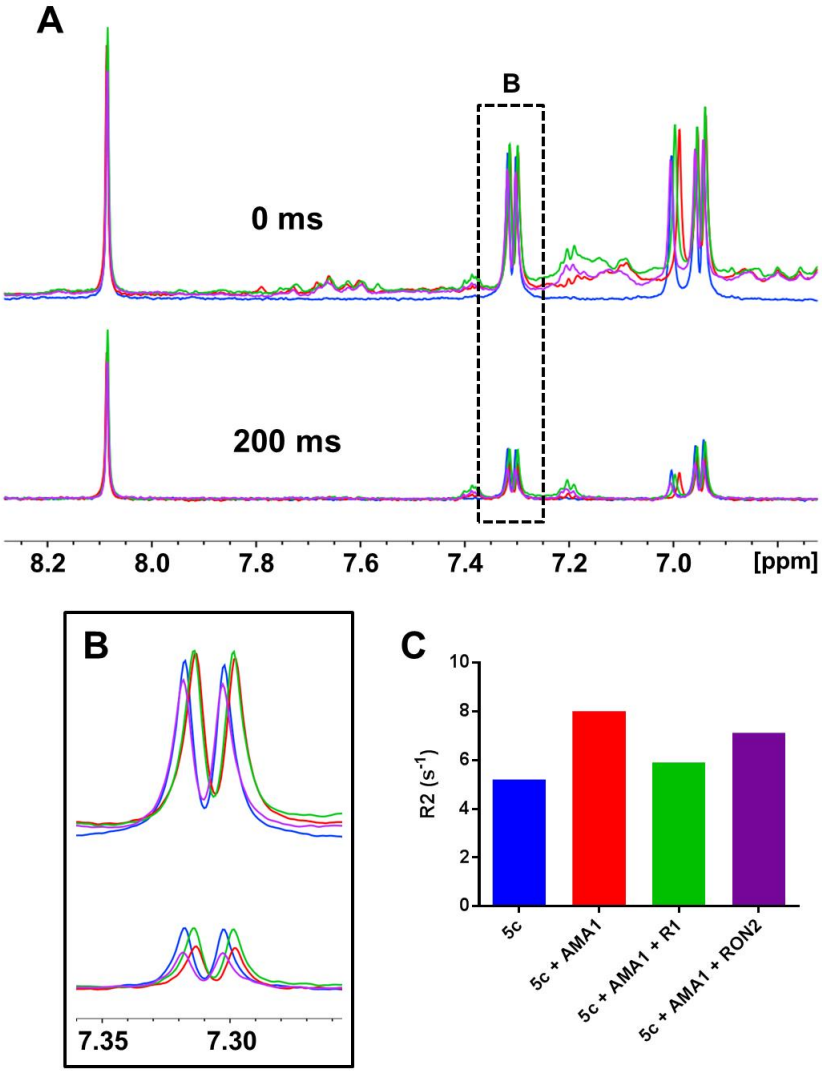
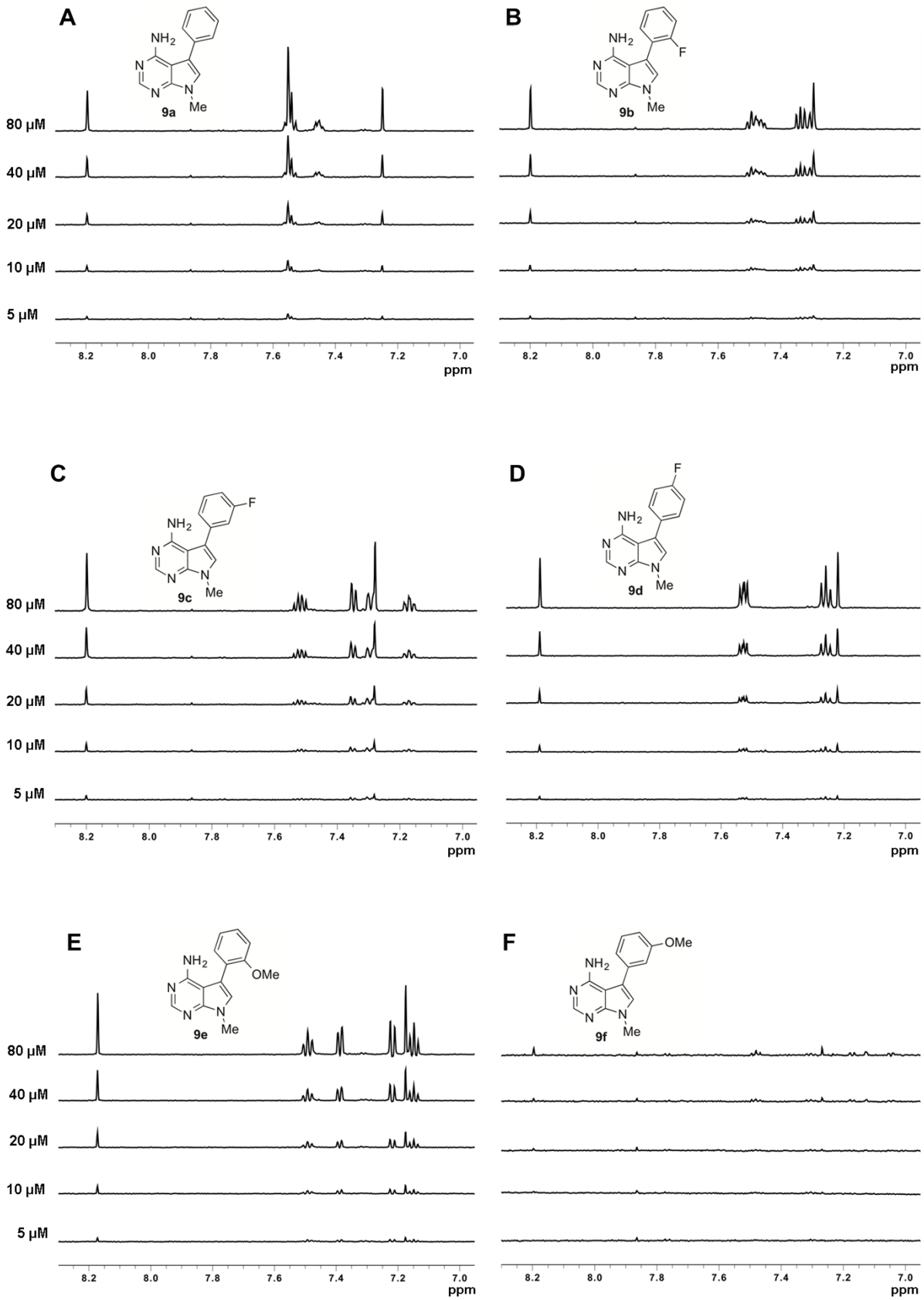


Figure S1. Binding monitored by CPMG. (A and B) All CPMG spectra were acquired with both 0 (top spectra) and 200 ms (bottom spectra) spin-relaxation filters in the presence of 0.05 % v/v Tween 20. Blue and red spectra were from samples containing **5c** in presence or absence of 3D7 *Pf*AMA1, respectively. Green and purple spectra correspond to samples containing **5c** and 3D7 *Pf*AMA1 with the additions of R1 and RON2 peptides, respectively. (C) Transverse relaxation rate (R2) for different samples used in the binding studies.

NMR Aggregation Studies of 9a−k

In a similar fashion to **5a−c**, the aggregation of **9a−k** was studied by 1D ^{1}H NMR.



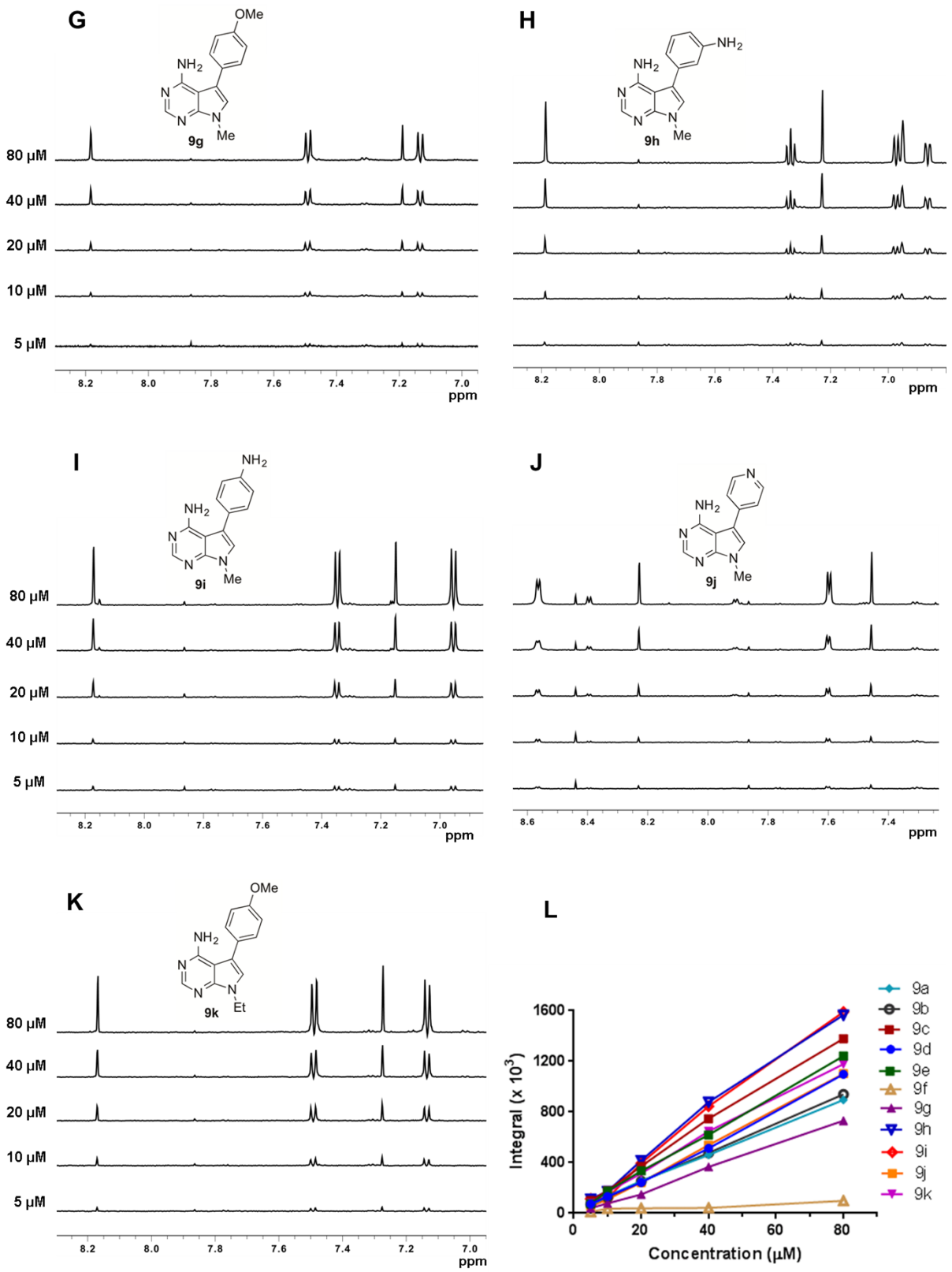


Figure S2. (A-K) 1D ^{1}H NMR spectra of pyrrolo[2,3-*d*]pyrimidine-4-amines (**9a**–**k**) at different concentrations. (L) Peak intensities of proton signals (~8.2 ppm (H2)) of **9a**–**k** at increasing compound concentrations. Note that **9f** precipitated out of solution, which caused the reduction in signal.

Parasite growth inhibition assays

Flow cytometry based growth inhibition assays were performed as described in detail elsewhere.[37,46,47] Compounds **5a**−**5c** and **9a**−**9k** were dissolved in 100% DMSO to make a 100 mM stock. These stocks and DMSO controls were then diluted in PBS to make the dilution series shown in Figure S3. Synchronised early trophozoite stage parasites (3D7 strain) were grown in the presence of each compound and DMSO controls. After two invasion cycles, early trophozoite stage parasites were stained, and parasitaemia measured by flow cytometry. Each dilution series was run in duplicate with the percent growth inhibition = (% parasitaemia in test well/mean % parasitaemia of DMSO control wells x100) -100) x-1).

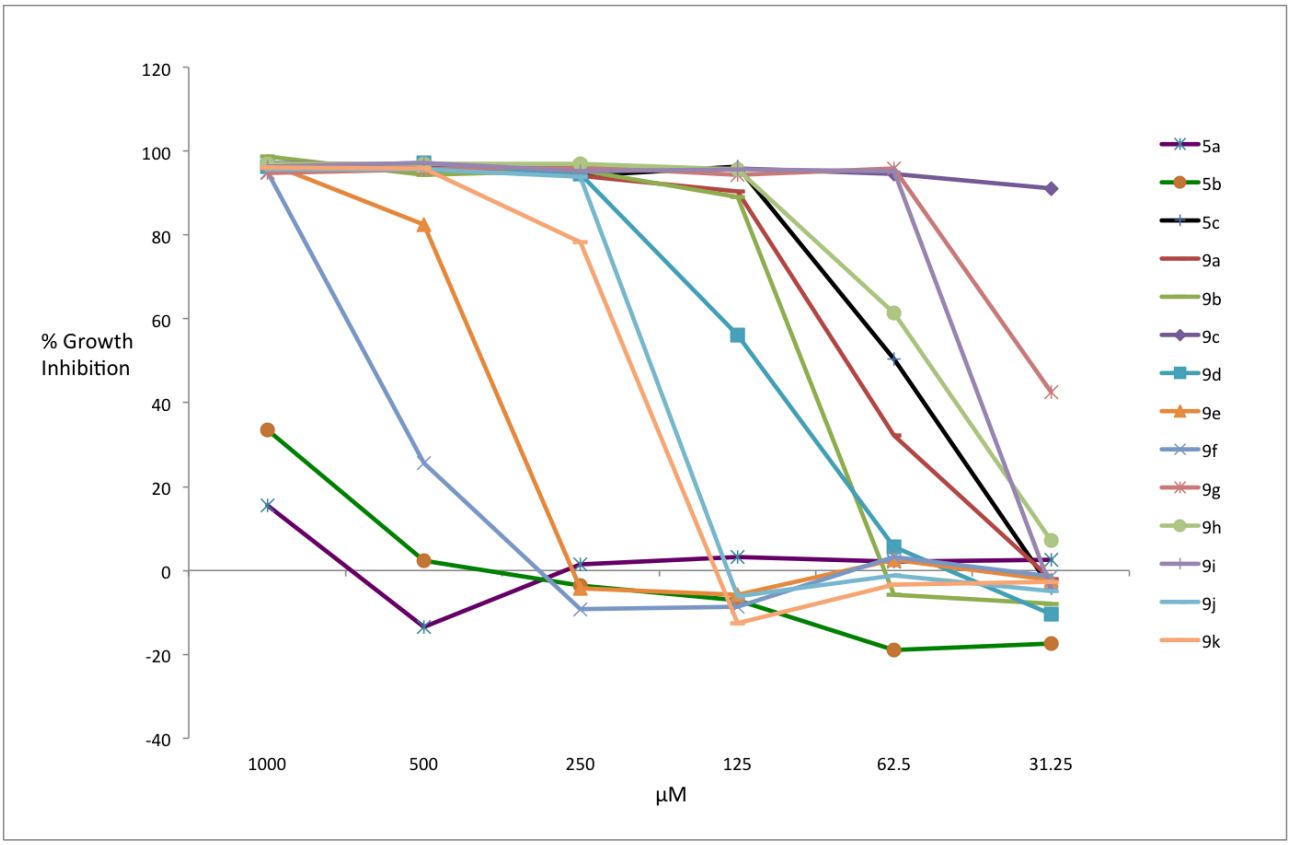


Figure S3. Growth inhibition assay of **9a**−**k**. *Pf*RON2 gave 100% inhibition in this assay over this concentration range.

References

45. H. E. Gottlieb, V. Kotlyar, and A. Nudelman, *J. Org. Chem.*, 1997, **62**, 7512–7515.
46. D. W. Wilson, B. S. Crabb, and J. G. Beeson, *Malar. J.*, 2010, **9**, 152.
47. K. E. M. Persson, C. T. Lee, K. Marsh, and J. G. Beeson, *J. Clin. Microbiol.*, 2006, **44**, 1665–1673.

Chapter 4:

Crystal Structure of FVO *Pf*AMA1

Submitted manuscript

Lim, S. S.; Yang, W.; Krishnarjuna, B.; Sivaraman, K. K.; Chandrashekaran, I. R.; Kass, I.; MacRaild, C. A.; Devine, S. M.; Debono, C. O.; Anders, R. F.; Scanlon, M. J.; Scammells, P. J.; Norton, R. S.; McGowan, S., Structure and dynamics of apical membrane antigen 1 from *Plasmodium falciparum* FVO. *Submitted to Biochemistry* **2014**.

4.1 Declaration for Thesis Chapter 4

Declaration by candidate

In the case of Chapter 4, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Designed and performed experiments, data analysis, manuscript preparation	70

The following co-authors contributed to the work. If co-authors are students at Monash University, the extent of their contribution in percentage terms must be stated:

Name	Nature of contribution	Extent of contribution (%) for student co-authors only
Wei Yang	Performed experiments, data analysis	10
Bankala Krishnarjuna	Performed experiments, data analysis	5
Komagal Kannan Sivaraman	Performed experiments	
Indu R. Chandrashekaran	Performed experiments	
Itamar Kass	Performed experiments, data analysis, manuscript preparation	
Christopher A. MacRaild	Manuscript preparation	
Shane M. Devine	Manuscript preparation	
Cael O. Debono	Performed experiments	2.5
Robin F. Anders	Intellectual input, manuscript preparation	
Martin J. Scanlon	Intellectual input, manuscript preparation	
Peter J. Scammells	Intellectual input, manuscript preparation	
Raymond S. Norton	Intellectual input, manuscript preparation	
Sheena McGowan	Performed experiments, intellectual	

	input, manuscript preparation	

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the candidate's and co-authors' contributions to this work.

Candidate's Signature	[redacted]	**Date 5/10/2014**
Main Supervisor's Signature	[redacted]	**Date 7/10/2014**

Structure and Dynamics of Apical Membrane Antigen 1 from *Plasmodium falciparum* FVO

San Sui Lim[1], Wei Yang[2], Bankala Krishnarjuna [1], Komagal Kannan Sivaraman[2], Indu R. Chandrashekaran[1], Itamar Kass [2,3], Christopher A. MacRaild[1], Shane M. Devine[1], Cael O. Debono[1], Robin F. Anders[1,4], Martin J. Scanlon[1,5], Peter J. Scammells[1], Raymond S. Norton[1], and Sheena McGowan[2*]*

[1] Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Victoria 3052, Australia.

[2] Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia.

[3] Victorian Life Sciences Computation Initiative Life Sciences Computation Centre, Monash University, Clayton, VIC 3800, Australia.

[4] Department of Biochemistry, La Trobe University, Bundoora, Victoria 3086, Australia.

[5] Centre of Excellence for Coherent X-Ray Science, Monash University, Parkville, Victoria 3052, Australia.

KEYWORDS: Malaria, AMA1, FVO, Structure, Molecular Dynamics.

ABSTRACT: Apical membrane antigen 1 (AMA1) interacts with RON2 to form a protein complex that plays a key role in the invasion of host cells by malaria parasites. Blocking this protein-protein interaction represents a potential route to control malaria and related parasitic diseases, but the polymorphic nature of AMA1 has proven to be a major challenge for vaccine-induced antibodies and peptide inhibitors to exert strain-transcending inhibitory effects. Here we present the X-ray crystal structure of AMA1 domains I and II from the *Plasmodium falciparum* strain FVO. We compare our new structure to those of AMA1 from *P. falciparum* 3D7 and *P. vivax*. A combination of normalized B-factor analysis and computational methods has been used to investigate the flexibility of the domain I loops and how this correlates with their roles in determining the strain specificity of human antibody responses and inhibitory peptides. We also investigated the domain II loop, a key region involved in inhibitor binding, by comparison of multiple AMA1 crystal structures. Collectively, these results provide valuable insights that should contribute to the design of strain-transcending agents targeting *Plasmodium falciparum* AMA1.

Malaria is one of the most widespread infections, with more than 40 % of the global population at risk of contracting the disease.[1, 2] Each year, there are approximately 250 million clinical cases of malaria that result in more than 600,000 deaths worldwide.[2] The majority of these deaths are due to *Plasmodium falciparum* infections occurring in young children in sub-Saharan Africa.[2] Although much less likely to cause death, *P. vivax* infections also contribute to a substantial malaria burden across the globe, with 70 to 80 million cases occurring annually.[3] Although current artemisinin combination therapies have been highly effective against *Plasmodium* parasites, signs of resistance have already emerged.[4] There is an urgent need to combat this threat using therapeutic agents that act against a broad range of parasite strains, especially those that have become resistant to available therapies.

Apical membrane antigen 1 (AMA1) forms part of the moving junction complex essential for erythrocyte invasion by *Plasmodium* merozoites, and ligands that disrupt AMA1 function inhibit the growth *in vitro* of *P. falciparum* asexual blood stages.[5-9] Further, a conditional knockdown of *Pf*AMA1 severely impaired the parasite's ability to invade red cells[10] and a complete gene knockout is not viable in *P. falciparum*.[11] AMA1 is a type I integral membrane protein[12] the extracellular region of which consists of three domains based on the connectivities of its eight intramolecular disulfide bonds: an N-terminal domain I, a central domain II, and a C-terminal domain III.[13, 14] Antibodies to AMA1 can block red cell invasion by *P. falciparum in vitro* and contribute to the adaptive immune response that partially protects exposed individuals against malaria. AMA1 has been a leading candidate for inclusion in a vaccine against *P. falciparum*[15] but in a Phase IIb clinical trial in Mali a 3D7 *Pf*AMA1 vaccine provided protection against only a subset of *P. falciparum* AMA1 genotypes, reflecting the extensive sequence polymorphisms in this antigen[16]. A bivalent 3D7 and FVO *Pf*AMA1 vaccine, also tested in Phase IIb trials, failed

to protect because of the poor immunogenicity of the alum formulation.[17] Currently, preclinical studies with multivalent vaccines (4 – 6 AMA1 alleles) show promise, inducing a more broadly cross-reactive antibody response.[18, 19]

The two most extensively studied forms of *Pf*AMA1, 3D7 and FVO, have 24 amino acid differences and these polymorphic residues have been grouped into domain I, II and III clusters based on their spatial proximity on the X-ray crystal structure of 3D7 *Pf*AMA1.[20] The domain I cluster is the most important of the three in mediating escape from inhibitory antibodies and was further classified into sub-clusters C1, C2 and C3. Within C1, the region termed C1L, is particularly important for immune escape.[18, 21] In the preclinical studies of *Pf*AMA1 vaccines it was noted that antisera to 3D7 *Pf*AMA1 were more strain specific than antisera to FVO *Pf*AMA1.

AMA1 has a hydrophobic cleft that is the site of interactions with its protein-binding partner RON2.[9, 22] The cleft is surrounded by six loops from domain I (loops Ia – If) and an extended loop from domain II (DII loop).[23] The DII loop appears to contain a strain-transcending epitope as the monoclonal antibody 4G2, which binds to the base of the DII loop, exhibits strain-independent inhibition of *P. falciparum*.[5] We and others have proposed that small molecules targeted to the hydrophobic cleft may interrupt the AMA1-RON2 protein-protein interaction and provide a route to novel therapeutics in the form of protein-protein interaction inhibitors (PPIs).[24-26] However, the polymorphic regions C1 and C3 surrounding one end of the cleft were shown to restrict the cross-reactivity of inhibitory antibodies and peptides such as IgNAR,[7] 1F9,[27, 28] and R1.[6, 9, 29] It has been postulated that sequence variation in the highly polymorphic C1L region (of the C1 sub-cluster) may result in local secondary structure changes.[15, 21] In particular, the presence of a two-turn helix in the Id loop has been questioned for the FVO allele due to the presence of a glycine residue at position 197.[15]

To answer fundamental questions regarding the impact of sequence diversity on AMA1, we have determined the first X-ray crystal structure of *Pf*AMA1 from FVO. We show through X-ray crystallography, all-atom molecular dynamics and nuclear magnetic resonance (NMR) spectroscopy that the sequence divergence does not result in structural changes that account for the strain-specific effects documented for inhibitory antibodies and peptides.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification. Domains I and II of the ectodomains of FVO and 3D7 *Pf*AMA1 (residues 108-438) were produced according to the protocol described by Lim *et al*.[24] except that the 6 x His-tag was removed using TEV protease. The pure His-tagged proteins were dialyzed using membrane tubing (Spectra/Por 3, MWCO 3.5 kDa) against 100 times volume of 50 mM Tris pH 8.0 under constant stirring at 4 °C overnight then TEV protease was added to the sample at a 1:30 ratio (w/w) and cleavage was allowed to proceed at 4 °C for 48 h with gentle agitation. The mixture was filtered (0.2 µm) and loaded onto a 5 mL CHT I ceramic hydroxyapatite column (Bio-Rad). The cleaved protein was eluted using a linear gradient of 10 to 150 mM phosphate buffers (Na_2HPO_4 and $NaH_2PO_4 \cdot H_2O$), pH 7.4, over 15 column volumes. The pooled fractions were concentrated and buffer exchanged into 20 mM Tris, pH 8.0, using an Amicon Ultra-4 centrifugal unit with an Ultracel-10 membrane (Millipore).

^{1}H NMR Spectroscopy & Size Exclusion Chromatography. FVO *Pf*AMA1 purified from hydroxyapatite chromatography was buffer exchanged into 20 mM phosphate buffer (Na_2HPO_4 and $NaH_2PO_4 \cdot H_2O$), 50 mM NaCl, pH 7.4, containing 10 % ^2H_2O using a PD-10 desalting column (GE Healthcare). The sample was subsequently concentrated as above to a final protein concentration of 50 µM. Part of this final product was used for ^{1}H NMR. A ^{1}H-detected pulse

program incorporating the excitation sculpting scheme for water suppression was employed to characterize FVO *Pf*AMA1. A total of 128 scans and 16 K data points was acquired at 600 MHz on a Bruker Avance III spectrometer at 35 °C. The data were processed in Topspin 3.2 using an exponential multiplication function with 2 Hz line broadening. The water signal was used to reference the ^{1}H NMR spectrum. The final product was loaded onto a Superdex 200 10/30 GL column and eluted isocratically with 20 mM phosphate buffer (Na_2HPO_4 and $NaH_2PO_4 \cdot H_2O$), 50 mM NaCl, pH 7.4, at flow rate of 0.5 mL/min.

$^{2}H_2O$ *E. coli* Adaptation. Fifty µL of competent *E. coli* BL21 (DE3) glycerol stock, previously frozen at -80 °C, was thawed on ice for 10 min prior to adding 1 µL of plasmid carrying the expression vector pPROEX HTb (Novagen) with FVO or 3D7 *Pf*AMA1$_{[104-438]}$ sequences. The mixture was left on ice for a further 30 min and then in water at 42 °C for 45 s. 1 mL of Luria-Broth (LB) was added to each sample and the culture was incubated at 37 °C with constant shaking at 225 rpm. After 45 min, 50 µL of culture was spread over a LB plate containing 50 % (v/v) $^{2}H_2O$ and 100 µg/mL of ampicillin, then incubated at 37 °C overnight. A single colony of the freshly transformed cells was inoculated into 10 mL of LB medium with 50 % (v/v) $^{2}H_2O$ and 100 µg/mL ampicillin. The culture was grown for 24 h at 37 °C with constant shaking at 225 rpm. The cell mixture was spread on a culture plate, and subsequently a single colony of cells was incubated in growth medium as described above, except that the LB plate and medium prepared with 75 % (v/v) $^{2}H_2O$ were used instead. The final cell culture containing 75 % (v/v) $^{2}H_2O$ was stored at -80 °C with 20 % (v/v) glycerol.

Isotopically-Labelled AMA1. A scrape of the glycerol stock of $^{2}H_2O$-adapted *E. coli* was inoculated into LB medium with 75 % (v/v) $^{2}H_2O$ and 100 µg/mL ampicillin. The culture was incubated overnight at 37 °C with constant shaking at 225 rpm. The overnight culture was then

centrifuged at 1500 *g* for 15 min. The supernatant was decanted and the cell pellets were resuspended in two volumes of optimized minimal medium[30] prepared with 100 % (v/v) $^{2}H_2O$, 1 g/L $^{15}NH_4Cl$ and 8 g/L ^{13}C-glucose. The cells were allowed to grow for 3 h with shaking at 37 °C before being induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 24 h. The protein was then purified as described above.

^{1}H-^{15}N HSQC and 3D HNCO Experiments. The ^{2}H-^{15}N-^{13}C-labelled 3D7 and FVO *Pf*AMA1 were dissolved at 300 and 75 µM, respectively, in 20 mM phosphate buffer (Na_2HPO_4 and $NaH_2PO_4 \cdot H_2O$), pH 7.0, containing 50 mM L-arginine, 50 mM L-glutamic acid, 0.2% (w/v) protease inhibitors cocktail (Roche), 0.01 % (w/v) sodium azide and 10 % (v/v) $^{2}H_2O$. The FVO *Pf*AMA1 spectrum was acquired on a Bruker Avance III 600 MHz spectrometer at 35 °C. A spectrum of 3D7 *Pf*AMA1 was acquired on a Bruker Avance III 800 MHz spectrometer at 30 °C. Both the ^{1}H-^{15}N HSQC and 3D HNCO experiments were carried out using pulse sequences with transverse relaxation optimized spectroscopy (TROSY) effects.[31, 32] ^{1}H-^{15}N HSQC spectra were acquired with 64 scans at 2048 and 256 data points for the ^{1}H and ^{15}N dimensions, respectively. A total of 32 scans was recorded for the 3D HNCO experiments, with 2048, 128 and 128 data points for ^{1}H, ^{15}N and ^{13}C dimensions, respectively. Both the direct and indirect dimensions of ^{1}H-^{15}N HSQC and 3D HNCO data were processed using a QSINE window with phase shift of 2 Hz. 32 linear prediction coefficients were applied to all indirect dimensions.

R1-FVO *Pf*AMA1 Interactions. A Biacore T200 biosensor was employed to measure the interaction between recombinant FVO *Pf*AMA1 DI + II and R1 peptide (GL Biochem). Surface plasmon resonance (SPR) experiments were conducted essentially the same as described previously[24] except that dimethyl sulfoxide (DMSO) was not included in the running buffer. Approximately 8000 RU of protein was coupled in a single flow cell (1000 RU = 1 ng of protein

per mm^2). The binding of R1 peptide[6, 29] to FVO *Pf*AMA1 was evaluated using a two-fold serial dilution ranging in concentration from 15.6 to 500 µM.

Sequence Alignment and Analysis. All sequence alignments and analyses were performed using the UniProt online tool (http://www.uniprot.org/). Accession numbers for AMA1 sequences used in this study are as follows: FVO *Pf*AMA1 (UniProt ID: Q9TY48) and 3D7 *Pf*AMA1 (UniProt ID: Q7KQK5). The residues defining the AMA1 hydrophobic cleft and polymorphic sites were obtained from published literature.[7, 23, 33-35]

Crystallization, X-Ray Data Collection, Structure Determination and Refinement. Crystallization conditions for FVO *Pf*AMA1 were identified following a robotic broad screen using the IndexHT (Hampton Research) and JCSGPlus (Molecular Dimensions) crystal screens. Optimization of a single initial hit from the Index screen used the hanging drop vapour diffusion method, with a 1:1 (v/v) ratio of protein to mother liquor (0.5 ml well volume). Small, stacked crystals appeared after three months in 25 % (v/v) polyethylene glycol 3350, 0.1 M HEPES (pH 7.5) and 0.2 M $MgCl_2$. A single crystal was separated from the stacked cluster and cryo-protected by the addition of 10 % glycerol prior to data collection.

3D7 *Pf*AMA1 crystals were grown in 12 - 15 % (v/v) polyethylene glycol 3350, 0.02 M MES (pH 6.0) and 10 mM $MnCl_2$ as detailed in Bai *et al.*[23] 3D7 *Pf*AMA1 crystals were dehydrated overnight in reservoir solution with increased (35 %) (v/v) polyethylene glycol 3350 before cryo-stabilization in 38 % (v/v) polyethylene glycol 3350, 0.088 M MES (pH 6.0) and 44 mM $MnCl_2$ for 6 – 8 h prior to data collection. For crystals used to test soaking solvents, 5 % (v/v) methanol or MilliQ water was added to the stabilization solution.

Data were collected at 100 K for all crystals using the Australian Synchrotron micro crystallography MX2 beamline 3ID1. Diffraction images were processed using XDS[36] and

AIMLESS[37] from the CCP4 suite.[38] 5 % of each dataset was flagged for calculation of R_{Free}[39] with neither a sigma nor a low-resolution cut-off applied to the data. A summary of data collection statistics is provided in Supplementary Table 1.

Structure determination proceeded using the Molecular Replacement method and the program PHASER.[40] A search model for FVO *Pf*AMA1 was constructed by removing the solvent and flexible loops from the crystal structure of a 3D7 *Pf*AMA1 (PDB ID 1Z40). A single clear peak in both the rotation and translation functions was evident and packed well within the asymmetric unit. Together with the unbiased features in the initial electron density maps, the correctness of the molecular replacement solution was confirmed. All subsequent model building and structural validation for FVO and 3D7 *Pf*AMA1 structures was done using Phenix[41, 42] and COOT.[43] Solvent molecules were added only if they had acceptable hydrogen-bonding geometry contacts of 2.5 to 3.5 Å with protein atoms or with existing solvent and were in good $2F_o$-F_c and F_o-F_c electron density. Hydrogen bonds (excluding water-mediated bonds) and salt bridges were calculated using PDBePISA.[44] The coordinates and structure factors are available from the Protein Data Bank (PDB IDs 4R1A, 4R19, 4R1B, 4R1C). Raw data and images are available from TARDIS[45] (www.tardis.edu.au).

B-Factor Analysis. The B-factors obtained from PDB files cannot be used directly, since the values may be on different scales owing to the application of different refinement procedures.[46] To compare the B-factors from different structures, the values were normalized as described by Parthasarathy *et al.*[47] The Cα B-factor values were extracted from the FVO *Pf*AMA1 (PDB ID 4R1A) and *P. vivax* AMA1 (PDB ID 1W81) as well as the three 3D7 *Pf*AMA1 crystal structures (PDB IDs 1Z40 4R19, 4R1B and 4R1C), and normalized using the following equation: $B_{normalized} = (C\alpha\ B\text{-factors} - B_{mean}) / \sigma(B)$, where B_{mean} and $\sigma(B)$ are, respectively, the mean value and standard

deviation of the distribution of observed thermal factors. The average values and standard deviations of normalized B-factors were calculated for α-helical and β-sheet regions, loops Ia – f and the DII loop. Average values were not calculated for regions where more than half the amino acid sequence was missing in the crystal structures.

Molecular Dynamic Simulations. The dynamics of three different AMA1 structures were studied. Atomic coordinates of *P. vivax* AMA1 (PDB ID 1W81; residues 49 – 383) and 3D7 *Pf*AMA1 (PDB ID 1Z40; residues 104 – 438) were obtained from the Protein Databank. Coordinates for FVO *Pf*AMA1 (residues 104 – 438) were obtained in this study. Missing atoms and residues were modelled using the MOE 2012.10.[48] Each protein was solvated in a water cubic box consisting of TIP3P water molecules[49] with Na^+ ions added to neutralize any charge. The minimum distance from the surface of each complex to the faces of the water box was set to 12 Å.

All-atom molecular dynamics (MD) simulations were performed using the NAMD 2.9 MD package[50] on the IBM Blue Gene/Q supercomputer of the Victorian Life Sciences Computation Initiative (VLSCI). Proteins were defined by the newly published and tested AMBER force field, FF12SB.[51-53] Equilibration was performed in three stages: first, potential steric clashes in the initial configurations were relieved with 50000 steps of energy minimization. Initial velocities for each system were assigned randomly according to Maxwell–Boltzmann distribution at 100 K. Each system was then heated to 300 K over 0.1 ns, with the protein harmonically restrained (10 kcal/mol/$Å^2$) under the canonical ensemble (NVT) conditions. Following this, each system was simulated for another 0.1 ns under the isothermal-isobaric ensemble (NPT) conditions with all heavy protein atoms harmonically restrained (10 kcal/mol/$Å^2$). Thereafter, each system was subjected to 250 ns of free simulation.

For all simulations, an integration time of 2 fs was used and the non-bonded cutoff length was set at 1 nm. All simulations were run at constant temperature (300 K) and pressure (1 atm), using a Langevin damping coefficient of 0.5 fs^{-1}. For each simulation system, periodic boundary conditions (PBC) were used together with the Particle-Mesh Ewald (PME) method for electrostatics interactions.[54] Electrostatics and VDW non-bonded forces were cut off at 1 nm. For each protein, 3 trajectories were run in parallel, differing only by their distribution of the initial velocities. System conformations were saved every 10 ps for subsequent analysis.

All the analyses were performed using the GROMACS 4.0.7 simulation software package.[55] Prior to MD analyses, translational and rotational motions were eliminated by superposition of each frame to the initial conformation. The Root-Mean-Square Deviations (RMSD) of the backbone heavy atoms in each system were calculated relative to their corresponding initial minimized structures. Backbone Root-Mean-Square Fluctuations (RMSF) were calculated for the productive phase (50 – 250 ns) of each simulation. All images were created by VMD 1.91.[56] PyMOL 1.3r2 (Schrodinger, LLC. 2010. The PyMOL Molecular Graphics System, Version 1.3r2).

Electrostatic surface potential calculations. Protein electrostatic potentials were calculated using APBS version 1.3.[57] Atom electrostatic-charges were taken from the FF12SB force-field. The electrostatic potential was visualized using PyMOL 1.3r2 (Schrodinger, LLC. 2010. The PyMOL Molecular Graphics System, Version 1.3r2) with positive potential in blue and negative potential in red over the range -3 and +3 k_bT/e_c, where k_b is the Boltzmann constant, T is the temperature (set to 300K˚) and e_c is electron charge.

Effect of mutations on the R1-3D7 *Pf*AMA1 stability. The contributions of specific mutations to the overall thermodynamic stability of the R1-3D7 *Pf*AMA1 complex structure[9] were

estimated *in silico* with FoldX using default settings.[58] The reported Gibbs free energies are the difference between those of wild-type and mutated 3D7 *Pf*AMA1 in the context of the complex.

RESULTS

X-Ray Crystal Structure of FVO *Pf*AMA1. To generate protein crystals of FVO *Pf*AMA1, an equivalent construct of FVO domain I + II (DI + II) was produced to that of 3D7 *Pf*AMA1.[23] The quality of our recombinant FVO *Pf*AMA1 protein was assessed using ^{1}H NMR spectroscopy and size-exclusion chromatography (Supplementary Fig. 1A). Overall, good signal dispersion was observed in the ^{1}H NMR spectrum, with methyl proton signals at -0.48 and -0.52 ppm as well as the amide proton signals beyond 9 ppm (Supplementary Fig. 1A); the spectrum was consistent with a single folded product. In size-exclusion chromatography, FVO *Pf*AMA1 eluted as a single peak consistent with a monomeric form of the protein (MW = 38 kDa) (Supplementary Fig. 1B).

The X-ray crystal structure of FVO *Pf*AMA1 was solved to 2.0 Å with final R and R_{Free} values of 19.5 and 25.5 %, respectively (Fig. 1A and Supplementary Table 1). FVO *Pf*AMA1 crystallized with one molecule in the asymmetric unit in space group $C2_1$. Seven α-helical and 16 β-sheet regions were identified in the final FVO *Pf*AMA1 structure. Similar to the 3D7 *Pf*AMA1 (Fig. 1B, PDB ID 1Z40) and *P. vivax* AMA1 structures (PDB ID 1W81), both DI and DII of FVO *Pf*AMA1 formed PAN folds that consist of a two-turn α-helix packed against a five-stranded β-sheet. The two PAN folds pack against each other to form the protein core as seen in the 3D7 *Pf*AMA1[23] (PDB ID 1Z40; 0.27 Å RMSD over 219 C-α atoms) and *P. vivax*[59] AMA1 (PDB ID 1Z40; 0.29 Å RMSD over 197 C-α atoms) structures. The DI + II sequences of 3D7 and FVO *Pf*AMA1 have 21 amino acid residue differences (sequence identity 94 %, Supplementary Fig. 2).

All residue differences found in the C1 (residues 187, 190, 196, 197, 200, 204, 206, and 225) and C2 (residues 242, 243, 282, 283 and 285) clusters are observable in our FVO *Pf*AMA1 structure (Supplementary Fig. 2). Structural analysis of regions in the vicinity of the C1L (residues 196, 197, 200, 204 and 206), C2 and DII clusters, as well as residues 167 and 300, did not reveal any significant structural differences between FVO and 3D7 *Pf*AMA1 DI + II (Supplementary Fig. 3). Structural differences were also not found in regions around residues 187 and 190 within the C1 cluster. Of the regions in close proximity to residue 225, residues 226 – 232 are disordered in FVO *Pf*AMA1, but ordered in 3D7 *Pf*AMA1. Structural comparisons could not be performed for C3 because residues 175 and 267, which constitute this cluster, are disordered in both structures. Nine of the 12 residues that define the hydrophobic cleft were observable in our FVO *Pf*AMA1 structure (V169, L176, F183, I190, Y202, V208, M224, Y251 and I252);[(7)] the remaining three hydrophobic residues (M273 from DI, and L357 and F367 from DII) are disordered in the FVO *Pf*AMA1 structure.

The structure of FVO *Pf*AMA1 showed numerous disordered loops in both domains. In DI, the disordered residues not observed in the density (160 – 163, 173 – 176, 226 – 232 and 258 – 273) correspond to loops Ib, If and part of the Ie loop (Fig. 1A). There are 38 residues missing in DII (351 – 388; Fig. 1B), which correspond to most of loop DII.

Flexibility of the Loops in Different Forms of AMA1. To ascertain whether the sequence polymorphisms might provide a structure-based 'escape' route from the host immune response and inhibitor binding, we analyzed loop flexibility of our FVO *Pf*AMA1 structure in comparison to the published 3D7 *Pf*AMA1 (PDB ID 1Z40) and *P. vivax* AMA1 (PDB ID 1W81) structures. We compared normalized B-factors from all Cα from each AMA1. The normalized B-values were expressed in units of standard deviations about the mean Cα B-factor for the corresponding

structure, and therefore regions that are more rigid in a protein would have low normalized B-factors, whereas flexible regions would have high normalized B-values. This analysis showed that loops Ib and If (polymorphic cluster C3) are highly flexible regions in all three structures (Fig. 2 and Supplementary Table 2). Loops Ia, Ic (residue 187 and 190 of C1 cluster) and Ie (residue 225 of C1 cluster) are mobile in FVO *Pf* and *P. vivax* AMA1, but exhibit limited mobility in 3D7 *Pf*AMA1. The Id loop (C1L cluster) appears to be more rigid in FVO *Pf*AMA1 than in 3D7 and *P. vivax* AMA1. However, inferring the biological relevance from this result is difficult because of the presence of extensive crystal contacts made by residues within the Id loop of FVO *Pf*AMA1 (Supplementary Fig. 4; see below).

In an effort to further assess the flexibility of the three AMA1 proteins, we undertook all-atom molecular dynamics simulations. Throughout our MD simulations, all systems were found to be stable following an initial structural rearrangement that took place early in the simulations (30 – 50 ns). Therefore, all subsequent analyses were carried out for the last 200 ns of each simulation (productive stage). Calculated RMSD values for the productive stage of MD simulations (3D7 *Pf*AMA1, 0.19 ± 0.04 nm; FVO *Pf*AMA1, 0.20 ± 0.03 nm and *P. vivax* AMA1, 0.30 ± 0.05 nm; Supplementary Fig. 5), indicated that all systems were stable and that *P. vivax* was slightly more flexible than the other proteins (particularly in the region of loop Ia, Supplementary Fig 6). The RMSF results (Supplementary Fig. 6) show that the fluctuation patterns of the three proteins were similar, with peak fluctuations occurring in the same loops of each AMA1 structure (loops Ib, Ic, Ie and If), which is consistent with the B-factor analysis. The Ie loop appears to be more flexible in FVO *Pf*AMA1 than in 3D7 *Pf*AMA1. This observation may explain why the Ie loop is disordered in the FVO *Pf*AMA1 crystal structure but not in 3D7 *Pf*AMA1. Based on the RMSF

results, the Id loop is flexible in all three AMA1 structures, which supports our earlier conclusions that the rigidity of the Id loop (Fig 2B) seen in FVO *Pf*AMA1 was a result of crystal contacts.

Crystallization artifacts, such as crystal contacts, can often complicate structural comparisons and implications drawn from static structures. The presence of stabilizing crystal contacts was identified in the original 3D7 *Pf*AMA1 structure by Bai *et al*.[23] In the search for suitable crystallization conditions of 3D7 *Pf*AMA1 for our fragment screening campaign,[24] we noticed that such crystal contacts could be modulated by soaking crystals in different solvents. This allowed us to examine the effect of these contacts on the conformation and/or flexibility of the surface loops in 3D7 *Pf*AMA1.We reproduced the original crystal conditions of Bai *et al*.,[23] solved the 1.8 Å X-ray crystal structure (PDB ID 4R19), and showed that it was identical to the published structure (PDB ID 1Z40; 0.118 Å RMSD over 275 Cα atoms). In these conditions, 3D7 *Pf*AMA1 crystallized in a $P3_1$ space group with two molecules per asymmetric unit, and showed crystal contacts that potentially stabilize the Ia – f and DII loops surrounding the hydrophobic cleft (Supplementary Table 3). Subjecting the crystals to 5 % methanol (PDB ID 4R1C) or water (PDB ID 4R1B) during their stabilization immediately prior to data collection produced changes in the space group and unit cell dimensions (Supplementary Table 1), and these two new 3D7 *Pf*AMA1 structures showed different degrees of flexibility within the Ia, Ic and Ie loops relative to the published 3D7 *Pf*AMA1 structure (Supplementary Table 2). The 3D7 *Pf*AMA1 structures from the MilliQ- and methanol-treated crystals have higher average normalized B-factors for the Ia, Ic and Ie loops when compared to the 1Z40 structure (Figs. 3 and 4). Unlike the Ie loop of FVO *Pf*AMA1, which is disordered, this loop is ordered in all the 3D7 *Pf*AMA1 structures. However, it is difficult to determine if there is a real difference in the flexibility of Ie loop as crystal packing is found in all the 3D7 *Pf*AMA1 structures (Supplementary Table 3). The Ib and If loops are

disordered in both the water- and methanol-treated 3D7 *Pf*AMA1 crystals, indicating that these regions are highly mobile in the protein (Figs. 3 and 4 and Supplementary Table 2).

Conformational Flexibility of the DII Loop. The DII loop in both *P. vivax* AMA1 and our FVO *Pf*AMA1 (Fig. 1A) structure is disordered. This is in contrast to the published 3D7 *Pf*AMA1 structure where only five residues are missing from the DII loop (residues 383 – 387) (Fig. 1B). However, there was missing density for DII loop residues 370 – 387 in our MilliQ-treated 3D7 *Pf*AMA1 structure and residues 351 – 387 in the methanol-treated 3D7 *Pf*AMA1 structure, similar to our FVO *Pf*AMA1 structure (Fig 1B and 3). MD simulations showed that, although the *N*- and *C*-termini of the DII loop are highly mobile for each FVO *Pf*AMA1, 3D7 *Pf*AMA1 and *P. vivax* AMA1 (Supplementary Fig. 6), large conformational changes or movements of the α-helix at the center of the DII loop were not observed in any of these proteins. This implies that, despite its flexibility, the DII loop undergoes slow conformational exchange, beyond the time scale used in our MD simulations.

The original 3D7 *Pf*AMA1 structure (PDB ID: 1Z40) provides support for the DII loop being ordered as a consequence of crystal contacts.[23] Our investigation of these contacts found that there were 9 and 13 residues from neighboring molecules in close proximity (< 4.0 Å) to the DII loop of the reference 3D7 *Pf*AMA1 chain A and E, respectively (Supplementary Fig. 7A). In particular, we found that both Glu354 (Oε1) and His356 (NH) of chain E formed hydrogen bonds with the main and side chains of Ser423 of a symmetry-related molecule (Supplementary Fig. 7A). The DII loop of chain A was stabilized by polar interaction between Lys363 (Nζ) and Asp317 of a symmetry-related molecule. In our FVO structure, there are no crystal contacts close to the DII loop, possibly allowing the loop to populate different positions within the crystal lattice (Supplementary Fig. 7B).

NMR studies of FVO and 3D7 *Pf*AMA1 were undertaken in order to further assess their flexibility in solution. Both 2D ^{1}H-^{15}N HSQC and 3D HNCO spectra were acquired. Of the 316 amide backbone NMR resonances expected in both forms of *Pf*AMA1, only 261 and 250 peaks were identified in the HSQC spectra of FVO and 3D7 *Pf*AMA1, respectively (Supplementary Fig. 8). The discrepancies between the expected and observed resonances are due to regions of the protein that undergo conformational exchange in the intermediate time scale (µs to ms).[60] NMR signals from regions with such motion are often broadened beyond detection in multi-dimensional NMR experiments. In the HNCO spectra, there were fewer resonances in 3D7 *Pf*AMA1 compared to FVO *Pf*AMA1 (248 and 216 peaks in FVO and 3D7 *Pf*AMA1, respectively, Supplementary Fig. 9). This difference suggests that 3D7 *Pf*AMA1 has a slightly greater number of backbone resonances that are broadened by intermediate conformational exchange than FVO *Pf*AMA1. This is in contrast to what is predicted from comparison of the published 3D7 *Pf*AMA1 structure (PDB ID 1Z40), which is highly ordered throughout the molecule, with our crystal structures of both FVO *Pf*AMA1 and 3D7 *Pf*AMA1 in different solvents. In summary, these data indicate that in solution both FVO *Pf*AMA1 and 3D7 *Pf*AMA1 contain significant regions of disordered structure, some of which are undergoing conformational exchange on an intermediate time scale which produces substantial broadening of NMR resonances.[60]

Mapping the Strain Variation of the Inhibitory Peptide R1. The R1 peptide,[6, 29] identified by phage display, inhibits red cell invasion by merozoites of 3D7 *P. falciparum* and related strains with a 50 % inhibitory concentration (IC_{50}) of ~ 0.1 µM. X-ray crystal structures[23, 25] show that R1 contacts three polymorphic residues (Tyr175, Met224 and Ile225) in 3D7 *Pf*AMA1, with residues at position 175 and 225 being important determinants of R1 strain specificity. Substitution of these residues as in W2mef (I225E) or HB3 (Y175D and I225N) significantly reduced the

peptide's inhibitory effect.[9] FVO *Pf*AMA1 also has Y175D and I225N substitutions (Supplementary Fig. 2). Accordingly, we found that R1 binds weakly to FVO *Pf*AMA1, with an estimated K_d of ~ 500 µM (Supplementary Fig. 10).

In order to investigate why R1 binds so weakly to FVO *Pf*AMA1, we estimated the effect of mutations on R1 binding by FVO *Pf*AMA1. The crystal structure of the 3D7 *Pf*AMA1-R1 complex (PDB ID 3SRJ) was used as a template and four FVO *Pf*AMA1 sequence variations found in the R1 binding cleft (Y175D, M190I, D204N and I225N) were generated by FoldX 3.0[61, 62] to mimic an apparent FVO *Pf*AMA1-R1 binding structure (Fig. 5). Single mutations were also generated in order to compare the individual effects on the binding energy. Analysis of point mutations shows that the changes at position 175 and 225 (Y to D and I to N, respectively) had the largest effect on stability, with increases in free energy of 6 and 3 kcal/mol, respectively (Supplementary Table 4). These energy changes are presumably due to the loss of two hydrogen bonds between R1 and AMA1 at these two positions (Fig. 5). The change M190I resulted in a small increase in free energy (1 kcal/mol), suggesting that the increase in cleft volume and hydrophobicity resulting from the M190I change has an effect on R1 binding. Distal to the R1 interaction sites, the D204N change has no effect on the complex. Taken together, the four changes contributed to a total increase in ΔG by 10 kcal/mol and significant changes in the hydrophobic cleft, creating an unfavorable environment for the R1 peptide (Supplementary Table 4).

DISCUSSION

AMA1 is implicated in the invasion of host cells by malaria parasites as well as other apicomplexan parasites.[22, 25, 63, 64] Sequence comparison of FVO and 3D7 *Pf*AMA1 domains I and II identify 21 amino acid differences that occur exclusively at the polymorphic face of AMA1.[23, 33, 35] However,

despite earlier hypotheses[15, 21], our FVO structure shows that these changes do not influence the overall fold of AMA1. The Ia – f loops that surround most of the hydrophobic cleft in AMA1 all display some level of disorder in the electron density from the FVO *Pf*AMA1, 3D7 *Pf*AMA1 and *P. vivax* structures. This apparent loop flexibility coupled with the polymorphic residues found in most of these regions provides AMA1 with an effective means of restricting cross-strain inhibitory activities of various ligands such as the R1 peptide and pose a challenge to efforts to design a vaccine or therapeutic agents effective against a broad range of strains and species of *Plasmodium* parasites.

Polymorphic residues within the loop Id (C1L cluster) are important in mediating escape against AMA1 antibodies induced by *P. falciparum* infections or in vaccine trials.[23, 25, 65] Residue 197 appears to be one of the most important residue in AMA1 responsible for immune escape.[65, 66] In this study we have shown that residue changes, including a Gly at position 197, do not result in structural changes in this region. This includes the two-turn helix within C1L. Residues 196, 204 and 206 are not engaged in any polar interactions and are unlikely to be important in stabilizing the structure of the Id loop (Supplementary Table 5). The main-chain atoms of residues 197 and 200 in both strains form polar interactions with main-chain atoms of Thr194 and Lys203, respectively. Substitution of Glu by Gly at position 197 in FVO *Pf*AMA1 prevents the polar interaction between Thr194 [O] and Glu197 [Oγ1] observed in 3D7. The H200D change at position 200 results in an additional side chain electrostatic interaction between Asp200 [Oδ1] and Lys203 [Nζ] in the FVO form of AMA1. The extensive pattern of hydrogen bonds stabilizing the 1d loop is largely conserved in both strains (Fig. 6 and Supplementary Table 5). These interactions appear to stabilize the structure and permit radical changes at polymorphic sites in the 1d loop without significant conformational change. This implies that the polymorphic nature of this region does

not affect the structure of the protein and that immune escape arises largely from changes in properties of individual side chains.

Unlike the published structures of 3D7 *Pf*AMA1 and one 1F9-3D7 *Pf*AMA1 complex (PDB IDs 1Z40 and 2Q8A), in which the conserved DII loop is partially ordered, the DII loop of FVO *Pf*AMA1 is completely disordered. This is similar to another crystal form of the 1F9-3D7 *Pf*AMA1 complex (PDB ID 2Q8B)[27] and the *P. vivax* AMA1 structure (PDB ID 1W81).[59] The flexibility of the DII loop is linked to a conformational change that allows the protein to interact with its protein-binding partner, RON2 (PDB ID 3ZWZ).[27] By sequence and structural superposition, it appears that the intra-protein contacts involved in stabilizing the local secondary structure of the DII loop are maintained in FVO *Pf*AMA1. The differences observed between the crystal structures of the two forms of AMA1 reflect different individual conformational states captured under the different crystallization conditions rather than an inherently greater disorder in the FVO *Pf*AMA1 DII loop. This view is further supported by our 3D7 *Pf*AMA1 structures, in which different extents of crystal contacts gave rise to different degrees of order in the DII loop.

In conclusion, our results show that the overall structure, including the flexible nature of the DII loop, is conserved between the FVO and 3D7 forms of AMA1. The interacting interfaces between DI and the DII loop consist of invariant residues across all *P. falciparum* strains. The structural conservation of DI and DII appears likely to be conserved across all allelic forms of AMA1, and hence represents an attractive site for strain-transcending therapeutic interventions. Given that DII loop displacement is associated with AMA1-RON2 complex formation, it is conceivable that stabilizing the DII loop in its ordered state would be inhibitory to AMA1 function.[9] Recently, we have undertaken a fragment-based screening campaign against AMA1 to identify chemical

scaffolds capable of inhibiting the protein-protein interactions.[24] The structure described here will help guide the design of small-molecule inhibitors of AMA1 with broad strain specificity.

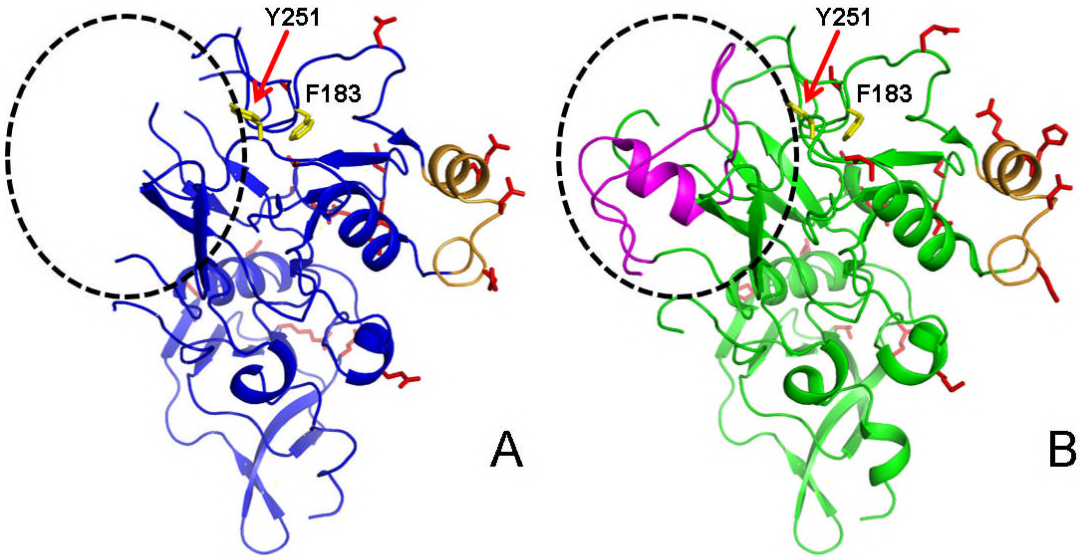


Figure 1. X-ray crystal structures of (A) FVO (PDB ID 4R1A) and (B) 3D7 (PDB ID 1Z40) *Pf*AMA1. Residues that vary between FVO and 3D7 *Pf*AMA1 proteins are shown as red sticks. The 1d loops are coloured orange in both structures. The DII loop is indicated by the dotted circle in both panels and the structure of the DII loop in 1Z40 is in magenta (B). Two residues critical for the AMA1-RON2 interaction (Phe183 and Tyr251) are shown as yellow sticks in both structures. The hydrophobic cleft runs across the top of the molecule in this view, from the Id loop at one end to the DII loop at the other.

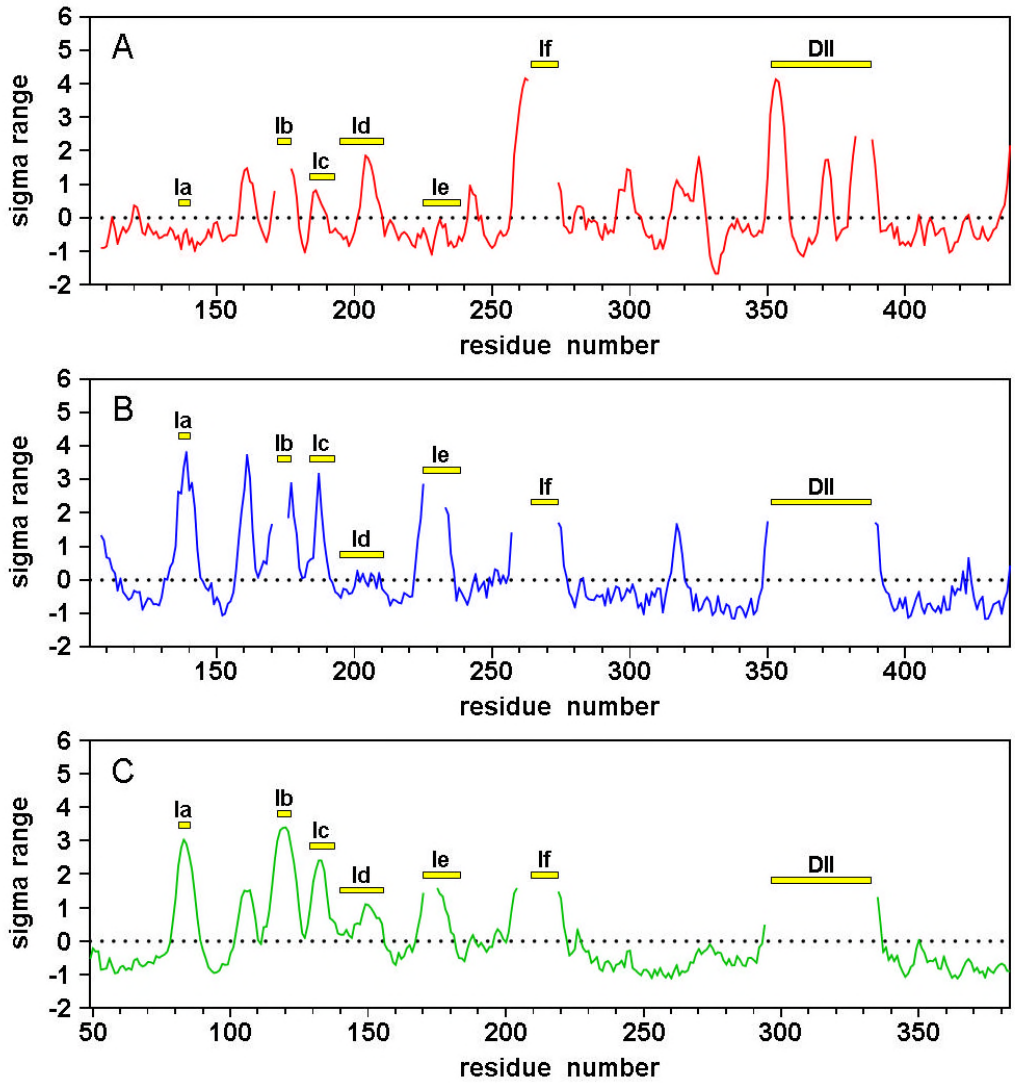


Figure 2. Flexibility of Ia – f and DII loops (identified with yellow bars). Normalized Cα B-factors of (A) 3D7 *Pf*AMA1 (red, PDB ID 1Z40), (B) FVO *Pf*AMA1 (blue, PDB ID 4R1A), and (C) *P. vivax* AMA1 (green, PDB ID 1W81). Residues 49 – 383 for *P. vivax* AMA1 sequence are equivalent to residues 104 – 438 in the FVO and 3D7 *Pf*AMA1 sequences.

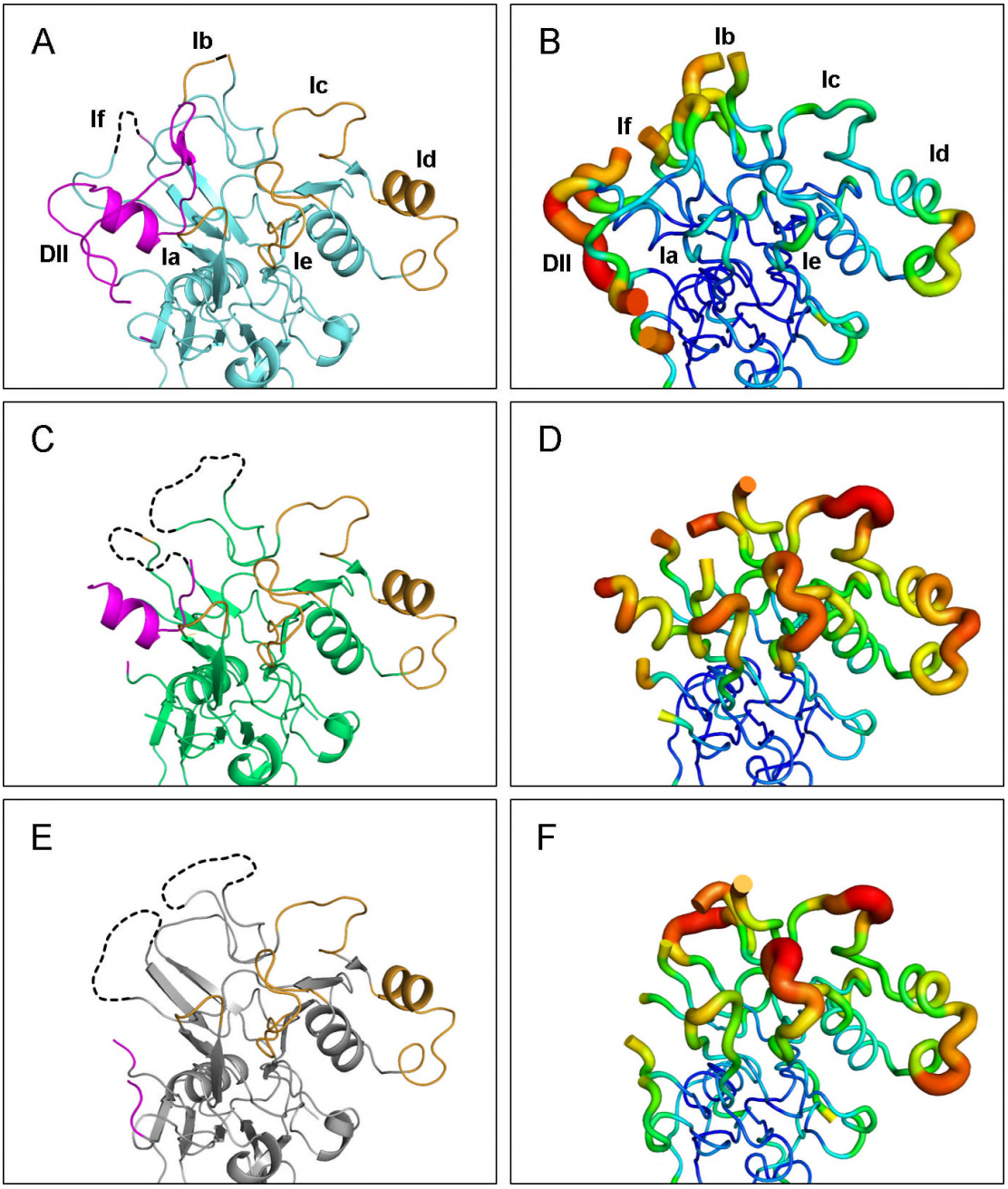


Figure 3. Cartoon model (left panel) and B-factor putty (right panel) of the X-ray crystal structures of 3D7 *Pf*AMA1. (A) and (B) 3D7 crystal structure from Bai *et al.*[23], (C) and (D) 3D7 crystal treated with MilliQ (PDB ID 4R1B), and (E) and (F) 3D7 crystal treated with methanol (PDB ID 4R1C).

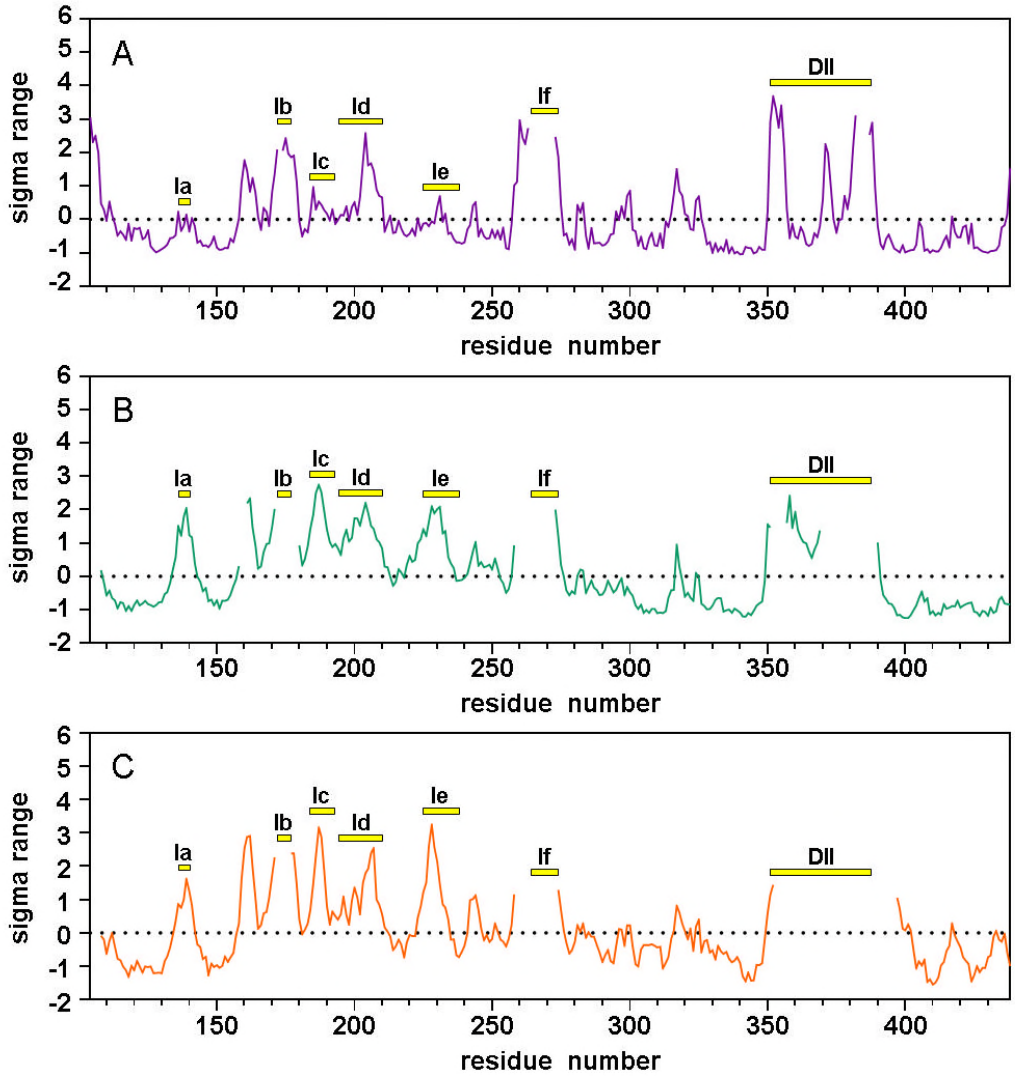


Figure 4. Flexibility of Ia – f and DII loops (identified with yellow bars). Normalized Cα B-factors of (A) 3D7 original condition (purple).[23], (B) 3D7 crystal treated with MilliQ (green-blue), and (C) 3D7 crystal treated with methanol (orange).

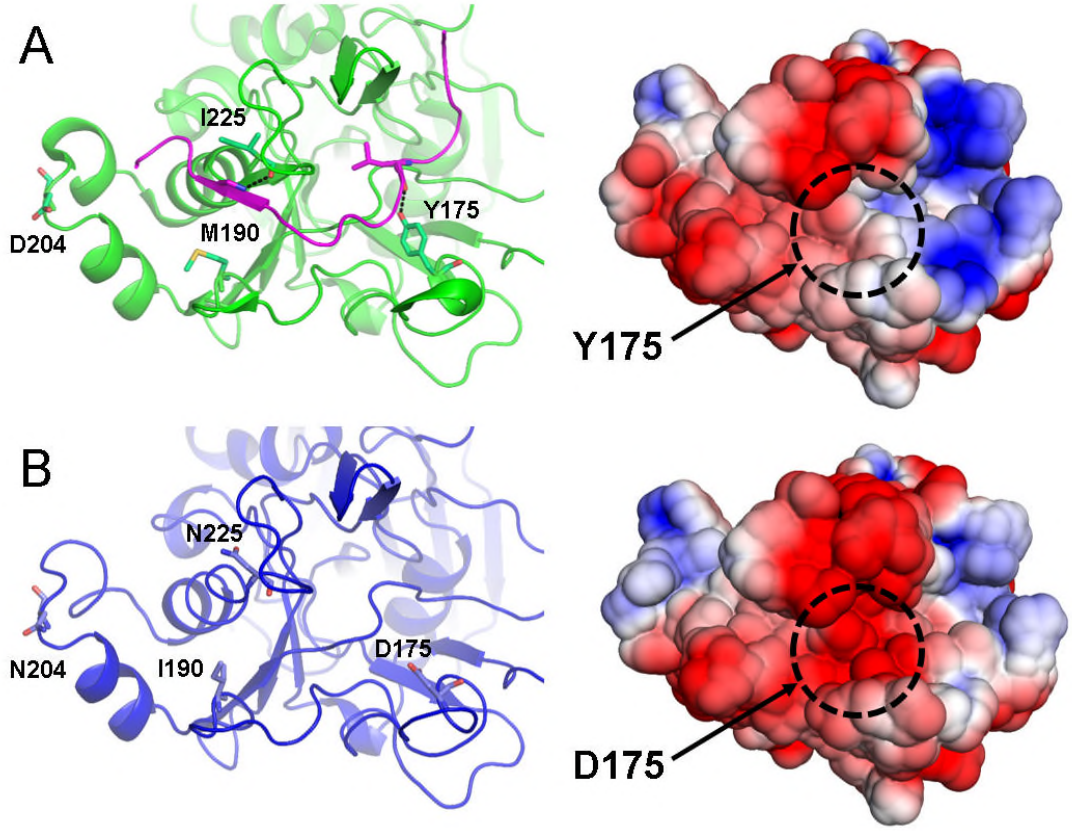


Figure 5. A comparison between R1-3D7 *Pf*AMA1 and free FVO *Pf*AMA1. (A) R1-3D7 *Pf*AMA1 crystal structure (PDB ID 3SRJ). (B) FVO *Pf*AMA1 model. Electrostatic potential mapped on the solvent accessible surface for the R1-bound *Pf*AMA1 structures are shown on the right side of the figure. Sequence variations between 3D7 and FVO *Pf*AMA1 are shown as sticks. R1 is shown as magenta ribbon. Protein surfaces are color-coded according to electrostatic potential gradient, where positively and negatively charged areas are represented in blue and red (iso-values from +3 k_bT/e_c to -3 k_bT/e_c), respectively.

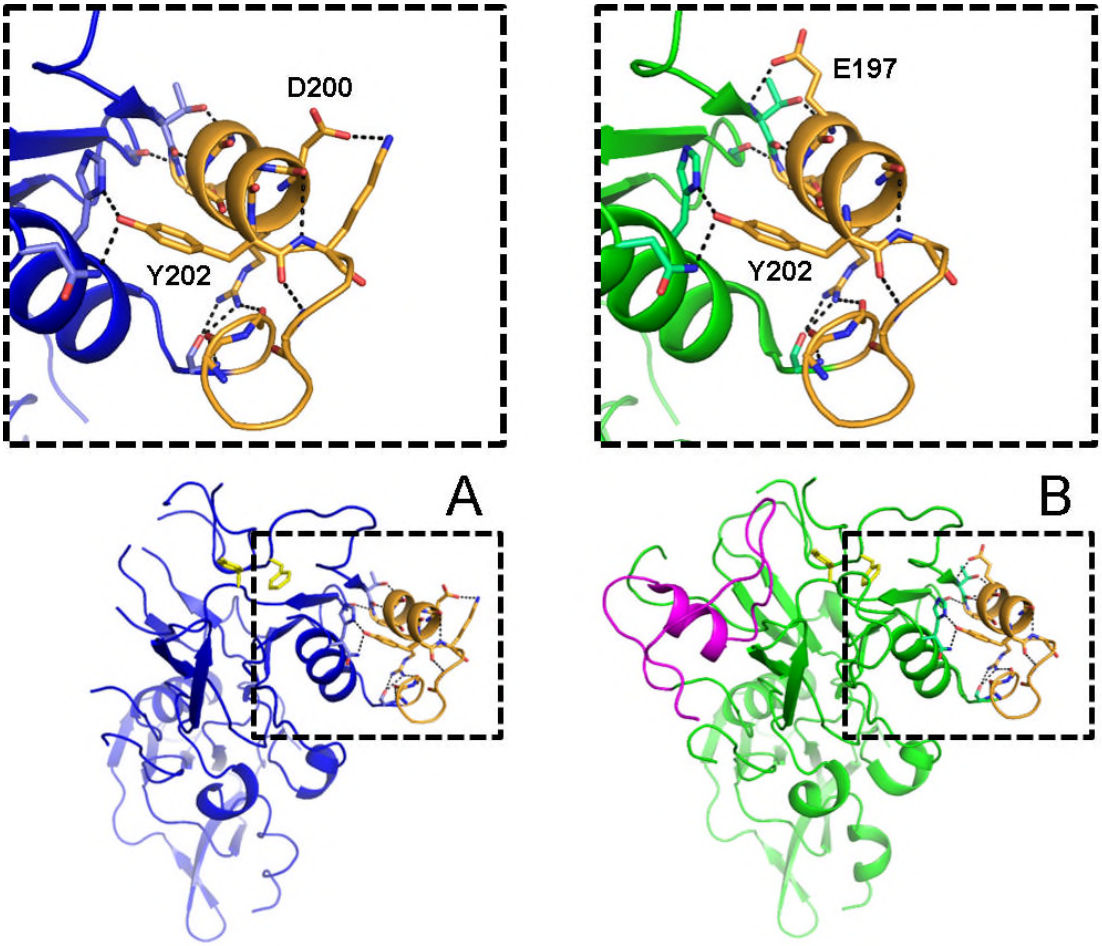


Figure 6. Cartoon showing polar interactions that stabilize the Id loop (orange) in (A) 3D7 (green) and (B) FVO (blue) *Pf*AMA1. The DII loop is coloured magenta. Residues Phe183 and Tyr251 are shown as yellow sticks.

ASSOCIATED CONTENT

Supporting Information

Supplementary material is accompanied with the publication. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Raymond S. Norton [redacted] and Sheena McGowan ([redacted]).

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

SM is an Australian Research Council Future Fellow (FT100100690). RSN acknowledges fellowship support from the National Health and Medical Research Council of Australia. This work was supported by the National Health and Medical Research (Project grant 1025150). This research was supported by the Victorian Life Sciences Computation Initiative Life Sciences Computation Centre (RAS Grant VR0316), a collaboration between Melbourne, Monash and La Trobe Universities and an initiative of the Victorian Government, Australia.

ACKNOWLEDGMENT

We thank Nyssa Drinkwater for technical assistance. We thank the Australian Synchrotron for beam-time and technical assistance.

ABBREVIATIONS

AMA1, apical membrane antigen 1; NMR, nuclear magnetic resonance; HSQC, heteronuclear single quantum coherence; Pf, *Plasmodium falciparum*; TROSY, Transverse relaxation optimized spectroscopy; MD, molecular dynamics; SPR, surface plasmon resonance; MWCO, molecular weight cut-off; TEV, Tobacco Etch Virus; LB, Luria broth; IPTG, Isopropyl β-D-1-thiogalactopyranoside; RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuation; DMSO, dimethyl sulfoxide; PEG, polyethylene glycol; MES, 2-(N-morpholino)ethanesulfonic acid; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; PME, Particle mesh Ewald.

REFERENCES

1. White, N. J., Dondorp, A. M., Faiz, A., Mishra, S., and Hien, T. T. (2012) New global estimates of malaria deaths, *Lancet 380*, 559-560.
2. World Health Organization (WHO). (2013) *World Malaria Report*.
3. Mendis, K., Sina, B. J., Marchesini, P., and Carter, R. (2001) The neglected burden of *Plasmodium vivax* malaria, *Am. J. Trop. Med. Hyg. 64*, 97-106.
4. Duffy, P. E., and Sibley, C. H. (2005) Are we losing artemisinin combination therapy already?, *Lancet 366*, 1908-1909.
5. Collins, C. R., Withers-Martinez, C., Bentley, G. A., Batchelor, A. H., Thomas, A. W., and Blackman, M. J. (2007) Fine mapping of an epitope recognized by an invasion-inhibitory monoclonal antibody on the malaria vaccine candidate apical membrane antigen 1, *J. Biol. Chem. 282*, 7431-7441.
6. Harris, K. S., Casey, J. L., Coley, A. M., Karas, J. A., Sabo, J. K., Tan, Y. Y., Dolezal, O., Norton, R. S., Hughes, A. B., Scanlon, D., and Foley, M. (2009) Rapid optimization of a peptide inhibitor of malaria parasite invasion by comprehensive N-methyl scanning, *J. Biol. Chem. 284*, 9361-9371.
7. Henderson, K. A., Streltsov, V. A., Coley, A. M., Dolezal, O., Hudson, P. J., Batchelor, A. H., Gupta, A., Bai, T., Murphy, V. J., Anders, R. F., Foley, M., and Nuttall, S. D. (2007) Structure of an IgNAR-AMA1 complex: targeting a conserved hydrophobic cleft broadens malarial strain recognition, *Structure 15*, 1452-1466.
8. Li, F., Dluzewski, A., Coley, A. M., Thomas, A., Tilley, L., Anders, R. F., and Foley, M. (2002) Phage-displayed peptides bind to the malarial protein apical membrane antigen-1 and inhibit the merozoite invasion of host erythrocytes, *J. Biol. Chem. 277*, 50303-50310.

9. Vulliez-Le Normand, B., Tonkin, M. L., Lamarque, M. H., Langer, S., Hoos, S., Roques, M., Saul, F. A., Faber, B. W., Bentley, G. A., Boulanger, M. J., and Lebrun, M. (2012) Structural and functional insights into the malaria parasite moving junction complex, *PLoS Pathog. 8*, e1002755.

10. Yap, A., Azevedo, M. F., Gilson, P. R., Weiss, G. E., O'Neill, M. T., Wilson, D. W., Crabb, B. S., and Cowman, A. F. (2014) Conditional expression of apical membrane antigen 1 in *Plasmodium falciparum* shows it is required for erythrocyte invasion by merozoites, *Cell. Microbiol. 16*, 642-656.

11. Triglia, T., Healer, J., Caruana, S. R., Hodder, A. N., Anders, R. F., Crabb, B. S., and Cowman, A. F. (2000) Apical membrane antigen 1 plays a central role in erythrocyte invasion by *Plasmodium* species, *Mol. Microbiol. 38*, 706-718.

12. Waters, A. P., Thomas, A. W., Deans, J. A., Mitchell, G. H., Hudson, D. E., Miller, L. H., McCutchan, T. F., and Cohen, S. (1990) A merozoite receptor protein from *Plasmodium knowlesi* is highly conserved and distributed throughout *Plasmodium*, *J. Biol. Chem. 265*, 17974-17979.

13. Hodder, A. N., Crewther, P. E., Matthew, M. L., Reid, G. E., Moritz, R. L., Simpson, R. J., and Anders, R. F. (1996) The disulfide bond structure of *Plasmodium* apical membrane antigen-1, *J. Biol. Chem. 271*, 29446-29452.

14. Nair, M., Hinds, M. G., Coley, A. M., Hodder, A. N., Foley, M., Anders, R. F., and Norton, R. S. (2002) Structure of domain III of the blood-stage malaria vaccine candidate, *Plasmodium falciparum* apical membrane antigen 1 (AMA1), *J. Mol. Biol. 322*, 741-753.

15. Anders, R. F., Adda, C. G., Foley, M., and Norton, R. S. (2010) Recombinant protein vaccines against the asexual blood stages of *Plasmodium falciparum*, *Hum. Vaccin. 6*, 39-53.

16. Thera, M. A., Doumbo, O. K., Coulibaly, D., Laurens, M. B., Ouattara, A., Kone, A. K., Guindo, A. B., Traore, K., Traore, I., Kouriba, B., Diallo, D. A., Diarra, I., Daou, M., Dolo, A., Tolo, Y., Sissoko, M. S., Niangaly, A., Sissoko, M., Takala-Harrison, S., Lyke, K. E., Wu, Y., Blackwelder, W. C., Godeaux, O., Vekemans, J., Dubois, M. C., Ballou, W. R., Cohen, J., Thompson, D., Dube, T., Soisson, L., Diggs, C. L., House, B., Lanar, D. E., Dutta, S., Heppner, D. G., Jr., and Plowe, C. V. (2011) A field trial to assess a blood-stage malaria vaccine, *N. Engl. J. Med. 365*, 1004-1013.

17. Ouattara, A., Mu, J. B., Takala-Harrison, S., Saye, R., Sagara, I., Dicko, A., Niangaly, A., Duan, J. H., Ellis, R. D., Miller, L. H., Su, X. Z., Plowe, C. V., and Doumbo, O. K. (2010) Lack of allele-specific efficacy of a bivalent AMA1 malaria vaccine, *Malaria J 9*.

18. Dutta, S., Dlugosz, L. S., Drew, D. R., Ge, X., Ababacar, D., Rovira, Y. I., Moch, J. K., Shi, M., Long, C. A., Foley, M., Beeson, J. G., Anders, R. F., Miura, K., Haynes, J. D., and Batchelor, A. H. (2013) Overcoming antigenic diversity by enhancing the immunogenicity of conserved epitopes on the malaria vaccine candidate apical membrane antigen-1, *PLoS Pathog. 9*, e1003840.

19. Miura, K., Herrera, R., Diouf, A., Zhou, H., Mu, J., Hu, Z., MacDonald, N. J., Reiter, K., Nguyen, V., Shimp, R. L., Jr., Singh, K., Narum, D. L., Long, C. A., and Miller, L. H. (2013) Overcoming allelic specificity by immunization with five allelic forms of *Plasmodium falciparum* apical membrane antigen 1, *Infect. Immun. 81*, 1491-1501.

20. Dutta, S., Lee, S. Y., Batchelor, A. H., and Lanar, D. E. (2007) Structural basis of antigenic escape of a malaria vaccine candidate, *Proc. Natl. Acad. Sci. USA 104*, 12488-12493.

21. Harris, K. S., Adda, C. G., Khore, M., Drew, D. R., Valentini-Gatt, A., Fowkes, F. J., Beeson, J. G., Dutta, S., Anders, R. F., and Foley, M. (2014) Immunodampening to overcome diversity in the malarial vaccine candidate apical membrane antigen 1, *Infect. Immun.*, In press.

22. Tonkin, M. L., Crawford, J., Lebrun, M. L., and Boulanger, M. J. (2013) *Babesia divergens* and *Neospora caninum* apical membrane antigen 1 structures reveal selectivity and plasticity in apicomplexan parasite host cell invasion, *Protein Sci. 22*, 114-127.

23. Bai, T., Becker, M., Gupta, A., Strike, P., Murphy, V. J., Anders, R. F., and Batchelor, A. H. (2005) Structure of AMA1 from *Plasmodium falciparum* reveals a clustering of polymorphisms that surround a conserved hydrophobic pocket, *Proc. Natl. Acad. Sci. USA 102*, 12736-12741.

24. Lim, S. S., Debono, C. O., MacRaild, C. A., Chandrashekaran, I. R., Dolezal, O., Anders, R. F., Simpson, J. S., Scanlon, M. J., Devine, S. M., Scammells, P. J., and Norton, R. S. (2013) Development of inhibitors of *Plasmodium falciparum* apical membrane antigen 1 based on fragment screening, *Aust. J. Chem. 66*, 1530-1536.

25. MacRaild, C. A., Anders, R. F., Foley, M., and Norton, R. S. (2011) Apical membrane antigen 1 as an anti-malarial drug target, *Curr. Top. Med. Chem. 11*, 2039-2047.

26. Srinivasan, P., Yasgar, A., Luci, D. K., Beatty, W. L., Hu, X., Andersen, J., Narum, D. L., Moch, J. K., Sun, H., Haynes, J. D., Maloney, D. J., Jadhav, A., Simeonov, A., and Miller, L. H. (2013) Disrupting malaria parasite AMA1–RON2 interaction with a small molecule prevents erythrocyte invasion, *Nat. Commun. 4*, 2261.

27. Coley, A. M., Gupta, A., Murphy, V. J., Bai, T., Kim, H., Foley, M., Anders, R. F., and Batchelor, A. H. (2007) Structure of the malaria antigen AMA1 in complex with a growth-inhibitory antibody, *PLoS Pathog. 3*, 1308-1319.

28. Coley, A. M., Parisi, K., Masciantonio, R., Hoeck, J., Casey, J. L., Murphy, V. J., Harris, K. S., Batchelor, A. H., Anders, R. F., and Foley, M. (2006) The most polymorphic residue on *Plasmodium falciparum* apical membrane antigen 1 determines binding of an invasion-inhibitory antibody, *Infect. Immun. 74*, 2628-2636.

29. Harris, K. S., Casey, J. L., Coley, A. M., Masciantonio, R., Sabo, J. K., Keizer, D. W., Lee, E. F., McMahon, A., Norton, R. S., Anders, R. F., and Foley, M. (2005) Binding hot spot for invasion inhibitory molecules on *Plasmodium falciparum* apical membrane antigen 1, *Infect. Immun. 73*, 6981-6989.

30. Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J., and Li, Q. (2009) Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*, *Protein Sci. 18*, 936-948.

31. Salzmann, M., Pervushin, K., Wider, G., Senn, H., and Wuthrich, K. (1998) TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins, *Proc. Natl. Acad. Sci. USA 95*, 13585-13590.

32. Riek, R., Pervushin, K., and Wuthrich, K. (2000) TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution, *Trends Biochem. Sci. 25*, 462-468.

33. Chesne-Seck, M. L., Pizarro, J. C., Vulliez-Le Normand, B., Collins, C. R., Blackman, M. J., Faber, B. W., Remarque, E. J., Kocken, C. H., Thomas, A. W., and Bentley, G. A. (2005) Structural comparison of apical membrane antigen 1 orthologues and paralogues in apicomplexan parasites, *Mol. Biochem. Parasitol. 144*, 55-67.

34. Gunasekera, A. M., Wickramarachchi, T., Neafsey, D. E., Ganguli, I., Perera, L., Premaratne, P. H., Hartl, D., Handunnetti, S. M., Udagama-Randeniya, P. V., and Wirth, D. F. (2007) Genetic diversity and selection at the *Plasmodium vivax* apical membrane antigen-1 (*Pv*AMA-1) locus in a Sri Lankan population, *Mol. Biol. Evol. 24*, 939-947.

35. Remarque, E. J., Faber, B. W., Kocken, C. H., and Thomas, A. W. (2008) Apical membrane antigen 1: a malaria vaccine candidate in review, *Trends Parasitol. 24*, 74-84.

36. Kabsch, W. (2010) Xds, *Acta Crystallogr. D Biol. Crystallogr. 66*, 125-132.

37. Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution?, *Acta Crystallogr. D Biol. Crystallogr. 69*, 1204-1214.

38. CCP4. (1994) The CCP4 suite: programs for protein crystallography, *Acta Crystallogr. D50*, 760-763.

39. Brunger, A. T. (1993) Assessment of phase accuracy by cross validation: the free *R* value. Methods and applications, *Acta Crystallogr. D Biol. Crystallogr. 49*, 24-36.

40. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C., and Read, R. J. (2005) Likelihood-enhanced fast translation functions, *Acta Crystallogr. D Biol. Crystallogr. 61*, 458-464.

41. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart., P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution., *Acta Crystallogr. D66*, 213-221.

42. Afonine P.V., Grosse-Kunstleve R. W., Echols N., Headd J. J., Moriarty N. W., Mustyakimov M., Terwilliger T. C., Urzhumtsev A., Zwart P. H., and D.., A. P. (2012)

Towards automated crystallographic structure refinement with phenix.refine., *Acta Cryst. D D68*, 352-367.

43. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr. D Biol. Crystallogr. 60*, 2126-2132.

44. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state, *J. Mol. Biol. 372*, 774-797.

45. Androulakis, S., Schmidberger, J., Bate, M. A., DeGori, R., Beitz, A., Keong, C., Cameron, B., McGowan, S., Porter, C. J., Harrison, A., Hunter, J., Martin, J. L., Kobe, B., Dobson, R. C., Parker, M. W., Whisstock, J. C., Gray, J., Treloar, A., Groenewegen, D., Dickson, N., and Buckle, A. M. (2008) Federated repositories of X-ray diffraction images, *Acta Crystallogr. D Biol. Crystallogr. D64*, 810-814.

46. Tronrud, D. E. (1996) Knowledge-based B-factor restraints for the refinement of proteins, *J. Appl. Crystallogr. 29*, 100-104.

47. Parthasarathy, S., and Murthy, M. R. N. (1997) Analysis of temperature factor distribution in high-resolution protein structures, *Protein Sci. 6*, 2561-2567.

48. Molecular Operating Environment (MOE), C. C. G. I., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2012.

49. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys. 79*, 926-935.

50. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD, *J. Comput. Chem. 26*, 1781-1802.

51. Case, D., Darden, T., Cheatham III, T., Simmerling, C., Wang, J., Duke, R., Luo, R., Walker, R., Zhang, W., and Merz, K. (2012) AMBER 12, *University of California, San Francisco.*

52. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins: Struct. Funct. Bioinform. 65*, 712-725.

53. Wickstrom, L., Okur, A., and Simmerling, C. (2009) Evaluating the performance of the ff99SB force field based on NMR scalar coupling data, *Biophys. J. 97*, 853-856.

54. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method, *J. Chem. Phys. 103*, 8577-8593.

55. Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation, *J. Chem. Theory Comput. 4*, 435-447.

56. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *J. Mol. Graphics 14*, 33-38.

57. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome, *Proc. Natl. Acad. Sci. USA 98*, 10037-10041.

58. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005) The FoldX web server: an online force field, *Nucleic Acids Res. 33*, W382-W388.

59. Pizarro, J. C., Vulliez-Le Normand, B., Chesne-Seck, M. L., Collins, C. R., Withers-Martinez, C., Hackett, F., Blackman, M. J., Faber, B. W., Remarque, E. J., Kocken, C. H.,

Thomas, A. W., and Bentley, G. A. (2005) Crystal structure of the malaria vaccine candidate apical membrane antigen 1, *Science 308*, 408-411.

60. Ge, X., MacRaild, C. A., Devine, S. M., Debono, C. O., Wang, G., Scammells, P. J., Scanlon, M. J., Anders, R. F., Foley, M., and Norton, R. S. (2014) Ligand-induced conformational change of *Plasmodium falciparum* AMA1 detected using ^{19}F NMR, *J Med. Chem. 57*, 6419-6427.

61. Guerois, R., Nielsen, J. E., and Serrano, L. (2002) Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations, *J. Mol. Biol. 320*, 369-387.

62. Schymkowitz, J. W. H., Rousseau, F., Martins, I. C., Ferkinghoff-Borg, J., Stricher, F., and Serrano, L. (2005) Prediction of water and metal binding sites and their affinities by using the Fold-X force field, *Proc. Natl. Acad. Sci. USA 102*, 10147-10152.

63. Hehl, A. B., Lekutis, C., Grigg, M. E., Bradley, P. J., Dubremetz, J. F., Ortega-Barria, E., and Boothroyd, J. C. (2000) *Toxoplasma gondii* homologue of *Plasmodium* apical membrane antigen 1 is involved in invasion of host cells, *Infect. Immun. 68*, 7078-7086.

64. Lamarque, M., Besteiro, S., Papoin, J., Roques, M., Vulliez-Le Normand, B., Morlon-Guyot, J., Dubremetz, J. F., Fauquenoy, S., Tomavo, S., Faber, B. W., Kocken, C. H., Thomas, A. W., Boulanger, M. J., Bentley, G. A., and Lebrun, M. (2011) The RON2-AMA1 interaction is a critical step in moving junction-dependent invasion by apicomplexan parasites, *PLoS Pathog. 7*, e1001276.

65. Takala, S. L., Coulibaly, D., Thera, M. A., Batchelor, A. H., Cummings, M. P., Escalante, A. A., Ouattara, A., Traore, K., Niangaly, A., Djimde, A. A., Doumbo, O. K., and Plowe,

C. V. (2009) Extreme polymorphism in a vaccine antigen and risk of clinical malaria: implications for vaccine development, *Sci. Transl. Med. 1*, 2ra5.

66. Ouattara, A., Takala-Harrison, S., Thera, M. A., Coulibaly, D., Niangaly, A., Saye, R., Tolo, Y., Dutta, S., Heppner, D. G., Soisson, L., Diggs, C. L., Vekemans, J., Cohen, J., Blackwelder, W. C., Dube, T., Laurens, M. B., Doumbo, O. K., and Plowe, C. V. (2013) Molecular basis of allele-specific efficacy of a blood-stage malaria vaccine: vaccine development implications, *J. Infect. Dis. 207*, 511-519.

Supplementary Information

Structure and Dynamics of Apical Membrane Antigen 1 from *Plasmodium falciparum* FVO

San Sui Lim[1], Wei Yang[2], Krishnarjuna Bankala[1], Komagal Kannan Sivaraman[2], Indu R. Chandrashekaran[1], Itamar Kass [2,3], Christopher A. MacRaild[1], Shane M. Devine[1], Cael O. Debono[1], Robin F. Anders[1,4], Martin J. Scanlon[1,5], Peter J. Scammells[1], Raymond S. Norton[1], and Sheena McGowan[2]

[1] Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Victoria 3052, Australia.

[2] Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia.

[3] Victorian Life Sciences Computation Initiative Life Sciences Computation Centre, Monash University, Clayton, VIC 3800, Australia.

[4] Department of Biochemistry, La Trobe University, Bundoora, Victoria 3086, Australia.

[5] Centre of Excellence for Coherent X-Ray Science, Monash University, Parkville, Victoria 3052, Australia.

CONTENTS

Supplementary Figure 8. ^{1}H-^{15}N HSQC spectra of FVO and 3D7 *Pf*AMA1 domain I + II.

Supplementary Figure 9. ^{1}H-^{13}C projection of HNCO spectra of FVO and 3D7 *Pf*AMA1 domain I + II.

Supplementary Figure 10. R1-FVO *Pf*AMA1 binding affinity determined by surface plasmon resonance.

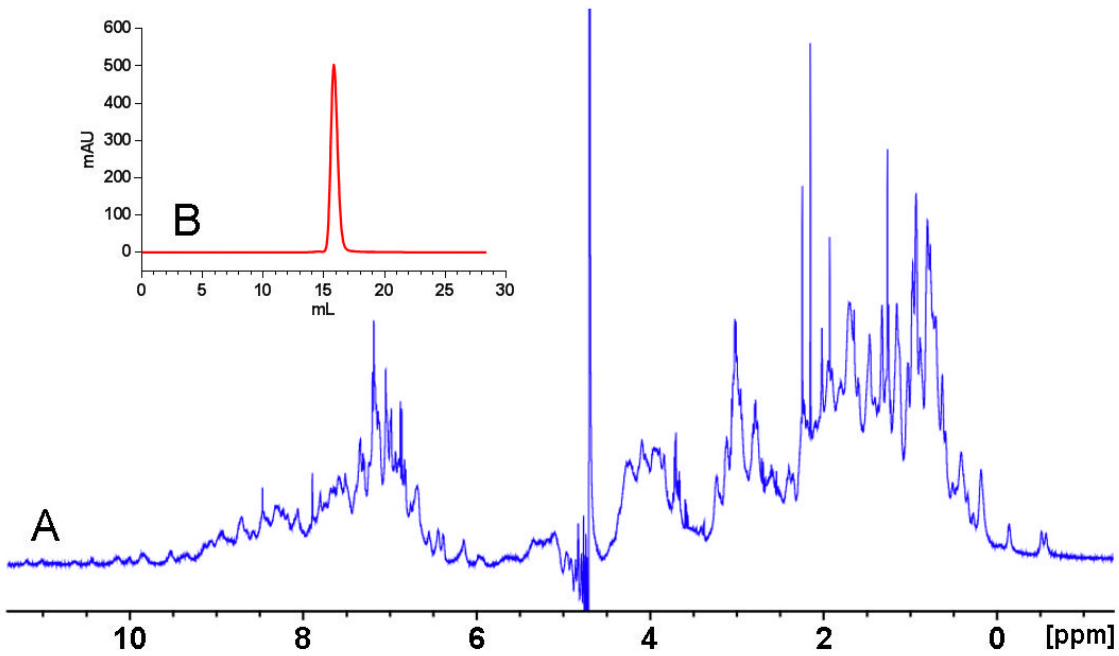
Supplementary Table 1. X-ray data collection and refinement statistics

Supplementary Table 2. Average values of normalized B-factors for 1a – f and DII loops

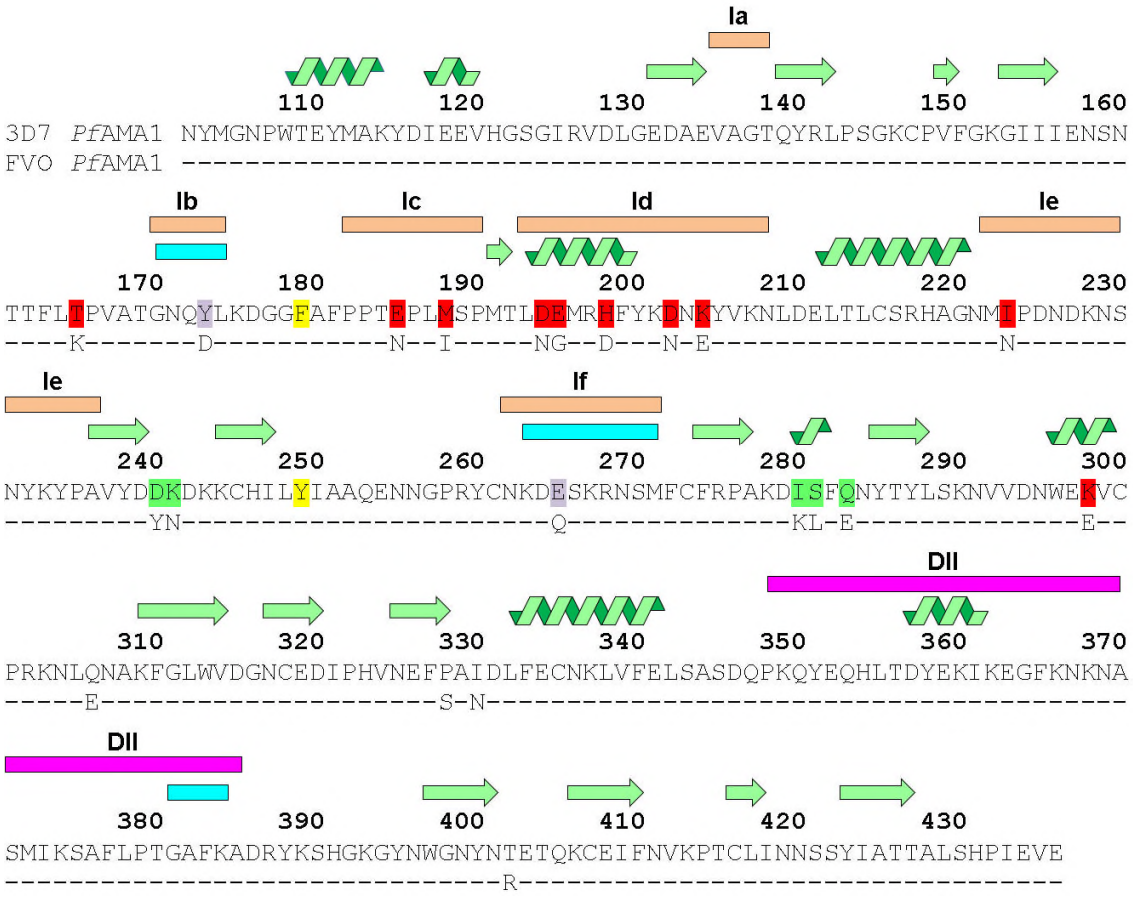
Supplementary Table 3. Crystal packing in AMA1 crystal structures

Supplementary Table 4. The FoldX calculated energy differences between the 3D7 *Pf*AMA1-R1 structure and the 5 mutate models

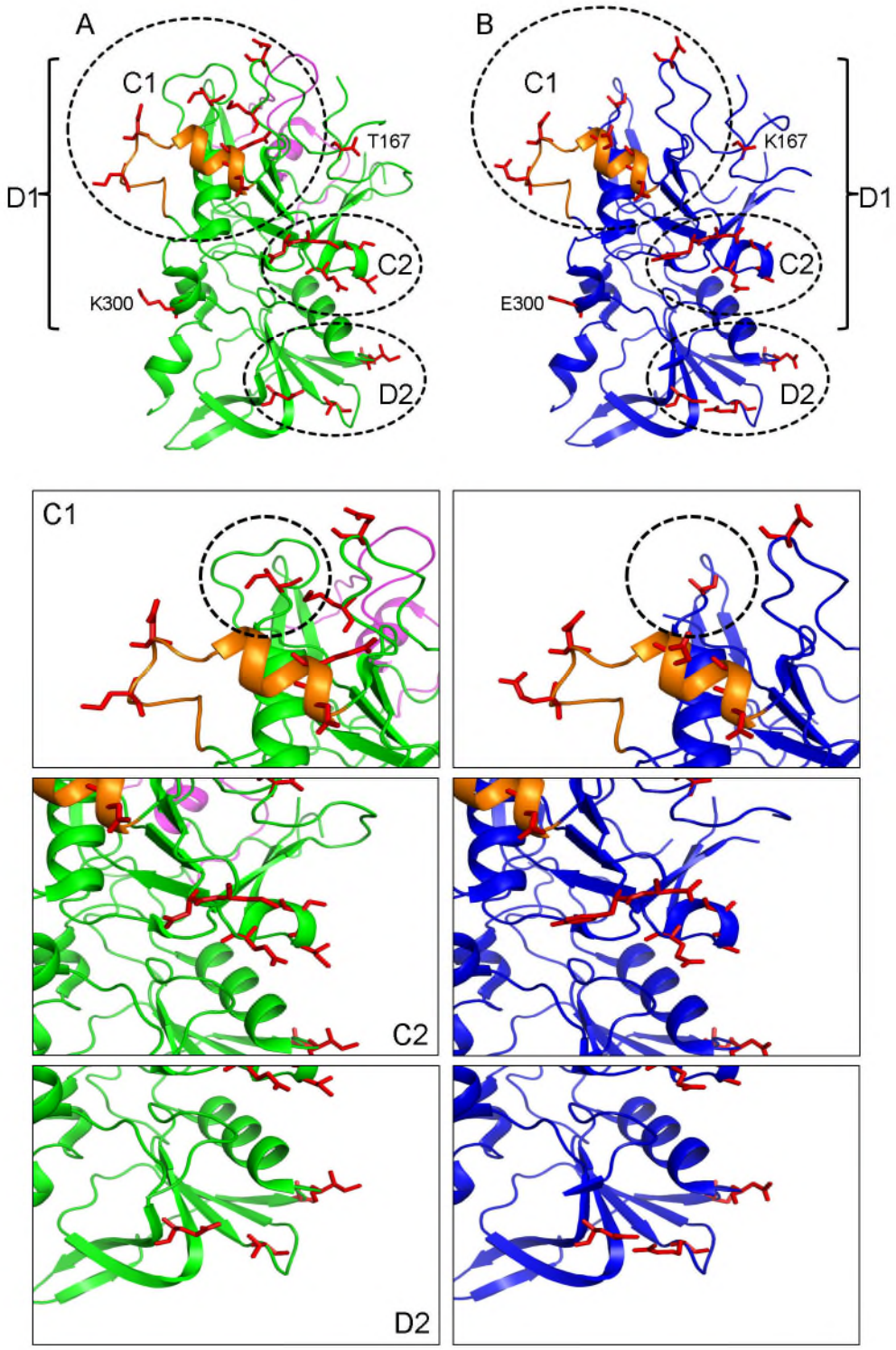
Supplementary Table 5. Polar interactions involved in stabilizing the 1d loops of 3D7 and FVO *Pf*AMA1



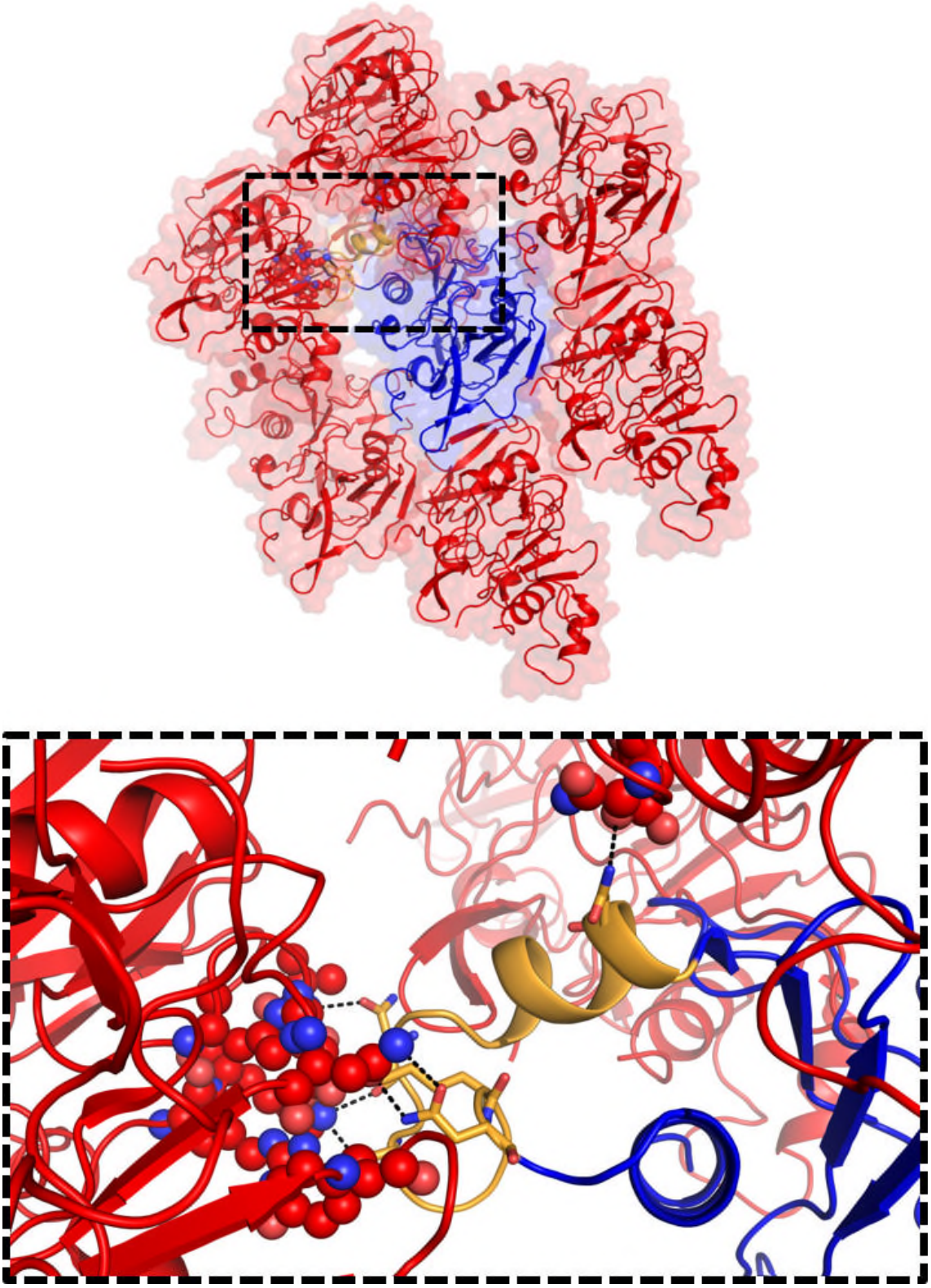
Supplementary Figure 1. Characterisation of recombinant FVO *Pf*AMA1 DI +II. (A) ^{1}H NMR spectrum acquired at 35 °C, pH 7.4 to evaluate the protein folding. (B) Size-exclusion chromatography confirmed that the protein is monomeric.



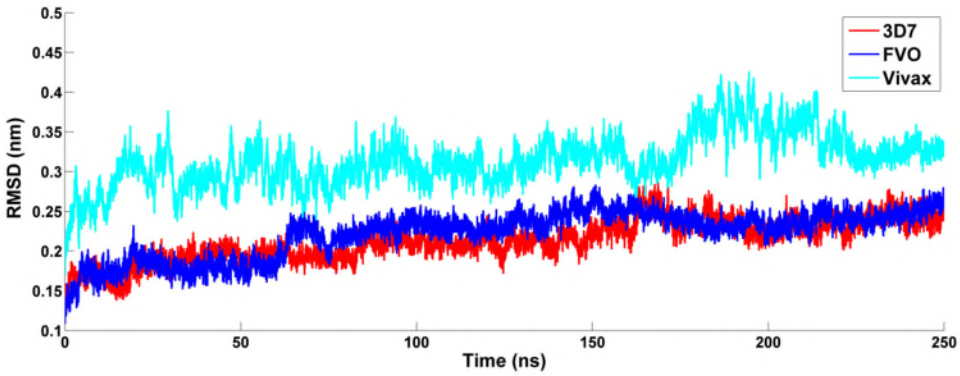
Supplementary Figure 2. Amino acid sequence of 3D7 *Pf*AMA1 (top) with positions of sequence variation in FVO shown directly underneath residue variant position. Secondary structure elements are shown above, with arrows depicting β-sheets, coils depicting α-helices and orange rectangles depicting loop regions. The disordered regions are coloured cyan. Phe183 and Tyr251 are highlighted in yellow. Polymorphic clusters. The domain I cluster is further sub-divided into C1 (residues 187, 190, 196, 197, 200, 204, 206, 225, boxed in red), C2 (residues 242, 243, 282, 283 and 285, boxed in green) and C3 (residues 175 and 267, boxed in lilac).



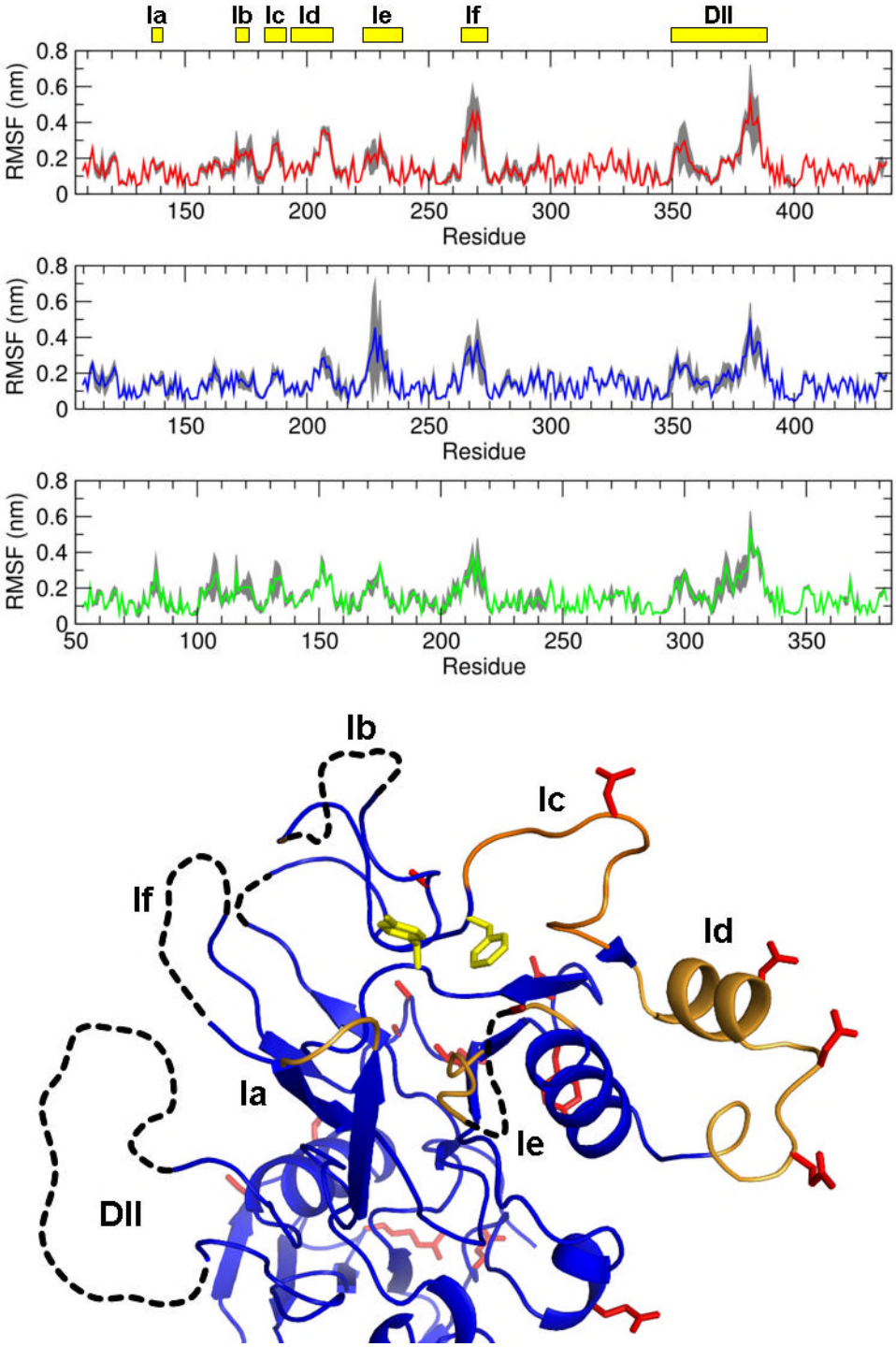
Supplementary Figure 3: Sequence variations between 3D7 (green) and FVO (blue) *Pf*AMA1 in the D1 and D2 polymorphic clusters. The D1 cluster consists of C1 and C2 clusters as well as residue 167 and 300. The Id loop (C1L) is coloured orange in both structures. Sequence variations are shown as red sticks. The DII loop is coloured magenta. Residues 226 – 232 are shown in dotted circles.



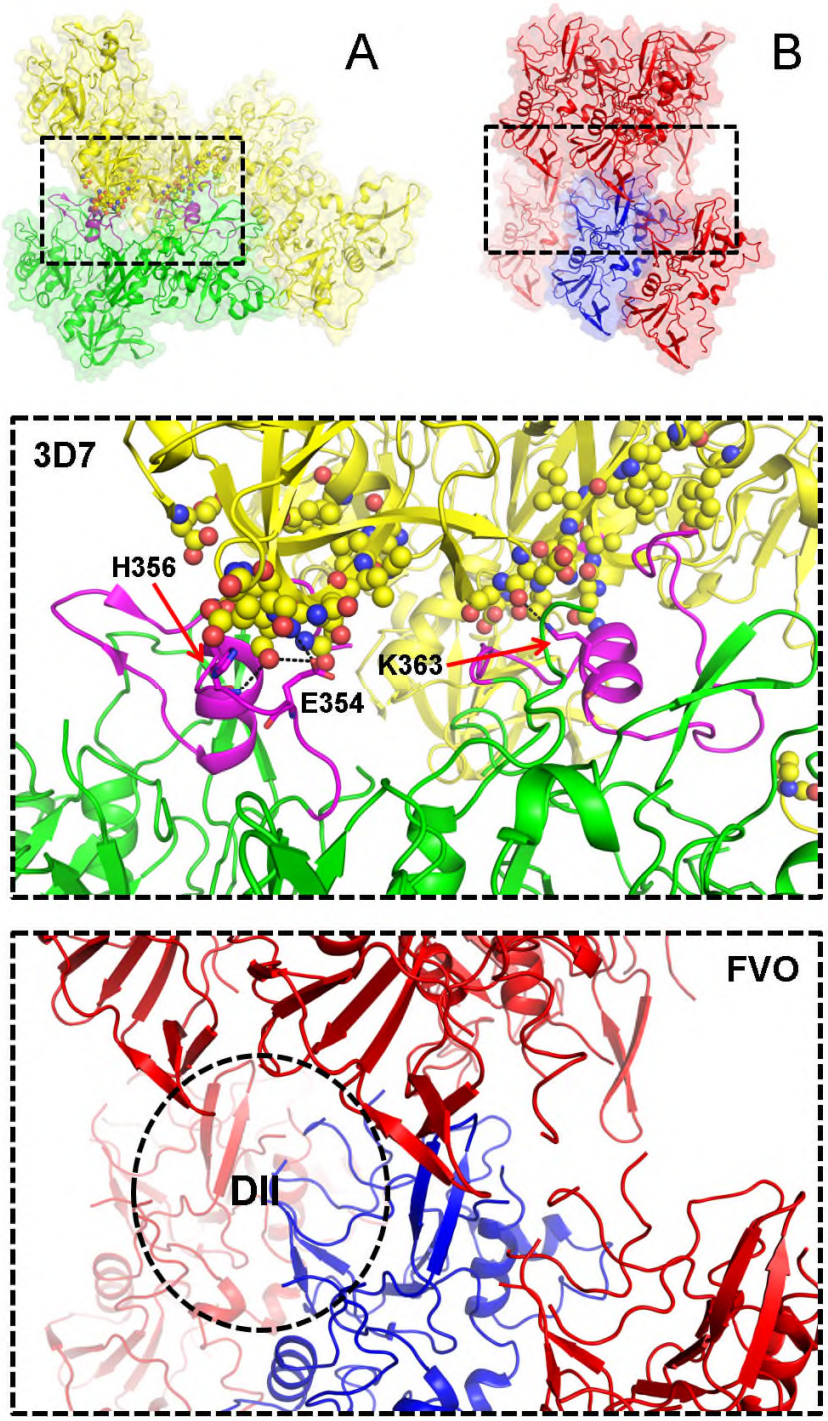
Supplementary Figure 4. Crystal packing of FVO *Pf*AMA1 Id loop (orange). Residues that are within 4 Å of the Id loop are shown as spheres.



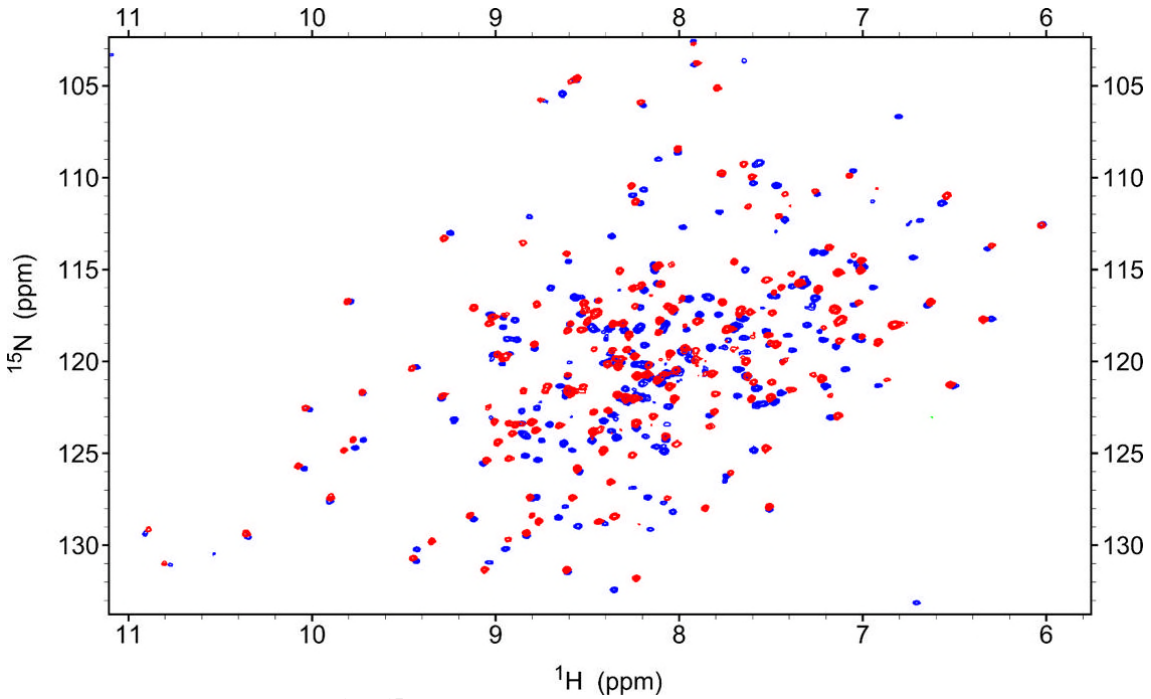
Supplementary Figure 5. Backbone Root-mean-square deviation (RMSD) of different systems studied. RMSD plot as a function of time of backbone atoms for representative trajectories of 3D7 *Pf*AMA1 (red), FVO *Pf*AMA1 (blue) and *P. vivax* AMA1 (cyan). This plot is produced by Matlab 2012b.[(2)]



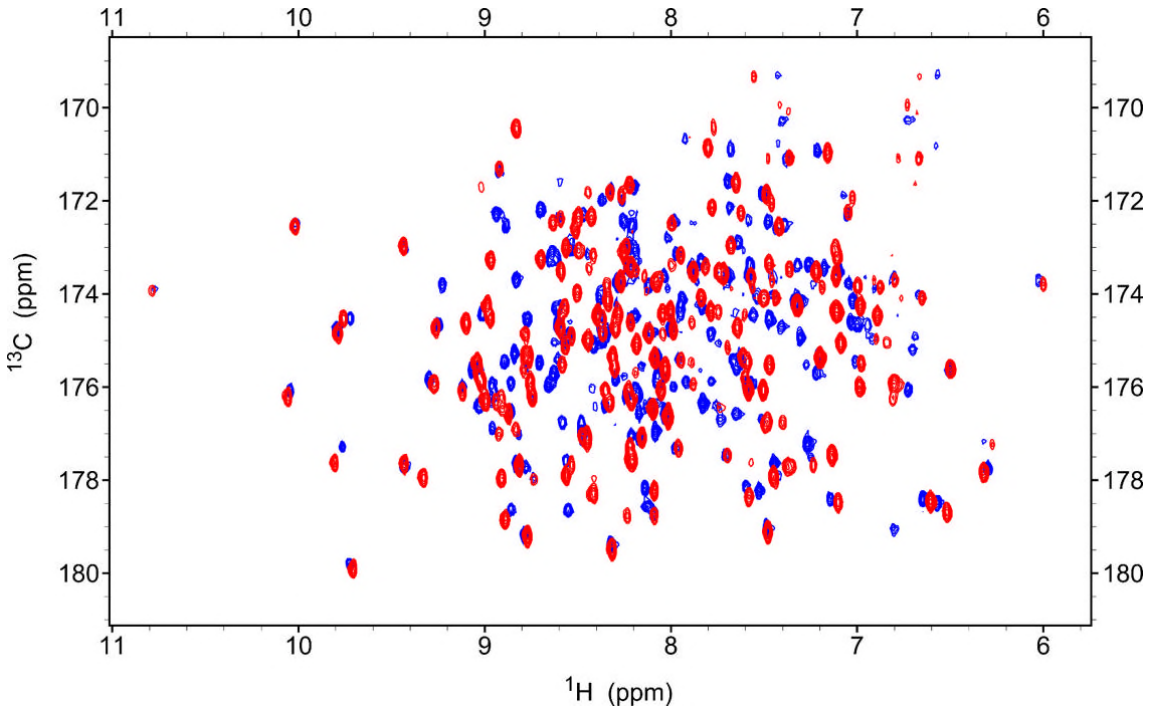
Supplementary Figure 6. Root-mean-square fluctuation (RMSF) of 3D7 *Pf*AMA1 (red), FVO *Pf*AMA1 (blue) and *P. vivax* AMA1 (green). All molecular dynamics (MD) simulations were performed in triplicate. The average values of RMSF – over three repeats of MD simulations – and its corresponding standard deviations are shown as coloured lines and grey shades, respectively. Crystal structure of FVO *Pf*AMA1with Ia – f and DII loops shown on the bottom. Red sticks represent sequence variations between FVO and 3D7 *Pf*AMA1. Yellow sticks correspond to Phe183 and Tyr 251. Dotted lines are disordered regions.



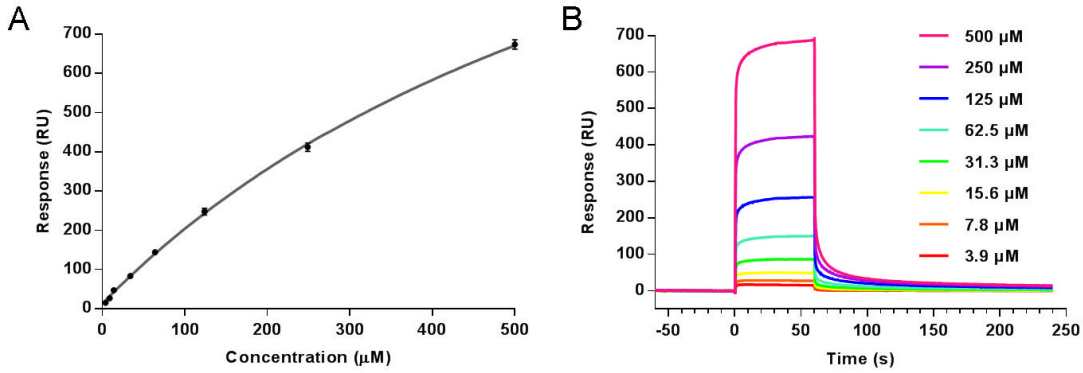
Supplementary Figure 7. Crystal packing against the *Pf*AMA1 DII loop. (A) The presence of crystal contacts in the 3D7 *Pf*AMA1 DII loop (PDB ID: 1Z40). The reference and neighbouring proteins are coloured green and yellow, respectively. The DII loops are highlighted in magenta and residues that are within 4 Å of the DII loops are shown as spheres. (B) The FVO *Pf*AMA1 DII loop is disordered in the absence of crystal packing (PDB ID 4R1A); no residues were found within 4 Å of the DII loops. There appears to be sufficient space in the lattice for the DII loop to adopt a closed conformation as in 3D7 *Pf*AMA1 (PDB ID: 1Z40) if this were its preferred conformation in FVO *Pf*AMA1. Blue and red cartoon structures represent the reference and neighbouring AMA1 molecules, respectively. Black dotted lines indicate the position of where the DII loop of FVO *Pf*AMA1 should be.



Supplementary Figure 8. ^{1}H-^{15}N HSQC spectra of FVO (blue) and 3D7 (red) *Pf*AMA1 domain I + II.



Supplementary Figure 9. ^{1}H-^{13}C projection of HNCO spectra of FVO (blue) and 3D7 (red) *Pf*AMA1 domain I + II.



Supplementary Figure 10. R1-FVO *Pf*AMA1 binding affinity determined by surface plasmon resonance. (A) Affinity fit for experiments performed in triplicate (B) SPR sensorgram. Although concentration-dependent responses were observed, super-stoichiometric interaction, with the ligand R_{max} (650 RU) exceeding the theoretical R_{max} (440 RU at immobilization level of ~ 8000 RU).

Supplementary Table 1. Data Collection and refinement statistics

Data collection	FVO	3D7 (original)	3D7 (5% MeOH)	3D7 (MQ)
Space Group	$C2_1$	$P3_1$	$C2_1$	$C2_1$
Cell dimensions (Å)	a=109.8 b=37.8 c=71.7, α = γ =90°; β = 94°	a=54.4 b=54.4 c=214.3, α = β = 90°; γ = 120°	a=156.5 b=54.5 c=67.9, α = γ = 90°; β = 91.4°	a=74.9 b=51.4 c=87.6, α = γ = 90°; β = 114.3°
Resolution (Å)	35.72 – 2.00 (2.05 – 2.00)	47.07 – 1.80 (1.84 – 1.80)	41.2 – 2.0 (2.05 – 2.0)	41.07 – 1.60 (1.63 – 1.60)
Total observations	75353	562849	182343	680213
Unique observations	19576	65700	38986	40241
Multiplicity	3.8 (4.0)	8.6 (8.5)	4.7 (4.7)	16.9 (9.1)
Data Completeness (%)	97.5 (96.4)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
$\langle I/\sigma_I \rangle$	5.1 (1.1)	6.7 (1.4)	9.7 (2.2)	11.7 (1.1)
CC(1/2)	0.987 (0.551)	0.994 (0.487)	0.996 (0.305)	0.997 (0.526)
R_{pim} (%)[b]	10.7 (67.9)	6.6 (80.2)	8.4 (119.0)	5.4 (80.7)
PDB	4R1A	4R19	4R1B	4R1C
Structure refinement				
Molecules in a.u.	1	2	2	1
Non hydrogen atoms				
Protein	2038	4959	4569	2188
Solvent (HOH)	145	452	317	300
R_{free} (%)	25.9	20.3	24.8	17.9
R_{cryst} (%)	20.9	17.2	20.3	20.7
Bond lengths (Å)	0.003	0.010	0.006	0.012

Bond angles (°)	0.65	1.17	0.98	1.30
Ramachandran plot	95.6	96.8	96.3	96.7
Favoured (%)	0.0	0.0	0.0	0.0
Outliers (%)				
B factors (Å^2)				
Mean protein	30.9	35.0	45.2	29.1
Mean water molecule	36.7	41.8	49.6	39.4
Molprobity Score[c]	1.10 (100th percentile (N= 12522, 2.0 Å ± 0.25 Å)	1.28 (98th percentile (N=11444, 1.8 Å ± 0.25 Å)	1.53 (96th percentile (N=12522, 2.0 Å ± 0.25 Å)	1.51 (89th percentile (N=7200, 1.6 Å ± 0.25 Å)

[a] Values in parentheses refer to the highest resolution shell.
[b] Agreement between intensities of repeated measurements of the same reflections and can be defined as: $\sum(I_{h,i} - \langle I_h \rangle)/\sum I_{h,i}$, where $I_{h,i}$ are individual values and $\langle I_h \rangle$ is the mean value of the intensity of reflection *h*.

Supplementary Table 2. Average values of normalized Cα B-factors for α-helical and β-sheet regions, and Ia – f and DII loops. D = disordered region with more than half the residues missing.

AMA1	α-helix	B-sheet	Ia (137 – 140)	Ib (172 – 176)	Ic (184 – 192)	Id (195 – 210)	Ie (224 – 238)	If (264 – 273)	DII (351 – 387)
3D7 (1Z40)	-0.30 ± 0.54	-0.37 ± 0.48	-0.64 ± 0.27	D	0.23 ± 0.47	0.34 ± 0.94	-0.58 ±0.30	D	0.75 ± 1.69
FVO (4R1A)	-0.36 ± 0.48	-0.16 ± 0.85	3.11 ± 0.58	D	1.12 ± 1.06	-0.12 ± 0.25	1.13 ± 1.26	D	D
***P. vivax* (1W81)**	-0.40 ± 0.42	-0.33 ± 0.76	2.82 ± 0.21	3.27 ± 0.17	1.62 ± 0.71	0.61 ± 0.32	0.72 ± 0.72	D	D
3D7 (4R19)	-0.27 ± 0.48	-0.51 ± 0.42	-0.19 ± 0.23	2.13 ± 0.20	0.34 ± 0.28	0.85 ± 0.74	-0.18 ± 0.38	D	0.78 ± 1.46
3D7 (4R1B)	-0.34 ± 0.68	-0.37 ± 0.58	1.16 ± 0.37	D	1.56 ± 1.05	1.18 ± 0.08	1.05 ± 1.22	D	D
3D7 (4R1C)	-0.06 ± 0.96	-0.43 ± 0.66	1.57 ± 0.41	D	1.81 ± 0.68	1.39 ± 0.43	1.14 ± 0.78	D	D

Supplementary Table 3. Crystal packing in AMA1 crystal structures. The number of residues from the neighbouring proteins in close vicinity (< 4.0 Å) of the Ia – f and DII loops of the respective AMA1 crystal structures. The numbers of polar contacts with symmetry-related molecules are indicated as bold numbers inside the brackets. Where no crystal packing was present, this is indicated as a long dash.

AMA1	**Ia (137 – 140)**	**Ib (172 – 176)**	**Ic (184 – 192)**	**Id (195 – 210)**	**Ie (224 – 238)**	**If (264 – 273)**	**DII (351 – 387)**
FVO (4R1A)	–	–	3 (**1**)	9 (**7**)	–	–	–
3D7 (1Z40)	–	–	5 (**1**)	7 (**1**)	5 (**2**)	–	9 (**1**)
***P. vivax* (1W81)**	–	1 (**0**)	1 (**0**)	3 (**0**)	3 (**2**)	–	–
3D7 (4R19)	–	–	6 (**0**)	4 (**1**)	5 (**2**)	–	12 (**2**)
3D7 (4R1B)	1 (**0**)	–	2 (**0**)	5 (**3**)	2 (**1**)	–	–
3D7 (4R1C)	–	–	–	3 (**0**)	1 (**0**)	–	–

Supplementary Table 4. The FoldX calculated energy differences between the 3D7-R1 structure and the 5 mutant models (Positive numbers indicate lower stability).

Model	Total energy change (kcal/mol)
Y175D	6
M190I	1
D204N	0
I225N	3
FVO model	10

* Insignificant change. Alteration is unlikely to affect stability

Supplementary Table 5. Polar interactions involved in stabilizing the α-helical and loop components of FVO and 3D7 *Pf*AMA1 1d loop (residues 195 to 210). Red text indicates residues that are different between the two strains.

FVO α-helix	**Interacting residues**	**Distance (Å)**	**3D7 α-helix**	**Interacting residues**	**Distance (Å)**
L195 [N]	K245 [O]	2.9	L195 [N]	K245 [O]	2.9
G197 [N]	T194 [Oγ1]	3.0	**E**197 [N]	T194 [O]	3.1
			E197 [N]	T194 [Oγ1]	3.2
			E197 [Oε1]	T194 [O]	3.0
M198 [N]	T194 [O]	3.0	M198 [N]	T194 [O]	2.9
R199 [NH1]	L211 [O]	2.8	R199 [NH1]	L211 [O]	2.7
R199 [NH1]	K209 [O]	2.7	R199 [NH1]	K209 [O]	2.8
R199 [NH2]	L211 [O]	2.9	R199 [NH2]	L211 [O]	3.0
D200 [O]	K203 [N]	3.1	**H**200 [O]	K203 [N]	3.2
D200 [Oδ1]	K203 [Nζ]	2.9			
Y202 [O]	N205 [N]	2.8	Y202 [O]	N205 [N]	2.9
Y202 [OH]	H220 [Nδ1]	2.7	Y202 [OH]	H220 [Nδ1]	2.6
Y202 [OH]	N223 [Oδ1]	3.3	Y202 [OH]	N223 [Oδ1]	3.0
FVO loop	**Interacting residues**	**Distance (Å)**	**3D7 loop**	**Interacting residues**	**Distance (Å)**
K203[N]	D200 [O]	3.1	K203[N]	H200 [O]	3.2
K203 [Nζ]	D200 [Oδ1]	2.9			
N205[N]	Y202 [O]	2.8	N205[N]	Y202 [O]	2.9
V208 [O]	L211 [N]	3.2	V208 [O]	L211 [N]	2.8
K209 [O]	R199 [Nη1]	2.7	K209 [O]	R199 [Nη1]	3.1

REFERENCE

1. Dutta, S., Lee, S. Y., Batchelor, A. H., and Lanar, D. E. (2007) Structural basis of antigenic escape of a malaria vaccine candidate, *Proc. Natl. Acad. Sci. USA 104*, 12488-12493.
2. Matheworks, T. (2012) Matlab 2012b, *Matheworks, Natick, MA.*

Chapter 5:

Mapping Specific Binding Site of Fragment on AMA1

Traditional Chapter

Mapping the fragment binding sites on apical membrane antigen 1 using high resolution NMR spectroscopy.

5.1 Introduction and Aims

Structural information of the fragment binding site may be obtained using either X-ray crystallography or NMR. The former is usually the method of choice as it provides detailed structural information about the fragment-protein interactions essential for designing potent AMA1 inhibitors.[82, 94] However, the study of fragment-protein crystal in a static environment may not always be fully informative given that the *in vivo* interactions occur in a dynamic solution environment.[95] ^{1}H-^{15}N HSQC chemical shift perturbation (CSP) of the target protein is a very useful NMR technique for studying compound binding sites in solution.[84, 96] In this method, the chemical shifts of each resonance follow changes in the local environment of the target protein upon fragment binding. This information is used to determine the specific binding sites of the compounds on the protein. CSP experiments can serve as a valuable complement to any crystallographic data. In situations where an accurate fragment binding pose is difficult to obtain, such as low resolution crystal structures or missing fragment electron densities, mapping the binding sites of fragment hits using ^{1}H-^{15}N HSQC perturbations will provide valuable support to the crystallographic data. Obviously, in cases where fragments can be neither soaked nor co-crystallised with the target protein, the ^{1}H-^{15}N HSQC perturbation experiments will be paramount in guiding the chemical elaborations of fragment hits.

The AMA1 hydrophobic cleft plays important role in the AMA1-RON2 complex formation.[37, 49, 54, 68] Ligands that target this region on AMA1 are capable of inhibiting red cell invasion by merozoites, but mostly in a strain-specific manner as numerous polymorphic residues are clustered at one end of the cleft.[34-36, 71, 73-75] Our fragment screening campaign have identified a range of scaffolds that bind to AMA1 hydrophobic cleft using R1 competition assay.[97] Whether these compounds interact with some polymorphic residues surrounding the cleft remains a question to be answered. Compounds that interact with polymorphic residues are generally poor starting points for chemical elaboration as further efforts are required to resolve this problem before a strain-transcending inhibitor of AMA1 can be developed. Therefore, it is critical to determine the specific binding sites of the R1-competing hits early in the lead discovery pipeline before investing large amount of time to develop a particular fragment scaffold. This part of the project aimed at mapping the specific binding sites of fragment hits using the ^{1}H-^{15}N HSQC perturbation experiments.

5.2 Methods

5.2.1 Protein Expression and Purification

Uniformly ^{15}N, ^{13}C, ^{2}H-labelled FVO *Pf*AMA1$_{[104-438]}$ was expressed and purified using a protocol described by Lim *et al*.[97] The uniformly ^{2}H and selectively ^{15}N-Lys labelled FVO *Pf*AMA1 were prepared by adding 200 mg/L of ^{15}N labelled Lys into the high-cell-density medium 1 h prior to IPTG induction. Subsequent work was similar to that for uniformly labelled sample.

5.2.2 Surface Plasmon Resonance

The binding activities of MIPS-0000865, MIPS-0001404, MIPS-0001160, MIPS-0000620, MIPS-0008939 and PJS-2156 were evaluated using Biacore T200 biosensor as described by Lim *et al*.[97]

5.2.3 NMR Spectroscopy

All NMR data were acquired with transverse relaxation optimised spectroscopy (TROSY) effects at 35 °C on a Bruker Avance III 600 MHz spectrometer. The buffer used for acquisition of NMR spectra was 20 mM phosphate buffer, pH 6.8, with 50 mM L-arginine, 50 mM L-glutamic acid, 0.2% (w/v) protease inhibitors cocktail (Roche), 0.01 % (w/v) sodium azide and 10 % (v/v) $^{2}H_2O$. ^{1}H-^{15}N HSQC spectra were acquired with 32 scans at 2048 and 256 data points for the ^{1}H and ^{15}N dimensions, respectively. HNCA and HN(CO)CA spectra were acquired with a sample concentration of 150 µM and 32 scans. HN(CA)CB and HN(COCA)CB spectra were acquired with a sample concentration of 250 µM and 64 scans. All 3D NMR experiments were recorded with 2048, 46 and 80 data points for ^{1}H, ^{15}N and ^{13}C dimensions, respectively. Data were processed using Topspin 3.2 and analysed using Sparky (v3.114, University of California, San Francisco). A QSINE window with a phase shift of 2 was applied to both the direct and indirect dimensions of ^{1}H-^{15}N HSQC and 3D NMR data. All indirect dimensions were linear predicted with 32 coefficients.

The stability of the sample at 35 °C was monitored using both ^{1}H NMR and ^{1}H-^{15}N HSQC experiments at days 0, 3 and 6. ^{1}H NMR spectra were acquired with 256 scans and 8192 data points using a pulse sequence with excitation sculpting for water suppression. ^{1}H-^{15}N HSQC spectra were acquired with 8 scans at 1024 and 64 data points for the ^{1}H and ^{15}N dimensions, respectively. The HSQC spectra were processed as above.

5.2.4 ^{1}H-^{15}N HSQC Perturbations

^{1}H-^{15}N HSQC spectra of ^{15}N, ^{2}H-labelled AMA1 in the presence and absence of the fragment were recorded at 35 °C with 128 scans. The concentrations of AMA1 and fragment were 25 µM and 2 mM, respectively. The buffer used in these experiments contained 20 mM phosphate buffer, pH 6.8, with 50 mM L-arginine, 50 mM L-glutamic acid, and 10 % (v/v) $^{2}H_2O$. Weighting of chemical shifts from ^{1}H and ^{15}N atoms was performed using the following equation[96]:

$$\text{Weighted CSP} = \sqrt{\frac{1}{2}[{\delta_{\mathrm{H}}}^2 + 0.14 \cdot {\delta_{\mathrm{N}}}^2]}$$

Where, δ_{H} and δ_{N} are the CSP (ppm) at ^{1}H and ^{15}N dimensions, respectively. Fragments were considered to be binding to a specific site on AMA1 only when more than one peak perturbation was observed.

5.3 Results

5.3.1 Backbone Resonance Assignments

In order to map the specific binding sites of the fragment hits with CSP, it was necessary to first assign the backbone amide resonances of *Pf*AMA1 on the HSQC spectrum. We have previously attempted the backbone amide resonance assignments of both 3D7 and FVO *Pf*AMA1 domain I and II. However, it was found that the 3D7 *Pf*AMA1 has a greater number of missing resonances that are broadened by intermediate conformational exchange than FVO *Pf*AMA1.[98] In addition, 3D7 *Pf*AMA1 exhibits poorer thermo stability than FVO *Pf*AMA1 in solution (personal communication with Bankala Krishnarjuna). These undesirable properties in 3D7 *Pf*AMA1 resulted in poor quality spectra of the 3D NMR experiments. Therefore, backbone resonance assignment was performed only for FVO *Pf*AMA1 in this study.

The FVO *Pf*AMA1 domain I and II construct consisted of 335 residues, including 19 proline residues, with a total mass of 38 kDa. A combination of protein deuteration[99-100] and TROSY scheme[101-102] in pulse sequences was used to improve sensitivity of all heteronuclear NMR experiments for AMA1 (Figure 1). FVO *Pf*AMA1 was stable in the NMR buffer over the acquisition period as shown by the ^{1}H NMR and ^{1}H-^{15}N HSQC spectra (Figure 2). The 2D ^{1}H-^{15}N HSQC spectrum displayed 261 of the 316 expected backbone amide resonances, corresponding to ~ 83 % of the total expected resonances (Figure 3). Overall, the ^{1}H-^{15}N HSQC spectrum showed well-dispersed spin-systems, indicating a single folded conformer. 3D HNCA, HN(CO)CA, HN(CA)CB and HN(COCA)CB spectra were acquired to obtain the

backbone resonance assignments (Figure 4). We observed 246 $C\alpha_i$ and 244 $C\alpha_{i-1}$ signals in the HNCA and HN(CO)CA, respectively. A total of 205 $C\beta_i$ and 202 $C\beta_{i-1}$ signals were observed in the HN(CA)CB and HN(COCA)CB, respectively. Fewer signals were observed in the Cβ-detected experiments (HN(CA)CB and HN(COCA)CB) because of the lower sensitivity of these experiments. The HSQC spectrum of ^{15}N-lysine labelled AMA1 was recorded to confirm unambiguous resonance assignments (Figure 5). Lysine was selected because it is the most abundant residues in AMA1 that can be selectively labelled using *E. coli* cells without the problem of metabolic scrambling.[103-106]

In addition to manual peak assignments, the assignments were also facilitated by the Resonance Assignment by chemical Shift Prediction (RASP) algorithm.[107] A total of 220 peaks could be assigned on the ^{1}H-^{15}N HSQC spectrum, which corresponds to 70 % of the expected peaks from AMA1 (Figure 3 and 6). Notably, 8 of the 12 residues in the AMA1 hydrophobic cleft, and 31 of the 40 polymorphic residues, including one of the most important residues in mediating immune escape against antibodies, Gly197, were assigned.[64-65] With the exception of Arg143 and Val151, the unassigned or undetectable residues correspond to loop regions in the FVO *Pf*AMA1 crystal structure[98] (104 – 108, 131 – 132, 144, 146 – 147, 159 – 167, 169, 227 – 236, 253 – 254, 257 – 259, 261 – 275, 291 – 292, 332, 349, 351 – 360, 365 – 380 and 382 – 393). These unassigned regions include the entire If loop, most of the Ie loop and the DII loops that surrounds the AMA1 hydrophobic cleft (Figure 6). Assignments could not be obtained in these regions because of weak or missing signals in the NMR spectra that were probably a result of signal broadening caused by intermediate conformational exchange. The full list of the assigned chemical shifts (HN, N, C_α and C_β) and the chemical shift index of C_α and C_β are shown in appendix II.

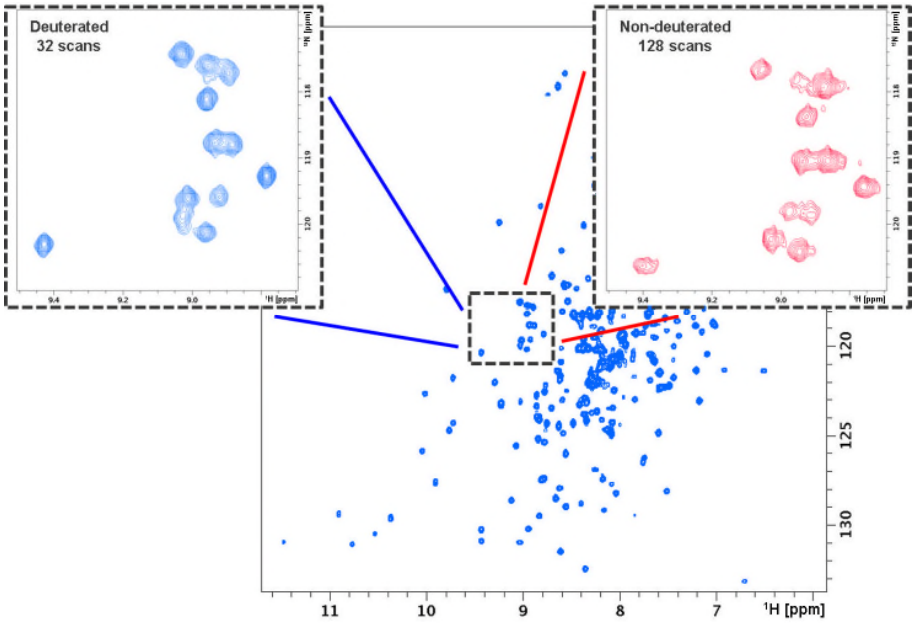


Figure 1. Comparison between ^{1}H-^{15}N HSQC spectra of deuterated (blue) and non-deuterated (red) FVO *Pf*AMA1.

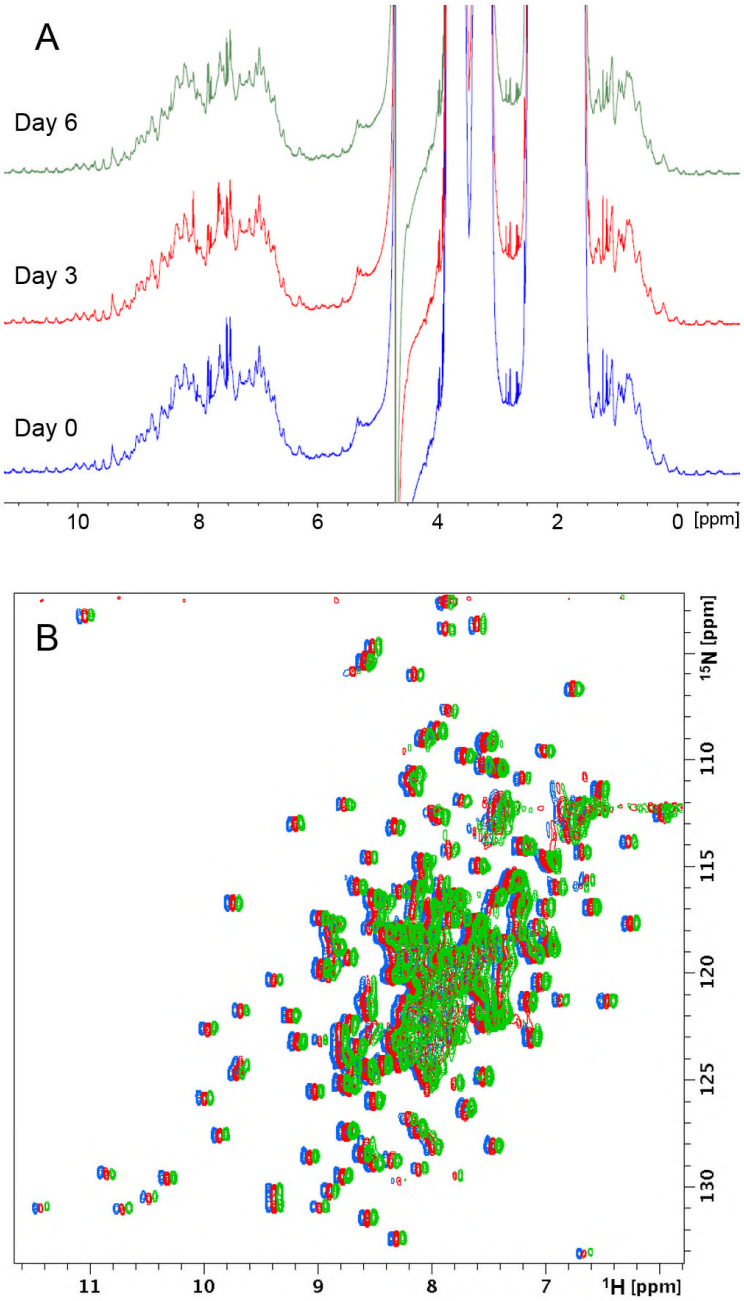


Figure 2. Stability studies of FVO *Pf*AMA1 at 35 °C using (A) ^{1}H NMR and (B) ^{1}H-^{15}N HSQC experiments. Spectra were acquired at Day 0 (blue), Day 3 (red) and Day 6 (green). For clarity, the HSQC spectra at Day 3 and 6 were offset by -0.1 and -0.2 ppm at ^{1}H dimension, respectively. Large signals at 1.5 to 4.0 ppm in the ^{1}H NMR spectrum correspond to those of L-arginine and L-glutamic acid found in the buffer. Sharp peaks at 7.0 to 8.0 ppm in (A) correspond to signals from the protease inhibitors cocktail (Roche).

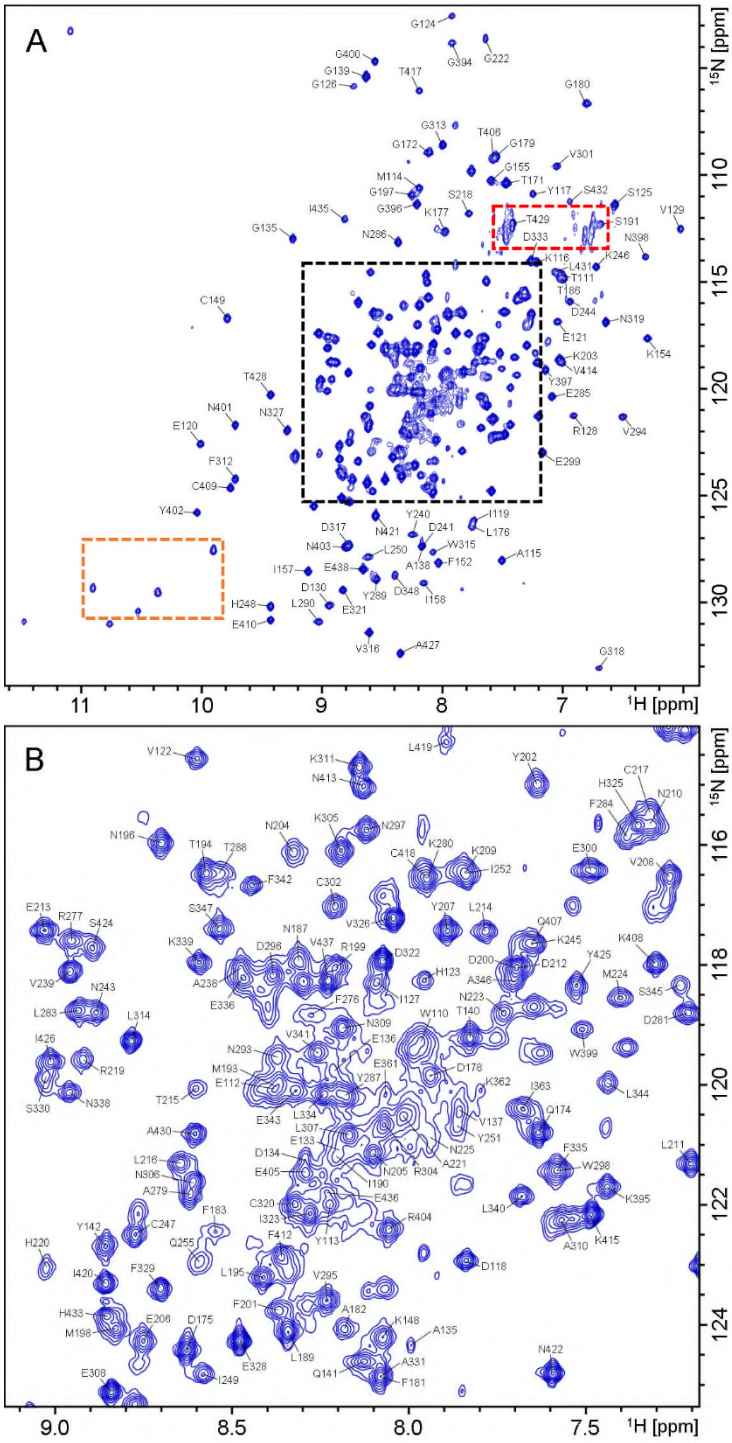


Figure 3. ^{1}H-^{15}N HSQC spectrum of FVO *Pf*AMA1 domain I and II with resonance assignments shown. (A) Full spectrum. (B) Enlarged view of the middle region (dotted box) in (A). The experiment was performed at 35 °C, pH 6.8 on a Bruker Avance III 600 MHz spectrometer. Trp Hε-Nε side chain peaks are indicated with orange dotted box. Asn and Gln NH_2 side chain peaks are indicated with red dotted box. Note that the backbone amide peaks of Thr429 and Ser191 are in the red dotted box.

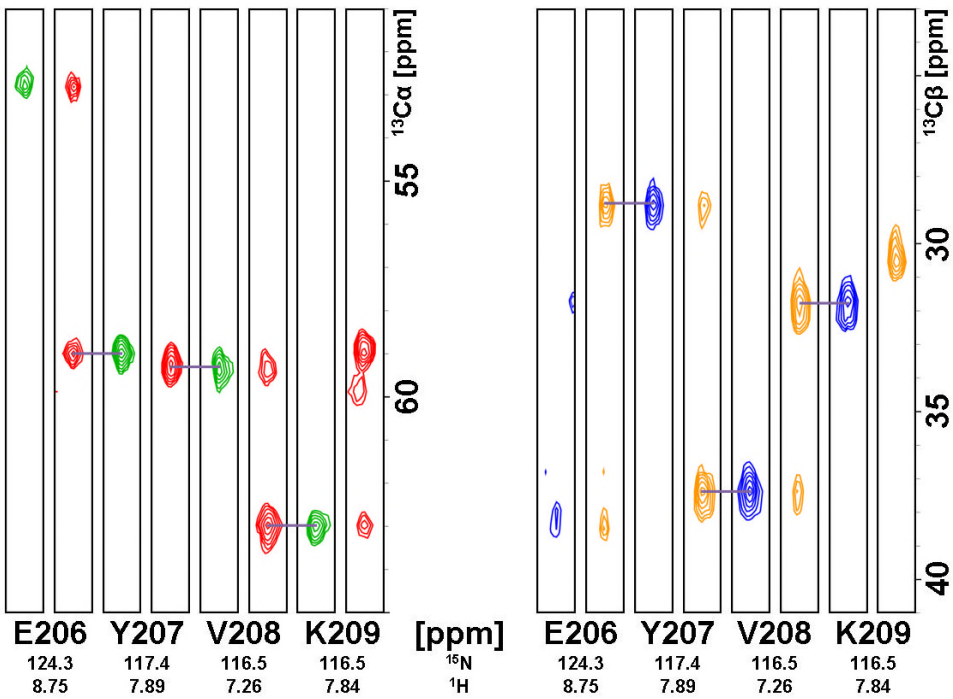


Figure 4. Strip plots for sequential assignment of FVO *Pf*AMA1. The sequential connectivities of C_α and C_β chemical shift from residues E206 to K209 are shown in left and right panels, respectively. HNCA, HN(CO)CA, HN(CA)CB and HN(COCA)CB spectra are indicated in red, green, orange and blue, respectively.

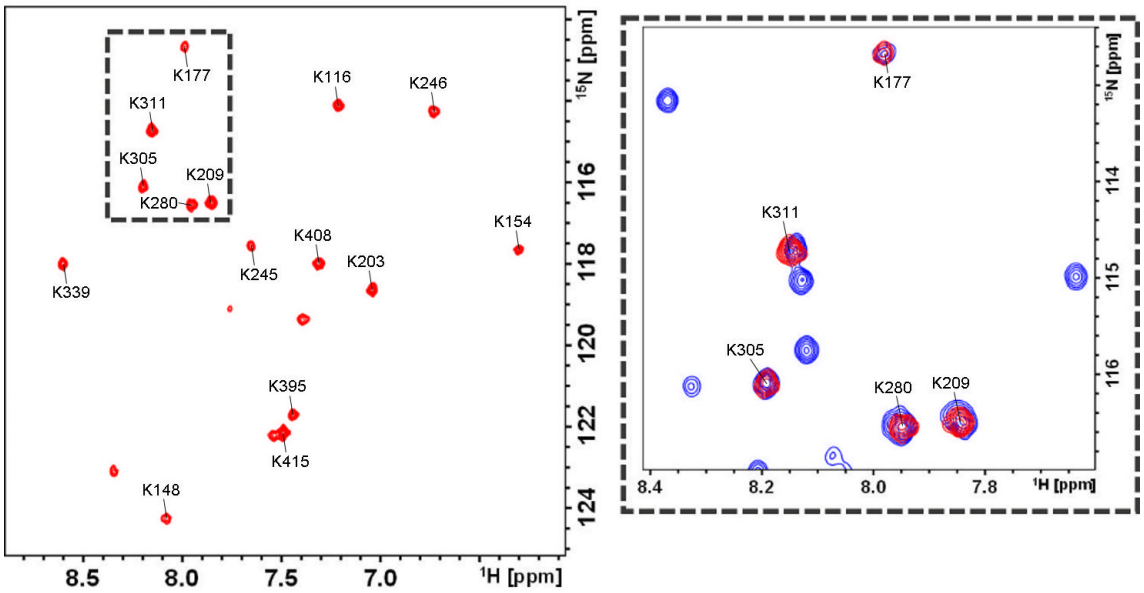


Figure 5. HSQC spectrum (red) of selectively ^{15}N-Lys labelled FVO *Pf*AMA1. Overlay of the HSQC spectra of uniformly (blue) and selectively ^{15}N-Lys (red) labelled samples.

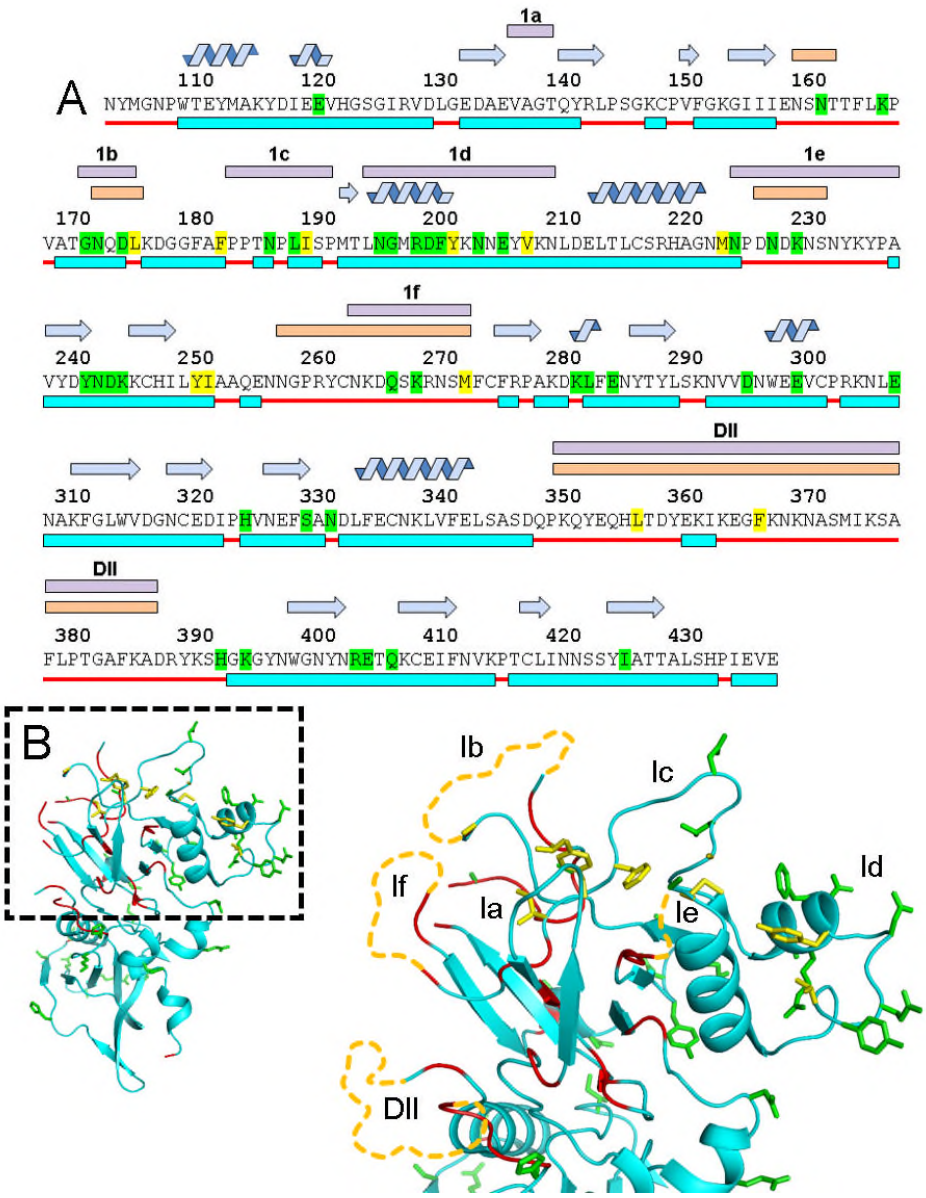


Figure 6. Backbone amide assignments of FVO *Pf*AMA1 domain I and II. (A) Amino acid sequence of FVO *Pf*AMA1 showing the positions of the secondary elements, Ia – f, and DII loops. The Ia – f and DII loops are defined with purple bars. The disordered regions are defined with orange bars. Residues that line the hydrophobic cleft are coloured yellow. The assigned and unassigned regions are shown as cyan bars and red lines, respectively. (B) The crystal structure of FVO *Pf*AMA1[98] (PDB ID 4R1A). Disordered regions are indicated with orange dotted lines. The residues that line the cleft are shown as yellow sticks. The polymorphic residues are shown as green sticks. The assigned and unassigned regions are coloured cyan and red, respectively.

5.3.2 ^{1}H-^{15}N HSQC Perturbations

Fragment classes that displayed structure-activity relationships (SAR) on AMA1 were chosen for further evaluation in this study. These scaffolds include benzimidazoles (MIPS-0000865 and MIPS-0001404), pyrazoles (MIPS-0001160 and PJS-2156) and 2-aminothiazoles (MIPS-0000620 and MIPS-0008939). All these compounds bind weakly to 3D7 *Pf*AMA1 with K_D values ranging from 0.5 to 4.0 mM, as determined by SPR (Table 1). Of the compounds tested in the ^{1}H-^{15}N HSQC experiments, all except for MIPS-0000620 induced chemical shift perturbations (CSPs) to more than one backbone amide resonance of AMA1 (Table 2). The numbers of peaks perturbed by MIPS-0000865, MIPS-0001404, MIPS-0001160, PJS-2156, and MIPS-0008939 range from 5 to 19 resonances (Table 2).

Table 1. SPR results for fragment hits.

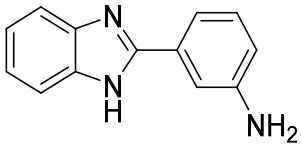
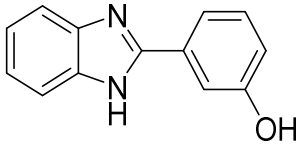
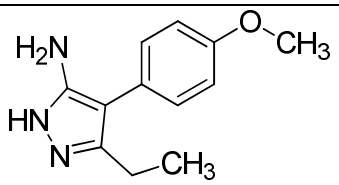
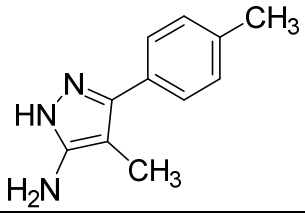
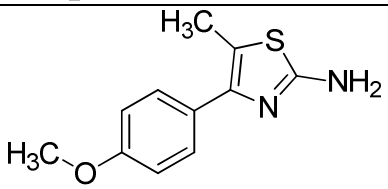
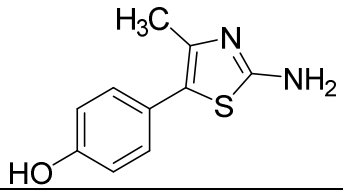
Compound ID	Structure	K_D (mM)
MIPS-0000865		0.5
MIPS-0001404		0.8
MIPS-0001160		4.1
PJS-2156		0.6
MIPS-0000620		1.0
MIPS-0008939		1.0

Table 2. Specific binding sites of fragment hits determined using ^{1}H-^{15}N HSQC perturbation. Site 1 and 2 correspond to the conserved and polymorphic ends of the AMA1 hydrophobic cleft, respectively. Site 3 – 5 are outside the cleft. Only one perturbation was observed for MIPS-0000620. Polymorphic residues are highlighted red. CSP values are shown in parentheses.

Binding regions	**MIPS 0000865**	**MIPS 0001404**	**MIPS 0001160**	**PJS 2156**	**MIPS 0000620**
Site 1	G180 (0.06166) G179 (0.04818) D178 (0.02671) K177 (0.02656) T171 (0.02381) I363 (0.02194) D175 (0.01848)	G180 (0.03598) K177 (0.02567) G179 (0.01752) I363 (0.01634) D178 (0.01501)	K177 (0.05669) G180 (0.03312) I363 (0.01893) G172 (0.01887) G179 (0.01880)	G172 (0.03304) G180 (0.02443) T171 (0.02253) I363 (0.02024)	G180 (0.04579) K177 (0.03717) G179 (0.03050) I363 (0.02695) Q255 (0.02455) N173 (0.02450) D178 (0.02053) D175 (0.01937) T171 (0.01873)
Site 2	V208 (0.02898) M224 (0.02140) N205 (0.01984) K209 (0.01984)	M224 (0.01762) V208 (0.01536)			F183 (0.01927) S191 (0.01760)
Site 3			I420 (0.02517) L419 (0.02063)		N422 (0.02879) I420 (0.01891) N421

					(0.01845)
Site 4					A310 (0.03765) E308 (0.02205)
Site 5					L307 (0.02038) N309 (0.01827)

MIPS-0000865 and MIPS-0001404 of the benzimidazoles series perturbed resonances that were mapped onto DII loop and polymorphic ends of the AMA1 hydrophobic cleft (Table 2). Larger number and magnitude of CSPs were found at the DII loop region for both compounds, suggesting that this region is the primary binding site. All resonances, except MIPS-0000865 interaction with polymorphic residue 175, perturbed by these two compounds in the DII loop region consist of conserved residues. The CSPs induced by MIPS-0000865 and MIPS-0001404 of the benzimidazoles gave the largest magnitude CSPs of the three series, and these results are exemplified in Figure 7. The perturbed residues are mapped on the crystal structure of FVO *Pf*AMA1 shown in Figure 8.

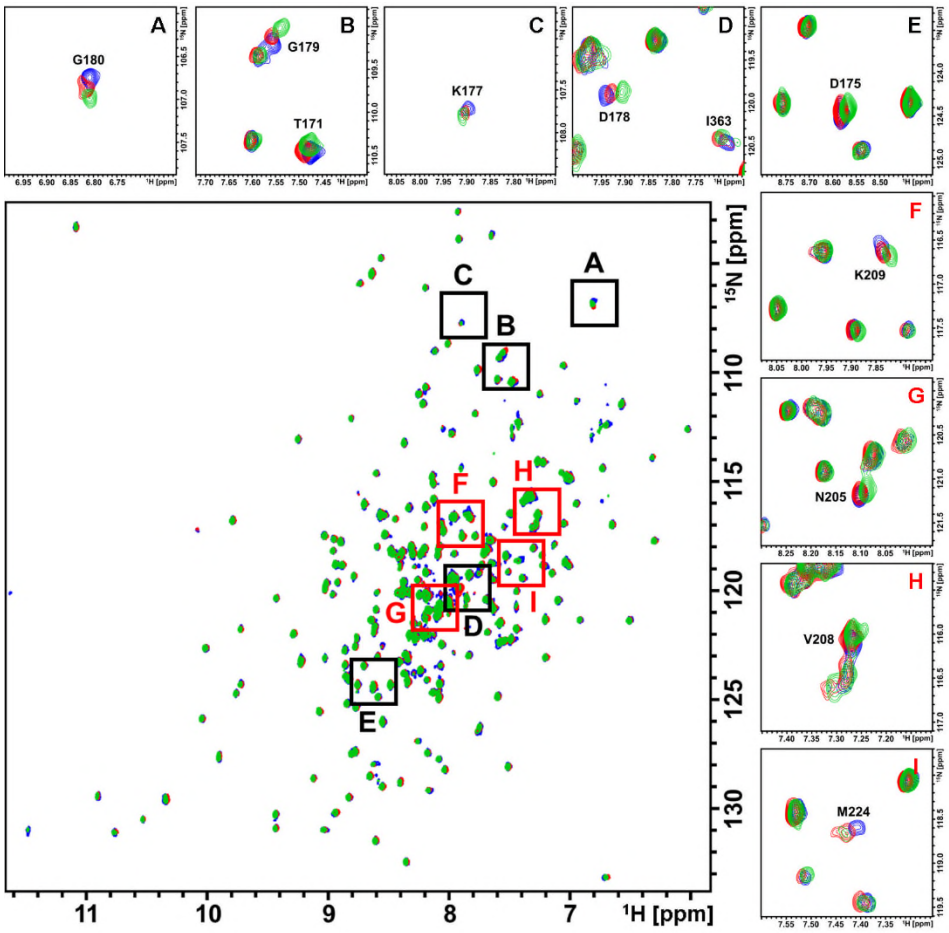


Figure 7. ^{1}H-^{15}N HSQC perturbations of FVO *Pf*AMA1 (blue) by MIPS-0001404 (green) and MIPS-0000865 (red). Perturbed residues in A – E and F – I were mapped onto the conserved and polymorphic ends of the hydrophobic cleft, respectively.

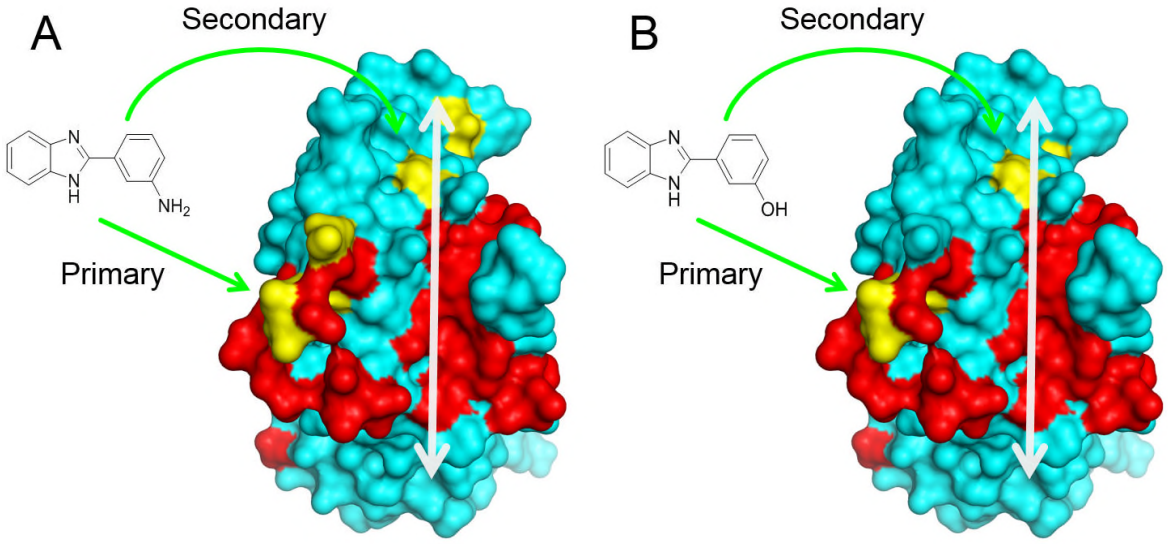


Figure 8. Mapping the ^{1}H-^{15}N HSQC perturbations of (A) MIPS-0000865 and (B) MIPS-0001404 on the FVO *Pf*AMA1 crystal structure (PDB ID 4R1A). The hydrophobic cleft is

indicated with white lines. The cyan and red regions correspond to assigned and unassigned regions, respectively. Residues that were perturbed by the compounds in the ^{1}H-^{15}N HSQC experiments are coloured yellow. Both compounds bind to two sites in the cleft, with primary and secondary binding sites correspond to conserved and polymorphic ends of the cleft, respectively.

MIPS-0001160 and PJS-2156 of the pyrazoles series also showed binding to the DII loop end of the cleft. Apart from their interactions with polymorphic residue 172, these fragments contact mostly conserved residues in this region on AMA1 similar to the case of benzimidazoles compounds (Table 2). It was noted that MIPS-0001160 interacts with residues 419 and 420 that are located outside the hydrophobic cleft.

MIPS-0000620 of the 2-aminothiazoles series induced CSPs at only one resonance, suggesting that there might be a possibility of non-specific interaction. This CSPs correspond to Thr171 at the DII end of the cleft that is also part of the binding site of R1 peptide. Previous study have found that MIPS-0000620 competes with R1 and this observation together with the ^{1}H-^{15}N HSQC perturbation results suggest a significant binding activity of MIPS-0000620 at the DII end of the hydrophobic cleft. Unlike MIPS-0000620, the other 2-aminothiazoles compound, MIPS-0008939, showed extensive CSPs from multiple sites on AMA1 that include both the DII loop and polymorphic ends of the cleft (Table 2). This observation is in agreement with our previous study which showed that the 2-aminothiazole compounds can be problematic promiscuous binders (see submitted manuscript in Appendix III).[108]

5.4 Discussion

In this study, we have established an approach to map the specific binding sites of the fragments on AMA1 in solution using ^{1}H-^{15}N HSQC perturbation experiments. This NMR method is a valuable tool in the FBLD approach in that it is a sensitive technique that can reliably detect compound with weak binding affinity with a low rate of false positives as well as providing structural information about the ligand-protein interactions.[84, 96] A relatively small amount of AMA1 protein (~ 0.3 mg) is required in this approach. The majority of the backbone amide resonances in the ^{1}H-^{15}N HSQC spectrum were assigned using 3D NMR and specific ^{15}N-lysine labelled HSQC experiments. Monitoring the perturbations of these assigned resonances that correspond to the conserved and polymorphic residues provides information about fragment binding across different alleles of AMA1.

The DII end of the cleft has been proposed as the site for small-molecule intervention owing to its largely conserved nature.[98] The benzimidazole and pyrazole series showed binding to this region, and therefore represent promising starting scaffolds for the design of AMA1 lead compounds. Although the 2-aminothiazole series also showed binding to the DII end of the cleft, their promiscuous binding as shown in previous[108] and current studies makes them poor starting points for subsequent fragment elaboration. MIPS-0000865 contacts residue 175, which is one of the very few polymorphic residues found at this end of the cleft. This site can vary between Asp and Tyr residues in AMA1 and is known to limit the cross-reactivity of R1 against different forms of AMA1 with Y175D substitution reduces the peptide binding activities.[19, 37] Asp is the predominant residue in the *Pf*AMA1 alleles and, as a result of this trait, R1 potent inhibitory activity is only limited to a small subset of *Plasmodium* strains. MIPS-0000865 was initially identified from screening against 3D7 *Pf*AMA1 which has Tyr residue at position 175.[97] In this study, the compounds were shown to bind to FVO *Pf*AMA1 with Asp175 in the HSQC perturbation experiments. Hence, these fragments are able to bind to *Pf*AMA1 with either Asp or Tyr residues at position 175, perhaps through their binding on the backbone. The benzimidazoles series also bind to the polymorphic end of the cleft. The numerous polymorphic residues surrounding this binding region would likely limit the possible chemical space that can be explored for the design of potent inhibitors with broad strain specificity. Future analogue synthesis would aim to introduce functional groups that favour binding to the DII end of the cleft and disfavour binding to the polymorphic end of the cleft. Both the pyrazole compounds contact a dimorphic residue at position 172 in the DII region that can vary between Gly and Glu, with these two amino acids occurring at approximately the same frequency (Gly = 46 % and Glu 54 %). Both the 3D7 and FVO *Pf*AMA1 have Gly at this position. It is unclear at this stage if MIPS-0001160 and PJS-2156 could still bind to *Pf*AMA1 when Gly is substituted by Glu, and would require further investigation with either AMA1 mutation experiment or binding study of AMA1 allele carrying Glu172.

Chapter 6:

Conclusion and Future Directions

6 Conclusion and Future Directions

Malaria is a major global health problem that imposes a substantial burden on the world's most vulnerable societies.[1] AMA1 plays important role in host cell invasion by malaria parasites, forming a complex with parasite RON proteins as part of the moving junction that develops between the host cell and the invading parasite.[24] A hydrophobic cleft is central to the function of AMA1, and ligands that bind to this site prevent host cell entry and thus represent a viable target for designing drugs to block the development of parasites in the bloodstream.

My project aimed at developing small-molecule inhibitors of AMA1 using a fragment-based approach. The first stage of the project involved screening a fragment library to identify efficient AMA1 binders. In order to carry out the screening campaign efficiently, it was necessary to improve protein yields of the *E. coli* expression system. My efforts in establishing the high-cell-density methodology, which allows the *E. coli* cells grow to high optical density (OD_{600} = 6 – 8), generated high yields recombinant AMA1. This method also laid the foundation for subsequent isotopic labelling for high-resolution NMR studies. Initial STD fragment cocktail screening of a 1140-compound library successfully identified a range of *Pf*AMA1-binding scaffolds. This was followed up by the STD and CPMG R1 competition assays, and 57 hits were found binding to the AMA1 hydrophobic cleft, corresponding to a hit rate of 5 %. The high hit rate observed support the existence of a druggable pocket within the cleft for developing potent small-molecule inhibitors. The R1-competing hits were further characterised using SPR experiments to rank the compounds as well as allowing early SAR studies of the hits. A normalization scheme was used to circumvent the problem of "SPR-undesirable-behaviours" at high compound concentrations.[109] In this scheme, the maximal binding responses (R_{max}) of fragments were normalised using R1 peptide as a control, and this made it possible to the determine the fragment binding affinity (K_D) and ligand efficiency at low non-saturating concentrations in the SPR experiments.

Whilst generating analogues of the fragment hits, we also evaluated the pyrrolo[2,3-*d*]pyrimidine-4-amine scaffolds described by Srinivasan *et al.*[92] as AMA1 inhibitors. Owing to their poor solubility in aqueous buffer, these compounds were evaluated using ^{1}H NMR spectroscopy in the presence or absence of surfactant. My experiments revealed that the reported compounds exist largely as aggregates in solution. A series of analogues with better solubility was generated to determine if a meaningful structure-activity relationships could be established for this scaffold. Despite showing substantial inhibitory effects in *in vitro*

merozoite invasion assays, the reported compounds and their analogues showed weak or no binding activity in the presence of surfactants in SPR experiments. These observations suggest that pyrrolo[2,3-*d*]pyrimidine-4-amines do not inhibit merozoite invasions through a direct interaction with AMA1. Rather, given that pyrrolo[2,3-*d*]pyrimidine-4-amines are known kinase inhibitors, it is possible that their inhibitory effects on merozoite invasion are a result of their activities on the kinases in *Plasmodium* parasites.

The first X-ray crystal structure of FVO *Pf*AMA1 was determined to gain insights into the impact of sequence diversity on AMA1 structure and help in the development of small-molecule inhibitors. This structure was compared to the other divergent AMA1 allele from *P. falciparum* 3D7 as well as AMA1 from the *P. vivax*. A combination of normalised B-factor analysis and MD simulations were used to investigate how the flexibility of the Ia – f and DII loops correlates with their roles in limiting the cross-reactivity of human antibody responses and inhibitory peptides against different forms of AMA1. My study revealed that the overall structure, including the flexible nature of the DII loop, is conserved between the FVO and 3D7 forms of *Pf*AMA1. Further analysis found that the interacting interfaces between DI and the DII loop consist of invariant residues across all *P. falciparum* strains. As this structural conservation at the DII loop end of the hydrophobic cleft appears likely to be conserved across all allelic forms of AMA1, this region represents an attractive site for strain-transcending small-molecule intervention. Collectively, these results have advanced our understanding on AMA1 structures that should contribute to the design of design of fragment analogues with broader strain specificity.

Subsequent study employed ^{1}H-^{15}N HSQC perturbation spectroscopy to obtain detailed information on fragment binding. To perform this experiment, the backbone amide resonances of ^{1}H-^{15}N HSQC spectrum was first assigned using TROSY 3D NMR spectroscopy of the ^{2}H, ^{15}N, ^{13}C labelled AMA1. In order to obtain high yields of isotopically labelled AMA1 using the high-cell-density methodology, the *E. coli* cells were adapted to grow in a deuterium oxide ($^{2}H_2O$) environment. This was achieved by culturing the bacteria in media prepared using increasing concentrations $^{2}H_2O$. This adaptation process has enabled the *E. coli* cells to grow to high optical density (OD_{600} = 6 – 8) in 100% $^{2}H_2O$ with high yield expression of isotopically labelled *Pf*AMA1. Significant efforts were also devoted to finding a suitable buffer condition for the 3D NMR experiments which require the protein sample to be stable at high temperature for prolonged period of time. The buffer conditions found from my experiments allowed the protein to remain stable in solution at high temperature with no proteolytic degradation up to a week. Besides, the AMA1 can be solubilised at lower pH

which reduces the amide proton exchange in solution, giving better quality spectra. Using HNCA, HN(CO)CA, HN(CA)CB and HN(COCA)CB, the majority of the backbone resonances of AMA1 were assigned. Specific ^{15}N-Lys labelled HSQC spectrum was also acquired to confirm the backbone amide assignment. Using these backbone resonance assignments, the ^{1}H-^{15}N HSQC perturbation studies have identified fragments that selectively interact with conserved sites in the hydrophobic cleft, and these compounds represent promising scaffolds for subsequent medicinal chemistry effort.

The project will continue to optimise the fragment hits with medicinal chemistry to give compounds with potent AMA1 binding affinity. This may be approached with either incremental installation of functional groups on a scaffold or through merging of different AMA1-binding classes found in the screening campaign. These approaches will be guided by both multidimensional NMR spectroscopy and X-ray crystallography. The ^{1}H-^{15}N HSQC perturbation experiment is a rapid and efficient tool in identifying fragments and its derivatives that bind to the DII end of the hydrophobic cleft for subsequent crystallisation trials. The crystallisation conditions found in my project serve as valuable starting points for obtaining structures of fragments bound the antigenically diverse forms of AMA1 (FVO and 3D7) through either soaking or co-crystallisation. Other biophysical tools established in the screening campaign will continue be the cornerstone in guiding the medicinal chemistry effort. The SPR experiments provides information about the SAR of fragment analogues that would help in subsequent generation of compounds with improved binding affinities. The ^{1}H NMR spectroscopy used to evaluate pyrrolo[2,3-*d*]pyrimidine-4-amines scaffolds will be useful to rapidly assess the solubility of the fragment analogues, avoiding any potential downfalls associated with compound aggregation. Potent AMA1 inhibitors derived from this FBLD project will be tested in *in vitro* merozoite invasion assay.

Chapter 7:

References

7 References

1. World Health Organization (WHO) Malaria, **2011;** All informations available from. http://www.who.int/mediacentre/factsheets/fs094/en/index.html.
2. Aponte, J. J.; Schellenberg, D.; Egan, A.; Breckenridge, A.; Carneiro, I.; Critchley, J.; Danquah, I.; Dodoo, A.; Kobbe, R.; Lell, B.; May, J.; Premji, Z.; Sanz, S.; Sevene, E.; Soulaymani-Becheikh, R.; Winstanley, P.; Adjei, S.; Anemana, S.; Chandramohan, D.; Issifou, S.; Mockenhaupt, F.; Owusu-Agyei, S.; Greenwood, B.; Grobusch, M. P.; Kremsner, P. G.; Macete, E.; Mshinda, H.; Newman, R. D.; Slutsker, L.; Tanner, M.; Alonso, P.; Menendez, C., Efficacy and safety of intermittent preventive treatment with sulfadoxine-pyrimethamine for malaria in African infants: a pooled analysis of six randomised, placebo-controlled trials. *Lancet* **2009,** *374* (9700), 1533-1542.
3. Hastings Ian, M.; Korenromp Eline, L.; Bloland Peter, B., The anatomy of a malaria disaster: drug policy choice and mortality in African children. *Lancet Infect. Dis.* **2007,** *7* (11), 739-748.
4. Mendis, K.; Sina, B. J.; Marchesini, P.; Carter, R., The neglected burden of *Plasmodium vivax* malaria. *Am. J. Trop. Med. Hyg*. **2001,** *64* (1-2 Suppl), 97-106.
5. Sachs, J.; Malaney, P., The economic and social burden of malaria. *Nature* **2002,** *415* (6872), 680-685.
6. Michalakis, Y.; Renaud, F., Malaria: Evolution in vector control. *Nature* **2009,** *462* (7271), 298-300.
7. Voza, T.; Miller, J. L.; Kappe, S. H.; Sinnis, P., Extrahepatic exoerythrocytic forms of rodent malaria parasites at the site of inoculation: clearance after immunization, susceptibility to primaquine, and contribution to blood-stage infection. *Infect. Immun.* **2012,** *80* (6), 2158-2164.
8. Cowman, A. F.; Crabb, B. S., Invasion of red blood cells by malaria parasites. *Cell* **2006,** *124* (4), 755-766.
9. Wells, T. N. C.; Alonso, P. L.; Gutteridge, W. E., New medicines to improve control and contribute to the eradication of malaria. *Nat. Rev. Drug Discov.* **2009,** *8* (11), 879-891.
10. Kappe, S. H. I.; Vaughan, A. M.; Boddey, J. A.; Cowman, A. F., That was then but this is now: malaria research in the time of an eradication agenda. *Science* **2010,** *328* (5980), 862-866.

11. Petersen, I.; Eastman, R.; Lanzer, M., Drug-resistant malaria: molecular mechanisms and implications for public health. *FEBS Lett.* **2011,** *585* (11), 1551-1562.
12. Craft, J. C., Challenges facing drug development for malaria. *Curr. Opin. Microbiol.* **2008,** *11* (5), 428-433.
13. Cammack, N., Exploiting malaria drug resistance to our advantage. *Science* **2011,** *333* (6043), 705-706.
14. Burrows, J. N.; Chibale, K.; Wells, T. N. C., The state of the art in anti-malarial drug discovery and development. *Curr. Top. Med. Chem.* **2011,** *11* (10), 1226-1254.
15. Duffy, P. E.; Sibley, C. H., Are we losing artemisinin combination therapy already? *Lancet* **2005,** *366* (9501), 1908-1909.
16. Hastings, I. M.; Ward, S. A., Coartem (artemether-lumefantrine) in Africa: The beginning of the end? *J. Infect. Dis.* **2005,** *192* (7), 1303-1304.
17. Noedl, H.; Se, Y.; Schaecher, K.; Smith, B. L.; Socheat, D.; Fukuda, M. M.; Consortium, A. R. C. S., Evidence of artemisinin-resistant malaria in Western Cambodia. *N. Engl. J. Med.* **2008,** *359* (24), 2619-2620.
18. Tyler, J. S.; Treeck, M.; Boothroyd, J. C., Focus on the ringleader: the role of AMA1 in apicomplexan invasion and replication. *Trends Parasitol.* **2011,** *27* (9), 410-420.
19. Chesne-Seck, M.-L.; Pizarro, J. C.; Normand, B. V.-L.; Collins, C. R.; Blackman, M. J.; Faber, B. W.; Remarque, E. J.; Kocken, C. H. M.; Thomas, A. W.; Bentley, G. A., Structural comparison of apical membrane antigen 1 orthologues and paralogues in apicomplexan parasites. *Mol. Biochem. Parasit.* **2005,** *144* (1), 55-67.
20. Hehl, A. B.; Lekutis, C.; Grigg, M. E.; Bradley, P. J.; Dubremetz, J. F.; Ortega-Barria, E.; Boothroyd, J. C., *Toxoplasma gondii* homologue of *Plasmodium* apical membrane antigen 1 is involved in invasion of host cells. *Infect. Immun.* **2000,** *68* (12), 7078-7086.
21. Waters, A. P.; Thomas, A. W.; Deans, J. A.; Mitchell, G. H.; Hudson, D. E.; Miller, L. H.; McCutchan, T. F.; Cohen, S., A merozoite receptor protein from *Plasmodium knowlesi* is highly conserved and distributed throughout *Plasmodium*. *J. Biol. Chem.* **1990,** *265* (29), 17974-17979.
22. Hodder, A. N.; Crewther, P. E.; Matthew, M.; Reid, G. E.; Moritz, R. L.; Simpson, R. J.; Anders, R. F., The disulfide bond structure of *Plasmodium* apical membrane antigen-1. *J. Biol. Chem.* **1996,** *271* (46), 29446-29452.
23. Pizarro, J. C.; Normand, B. V.-L.; Chesne-Seck, M.-L.; Collins, C. R.; Withers-Martinez, C.; Hackett, F.; Blackman, M. J.; Faber, B. W.; Remarque, E. J.; Kocken, C.

H. M.; Thomas, A. W.; Bentley, G. A., Crystal structure of the malaria vaccine candidate apical membrane antigen 1. *Science* **2005,** *308* (5720), 408-411.

24. MacRaild, C. A.; Anders, R. F.; Foley, M.; Norton, R. S., Apical membrane antigen 1 as an anti-malarial drug target. *Curr. Top. Med. Chem.* **2011,** *11* (16), 2039-2047.
25. Remarque, E. J.; Faber, B. W.; Kocken, C. H. M.; Thomas, A. W., Apical membrane antigen 1: a malaria vaccine candidate in review. *Trends Parasitol.* **2008,** *24* (2), 74-84.
26. Harris, P. K.; Yeoh, S.; Dluzewski, A. R.; O'Donnell, R. A.; Withers-Martinez, C.; Hackett, F.; Bannister, L. H.; Mitchell, G. H.; Blackman, M. J., Molecular identification of a malaria merozoite surface sheddase. *PLoS Pathog.* **2005,** *1* (3), e29.
27. Giovannini, D.; Späth, S.; Lacroix, C.; Perazzi, A.; Bargieri, D.; Lagal, V.; Lebugle, C.; Combe, A.; Thiberge, S.; Baldacci, P.; Tardieux, I.; Ménard, R., Independent roles of apical membrane antigen 1 and rhoptry neck proteins during host cell invasion by apicomplexa. *Cell. Host Microbe* **2011,** *10* (6), 591-602.
28. Bargieri, D. Y.; Andenmatten, N.; Lagal, V.; Thiberge, S.; Whitelaw, J. A.; Tardieux, I.; Meissner, M.; Menard, R., Apical membrane antigen 1 mediates apicomplexan parasite attachment but is dispensable for host cell invasion. *Nat. Commun.* **2013,** *4*, 2552.
29. Yap, A.; Azevedo, M. F.; Gilson, P. R.; Weiss, G. E.; O'Neill, M. T.; Wilson, D. W.; Crabb, B. S.; Cowman, A. F., Conditional expression of apical membrane antigen 1 in *Plasmodium falciparum* shows it is required for erythrocyte invasion by merozoites. *Cell. Microbiol.* **2014,** *16* (5), 642-656.
30. Triglia, T.; Healer, J.; Caruana, S. R.; Hodder, A. N.; Anders, R. F.; Crabb, B. S.; Cowman, A. F., Apical membrane antigen 1 plays a central role in erythrocyte invasion by *Plasmodium* species. *Mol. Microbiol.* **2000,** *38* (4), 706-18.
31. Lamarque, M. H.; Roques, M.; Kong-Hap, M.; Tonkin, M. L.; Rugarabamu, G.; Marq, J. B.; Penarete-Vargas, D. M.; Boulanger, M. J.; Soldati-Favre, D.; Lebrun, M., Plasticity and redundancy among AMA-RON pairs ensure host cell entry of *Toxoplasma* parasites. *Nat. Commun.* **2014,** *5*.
32. Bargieri, D.; Lagal, V.; Andenmatten, N.; Tardieux, I.; Meissner, M.; Ménard, R., Host cell invasion by apicomplexan parasites: the junction conundrum. *PLoS Pathog.* **2014,** *10* (9), e1004273.
33. Collins, C. R.; Withers-Martinez, C.; Bentley, G. A.; Batchelor, A. H.; Thomas, A. W.; Blackman, M. J., Fine mapping of an epitope recognised by an invasion-inhibitory

monoclonal antibody on the malaria vaccine candidate apical membrane antigen 1. **2007,** *282* (10), 7431-41.

34. Harris, K. S.; Casey, J. L.; Coley, A. M.; Karas, J. A.; Sabo, J. K.; Tan, Y. Y.; Dolezal, O.; Norton, R. S.; Hughes, A. B.; Scanlon, D.; Foley, M., Rapid optimisation of a peptide inhibitor of malaria parasite invasion by comprehensive N-methyl scanning. *J. Biol. Chem.* **2009,** *284* (14), 9361-9371.
35. Henderson, K. A.; Streltsov, V. A.; Coley, A. M.; Dolezal, O.; Hudson, P. J.; Batchelor, A. H.; Gupta, A.; Bai, T.; Murphy, V. J.; Anders, R. F.; Foley, M.; Nuttall, S. D., Structure of an IgNAR-AMA1 complex: targeting a conserved hydrophobic cleft broadens malarial strain recognition. *Structure* **2007,** *15* (11), 1452-1466.
36. Li, F.; Dluzewski, A.; Coley, A. M.; Thomas, A.; Tilley, L.; Anders, R. F.; Foley, M., Phage-displayed peptides bind to the malarial protein apical membrane antigen 1 and inhibit the merozoite invasion of host erythrocytes. *J. Biol. Chem.* **2002,** *277* (52), 50303-50310.
37. Vulliez-Le Normand, B.; Tonkin, M. L.; Lamarque, M. H.; Langer, S.; Hoos, S.; Roques, M.; Saul, F. A.; Faber, B. W.; Bentley, G. A.; Boulanger, M. J.; Lebrun, M., Structural and functional insights into the malaria parasite moving junction complex. *PLoS Pathog.* **2012,** *8* (6), e1002755.
38. Fowkes, F. J. I.; Richards, J. S.; Simpson, J. A.; Beeson, J. G., The relationship between anti-merozoite antibodies and incidence of *Plasmodium falciparum* malaria: a systematic review and meta-analysis. *PLoS Med.* **2010,** *7* (1), e10000218.
39. Hodder, A. N.; Crewther, P. E.; Anders, R. F., Specificity of the protective antibody response to apical membrane antigen 1. *Infect. Immun.* **2001,** *69* (5), 3286-3294.
40. Collins, W. E.; Anders, R. F.; Pappaioanou, M.; Campbell, G. H.; Brown, G. V.; Kemp, D. J.; Coppel, R. L.; Skinner, J. C.; Andrysiak, P. M.; Favaloro, J. M.; Corcoran, L. M.; Broderson, J. R.; Mitchell, G. F.; Campbell, C. C., Immunisation of Aotus monkeys with recombinant proteins of an erythrocyte surface antigen of *Plasmodium falciparum*. *Nature* **1986,** *323* (6085), 259-262.
41. Kennedy, M. C.; Wang, J.; Zhang, Y.; Miles, A. P.; Chitsaz, F.; Saul, A.; Long, C. A.; Miller, L. H.; Stowers, A. W., *In vitro* studies with recombinant *Plasmodium falciparum* apical membrane antigen 1 (AMA1): production and activity of an AMA1 vaccine and generation of a multiallelic response. *Infect. Immun.* **2002,** *70* (12), 6948-6960.

42. Anders, R. F.; Adda, C. G.; Foley, M.; Norton, R. S., Recombinant protein vaccines against the asexual blood stages of *Plasmodium falciparum*. *Hum. Vaccin.* **2010,** *6* (1), 39-53.

43. Thera, M. A.; Doumbo, O. K.; Coulibaly, D.; Laurens, M. B.; Ouattara, A.; Kone, A. K.; Guindo, A. B.; Traore, K.; Traore, I.; Kouriba, B.; Diallo, D. A.; Diarra, I.; Daou, M.; Dolo, A.; Tolo, Y.; Sissoko, M. S.; Niangaly, A.; Sissoko, M.; Takala-Harrison, S.; Lyke, K. E.; Wu, Y.; Blackwelder, W. C.; Godeaux, O.; Vekemans, J.; Dubois, M. C.; Ballou, W. R.; Cohen, J.; Thompson, D.; Dube, T.; Soisson, L.; Diggs, C. L.; House, B.; Lanar, D. E.; Dutta, S.; Heppner, D. G., Jr.; Plowe, C. V., A field trial to assess a blood-stage malaria vaccine. *N. Engl. J. Med.* **2011,** *365* (11), 1004-1013.

44. Malkin, E. M.; Diemert, D. J.; McArthur, J. H.; Perreault, J. R.; Miles, A. P.; Giersing, B. K.; Mullen, G. E.; Orcutt, A.; Muratova, O.; Awkal, M.; Zhou, H.; Wang, J.; Stowers, A.; Long, C. A.; Mahanty, S.; Miller, L. H.; Saul, A.; Durbin, A. P., Phase 1 clinical trial of apical membrane antigen 1: An asexual blood-stage vaccine for *Plasmodium falciparum* malaria. *Infect. Immun.* **2005,** *73* (6), 3677-3685.

45. Ouattara, A.; Mu, J. B.; Takala-Harrison, S.; Saye, R.; Sagara, I.; Dicko, A.; Niangaly, A.; Duan, J. H.; Ellis, R. D.; Miller, L. H.; Su, X. Z.; Plowe, C. V.; Doumbo, O. K., Lack of allele-specific efficacy of a bivalent AMA1 malaria vaccine. *Malaria J.* **2010,** *9* (175).

46. Sagara, I.; Dicko, A.; Ellis, R. D.; Fay, M. P.; Diawara, S. I.; Assadou, M. H.; Sissoko, M. S.; Kone, M.; Diallo, A. I.; Saye, R.; Guindo, M. A.; Kante, O.; Niambele, M. B.; Miura, K.; Mullen, G. E.; Pierce, M.; Martin, L. B.; Dolo, A.; Diallo, D. A.; Doumbo, O. K.; Miller, L. H.; Saul, A., A randomized controlled phase 2 trial of the blood stage AMA1-C1/Alhydrogel malaria vaccine in children in Mali. *Vaccine* **2009,** *27* (23), 3090-3098.

47. Dutta, S.; Dlugosz, L. S.; Drew, D. R.; Ge, X.; Ababacar, D.; Rovira, Y. I.; Moch, J. K.; Shi, M.; Long, C. A.; Foley, M.; Beeson, J. G.; Anders, R. F.; Miura, K.; Haynes, J. D.; Batchelor, A. H., Overcoming antigenic diversity by enhancing the immunogenicity of conserved epitopes on the malaria vaccine candidate apical membrane antigen 1. *PLoS Pathog.* **2013,** *9* (12), e1003840.

48. Miura, K.; Herrera, R.; Diouf, A.; Zhou, H.; Mu, J.; Hu, Z.; MacDonald, N. J.; Reiter, K.; Nguyen, V.; Shimp, R. L., Jr.; Singh, K.; Narum, D. L.; Long, C. A.; Miller, L. H., Overcoming allelic specificity by immunization with five allelic forms of *Plasmodium falciparum* apical membrane antigen 1. *Infect. Immun.* **2013,** *81* (5), 1491-1501.

49. Srinivasan, P.; Beatty, W. L.; Diouf, A.; Herrera, R.; Ambroggio, X.; Moch, J. K.; Tyler, J. S.; Narum, D. L.; Pierce, S. K.; Boothroyd, J. C.; Haynes, J. D.; Miller, L. H., Binding of *Plasmodium* merozoite proteins RON2 and AMA1 triggers commitment to invasion. *Proc. Natl. Acad. Sci. U. S. A.* **2011**.

50. Alexander, D. L.; Mital, J.; Ward, G. E.; Bradley, P.; Boothroyd, J. C., Identification of the moving junction complex of *Toxoplasma gondii*: a collaboration between distinct secretory organelles. *PLoS Pathog.* **2005,** *1* (2), e17.

51. Baum, J.; Cowman, A. F., Revealing a parasite's invasive trick. *Science* **2011,** *333* (6041), 410-411.

52. Richard, D.; MacRaild, C. A.; Riglar, D. T.; Chan, J.-A.; Foley, M.; Baum, J.; Ralph, S. A.; Norton, R. S.; Cowman, A. F., Interaction between *Plasmodium falciparum* apical membrane antigen 1 and the rhoptry neck protein complex defines a key step in the erythrocyte invasion process of malaria parasites. *J. Biol. Chem.* **2010,** *285* (19), 14815-14822.

53. Cao, J.; Kaneko, O.; Thongkukiatkul, A.; Tachibana, M.; Otsuki, H.; Gao, Q.; Tsuboi, T.; Torii, M., Rhoptry neck protein RON2 forms a complex with microneme protein AMA1 in *Plasmodium falciparum* merozoites. *Parasitol. Int.* **2009,** *58* (1), 29-35.

54. Lamarque, M.; Besteiro, S.; Papoin, J.; Roques, M.; Normand, B. V.-L.; Morlon-Guyot, J.; Dubremetz, J.-F.; Fauquenoy, S.; Tomavo, S.; Faber, B. W.; Kocken, C. H.; Thomas, A. W.; Boulanger, M. J.; Bentley, G. A.; Lebrun, M., The RON2-AMA1 interaction is a critical step in moving junction-dependent invasion by apicomplexan parasites. *PLoS Pathog.* **2011,** *7* (2), e1001276.

55. Derbyshire, E. R.; Mota, M. M.; Clardy, J., The next opportunity in anti-malaria drug discovery: The liver stage. *PLoS Pathog.* **2011,** *7* (9), e1002178.

56. Mazier, D.; Renia, L.; Snounou, G., A pre-emptive strike against malaria's stealthy hepatic forms. *Nat. Rev. Drug Discov.* **2009,** *8* (11), 854-864.

57. Silvie, O.; Franetich, J.-F.; Charrin, S.; Mueller, M. S.; Siau, A.; Bodescot, M.; Rubinstein, E.; Hannoun, L.; Charoenvit, Y.; Kocken, C. H.; Thomas, A. W.; van Gemert, G.-J.; Sauerwein, R. W.; Blackman, M. J.; Anders, R. F.; Pluschke, G.; Mazier, D., A role for apical membrane antigen 1 during invasion of hepatocytes by *Plasmodium falciparum* sporozoites. *J. Biol. Chem.* **2004,** *279* (10), 9490-9496.

58. Feng, Z.-P.; Keizer, D. W.; Stevenson, R. A.; Yao, S.; Babon, J. J.; Murphy, V. J.; Anders, R. F.; Norton, R. S., Structure and inter-domain interactions of domain II from

the blood-stage malarial protein, apical membrane antigen 1. *J. Mol. Biol.* **2005,** *350* (4), 641-656.

59. Nair, M.; Hinds, M. G.; Coley, A. M.; Hodder, A. N.; Foley, M.; Anders, R. F.; Norton, R. S., Structure of domain III of the blood-stage malaria vaccine candidate, *Plasmodium falciparum* apical membrane antigen 1 (AMA1). *J. Mol. Biol.* **2002,** *322* (4), 741-753.
60. Bai, T.; Becker, M.; Gupta, A.; Strike, P.; Murphy, V. J.; Anders, R. F.; Batchelor, A. H., Structure of AMA1 from *Plasmodium falciparum* reveals a clustering of polymorphisms that surround a conserved hydrophobic pocket. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102* (36), 12736-12741.
61. Crawford, J.; Tonkin, M. L.; Grujic, O.; Boulanger, M. J., Structural characterisation of apical membrane antigen 1 (AMA1) from *Toxoplasma gondii*. *J. Biol. Chem.* **2010,** *285* (20), 15644-15652.
62. Healer, J.; Murphy, V.; Hodder, A. N.; Masciantonio, R.; Gemmill, A. W.; Anders, R. F.; Cowman, A. F.; Batchelor, A., Allelic polymorphisms in apical membrane antigen 1 are responsible for evasion of antibody-mediated inhibition in *Plasmodium falciparum*. *Mol. Microbiol.* **2004,** *52* (1), 159-168.
63. Dutta, S.; Lee, S. Y.; Batchelor, A. H.; Lanar, D. E., Structural basis of antigenic escape of a malaria vaccine candidate. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104* (30), 12488-12493.
64. Ouattara, A.; Takala-Harrison, S.; Thera, M. A.; Coulibaly, D.; Niangaly, A.; Saye, R.; Tolo, Y.; Dutta, S.; Heppner, D. G.; Soisson, L.; Diggs, C. L.; Vekemans, J.; Cohen, J.; Blackwelder, W. C.; Dube, T.; Laurens, M. B.; Doumbo, O. K.; Plowe, C. V., Molecular basis of allele-specific efficacy of a blood-stage malaria vaccine: Vaccine development implications. *J. Infect. Dis.* **2013,** *207* (3), 511-519.
65. Takala, S. L.; Coulibaly, D.; Thera, M. A.; Batchelor, A. H.; Cummings, M. P.; Escalante, A. A.; Ouattara, A.; Traore, K.; Niangaly, A.; Djimde, A. A.; Doumbo, O. K.; Plowe, C. V., Extreme polymorphism in a vaccine antigen and risk of clinical malaria: implications for vaccine development. *Sci. Transl. Med.* **2009,** *1* (2), 2ra5.
66. Collins, C. R.; Withers-Martinez, C.; Hackett, F.; Blackman, M. J., An inhibitory antibody blocks interactions between components of the malarial invasion machinery. *PLoS Pathog.* **2009,** *5* (1), e1000273.
67. Wells, J. A.; McClendon, C. L., Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **2007,** *450* (7172), 1001-1009.

68. Tonkin, M. L.; Roques, M.; Lamarque, M. H.; Pugnière, M.; Douguet, D.; Crawford, J.; Lebrun, M.; Boulanger, M. J., Host cell invasion by apicomplexan parasites: insights from the co-structure of AMA1 with a RON2 peptide. *Science* **2011,** *333* (6041), 463-467.

69. Coley, A. M.; Parisi, K.; Masciantonio, R.; Hoeck, J.; Casey, J. L.; Murphy, V. J.; Harris, K. S.; Batchelor, A. H.; Anders, R. F.; Foley, M., The most polymorphic residue on *Plasmodium falciparum* apical membrane antigen 1 determines binding of an invasion-inhibitory antibody. *Infect. Immun.* **2006,** *74* (5), 2628-2636.

70. Henderson, K. A.; Streltsov, V. A.; Coley, A. M.; Dolezal, O.; Hudson, P. J.; Batchelor, A. H.; Gupta, A.; Bai, T.; Murphy, V. J.; Anders, R. F.; Foley, M.; Nuttall, S. D., Structure of an IgNAR-AMA1 Complex: targeting a conserved hydrophobic cleft broadens malarial strain recognition. *Structure (Cambridge, MA, U. S.)* **2007,** *15* (11), 1452-1466.

71. Coley, A. M.; Gupta, A.; Murphy, V. J.; Bai, T.; Kim, H.; Anders, R. F.; Foley, M.; Batchelor, A. H., Structure of the malaria antigen AMA1 in complex with a growth-inhibitory antibody. *PLoS Pathog.* **2007,** *3* (9), 1308-1319.

72. Collins, C. R.; Withers-Martinez, C.; Bentley, G. A.; Batchelor, A. H.; Thomas, A. W.; Blackman, M. J., Fine mapping of an epitope recognized by an invasion-inhibitory monoclonal antibody on the malaria vaccine candidate apical membrane antigen 1. *J. Biol. Chem.* **2007,** *282* (10), 7431-7441.

73. Harris, K. S.; Casey, J. L.; Coley, A. M.; Masciantonio, R.; Sabo, J. K.; Keizer, D. W.; Lee, E. F.; McMahon, A.; Norton, R. S.; Anders, R. F.; Foley, M., Binding hot spot for invasion inhibitory molecules on *Plasmodium falciparum* apical membrane antigen 1. *Infect. Immun.* **2005,** *73* (10), 6981-9.

74. Keizer, D. W.; Miles, L. A.; Li, F.; Nair, M.; Anders, R. F.; Coley, A. M.; Foley, M.; Norton, R. S., Structures of phage-display peptides that bind to the malarial surface protein, apical membrane antigen 1, and block erythrocyte invasion. *Biochemistry* **2003,** *42* (33), 9915-9923.

75. Lee, E. F.; Yao, S.; Sabo, J. K.; Fairlie, W. D.; Stevenson, R. A.; Harris, K. S.; Anders, R. F.; Foley, M.; Norton, R. S., Peptide inhibitors of the malaria surface protein, apical membrane antigen 1: identification of key binding residues. *Biopolymers* **2011,** *95* (5), 354-364.

76. Wang, G.; MacRaild, C. A.; Mohanty, B.; Mobli, M.; Cowieson, N. P.; Anders, R. F.; Simpson, J. S.; McGowan, S.; Norton, R. S.; Scanlon, M. J., Molecular insights into the

interaction between *Plasmodium falciparum* apical membrane antigen 1 and an invasion-inhibitory peptide. *Submited to PloS ONE* **2014**.

77. Vulliez-Le Normand, B.; Tonkin, M. L.; Lamarque, M. H.; Langer, S.; Hoos, S.; Roques, M.; Saul, F. A.; Faber, B. W.; Bentley, G. A.; Boulanger, M. J.; Lebrun, M., Structural and functional insights into the malaria parasite moving junction complex. *PLoS Pathog.* **2012,** *8* (6), e1002755.
78. Fischer, M.; Hubbard Roderick, E., Fragment-based ligand discovery. *Mol. Interv.* **2009,** *9* (1), 22-30.
79. Hajduk, P.; Greer, J., A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discov.* **2007,** *6* (3), 211-219.
80. Macarron, R., Critical review of the role of HTS in drug discovery. *Drug Discov. Today* **2006,** *11* (7-8), 277-279.
81. Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R., Fragment-based lead discovery. *Nat. Rev. Drug Discov.* **2004,** *3* (8), 660-672.
82. Congreve, M.; Chessari, G.; Tisi, D.; Woodhead, A. J., Recent developments in fragment-based drug discovery. *J. Med. Chem.* **2008,** *51* (13), 3661-3680.
83. Erlanson, D. A.; McDowell, R. S.; O'Brien, T., Fragment-based drug discovery. *J. Med. Chem.* **2004,** *47* (14), 3463-3482.
84. Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W., Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996,** *274* (5292), 1531-1534.
85. van Deursen, R.; Reymond, J.-L., Chemical space travel. *ChemMedChem* **2007,** *2* (5), 636-640.
86. Reymond, J.-L.; Awale, M., Exploring chemical space for drug discovery using the chemical universe database. *ACS Chem. Neurosci.* **2012,** *3* (9), 649-657.
87. Hopkins, A. L.; Groom, C. R.; Alex, A., Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* **2004,** *9* (10), 430-431.
88. Antonysamy, S. S.; Aubol, B.; Blaney, J.; Browner, M. F.; Giannetti, A. M.; Harris, S. F.; Hébert, N.; Hendle, J.; Hopkins, S.; Jefferson, E.; Kissinger, C.; Leveque, V.; Marciano, D.; McGee, E.; Nájera, I.; Nolan, B.; Tomimoto, M.; Torres, E.; Wright, T., Fragment-based discovery of hepatitis C virus NS5b RNA polymerase inhibitors. *Biorg. Med. Chem. Lett.* **2008,** *18* (9), 2990-2995.
89. Artis, D.; Lin, J.; Zhang, C.; Wang, W.; Mehra, U.; Perreault, M.; Erbe, D.; Krupka, H.; England, B.; Arnold, J.; Plotnikov, A.; Marimuthu, A.; Nguyen, H.; Will, S.; Signaevsky, M.; Kral, J.; Cantwell, J.; Settachatgull, C.; Yan, D.; Fong, D.; Oh, A.; Shi,

S.; Womack, P.; Powell, B.; Habets, G.; West, B.; Bollag, G.; Milburn, M.; Vlasuk, G.; Hirth, K. P.; Nolop, K.; Ibrahim, P.; Tobin, J., Scaffold-based discovery of indeglitazar, a PPAR pan-active anti-diabetic agent. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (1), 262-267.

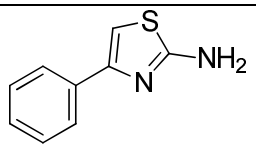
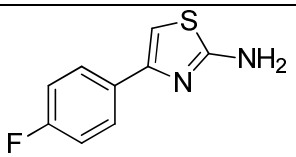
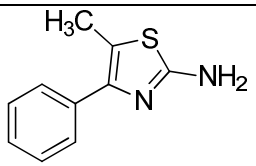
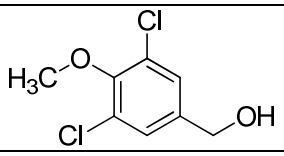
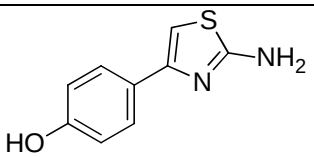
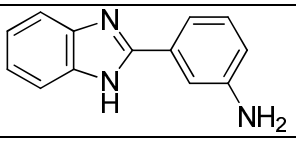
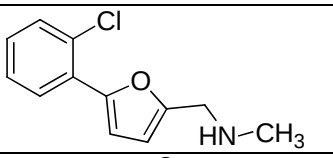
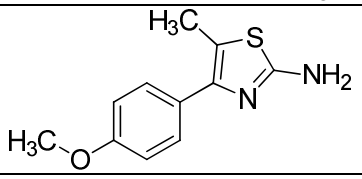
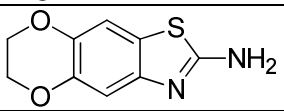
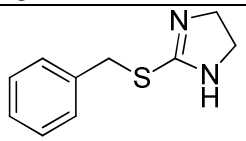
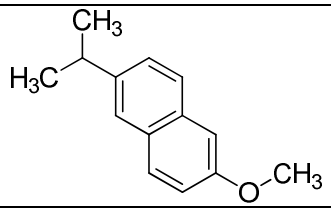
90. Howard, S.; Berdini, V.; Boulstridge, J. A.; Carr, M. G.; Cross, D. M.; Curry, J.; Devine, L. A.; Early, T. R.; Fazal, L.; Gill, A. L.; Heathcote, M.; Maman, S.; Matthews, J. E.; McMenamin, R. L.; Navarro, E. F.; O'Brien, M. A.; O'Reilly, M.; Rees, D. C.; Reule, M.; Tisi, D.; Williams, G.; Vinković, M.; Wyatt, P. G., Fragment-based discovery of the pyrazol-4-yl urea (AT9283), a multitargeted kinase inhibitor with potent aurora kinase activity. *J. Med. Chem.* **2008,** *52* (2), 379-388.
91. Sanchez, C. P.; Dave, A.; Stein, W. D.; Lanzer, M., Transporters as mediators of drug resistance in *Plasmodium falciparum*. *Int. J. Parasitol.* **2010,** *40* (10), 1109-1118.
92. Srinivasan, P.; Yasgar, A.; Luci, D. K.; Beatty, W. L.; Hu, X.; Andersen, J.; Narum, D. L.; Moch, J. K.; Sun, H.; Haynes, J. D.; Maloney, D. J.; Jadhav, A.; Simeonov, A.; Miller, L. H., Disrupting malaria parasite AMA1–RON2 interaction with a small molecule prevents erythrocyte invasion. *Nat. Commun.* **2013,** *4*.
93. Kar, S.; Kar, S., Control of malaria. *Nat. Rev. Drug Discov.* **2010,** *9* (7), 511-512.
94. Arkin, M. R.; Wells, J. A., Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat. Rev. Drug Discov.* **2004,** *3* (4), 301-317.
95. Ishima, R.; Torchia, D. A., Protein dynamics from NMR. *Nat. Struct. Mol. Biol.* **2000,** *7* (9), 740-743.
96. Williamson, M. P., Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Mag. Res. Sp.* **2013,** *73* (0), 1-16.
97. Lim, S. S.; Debono, C. O.; MacRaild, C. A.; Chandrashekaran, I. R.; Dolezal, O.; Anders, R. F.; Simpson, J. S.; Scanlon, M. J.; Devine, S. M.; Scammells, P. J.; Norton, R. S., Development of inhibitors of *Plasmodium falciparum* apical membrane antigen 1 based on fragment screening. *Aust. J. Chem.* **2013,** *66* (12), 1530-1536.
98. Lim, S. S.; Yang, W.; Krishnarjuna, B.; Sivaraman, K. K.; Chandrashekaran, I. R.; Kass, I.; MacRaild, C. A.; Devine, S. M.; Debono, C. O.; Anders, R. F.; Scanlon, M. J.; Scammells, P. J.; Norton, R. S.; McGowan, S., Structure and dynamics of apical membrane antigen 1 from *Plasmodium falciparum* FVO. *Submitted to Biochemistry* **2014**.
99. Sattler, M.; Fesik, S. W., Use of deuterium labeling in NMR: overcoming a sizeable problem. *Structure 4* (11), 1245-1249.

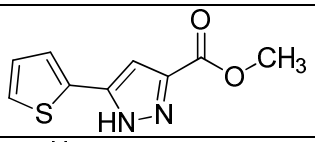
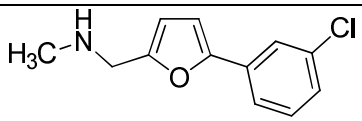
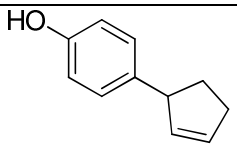
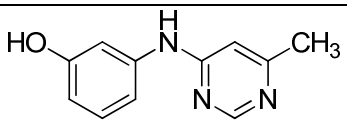
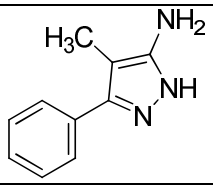
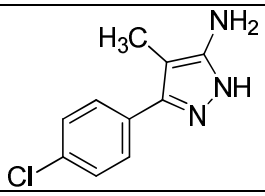
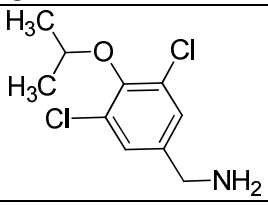
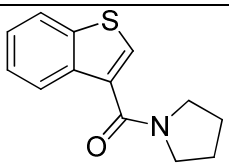
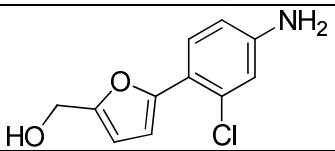
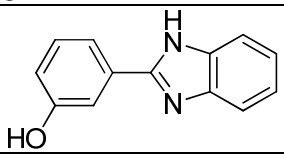
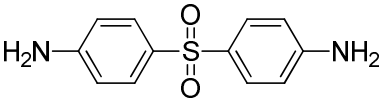
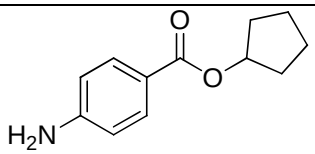
100. Gardner, K. H.; Kay, L. E., The use of ^{2}H, ^{13}C, ^{15}N multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.* **1998,** *27* (1), 357-406.
101. Riek, R.; Pervushin, K.; Wüthrich, K., TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution. *Trends Biochem. Sci.* *25* (10), 462-468.
102. Salzmann, M.; Pervushin, K.; Wider, G.; Senn, H.; Wüthrich, K., TROSY in triple-resonance experiments: New perspectives for sequential NMR assignment of large proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95* (23), 13585-13590.
103. Goto, N. K.; Kay, L. E., New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Curr. Opin. Struct. Biol.* **2000,** *10* (5), 585-592.
104. Krishnarjuna, B.; Jaipuria, G.; Thakur, A.; D'Silva, P.; Atreya, H., Amino acid selective unlabeling for sequence specific resonance assignments in proteins. *J. Biomol. NMR* **2011,** *49* (1), 39-51.
105. Muchmore, D. C.; McIntosh, L. P.; Russell, C. B.; Anderson, D. E.; Dahlquist, F. W., Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. In *Methods Enzymol.*, Norman, J. O.; Thomas, L. J., Eds. Academic Press: 1989; Vol. Volume 177, pp 44-73.
106. Ohki, S.-y.; Kainosho, M., Stable isotope labeling methods for protein NMR spectroscopy. *Prog. Nucl. Mag. Res. Sp.* **2008,** *53* (4), 208-226.
107. MacRaild, C.; Norton, R., RASP: rapid and robust backbone chemical shift assignments from protein structure. *J. Biomol. NMR* **2014,** *58* (3), 155-163.
108. Devine, S. M.; Mulcair, M. D.; Debono, C. O.; Leung, E. W. W.; Nissink, J. W. M.; Lim, S. S.; Chandrashekaran, I. R.; Simpson, J. S.; Baell, J. B.; Scammells, P. J.; Norton, R. S.; Scanlon, M. J., Promiscuous 2-aminothiazoles (PrATs): A frequent hitting scaffold. *Submitted to J. Med. Chem.* **2014**.
109. Giannetti, A. M.; Koch, B. D.; Browner, M. F., Surface plasmon resonance based assay for the detection and characterization of promiscuous inhibitors. *J. Med. Chem.* **2008,** *51* (3), 574-580.

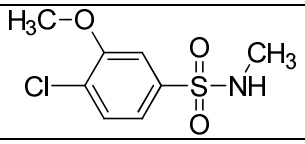
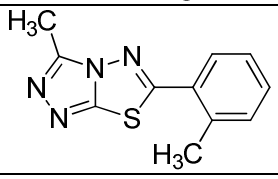
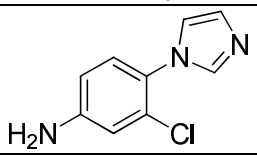
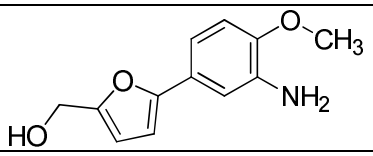
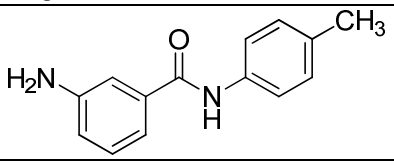
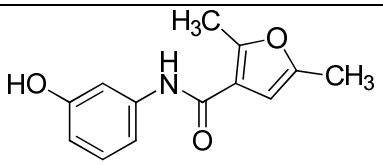
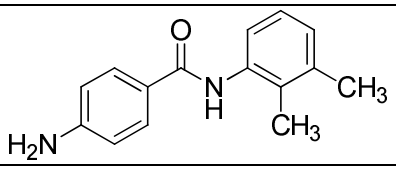
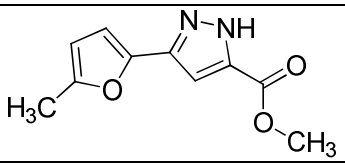
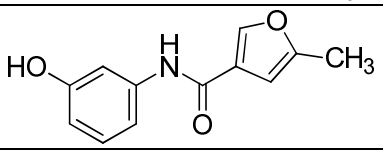
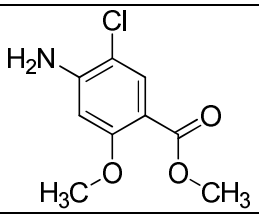
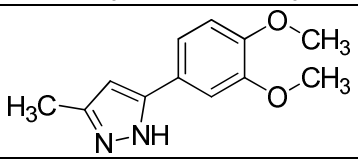
Appendix I

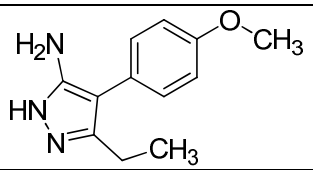
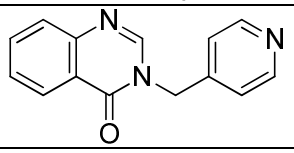
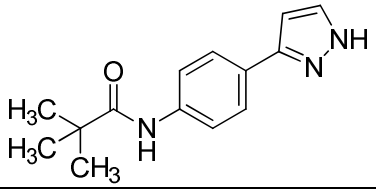
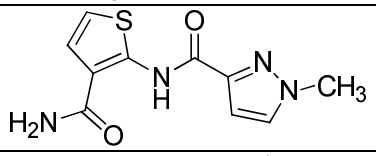
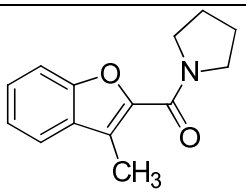
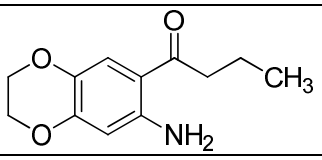
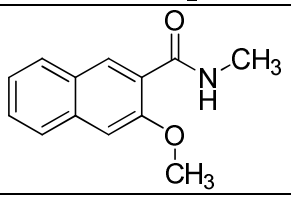
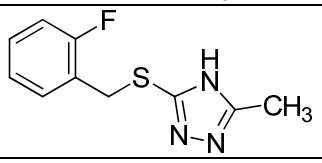
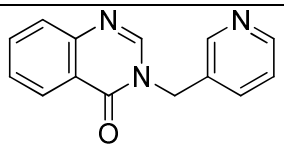
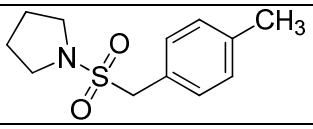
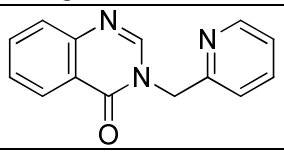
Fragment Hits

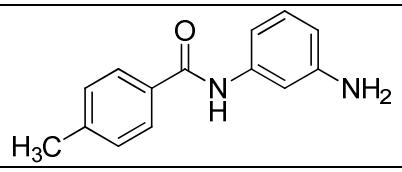
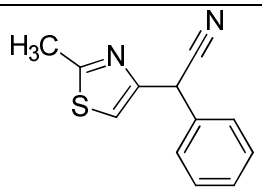
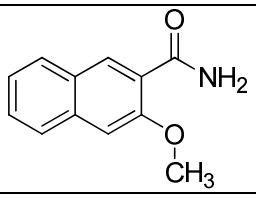
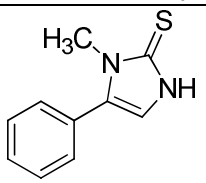
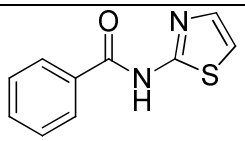
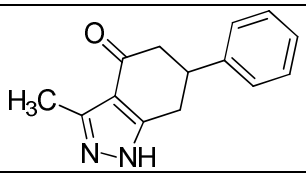
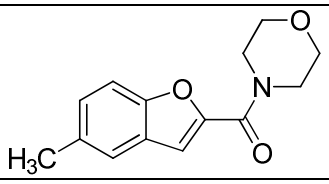
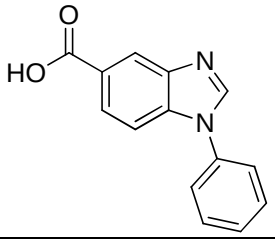
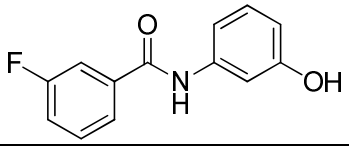
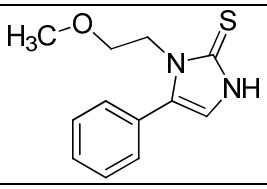
All R1-competing hits found in the screening campaign against AMA1 are shown. The compounds are shown in order from the highest to lowest ligand efficiency.

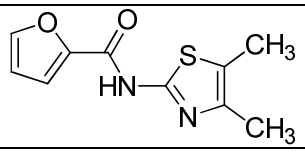
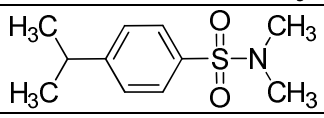
Rank	Structure	Log P	Molecular weight (MW)	Heavy atom (HA)	K_D (mM)	Ligand efficiency (kcal/mol/HA)
1		2.52	176	12	1.6	0.32
2		2.66	194	13	1.0	0.31
3		3.17	190	13	1.1	0.31
4		2.26	207	12	2.6	0.29
5		2.22	192	13	1.9	0.29
6		2.46	209	16	0.5	0.28
7		2.76	222	15	1.0	0.27
8		3.01	220	15	1.0	0.27
9		1.48	208	14	1.6	0.27
10		2.36	192	13	2.6	0.27
11		2.35	200	15	1.1	0.27

12		1.81	208	14	1.9	0.26
13		2.76	222	15	1.3	0.26
14		2.98	160	12	4.8	0.26
15		1.91	201	15	1.3	0.26
16		2.08	173	13	3.3	0.26
17		2.69	208	14	2.1	0.26
18		2.92	234	14	2.1	0.26
19		2.55	231	16	1.1	0.25
20		1.61	224	15	1.7	0.25
21		2.98	210	16	1.2	0.25
22		1.27	248	17	1.1	0.24
23		2.50	205	15	2.9	0.23

24		1.25	236	14	4.5	0.23
25		2.31	230	16	2.1	0.23
26		1.00	193	13	6.8	0.23
27		0.85	219	16	2.3	0.22
28		2.75	226	17	1.7	0.22
29		2.30	231	17	2.0	0.22
30		3.26	240	18	1.4	0.22
31		1.49	206	15	4.4	0.21
32		2.10	217	16	3.3	0.21
33		1.59	216	14	6.8	0.21
34		1.66	218	16	3.7	0.21

35		1.86	217	16	4.1	0.20
36		1.46	237	18	2.7	0.19
37		3.35	243	18	2.8	0.19
38		1.33	250	17	4.4	0.19
39		2.27	229	17	5.0	0.18
40		2.01	221	16	7.3	0.18
41		1.88	215	16	7.6	0.18
42		3.25	223	15	>10	0.18
43		1.46	237	18	4.4	0.18
44		1.71	239	16	9.3	0.17
45		1.54	237	18	5.5	0.17

46		2.75	226	18	5.6	0.17
47		2.51	214	15	>10	0.17
48		1.66	201	15	>10	0.17
49		2.04	190	13	>10	0.13
50		2.41	204	14	>10	0.13
51		1.90	226	17	>10	0.11
52		1.65	245	18	>10	0.09
53		0.92	238	18	>10	-0.30
54		2.90	231	17	>10	-0.32
55		2.00	234	16	>10	-0.34

56		2.25	222	15	>10	-0.36
57		2.27	227	15	>10	-0.36

Appendix II

Results for Chapter 5

Table showing the backbone amide resonance assignments and chemical shift index (CSI) of C_α and C_β for FVO *Pf*AMA1 domain I and II. Blank in the table indicates that assignments or CSI are unavailable. CSI was obtained using online tool http://randomcoilindex.com/cgi-bin/rci_cgi_current.py

Residue number	HN	N	C_α	C_β	C_α CSI	C_β CSI
104						
105						
106						
107						
108						
109						
110	7.96	119.3	57.5	29.1	0	1
111	7.02	114.7	67.2	68.7	1	0
112	8.38	120.1	59.4	28.2	1	-1
113	8.22	122.0	60.7	38.9	1	0
114	8.19	110.7	53.2	28.3	-1	-1
115	7.50	128.1	55.9	17.3	1	-1
116	7.21	114.1	56.7	31.0	0	-1
117	7.25	111.0	56.8	37.7	-1	-1
118	7.83	123.0	51.7	38.0	-1	-1
119	7.74	126.2	64.4	38.0	1	0
120	10.01	122.6	60.4	29.7	1	0
121	7.04	116.9	57.6	30.0	1	0
122	8.60	114.6	64.1	31.4	1	0
123	7.95	118.3	58.0	34.2	1	1
124	7.92	102.6	46.3	0.0	1	-1
125	6.56	111.4	54.0	64.7	-1	1
126	8.73	105.9	44.6	0.0	0	0
127	8.09	118.3	63.4	38.4	1	1
128	6.91	121.3	59.1	28.3	1	-1
129	6.02	112.6	61.7	32.9	-1	1
130	8.94	130.2	51.6	41.2	-1	0
131						
132						
133	8.19	121.2	56.5	29.5	0	0
134	8.29	121.3	54.1	40.7	0	0
135	7.99	124.4	52.6	18.7	0	0
136	8.21	119.8	56.6	29.4	0	0
137	7.86	120.5	62.1	31.7	-1	0
138	8.18	127.4	52.5		0	
139	8.63	105.4	45.2		0	
140	7.82	119.3	61.6	70.2	-1	1
141	8.12	124.6	55.2	28.9	-1	-1
142	8.86	122.7	56.7	41.2	-1	1
143						
144						

145						
146						
147						
148	8.07	124.2	58.6	31.8	1	0
149	9.79	116.8	51.4	39.9	-1	-1
150						
151						
152	8.03	128.2	60.8	37.8	1	-1
153	9.24	113.0	44.7	0.0	0	0
154	6.29	117.7	54.7	33.1	-1	1
155	7.59	110.3	43.1		-1	
156	3.42	116.5	57.4		-1	
157	9.12	128.6	60.6	38.5	-1	1
158	8.15	129.1	60.6	37.0	-1	0
159						
160						
161						
162						
163						
164						
165						
166						
167						
168						
169						
170						
171	7.47	110.4	59.5	72.0	-1	1
172	8.11	109.0	46.3		1	
173						
174	7.63	120.8	55.3	30.2	-1	0
175	8.63	124.5	53.8	42.2	0	1
176	7.75	126.4	58.2		1	
177	7.98	112.7	57.5	30.6	1	-1
178	7.94	119.9	54.5	41.1	0	0
179	7.55	109.2	43.7		-1	
180	6.80	106.7	44.6		0	
181	8.08	124.9				
182	8.18	124.1	52.7	20.6	0	1
183	8.54	122.5	56.9	38.6	-1	0
184						
185						
186	6.99	114.9	59.2	72.7	-1	1

187	8.31	118.0	50.2		-1	
188						
189	8.34	124.2	55.8	41.1	0	-1
190	8.20	121.2	59.3		-1	
191	6.68	112.3	55.7	63.0	-1	0
192						
193	8.37	120.0	54.9	36.5	-1	1
194	8.57	116.5	60.6	70.9	-1	1
195	8.41	123.3	58.5	40.5	1	-1
196	8.70	116.0	55.9	37.0	1	-1
197	8.25	111.0	46.5		1	
198	8.83	124.1	59.8	34.0	1	1
199	8.21	118.1	60.7	28.2	1	-1
200	7.69	118.1	56.8	41.8	1	1
201	8.37	123.8	60.7	39.1	1	0
202	7.63	115.0	58.1	37.0	0	-1
203	7.03	118.7	59.0	31.0	1	-1
204	8.33	116.2	52.8		-1	
205	8.10	121.2	52.0	38.5	-1	0
206	8.75	124.3	59.0	28.8	1	-1
207	7.89	117.5	59.3	37.6	0	-1
208	7.26	116.5	63.5	31.9	0	0
209	7.84	116.5	58.9	30.5	1	-1
210	7.31	115.7	52.7	39.4	-1	0
211	7.20	121.3	54.8	43.0	-1	1
212	7.69	118.1	53.6	41.8	0	1
213	9.03	117.5	59.1	28.2	1	-1
214	7.78	117.5	56.8	38.7	1	-1
215	8.60	120.1	66.7	66.3	1	-1
216	8.64	121.3	58.2	40.6	1	-1
217	7.31	115.5	56.4	33.8	-1	-1
218	7.78	111.8	60.6	63.1	1	0
219	8.92	119.6	56.6	27.4	0	-1
220	9.02	123.1	59.2	30.6	1	-1
221	8.01	120.6	54.2	19.4	1	0
222	7.64	103.7	45.8		1	
223	7.73	118.8	52.8	40.0	-1	1
224	7.40	118.6	53.9	34.1	-1	1
225	7.94	120.8	53.5		0	
226						
227						
228						

229						
230						
231						
232						
233						
234						
235						
236						
237						
238	8.47	118.1	50.5	24.6	-1	1
239	8.96	118.1	59.4	35.5	-1	1
240	8.24	126.9	56.4	40.7	-1	1
241	8.17	127.4	51.2	41.6	-1	1
242	9.22	123.2	61.8	38.9	1	0
243	8.88	118.8	56.5	38.7	1	0
244	6.94	115.9	52.8	40.6	-1	0
245	7.64	117.7	56.8	28.2	0	-1
246	6.72	114.3	54.7	37.0	-1	1
247	8.77	122.5	53.7	40.2	-1	-1
248	9.43	130.2	54.4	32.4	-1	0
249	8.58	124.8	61.8	36.5	-1	-1
250	8.62	127.9	53.9	40.5	-1	-1
251	7.86	120.7	61.1		1	
252	7.85	116.5	59.9		-1	
253						
254						
255	8.59	122.9	55.2	30.7	-1	0
256	8.77	125.3	61.5	33.5	1	1
257						
258						
259						
260						
261						
262						
263						
264						
265						
266						
267						
268						
269						
270						

271						
272						
273						
274						
275						
276	8.28	118.8	55.1	39.8	-1	0
277	8.95	117.6	52.7	29.4	-1	-1
278						
279	8.62	121.8	52.2	23.0	0	1
280	7.95	116.6	55.7	31.8	-1	0
281	7.21	118.8	53.2	46.4	-1	1
282						
283	8.93	118.8	56.5	40.6	1	-1
284	7.38	115.9	52.0	39.3	-1	0
285	7.09	120.4	60.0	29.3	1	0
286	8.37	113.2	52.8	39.2	-1	0
287	8.17	120.2	53.6	37.7	-1	-1
288	8.53	116.5	60.9		-1	
289	8.55	128.9	58.9	40.4	0	1
290	9.03	130.9	53.0	42.9	-1	1
291						
292						
293	8.37	119.6	50.9	38.9	-1	0
294	6.50	121.4	64.1	30.7	1	-1
295	8.23	123.6	61.0	33.0	-1	1
296	8.38	118.2	55.1	39.9	1	-1
297	8.12	115.8	51.1	35.8	-1	-1
298	7.58	121.5	59.9	27.2	1	-1
299	7.17	123.0	59.3	27.8	1	-1
300	7.47	116.5	57.1	30.6	0	1
301	7.05	109.6	61.7	32.7	-1	1
302	8.21	117.0	54.1	42.8	-1	1
303						
304	8.07	120.7	58.4	31.6	1	1
305	8.19	116.1	55.7	33.0	-1	1
306	8.61	121.6	54.0	39.9	0	1
307	8.17	120.9	53.0	42.1	-1	0
308	8.84	125.1	56.7	29.9	0	0
309	8.19	119.1	54.7	36.2	1	-1
310	7.57	122.3	50.5	22.1	-1	1
311	8.14	114.7	52.2	35.7	-1	1
312	9.72	124.3	57.9	40.3	0	1

313	8.00	108.6	43.9		-1	
314	8.78	119.3	52.4	44.7	-1	1
315	8.08	127.7	56.9	29.6	-1	1
316	8.61	131.5	61.2	34.0	-1	1
317	8.78	127.4	54.9	39.3	1	-1
318	6.70	133.1	45.8		1	
319	6.64	116.9	51.1	42.0	-1	1
320	8.32	122.0	54.2	39.3	-1	-1
321	8.83	129.5	54.1	31.2	-1	1
322	8.07	118.0		40.6		0
323	8.28	122.2	57.2	35.4	-1	-1
324						
325	7.35	115.7	54.9	29.5	-1	-1
326	8.04	117.3	59.3	34.0	-1	1
327	9.29	122.0	51.9	39.6	-1	0
328	8.48	124.3	55.6	31.8	-1	1
329	8.70	123.4	55.9	43.5	-1	1
330	9.03	119.9	59.5	63.3	1	0
331	8.08	124.9	51.5	20.9	-1	1
332						
333	7.26	114.1	51.9	41.6	-1	1
334	8.24	120.2	57.3	40.7	1	-1
335	7.57	121.5	59.9	37.9	1	-1
336	8.46	118.2	58.8	28.6	1	-1
337						
338	8.96	120.2	59.2	39.5	1	0
339	8.59	118.0	60.0	31.7	1	0
340	7.68	121.9	57.6	41.7	1	0
341	8.26	119.5	65.9	30.5	1	-1
342	8.44	116.7	62.2	38.6	1	-1
343	8.32	120.1	58.6	28.2	1	-1
344	7.44	120.0	54.5	41.2	-1	-1
345	7.23	118.4	59.2	65.1	1	1
346	7.70	118.3	52.9	17.8	0	-1
347	8.53	117.4	55.7	64.3	-1	1
348	8.40	128.8	55.6		1	
349						
350						
351						
352						
353						
354						

355						
356						
357						
358						
359						
360						
361	8.06	120.2	57.5	29.2	1	0
362	7.79	120.1	57.3	31.7	0	0
363	7.67	120.5	62.0	37.5	0	0
364						
365						
366						
367						
368						
369						
370						
371						
372						
373						
374						
375						
376						
377						
378						
379						
380						
381						
382						
383						
384						
385						
386						
387						
388						
389						
390						
391						
392						
393						
394	7.92	103.9	45.0		0	
395	7.44	121.7	56.5	31.9	0	0
396	8.21	111.4	44.9		0	

397	7.14	119.2	57.0	34.7	-1	-1
398	6.31	113.9	52.6	39.4	-1	0
399	7.51	119.2	57.7	34.1	0	1
400	8.56	104.7	44.8		0	
401	9.72	121.7	51.3	41.3	-1	1
402	10.04	125.8	55.7	38.8	-1	0
403	8.81	127.5	50.9	37.6	-1	-1
404	8.05	122.5	56.2	29.7	0	0
405	8.29	121.5	57.9	29.3	1	0
406	7.58	109.3	60.8	68.6	-1	0
407	7.66	117.7		40.7		1
408	7.30	118.0	53.3	35.0	-1	1
409	9.76	124.7	53.4	37.7	-1	-1
410	9.43	130.8	55.2	28.8	-1	-1
411	9.07	125.5	62.7	39.4	0	1
412	8.36	122.9	56.4	41.5	-1	1
413	8.13	115.0	52.2	37.0	-1	-1
414	7.00	118.8	58.7	33.3	-1	1
415	7.48	122.2	52.5	30.5	-1	-1
416						
417	8.19	106.1	59.7	69.8	-1	1
418	7.96	116.5	56.0	44.5	-1	1
419	7.89	114.3	52.5	0.0	-1	1
420	8.86	123.3	59.9	40.3	-1	1
421	8.55	126.0	53.4	38.1		-1
422	7.59	124.8	53.9	39.9		1
423	8.76	122.2	60.4		1	
424	8.89	117.7	59.7	62.9	1	0
425	7.52	118.4	57.7	44.1	-1	1
426	9.01	119.6	60.4	42.1	-1	1
427	8.35	132.4	51.5	17.1	-1	-1
428	9.43	120.3	60.0	72.3	-1	1
429	7.42	112.3	59.5	73.3	-1	1
430	8.60	120.8	54.6	20.2	1	1
431	7.07	114.6	56.5	43.4	1	1
432	6.94	111.3	58.8	63.4	0	0
433	8.85	123.9	56.6	31.6	1	0
434						
435	8.81	112.1	60.5	41.6	-1	1
436	8.23	121.8	58.0	30.7	1	1
437	8.23	118.3	59.1	35.4	-1	1
438	8.66	128.5	57.2	31.3	0	0

Appendix III

Submitted paper

Devine, S. M.; Mulcair, M. D.; Debono, C. O.; Leung, E. W. W.; Nissink, W. M.; **Lim, S. S.**; Chandrashekaran, I. R.; Simpson, J. S.; Baell, J. B.; Scammells, P. J.; Norton, R. S.; Scanlon, M. J., Promiscuous 2-aminothiazoles (PrATs): A frequent hitting scaffold. *Submitted to J. Med. Chem.* **2014**.

Promiscuous 2-Aminothiazoles (PrATs): A Frequent Hitting Scaffold

Shane M. Devine,[A] Mark D. Mulcair,[A] Cael O. Debono,[A] Eleanor W. W. Leung,[A] J. Willem M. Nissink,[B] San Sui Lim,[A] Indu R. Chandrashekaran,[A] Jamie S. Simpson,[A] Jonathan B. Baell,[A] Peter J. Scammells,[A] Raymond S. Norton,[A] Martin J. Scanlon[A*]

[A] Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, VIC 3052, Australia

[B] Oncology iMed, AstraZeneca, Alderley Park, Nether Alderley, Macclesfield SK10 4TG, United Kingdom

* Corresponding author. [redacted]

ABSTRACT

We have identified a class of molecules, known as 2-aminothiazoles (2-ATs), as frequent-hitting fragments in biophysical binding assays. This was exemplified by 4-phenylthiazol-2-amine being identified as a hit in 14/14 screens against a diverse range of protein targets, suggesting that this scaffold is a poor starting point for fragment-based drug discovery. This prompted us to undertake an analysis of this scaffold in the context of both an academic fragment library used for fragment-based drug discovery (FBDD) and two larger compound libraries used for high-throughput screening (HTS). This analysis revealed that such "Promiscuous 2-AminoThiazoles" (PrATs) were found to behave as frequent hitters under both FBDD and typical HTS settings, although the problem was more pronounced in the fragment-based studies. As 2-ATs are present in known drugs, they cannot necessarily be deemed undesirable, but the combination of their promiscuity and difficulties associated with optimizing them into a lead compound, make them, in our opinion, poor scaffolds for fragment libraries. In the context of HTS libraries where 2-ATs are found to hit a

target at much lower concentration there is a reasonable likelihood that they can be progressed successfully, although their observed promiscuity may constitute a greater developmental risk.

INTRODUCTION

Fragment-based drug discovery (FBDD) is becoming a widely-used technique in drug discovery as part of a medicinal chemists arsenal.[1,2] FBDD strategies utilise small (typically ~200 Da) molecules, known as 'fragments', that these typically have relatively low affinities for the target of interest. Nonetheless, numerous FBDD campaigns have demonstrated that it is feasible to elaborate fragment hits to achieve clinically useful compounds.[3] Furthermore, there are published examples where FBDD has enabled the development of potent compounds against targets where HTS of a large library did not yield any useful hits[4] and a review of internal projects by scientists at AstraZeneca revealed that FBDD represents a powerful tool to assess the likelihood of finding highly potent ligands for any given target.[5] This highlights one attraction of FBDD, which is that FBDD screening libraries typically contain relatively small numbers of compounds (a few hundred to thousands) but are very effective at finding hits because they are able to sample chemical space more efficiently than the larger compounds that are generally found in HTS libraries.[6] One way of illustrating this is to consider that a library of drug-like molecules of 30 heavy atoms might need up to 10^{60} members to efficiently cover chemical space, whereas this number is vastly smaller – around 10^{7} – for a library of fragments with around 12 heavy atoms.[7,8] As fragments are smaller, they typically bind with lower affinity to the relevant target, with K_D values often ranging from high micromolar to millimolar. In contrast drug-like hits from HTS typically have K_D values in the high nanomolar to low micromolar range, although in both cases the binding energy per heavy atom (or "ligand efficiency") can be comparable. Fragments can also be developed into lead compounds that are smaller and less lipophilic than those generated from HTS.[9] A highlight of FBDD to date was the discovery of Vemurafenib, a BRAF kinase inhibitor used for the treatment of late-stage melanoma, which, via medicinal chemistry optimisation, became the first FBDD-derived compound

to reach the clinic.[10,11] Since FBDD libraries are typically small it is essential to ensure that the library is populated with high quality fragments. Some of the key considerations in the design of fragment libraries have been discussed previously.[12,13]

PAINS

A matter of growing concern with screening libraries, for FBDD or otherwise, is the inclusion of compounds that may act promiscuously and display activity both across different target classes and via a number of different assays or biological readouts. These types of compounds, which are widely referred to as Pan Assay INterference CompoundS (PAINS), were first described by Baell and Holloway[14] and are identified by the presence of substructural features that promote frequent-hitting behavior. It has been suggested that they should be excluded from screening libraries for that reason. Whilst in many cases PAINS may appear to give optimizable hits, elaboration often results in flat or confusing structure-activity relationships (SAR).[15] The reasons for their promiscuity are varied and include the presence of potential Michael acceptors, chelation, redox activity, and strong chromophoric interference.[16] Some PAINS scaffolds have easily identifiable problems associated with them, but the chemical basis of the observed promiscuity for many PAINS is unknown. Confounding the complexity of PAINS identification is the fact that there are a small number of cases where seemingly "unprogressable" PAINS have in fact been developed into potent and selective molecules. One example is the PI3Kγ inhibitor (AS-604580) which is based on an alkylidene rhodanine hit containing this known PAINS motif.[17] However, the fact that certain PAINS or PAINS-like motifs are present in some potent and selective molecules does not imply that the PAIN is a viable starting point. In general, the chances that a PAIN will be progressed to a useful lead compound are overwhelmingly smaller than the chances it will not.[16]

There are fourteen sub-classes of 2-ATs that have been categorized as PAINS.[14,16] There are a number of possible reasons for their reported promiscuity, such as their potential photoreactivity or

the presence of impurities generated by their chemical precursors, for example where 2-ATs are prepared from bromomethylketones.[14] 2-ATs have also been identified as potentially thiol-reactive, which is another mechanism that produces problematic screening hits. For example a subclass of 2-ATs was identified in the ALARM-NMR assay as being thiol-reactive.[18,19] However, in this case it is possible that their reactive precursors were responsible.[20]

That 2-ATs can be progressed to generate useful compounds is testified by a number of marketed drugs: antibiotics such as carumonam, cefcapene, cefdinir, cefditoren, cefepime, cefetamet, cefoselis, cefotaxime, cefotiam, cefpodoxime, cefpirome, ceftazidime, ceftibuten, ceftriaxone; talipexole and pramipexole, dopamine agonists for the treatment of Parkinson's disease; mirabegron, a β_3-adrenoceptor agonist used to treat overactive bladder; and riluzole, a 2-aminobenzothiazole used to treat acute myeloid leukaemia (AML). Conversely, 2-ATs have displayed cytotoxicity and metabolic instability as antimycobacterial and antiplasmodial agents.[21] Thus, the value of 2-ATs in screening collections is currently unclear. Herein, we describe our efforts to determine if 2-ATs are promiscuous binders by both FBDD and HTS techniques.

RESULTS

FBDD Screening

The FBDD screening was undertaken at the Monash Institute of Pharmaceutical Sciences (MIPS) using a library of 1137 fragments comprising molecules that pass both biophysical and chemical filters, are chemically and structurally diverse, are soluble at 1 mM in aqueous buffer and can be chemically elaborated from readily accessible precursors.[12,22] Chemical filters that were applied in assembling the fragment library include the removal of PAINS,[14] unwanted functionality[13] and reactive groups.[23] A review of 14 fragment screening campaigns with this library revealed that at least one 2-AT from the library had been identified as a hit in every case.[12] This led to our investigation of their role as potential promiscuous binders via FBDD. We report here the results of

our study and analysis of all 2-ATs in our fragment library (Figure 1) by surface plasmon resonance (SPR) against 6 different protein targets. Target proteins included in this study were the *Plasmodium falciparum* apical membrane antigen AMA1,[22] the E3 ubiquitin ligase adapter protein SPSB2,[24] two DsbA oxidoreductases from different bacterial species (oxidoreductase 1 and 2), carbonic anhydrase II, and a kinase. These were selected as the proteins exhibit diversity in function and have little structural similarity across their binding sites. All except oxidoreductase 2 had known positive controls that were used in the SPR assays. In addition, the fragment screen contained a number of internal controls. Four fragments used in this study were previously identified as binders to SPSB2 by SPR (**1**, **3**, **15**, **17**; unpublished data), carbonic anhydrase II was expected to show a very strong preference for compounds containing a terminal sulphonamide and a prior fragment-screen of oxidoreductase 2 (unpublished data) had revealed only very weak binding for any fragment. Thus, oxidoreductase 2 was included as a negative control.

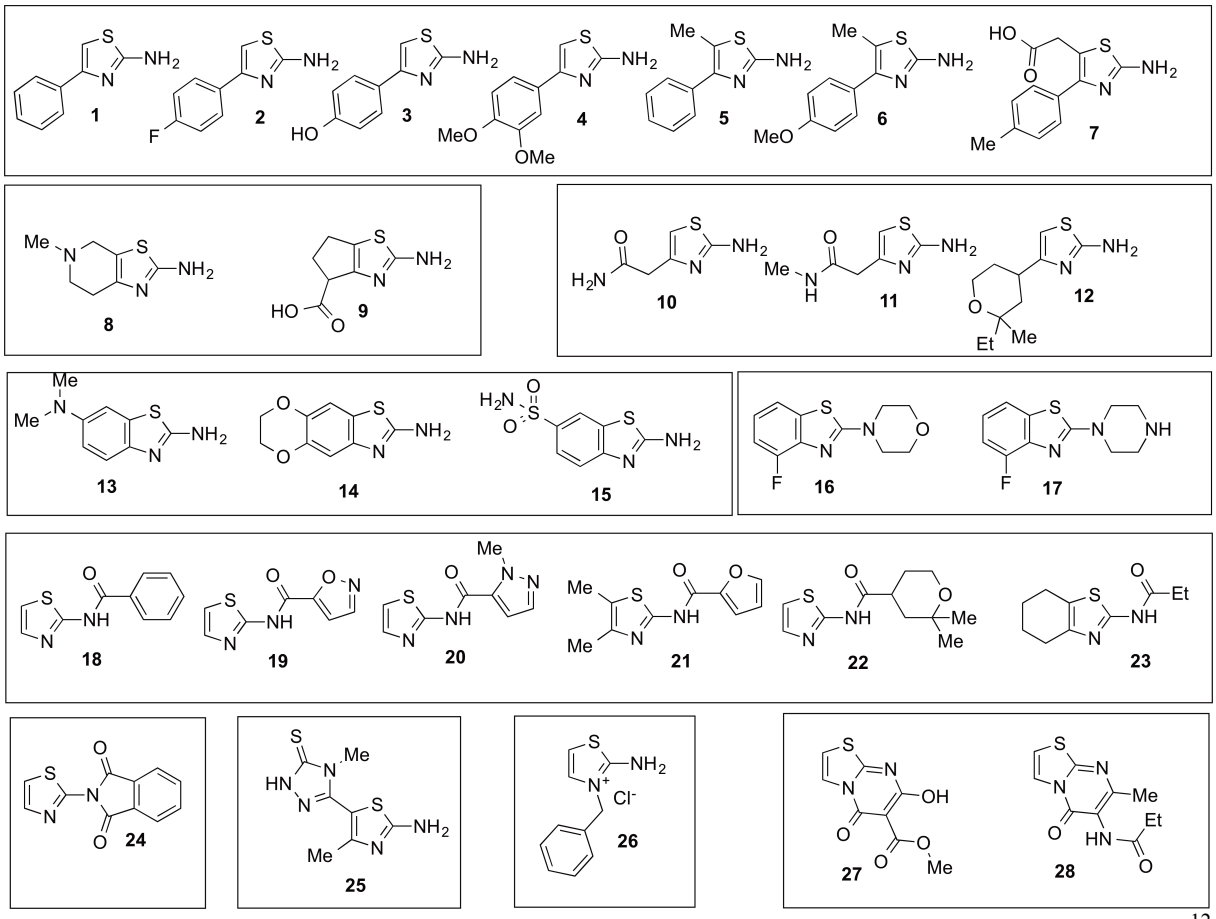


Figure 1. 2-Aminothiazole (2-AT) and 2-AT-like compounds in the MIPS Fragment Library[12]

Dose-response determinations for positive controls gave K_D values that were all within the expected ranges for each target protein (Supplementary Figure S1 and S2), confirming that all proteins were stable and active on the sensor chip under the conditions used. Binding of 2-ATs was carried out at a single concentration (200 μM) in duplicate, and the average raw response was converted to percentage occupancy of binding, corrected for the molecular weight of each compound (Figure 2). The sensorgrams for selected 2-ATs can be found in Supplementary Figures S3–5. Injections of buffer blank were included between each compound, both to provide for double referencing and to minimise carry-over problems from poorly behaved compounds.

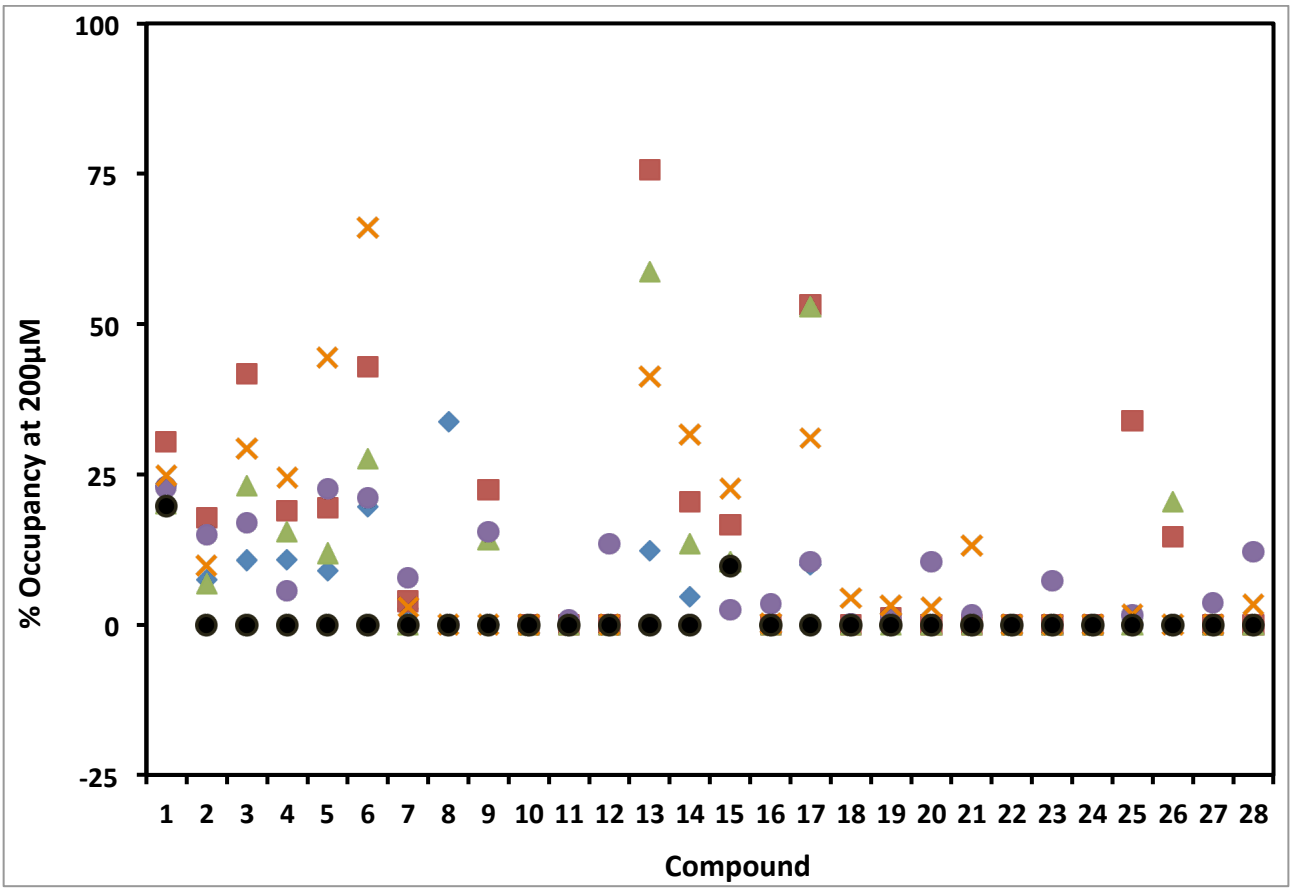


Figure 2. Binding of 2-ATs to AMA1 (green triangles), SPSB2 (red squares), oxidoreductase 1 (purple circles), oxidoreductase 2 (black circles), kinase (orange crosses), and CAII (blue diamonds). Compounds were tested at a concentration of 200 μM, and the raw response converted to percentage occupancy. Note that the data points for **8** (AMA1 and SPSB2) and **15** (CAII) are off the Y-axis scale and occur at 125% and 175%, respectively.

Figure 2 reveals a consistency in the pattern of protein binding. Individual fragments displayed very little discrimination in binding to five of the six proteins; excluding oxidoreductase 2, fragments generally bound either all five proteins or none. While there were some exceptions, there was also a general tendency for fragments to bind each protein at similar occupancy, suggesting approximately the same affinity and thus providing very little clear SAR. Further, only two 2-ATs (**1** and **15**) demonstrated any binding to oxidoreductase 2, consistent with its observed limited ability to bind fragments of any description.

To examine the SAR in a more systematic manner, 2-ATs from the MIPS library were clustered based on linear fingerprints and key chemical features, and the screening results examined for each grouping. Compounds **1** – **7** feature 4-aryl substitutions with a free 2-amino group, with or without a substituent at the 5-position. Fragments **8** – **12** maintain the free 2-amino group, with **8** and **9** containing fused aliphatic rings, whereas **10** – **12** have aliphatic substituents at the 4-position. Fragments **13** – **15** integrate benzothiazoles with their 2-amino group intact, while fragments **16** and **17** have morpholino or piperazino attachments through their 2-amino group, respectively. Fragments **18** – **24** are 2-amido containing thiazoles and **27** and **28** are thiazolo[3,2-*a*]pyrimidines, in which the 2-amino group is incorporated in the heterocycle. The triazolothione (**25**) and thiazolium (**26**) complete the selection of 2-ATs from the MIPS library. These molecules contain diverse substituents around the 2-AT core. Despite the chemical diversity in the 28 2-ATs in the library, evaluation of their binding to the different targets provided little clear SAR.

As an example, 4-phenylthiazol-2-amine (**1**), highlights the problem with this class of molecule, showing binding to all six proteins examined, all with similar occupancy levels (~25%). As this fragment contains 12 heavy atoms this level of binding corresponds to a favourable ligand efficiency (i.e. binding energy per heavy atom) in each case.[25] Thus, if this fragment were tested in an isolated setting against one particular protein, the likelihood of follow up by medicinal chemists

is high, demonstrating the insidious behavior associated with this class. To rule out an impurity in the commercial preparation of fragment **1**, it was resynthesised and purified. This gave similar occupancy levels to the commercial product across the protein targets, indicating an inherent problem with the molecule itself as opposed to reactive precursors or side-products potentially present in the commercial source.

Fragments **2** – **6** show binding to all of the proteins examined except oxidoreductase 2. The addition of the 5-methyl group appears to have little effect, as does the nature of the substituents on the 4-aryl ring, which include hydroxy, methoxy and fluoro moieties. The addition of carboxylic or amido containing chains (**7**, **10**, **11**) at the 4- or 5-position, coupled with saturation of the 4-aryl ring (**12**), seems to abrogate promiscuity, although, of these fragments, only **12** is observed to bind to any of the proteins at an occupancy > 10% and **12** binds only to oxidoreductase 1. Of the benzothiazole fragments with a 2-amino group, **13** and **14** do not bind CAII or oxidoreductase 2, **15** binds all six proteins and in the case of bicyclic **8**, occupancy levels > 100% of the theoretical R_{max} are observed for two proteins. Fragment **15** is the only sulfonamide in the series and demonstrates affinity to CAII, as expected for this protein since it specifically binds terminal sulfonamides, but a number of other fragments (example **3**, **8**, **13**) also demonstrate binding to CAII.[26] The 2-substituted aminothiazoles are an interesting test case; morpholino (**16**) shows little to no binding to any protein tested, whereas piperazino (**17**) binds to five of the six proteins. The 2-amido and heterocyclic thiazoles (**18** – **24**, **27** and **28**) show little or no binding to any of the proteins and do not appear to be inherently promiscuous. Triazolo thione **25** and thiazolium **26** showed no binding to 5 and 4 of the targets, respectively. In summary, 2-ATs appear to be generally promiscuous at fragment screening concentrations whereas the corresponding amides are not. Within the set of 2-ATs, there are no clear characteristics that distinguish the more promiscuous binders from their less promiscuous counterparts, and thus there is no clear SAR. We have so far been unsuccessful in generating a crystal structure with any 2-AT fragment, although 2-AT crystal complexes have been

reported by others from FBDD programs targeting prostaglandin D2 synthase (PGDS)[27] and leukotriene A4 hydrolase (LTA4H).[28] In the case of LTA4H, two molecules were bound at different sites on the protein, whereas in the case of PGDS the 2-AT was not itself a screening hit, but a compound designed as an analogue of the original hit.

The evidence of non-specific binding of 2-ATs to SPSB2 is supported by protein-detected NMR data. For example, ^{19}F-NMR studies on 5-F-Trp-SPSB2 have proven effective as an analytical tool for determining binding to the active site of the protein.[29] Specific binding of peptides and ligands to the active site is characterised by a downfield shift of the fluorine peak corresponding to residue W207 (Figure 3). In contrast, no downfield shift of this peak was observed for the 2-ATs that were observed to bind SPSB2 by SPR, despite their reasonable binding occupancy at 200 μM. This suggests that these 2-ATs do not bind to the active site of SPSB2, and that binding probably occurs at either a secondary or non-specific site on the protein.

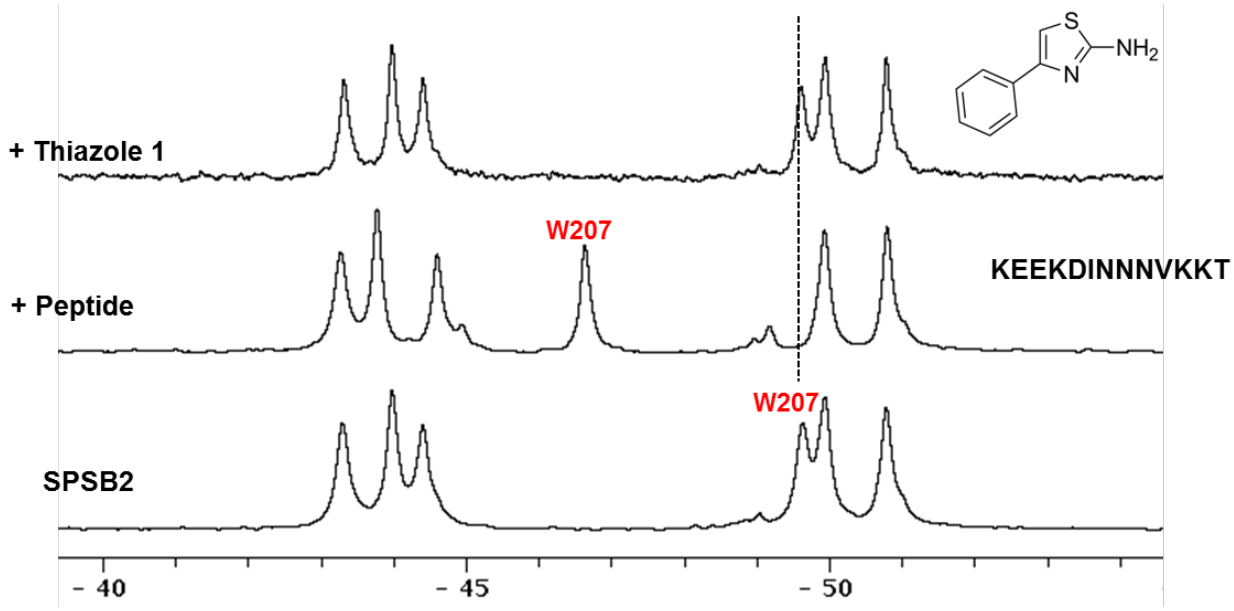


Figure 3. ^{19}F-NMR spectra of 5F-Trp SPSB2 alone (bottom), with 500 μM control peptide (middle), and 3 mM **1** (top). Specific binding of the control peptide to the active site is characterised by a downfield shift of the peak corresponding to W207.[24] No such shift is observed for the thiazole fragment, suggesting that binding to SPSB2 is non-specific.

HTS Screening – Academic and Corporate Collections

We were interested in assessing the apparent promiscuity of 2-AT derivatives in fragment-based *vs* high-throughput screening. To analyse the latter, we adopted two approaches. The first was to analyse the relative prevalence of 2-AT-based PAINS in a HTS library of 93,000 compounds housed at the Walter and Eliza Hall Institute (WEHI).[30] The second was to analyse the full AstraZeneca HTS screening deck. With respect to the first approach, a search of the WEHI HTS library revealed 989 2-AT-based compounds and an additional 1012 compounds where the amine was acylated. Amongst these are 14 sub-classes of PAINS, the structural definitions of which are given in Supplementary Figure 6. This analysis was concordant with data from our fragment library in that no PAINS class contains 2-ATs in which the 2-amino group is acylated.

The incidence of frequent hitting 2-ATs was then investigated with greater statistical power using a large corporate collection (AstraZeneca, January 2014). These compounds have been tested in a range of HTS campaigns with concentrations typically around 10 μM. To assess the promiscuity of 2-ATs we employed a descriptor that has been designed to flag potential frequent-hitter behavior for all compounds with sufficient data.[31] The descriptor is calculated by first designating a compound as active/inactive in each available screen using data from all historical AstraZeneca HTS campaigns. The body of HTS data that is available varies from compound to compound for a variety of reasons (age of compound, sample availability, membership of screen sets, manual collation), so potentially anomalous binding behavior was designated by comparing observed incidence of activity for a given compound to the expected activity for an 'average' compound. In this way a descriptor (termed the pBSF score) was derived that indicates whether the compound is more active than expected based on historical observations.[31] This knowledge-based descriptor is calculated as the negative logarithm of the probability of encountering the observed level of activity for a compound that behaves normally, by chance. A low probability, and hence a high score, indicates that it is very unlikely that the observed level of activity would be observed for a well-

behaved compound, which therefore suggests that the compound is a promiscuous hitter. We first analysed the number of 2-ATs that had a pBSF score above a threshold score of 2.0. Secondly we divided the 2-ATs into classes based on substructures, and determined the proportion of each substructure that had pBSF scores > 2.0.

A comparison was made of the pBSF score for the 61,040 2-ATs within the AstraZeneca collection. For comparison, a similar number of compounds were selected at random from the library, as documented in Figure 4. For the 2-ATs there were 6,122 compounds with a score > 2, whereas the random set had 3,704 compounds with a score > 2. It is evident that the incidence of frequent hitters in the random set is lower than for the 2-ATs, which suggests that the 2-ATs are somewhat promiscuous.

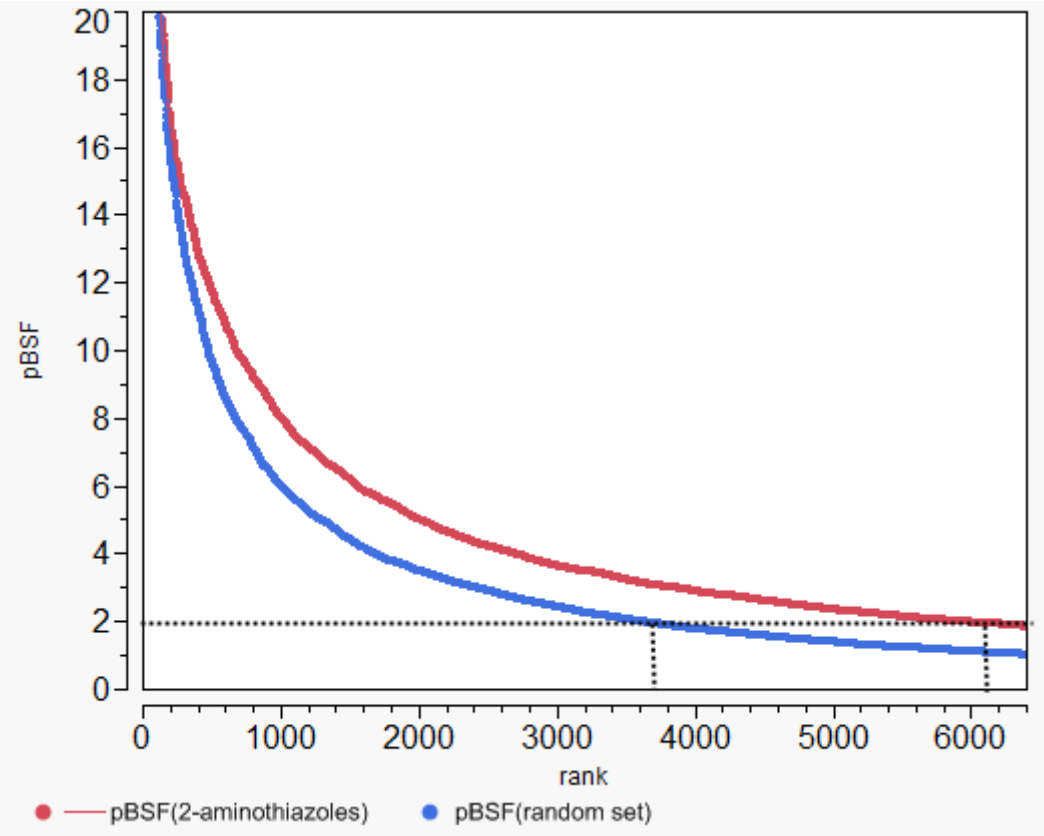
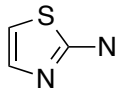
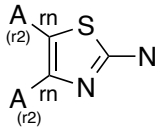
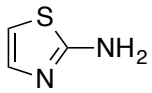
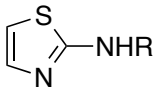
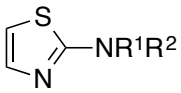
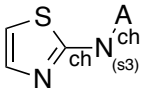
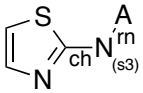
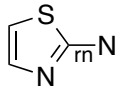
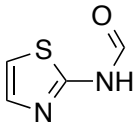


Figure 4. Rank-ordered frequent hitter scores for the first 6,122 2-ATs (red line) from a total set of 61,040 2-ATs retrieved by substructure from the AstraZeneca screening collection. Dotted lines show the numbers of compounds at a threshold value > 2, which is indicative of anomalous binding behavior.

Next we divided the 2-AT structural classes using the substructures shown in Table 1 and counted the incidence of suspicious compounds using a threshold of > 2 to distinguish potential frequent hitters from 'clean' compounds and those lacking sufficient data.

Table 1. Nine simplified 2-AT classes were inspected for promiscuous behavior across the AstraZeneca corporate collection. Counts are shown for subsets of small molecules (MW < 300) and larger compounds (MW ≥ 300).

Class		Substructure [a]	N	N with data MW<300	N with data MW>=300	N(FH) MW< 300	N(FH) MW>= 300	fraction pBSF>2 LowMW	fraction pBSF>2 highMW
1	*aminothiazoles*		77,826	8,070	52,970	559	5,563	6.93	10.50
2	*fused aminothiazoles*		3,259	284	2,190	17	144	5.99	6.58
3	*primary amines*		8,678	1,781	4,425	143	564	8.03	12.75
4	*secondary amines*		55,445	5,336	38,337	372	3,970	6.97	10.36
5	*tertiary amines*		13,864	978	10,312	53	1,068	5.42	10.36
6	*tertiary amines, acyclic*		3,517	304	2.795	15	108	4.93	3.86
7	*tertiary amines, cyclic*		10,307	661	7,495	37	960	5.60	12.81
8	*thiazole cyclic amines ('embedded')*		59	21	25	0	0	0.00	0.00
9	*acylated thiazole amines*		29,816	2,613	21,447	99	1,657	3.79	7.73

[a] Structure legend: A, any atom; rn, ring bond; ch, chain bond; s*n*, substitution count *n*; r*n*, number of ring bonds

The data in Table 1 are consistent with the patterns observed with the MIPS FBDD and WEHI HTS data, namely that 2-ATs show a higher incidence of anomalous binding behavior across a number of substructure classes. Typically, structures across the AstraZeneca collection that have on average ≤ 6.0% of the compounds within the class with a pBSF score > 2.0 are not considered to be frequent hitters. Here, an elevated incidence (approx. 10%) of such behavior can be seen for the overall class of 2-ATs and for subclasses 1, 3, 4, 5 and 7, which is in line with observations for the WEHI HTS library.

It is notable that promiscuity can be observed in the HTS data even within sets of low-MW compounds For example within class 3, there are 1781 primary amines with MW < 300, of which 8% are classified as frequent hitters. Although these compounds are "fragment-like" in terms of their size, they are observed to be promiscuous in the HTS data despite being screened at a typical HTS concentration of 10 μM, which suggests that their promiscuity in the FBDD data is not due solely to the higher concentrations used in the fragment screen.

Subclass 9 comprises the acylated 2-ATs, which show a lower incidence of frequent-hitter behaviour than the entire class of 2-ATs both for the low-MW and high-MW set. Nonetheless, the incidence of frequent-hitter behavior in the high-MW set is somewhat higher than the expected "normal" level of 6.0% but this can be explained by biases that are apparent in target-specific analyses of the screening data. Inspection of a target-specific frequent hitter descriptor (results not shown) reveals that subclass 9 shows an increased incidence of kinase activity. As kinases are typical drug targets pursued by pharma companies, a bias may be observed in the data where kinase-like motifs have been screened preferentially in kinase-targeted screens. This skews the descriptor results based on such data to some extent, i.e. some of the frequent hitters identified may actually be kinase-frequent hitters, and not necessarily problematic. Within the 1,657 potential frequent hitters of kinase-like class 9 (MW≥300), a subset of 556 molecules appears to be kinase-

frequent hitters. Subclass 7 contains a structural motif less likely to hit kinases, yet appears to have a high frequent-hitter incidence and contains 960 suspicious compounds (MW≥300). Of these, a smaller fraction of only 136 display kinase-related frequent hitter behavior, suggesting that the origin of frequent-hitting behavior in the remaining compounds is mostly not related to kinases. We suggest therefore that the somewhat increased incidence of frequent hitters we observe for the acylated 2-ATs in this data set is an artefact resulting from a kinase-activity bias, and that the acylated 2-ATs as a whole do not meet the criteria to be categorised as frequent hitters.

Using the AstraZeneca 2-AT set of 61,040 compounds and data derived from the AstraZeneca corporate collection, we investigated overall correlations of frequent hitter incidence with lipophilicity (experimental logD, clogP), experimental solubility (pSol, calculated as log[Sol/uM]), calculated polar and non-polar surface area (PSA and NPSA, in $Å^2$ and %PSA in %), and molecular volume.[32,33] Correlations are observed with ion class, PSA, %PSA, as well as donor and acceptor counts (Supplementary Figure S7), suggesting that there could be a relation with polarity (each of these properties reflects polarity and they are interrelated). A weak trend with logD (experimental octanol-water partitioning coefficient) was observed, with low-logD compounds somewhat more likely to be frequent hitters. No relation is observed with experimental aqueous solubility, but a trend is seen with increasing clogP. Although there are no categorical reasons for promiscuity in these compounds, the observation that ionisable groups and high lipophilicity may increase anomalous behavior in the 2-AT class is in line with recent observations by Tarcsay and Keserű,[34] who observed similar trends.

DISCUSSION

We have demonstrated that certain members of the class of compounds containing a 2-AT substructure are frequent hitting and promiscuous fragments in the context of FBDD, where screening is undertaken using biophysical binding assays. We have dubbed these fragments Promiscuous 2-AminoThiazoles (PrATs). Analysis of binding of the 28 2-ATs in our fragment

library was undertaken by SPR and NMR against six unrelated protein targets. Although some patterns have emerged, such as the presence of the free 2-amino group seeming to promote promiscuity, a clear mechanism of action has not been identified at this point. Nonetheless, our SPR analysis produced flat and confusing SAR against several of the targets, which has previously been noted in the characterisation of PAINS identified in HTS.

As several approved drugs contain a 2-AT, it is clearly possible to accommodate this structure in a therapeutically useful compound. However, it is noteworthy that a number of PAINS subclasses also contain the 2-AT chemotype, which suggests that the 2-AT may carry some risks if selected for development. For example, in common with many aromatic amines, 2-ATs can be Ames positive dependent on the substitution pattern. Further, aminothiazoles have been associated with liver toxicity via bioactivation of the double bond, leading to formation of thioureas that are further bioactivated. However, problems relating to reactivity and covalent modification are unlikely to contribute to the observed promiscuity in biophysical binding assays, since the ligand-detected NMR assays used in FBDD typically provide evidence of compound identity and binding simultaneously, whilst both NMR and SPR discriminate between covalent and non-covalent interactors.

Analysis of the AstraZeneca HTS data reveals that 2-ATs show elevated frequency as screening hits relative to the compound library as a whole. However, the percentage (10%) is significantly lower than the corresponding value for other known PAINS classes, which can be around 15-20%.[31] The AstraZeneca data also suggest that acylation of the 2-amino group ameliorates the problem of frequent hitting, which indicates that not all 2-ATs are problematic in the context of HTS. Similarly, only 3.2% of 2-AT-containing compounds in the WEHI HTS library are defined as PAINS, which does not provide a strong case for exclusion of all 2-ATs from HTS collections.[14] Thus, the two HTS analyses are in broad agreement with each other and suggest that the majority of

2-ATs do not seem to be generally promiscuous at HTS-relevant concentrations in the 10 – 25 μM range. Even at HTS-relevant concentrations, some optimizable hits can be relatively promiscuous and we have previously identified a 2-aminobenzothiazole derivative[16] that hit four out of the six HTS assays selected for the PAINS analysis,[14] yet led to a highly selective and potent compound.[35]

Nonetheless, many 2-ATs were observed to be problematic in both HTS and FBDD, suggesting that the promiscuity observed in FBDD is not solely due to the higher concentration used in the fragment screens. Whilst the higher hit rates that are expected in FBDD dictate that many fragments are likely to hit more than one target, these hits are only useful if they can be elaborated into more potent compounds. On this basis, it is possible that certain substructures such as 2-ATs, which show some limited promiscuity in HTS assays, but would not be flagged for exclusion from a fragment library by analysis of chemical or physical properties, are in fact undesirable as members of a fragment library based on a retrospective analysis of their behaviour in screening assays. A similar approach has previously been described in the HTS setting for deprioritising the use of compounds that are known to be promiscuous.[36]

CONCLUSION

We have identified 2-ATs as a promiscuous substructure in screens of our fragment library. Each of the 28 fragments containing a 2-AT substructure passed all of the biophysical and chemical filters that were used in designing the library and were demonstrated to have appropriate purity and aqueous solubility. However, based on our findings reported here and our unsuccessful attempts to optimize these fragments against different targets, we have removed 2-ATs from the fragment library at MIPS, on the basis that we judge the prospects for such compounds to be progressable as much smaller than the chances that they will not.

EXPERIMENTAL SECTION

General procedures. Biacore sensor chips, *N*-hydroxysuccinimide (NHS), *N*-ethyl-*N*′-(3-dimethylaminopropyl) carbodiimide (EDC), ethanolamine HCl, HBS-P running buffer, and P-20 surfactant were obtained from GE Healthcare. Carbonic anhydrase II and 4-carboxybenzyl sulfonamide were obtained from Sigma-Aldrich. Dr Isabelle Lucet (WEHI) provided the kinase protein and positive control H4. Doubly His-tagged oxidoreductases 1 and 2 were provided by Dr Martin Williams (MIPS). All 2-aminothiazole fragments were obtained from Chembridge or Life.

Surface Plasmon Resonance (SPR) conditions

Expression and purification of hexahistidine-tagged AMA1,[37] and GST-tagged SPSB2,[24] were as described previously. Binding of 2-aminothiazoles to the target proteins was measured on a Biacore T200 instrument (GE Healthcare). AMA1, SPSB2, and CAII were immobilized onto a CM5 sensor chip (Biacore) by standard amine coupling chemistry using sodium acetate at pH 4.5, in running buffer A. Doubly-His_6-tagged oxidoreductase proteins were immobilised onto an NTA chip (Biacore) charged with Ni^{2+} according to manufacturer's instructions, in running buffer B. Singly-His_6-tagged Kinase was immobilized to an NTA sensor chip charged with Ni^{2+} using the Capture-Couple method,[38] in running buffer C. Immobilisation levels were typically 9000 RU for AMA1, 6000 RU for SPSB2 and CAII, and 2000 – 3000 RU for both oxidoreductases, and the kinase.

Running buffer A: 25mM HEPES, 150mM NaCl, 5% DMSO, 0.005% P-20, pH 7.4.

Running buffer B: 25mM HEPES, 200mM NaCl, 5% DMSO, 0.005% Tween-20, pH 8.0

Running buffer C: 25mM HEPES, 150mM NaCl, 4mM $MgCl_2$, 2mM TCEP, 3% DMSO, 0.005% Tween-20, pH 7.5.

SPR Screening of 2-Aminothiazoles

Binding assays were carried out at 25°C using a flow-rate of 100 μL/min in running buffer A (AMA1, SPSB2, CAII), running buffer B (oxidoreductase 1 and 2), or running buffer C (kinase). 200 mM fragment stocks were diluted in the appropriate running buffer to obtain 200 μM working concentrations. The association and dissociation phases of binding were each followed for 30 s. An

identical injection of running buffer was carried out between each fragment injection, and the average of adjacent blanks was subtracted from the raw fragment response. Raw sensorgram data were reduced, solvent-corrected, and double-referenced using BIAEvaluation Software (GE Healthcare). Raw responses were converted to percentage occupancy by the method of Giannetti[39] using an R_{max} value based on the response measured with a saturating concentration of the appropriate control, and the molecular weight of each fragment. For oxidoreductase 2, there is no positive control available, and binding is reported as percentage of the theoretical R_{max}.

SPR Positive Controls

The R1 peptide[40] was used as a control for AMA1. A peptide corresponding to residues within the N-terminus of inducible nitric oxide synthase was used as a control for SPSB2.[24] 4-Carboxybezyl sulfonamide was used as a control for CAII. An elaborated fragment "H4" was used as a control for the kinase. An elaborated fragment "LA010" was used as a control for oxidoreductase 1. No positive control was available for oxidoreductase 2. Dose-response determinations were carried out for all positive controls to confirm protein activity. A concentration series of 2-fold dilutions in running buffer was used, with a top concentration of 1 μM (R1), 5 μM (iNOS peptide), 10 μM (4-CBS), 20 μM (H4) or 200 μM (LA010) (see Supplementary Figure S1).

ACKNOWLEDGMENTS

The authors would like to thank Martin Williams for the oxidoreductases 1 and 2, Luke Adams for the oxidoreductase 1 positive control, Hiromasa Yagi for assistance on the ^{19}F NMR work and Isabelle Lucet for the kinase protein sample and positive control. This work was supported in part by project grants from the Australian National Health and Medical Research Council (NHMRC) (1045995, 1009785, 1022693, 1025150). J.B.B. and R.S.N. acknowledge fellowship support from the NHMRC.

ABBREVIATIONS USED

2-AT, 2-aminothiazole; AMA1, apical membrane antigen 1; CAII, carbonic anhydrase II; FBDD, fragment-based drug discovery; HTS, high-throughput screening; LTA4H, Leukotriene A4 hydrolase; MIPS, Monash Institute of Pharmaceutical Sciences; NMR, nuclear magnetic resonance; PAINS, Pan Assay INterference compoundS; PrATs, Promiscuous 2-AminoThiazoles; PGDS, Prostaglandin D2 synthase; PPIs, protein-protein interactions; SPSB2, SPRY domain-containing SOCS box protein 2; SAR, structure-activity relationships; Walter and Eliza Hall Institute, WEHI.

REFERENCES

(1) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **1996**, *274*, 1531–1534.

(2) Erlanson, D. A.; McDowell, R. S.; O'Brien, T. Fragment-Based Drug Discovery. *J. Med. Chem.* **2004**, *47*, 3463–3482.

(3) Tsai, J.; Lee, J. T.; Wang, W.; Zhang, J.; Cho, H.; Mamo, S.; Bremer, R.; Gillette, S.; Kong, J.; Haass, N. K.; Sproesser, K.; Li, L.; Smalley, K. S. M.; Fong, D.; Zhu, Y.-L.; Marimuthu, A.; Nguyen, H.; Lam, B.; Liu, J.; Cheung, I.; Rice, J.; Suzuki, Y.; Luu, C.; Settachatgul, C.; Shellooe, R.; Cantwell, J.; Kim, S.-H.; Schlessinger, J.; Zhang, K. Y. J.; West, B. L.; Ben Powell; Habets, G.; Zhang, C.; Ibrahim, P. N.; Hirth, P.; Artis, D. R.; Herlyn, M.; Bollag, G. Discovery of a Selective Inhibitor of Oncogenic B-Raf Kinase with Potent Antimelanoma Activity. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 3041–3046.

(4) LaPlante, S. R.; Padyana, A. K.; Abeywardane, A.; Bonneau, P.; Cartier, M.; Coulombe, R.; Jakalian, A.; Wildeson-Jones, J.; Li, X.; Liang, S.; McKercher, G.; White, P.; Zhang, Q.; Taylor, S. J. Integrated Strategies for Identifying Leads That Target the NS3 Helicase of the Hepatitis C Virus. *J. Med. Chem.* **2014**, *57*, 2074–2090.

(5) Edfeldt, F. N. B.; Folmer, R. H. A.; Breeze, A. L. Fragment Screening to Predict Druggability (Ligandability) and Lead Discovery Success. *Drug Discov. Today* **2011**, *16*, 284–287.

(6) Fink, T.; Bruggesser, H.; Reymond, J.-L. Virtual Exploration of the Small-Molecule Chemical Universe Below 160 Daltons. *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 1504–1508.

(7) Congreve, M.; Chessari, G.; Tisi, D.; Woodhead, A. J. Recent Developments in Fragment-Based Drug Discovery. *J. Med. Chem.* **2008**, *51*, 3661–3680.

(8) Fink, T.; Reymond, J.-L. Virtual Exploration of the Chemical Universe Up to 11 Atoms of C, N, O, F: Assembly of 26.4 Million Structures (110.9 Million Stereoisomers) and Analysis for New Ring Systems, Stereochemistry, Physicochemical Properties, Compound Classes, and Drug Discovery. *J. Chem. Inf. Model.* **2007**, *47*, 342–353.

(9) Murray, C. W.; Carr, M. G.; Callaghan, O.; Chessari, G.; Congreve, M.; Cowan, S.; Coyle, J. E.; Downham, R.; Figueroa, E.; Frederickson, M.; Graham, B.; McMenamin, R.; O'Brien, M. A.; Patel, S.; Phillips, T. R.; Williams, G.; Woodhead, A. J.; Woolford, A. J.-A. Fragment-Based Drug Discovery Applied to Hsp90. Discovery of Two Lead Series with High Ligand Efficiency. *J. Med. Chem.* **2010**, *53*, 5942–5955.

(10) Bollag, G.; Hirth, P.; Tsai, J.; Zhang, J.; Ibrahim, P. N.; Cho, H.; Spevak, W.; Zhang, C.; Zhang, Y.; Habets, G.; Burton, E. A.; Wong, B.; Tsang, G.; West, B. L.; Powell, B.; Shellooe, R.; Marimuthu, A.; Nguyen, H.; Zhang, K. Y. J.; Artis, D. R.; Schlessinger, J.; Su, F.; Higgins, B.; Iyer, R.; D'Andrea, K.; Koehler, A.; Stumm, M.; Lin, P. S.; Lee, R. J.; Grippo, J.; Puzanov, I.; Kim, K. B.; Ribas, A.; McArthur, G. A.; Sosman, J. A.; Chapman, P. B.; Flaherty, K. T.; Xu, X.; Nathanson, K. L.; Nolop, K. Clinical Efficacy of a RAF Inhibitor Needs Broad Target Blockade in BRAF-Mutant Melanoma. *Nature* **2010**, *467*, 596–599.

(11) Bollag, G.; Tsai, J.; Zhang, J.; Zhang, C.; Ibrahim, P.; Nolop, K.; Hirth, P. Vemurafenib: the First Drug Approved for BRAF-Mutant Cancer. *Nat. Rev. Drug Discov.* **2012**, *11*, 873–886.

(12) Doak, B. C.; Morton, C. J.; Simpson, J. S.; Scanlon, M. J. Design and Evaluation of the Performance of an NMR Screening Fragment Library. *Aust. J. Chem.* **2013**, *66*, 1465.

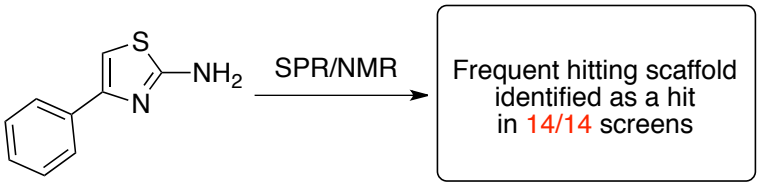
(13) Baurin, N.; Aboul-Ela, F.; Barril, X.; Ben Davis; Drysdale, M.; Dymock, B.; Finch, H.; Fromont, C.; Richardson, C.; Simmonite, H.; Hubbard, R. E. Design and Characterization of Libraries of Molecular Fragments for Use in NMR Screening Against Protein Targets. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2157–2166.

(14) Baell, J. B.; Holloway, G. A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) From Screening Libraries and for Their Exclusion in

Bioassays. *J. Med. Chem.* **2010**, *53*, 2719–2740.
(15) Hansch, C.; Fujita, T. P-Σ-Π Analysis. a Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616–1626.
(16) Baell, J. B.; Ferrins, L.; Falk, H.; Nikolakopoulos, G. PAINS: Relevance to Tool Compound Discovery and Fragment-Based Screening. *Aust. J. Chem.* **2013**, *66*, 1483.
(17) Camps, M.; Rückle, T.; Ji, H.; Ardissone, V.; Rintelen, F.; Shaw, J.; Ferrandi, C.; Chabert, C.; Gillieron, C.; Françon, B.; Martin, T.; Gretener, D.; Perrin, D.; Leroy, D.; Vitte, P.-A.; Hirsch, E.; Wymann, M. P.; Cirillo, R.; Schwarz, M. K.; Rommel, C. Blockade of PI3Kgamma Suppresses Joint Inflammation and Damage in Mouse Models of Rheumatoid Arthritis. *Nat. Med.* **2005**, *11*, 936–943.
(18) Huth, J. R.; Mendoza, R.; Olejniczak, E. T.; Johnson, R. W.; Cothron, D. A.; Liu, Y.; Lerner, C. G.; Chen, J.; Hajduk, P. J. ALARM NMR: a Rapid and Robust Experimental Method to Detect Reactive False Positives in Biochemical Screens. *J. Am. Chem. Soc.* **2005**, *127*, 217–224.
(19) Huth, J. R.; Song, D.; Mendoza, R. R.; Black-Schaefer, C. L.; Mack, J. C.; Dorwin, S. A.; Ladror, U. S.; Severin, J. M.; Walter, K. A.; Bartley, D. M.; Hajduk, P. J. Toxicological Evaluation of Thiol-Reactive Compounds Identified Using a La Assay to Detect Reactive Molecules by Nuclear Magnetic Resonance. *Chem. Res. Toxicol.* **2007**, *20*, 1752–1759.
(20) Metz, J. T.; Huth, J. R.; Hajduk, P. J. Enhancement of Chemical Rules for Predicting Compound Reactivity Towards Protein Thiol Groups. *J. Comput. Aided Mol. Des.* **2007**, *21*, 139–144.
(21) Mjambili, F.; Njoroge, M.; Naran, K.; De Kock, C.; Smith, P. J.; Mizrahi, V.; Warner, D.; Chibale, K. Synthesis and Biological Evaluation of 2-Aminothiazole Derivatives as Antimycobacterial and Antiplasmodial Agents. *Bioorg. Med. Chem. Lett.* **2014**, *24*, 560–564.
(22) Lim, S. S.; Debono, C. O.; MacRaild, C. A.; Chandrashekaran, I. R.; Dolezal, O.; Anders, R. F.; Simpson, J. S.; Scanlon, M. J.; Devine, S. M.; Scammells, P. J.; Norton, R. S. Development of Inhibitors of Plasmodium Falciparum Apical Membrane Antigen 1 Based on Fragment Screening. *Aust. J. Chem.* **2013**, *66*, 1530–1536.
(23) Chuprina, A.; Lukin, O.; Demoiseaux, R.; Buzko, A.; Shivanyuk, A. Drug- and Lead-Likeness, Target Class, and Molecular Diversity Analysis of 7.9 Million Commercially Available Organic Compounds Provided by 29 Suppliers. *J. Chem. Inf. Model.* **2010**, *50*, 470–479.
(24) Kuang, Z.; Lewis, R. S.; Curtis, J. M.; Zhan, Y.; Saunders, B. M.; Babon, J. J.; Kolesnik, T. B.; Low, A.; Masters, S. L.; Willson, T. A.; Kedzierski, L.; Yao, S.; Handman, E.; Norton, R. S.; Nicholson, S. E. The SPRY Domain-Containing SOCS Box Protein SPSB2 Targets iNOS for Proteasomal Degradation. *J. Cell Biol.* **2010**, *190*, 129–141.
(25) Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand Efficiency: a Useful Metric for Lead Selection. *Drug Discov. Today* **2004**, *9*, 430–431.
(26) Supuran, C. T.; Scozzafava, A. Carbonic Anhydrase Inhibitors and Their Therapeutic Potential. *Exp. Opin. Ther. Patents* **2000**, *10*, 575–600.
(27) Hohwy, M.; Spadola, L.; Lundquist, B.; Hawtin, P.; Dahmen, J.; Groth-Clausen, I.; Nilsson, E.; Persdotter, S.; Wachenfeldt, Von, K.; Folmer, R. H. A.; Edman, K. Novel Prostaglandin D Synthase Inhibitors Generated by Fragment-Based Drug Design. *J. Med. Chem.* **2008**, *51*, 2178–2186.
(28) Davies, D. R.; Mamat, B.; Magnusson, O. T.; Christensen, J.; Haraldsson, M. H.; Mishra, R.; Pease, B.; Hansen, E.; Singh, J.; Zembower, D.; Kim, H.; Kiselyov, A. S.; Burgin, A. B.; Gurney, M. E.; Stewart, L. J. Discovery of Leukotriene A4 Hydrolase Inhibitors Using Metabolomics Biased Fragment Crystallography. *J. Med. Chem.* **2009**, *52*, 4694–4715.
(29) Leung, E. W. W.; Yagi, H.; Harjani, J. R.; Mulcair, M. D.; Scanlon, M. J.; Baell, J. B.; Norton, R. S. (19) F NMR as a Probe of Ligand Interactions with the iNOS Binding Site of SPRY Domain-Containing SOCS Box Protein 2. *Chem. Biol. Drug Des.* **2014**, in press.

doi: 10.1111/cbdd.12355.
(30) Baell, J. B. Broad Coverage of Commercially Available Lead-Like Screening Space with Fewer Than 350,000 Compounds. *J. Chem. Inf. Model.* **2013**, *53*, 39–55.
(31) M Nissink, J. W.; Blackburn, S. Quantification of Frequent-Hitter Behavior Based on Historical High-Throughput Screening Data. *Future Med. Chem.* **2014**, *6*, 1113–1126.
(32) Bruneau, P. Search for Predictive Generic Model of Aqueous Solubility Using Bayesian Neural Nets. *J. Chem. Inf. Model.* **2001**, *41*, 1605–1616.
(33) Bruneau, P.; McElroy, N. R. logD7.4 Modeling Using Bayesian Regularized Neural Networks. Assessment and Correction of the Errors of Prediction. *J. Chem. Inf. Model.* **2006**, *46*, 1379–1387.
(34) Tarcsay, Á.; Keserü, G. M. Contributions of Molecular Properties to Drug Promiscuity. *J. Med. Chem.* **2013**, *56*, 1789–1795.
(35) Lessene, G.; Czabotar, P. E.; Sleebs, B. E.; Zobel, K.; Lowes, K. N.; Adams, J. M.; Baell, J. B.; Colman, P. M.; Deshayes, K.; Fairbrother, W. J.; Flygare, J. A.; Gibbons, P.; Kersten, W. J. A.; Kulasegaram, S.; Moss, R. M.; Parisot, J. P.; Smith, B. J.; Street, I. P.; Yang, H.; Huang, D. C. S.; Watson, K. G. Structure-Guided Design of a Selective BCL-X(L) Inhibitor. *Nat. Chem. Biol.* **2013**, *9*, 390–397.
(36) Nissink, J. W. M.; Schmitt, S.; Blackburn, S.; Peters, S. Stratified High-Throughput Screening Sets Enable Flexible Screening Strategies From a Single Plated Collection. *J. Biomol. Screen.* **2014**, *19*, 369–378.
(37) Hodder, A. N.; Crewther, P. E.; Anders, R. F. Specificity of the Protective Antibody Response to Apical Membrane Antigen 1. *Infect. Immun.* **2001**, *69*, 3286–3294.
(38) Rich, R. L.; Errey, J.; Marshall, F.; Myszka, D. G. Biacore Analysis with Stabilized G-Protein-Coupled Receptors. *Anal. Biochem.* **2011**, *409*, 267–272.
(39) Giannetti, A. M. From Experimental Design to Validated Hits a Comprehensive Walk-Through of Fragment Lead Identification Using Surface Plasmon Resonance. *Meth. Enzymol.* **2011**, *493*, 169–218.
(40) Harris, K. S.; Casey, J. L.; Coley, A. M.; Masciantonio, R.; Sabo, J. K.; Keizer, D. W.; Lee, E. F.; McMahon, A.; Norton, R. S.; Anders, R. F.; Foley, M. Binding Hot Spot for Invasion Inhibitory Molecules on Plasmodium Falciparum Apical Membrane Antigen 1. *Infect. Immun.* **2005**, *73*, 6981–6989.

Table of Contents Graphic



Promiscuous 2-AminoThiazoles

Supplementary Information

Promiscuous 2-Aminothiazoles (PrATs): A Frequent Hitting Scaffold

Shane M. Devine,[A] Mark D. Mulcair,[A] Cael O. Debono,[A] Eleanor W. W. Leung,[A] J. Willem M. Nissink,[B] San Sui Lim,[A] Indu R. Chandrashekaran,[A] Jamie S. Simpson,[A] Jonathan B. Baell,[A] Peter J. Scammells,[A] Raymond S. Norton,[A] Martin J. Scanlon[A*]

[A] Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, VIC 3052, Australia

[B] Oncology iMed, AstraZeneca, Alderley Park, Nether Alderley, Macclesfield SK10 4TG, United Kingdom

* Corresponding author.[redacted]

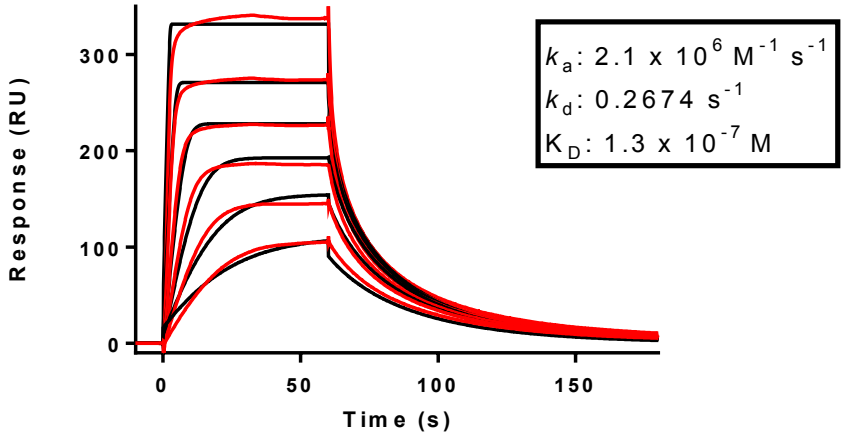
Contents

SPR Positive Controls

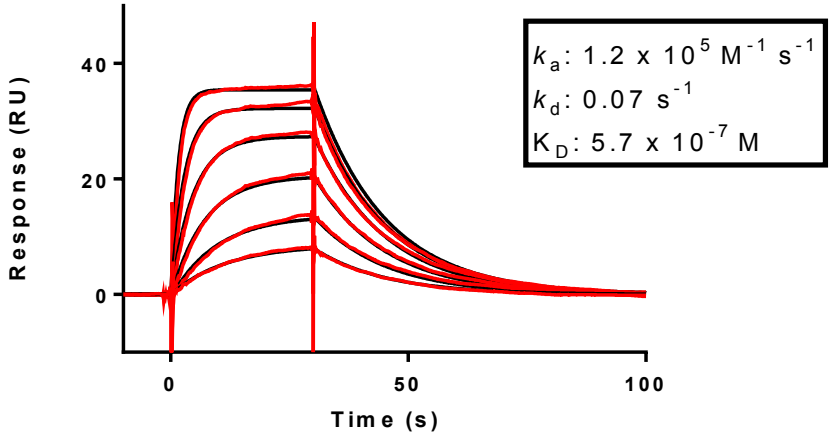
PROTEIN	CONTROL	EXPECTED K_D (μM)	K_D BY SPR (μM)
AMA1	R1 peptide	0.10	0.13
SPSB2	iNOS peptide	0.35	0.57
CAII	4-Carboxybenzyl Sulfonamide	1.0	0.89
Oxidoreductase 1	Fragment “LA010”	165	205
Kinase	Fragment "H4"	4	7.5

Figure S1. Binding of positive controls to target proteins by SPR. Control compounds were injected in a concentration series of 2-fold dilutions in running buffer, with a top concentration of 1 μM (R1), 5 μM (iNOS peptide), 10 μM (4-CBS), 20 μM (H4) or 200 μM (LA10)

A



B



C

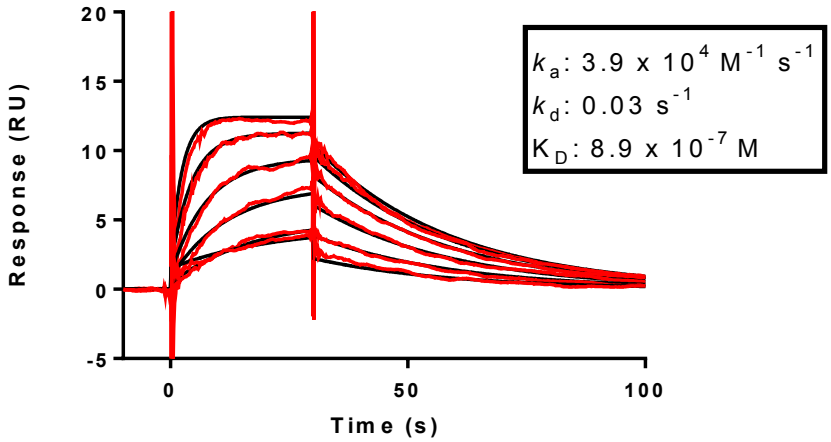


Figure S2. Raw data (red) and 1:1 kinetic fit from Biacore T200 Evaluation Software (black) for (A) R1 peptide binding to AMA1, (B) iNOS peptide binding to SPSB2, and (C) 4-CBS binding to CAII.

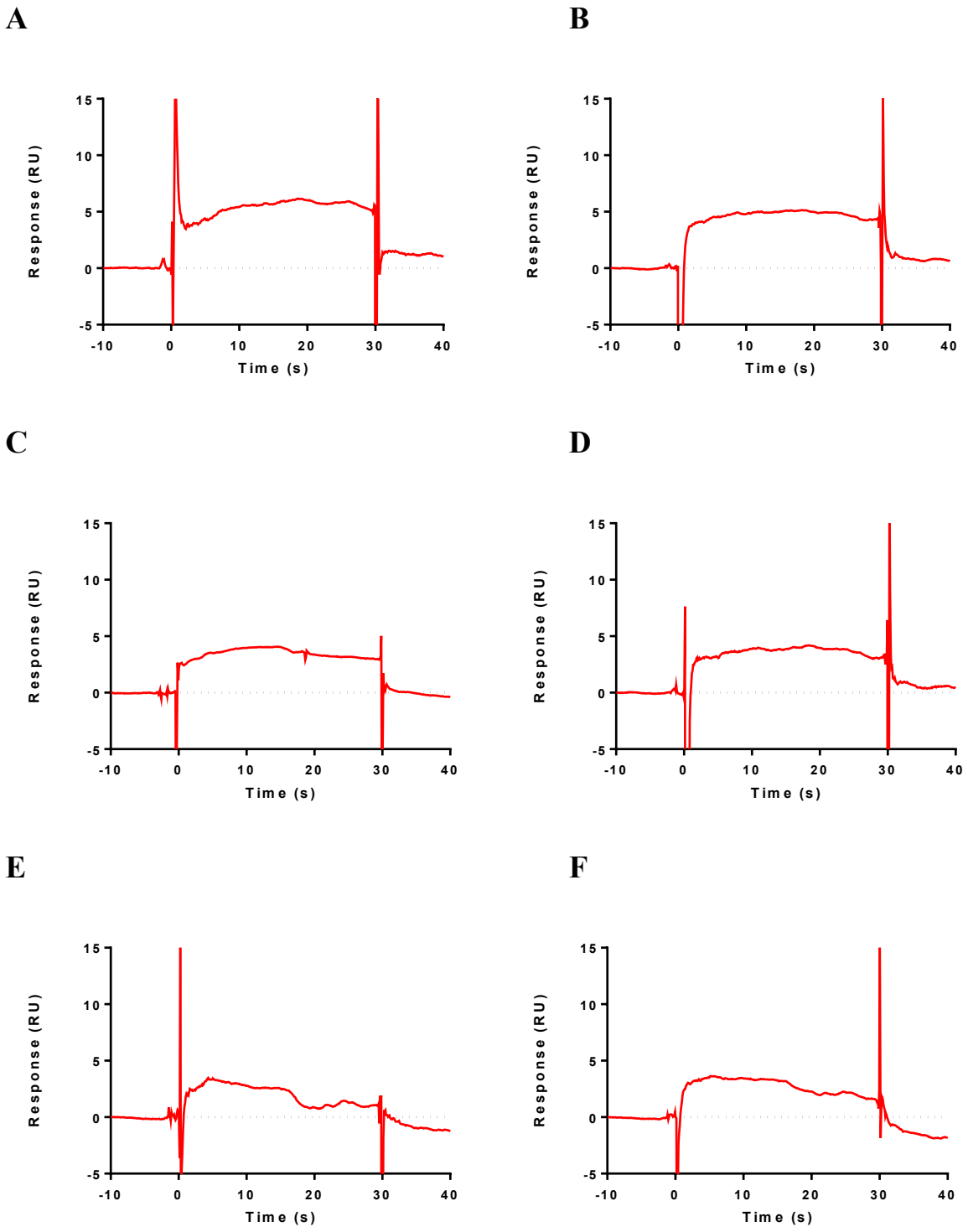


Figure S3. Binding of 2-AT **1** to immobilized AMA1 (A), SPSB2 (B), CAII (C), oxidoreductase 1 (D), kinase (E), and oxidoreductase 2 (F) by SPR. All compounds were screened at 200 µM.

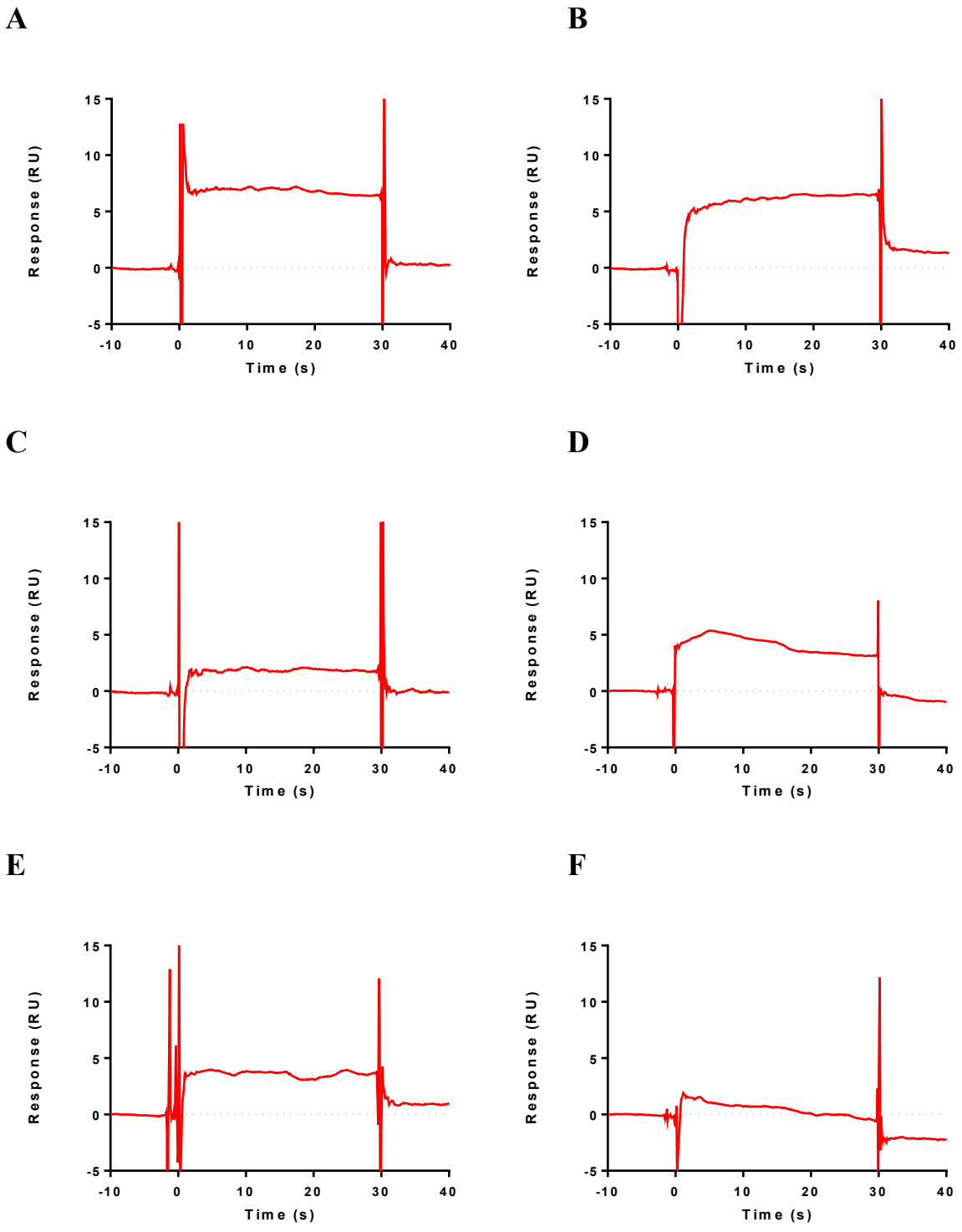


Figure S4. Binding of 2-AT **3** to immobilized AMA1 (A), SPSB2 (B), CAII (C), oxidoreductase 1 (D), kinase (E), and oxidoreductase 2 (F) by SPR. All compounds were screened at 200 µM.

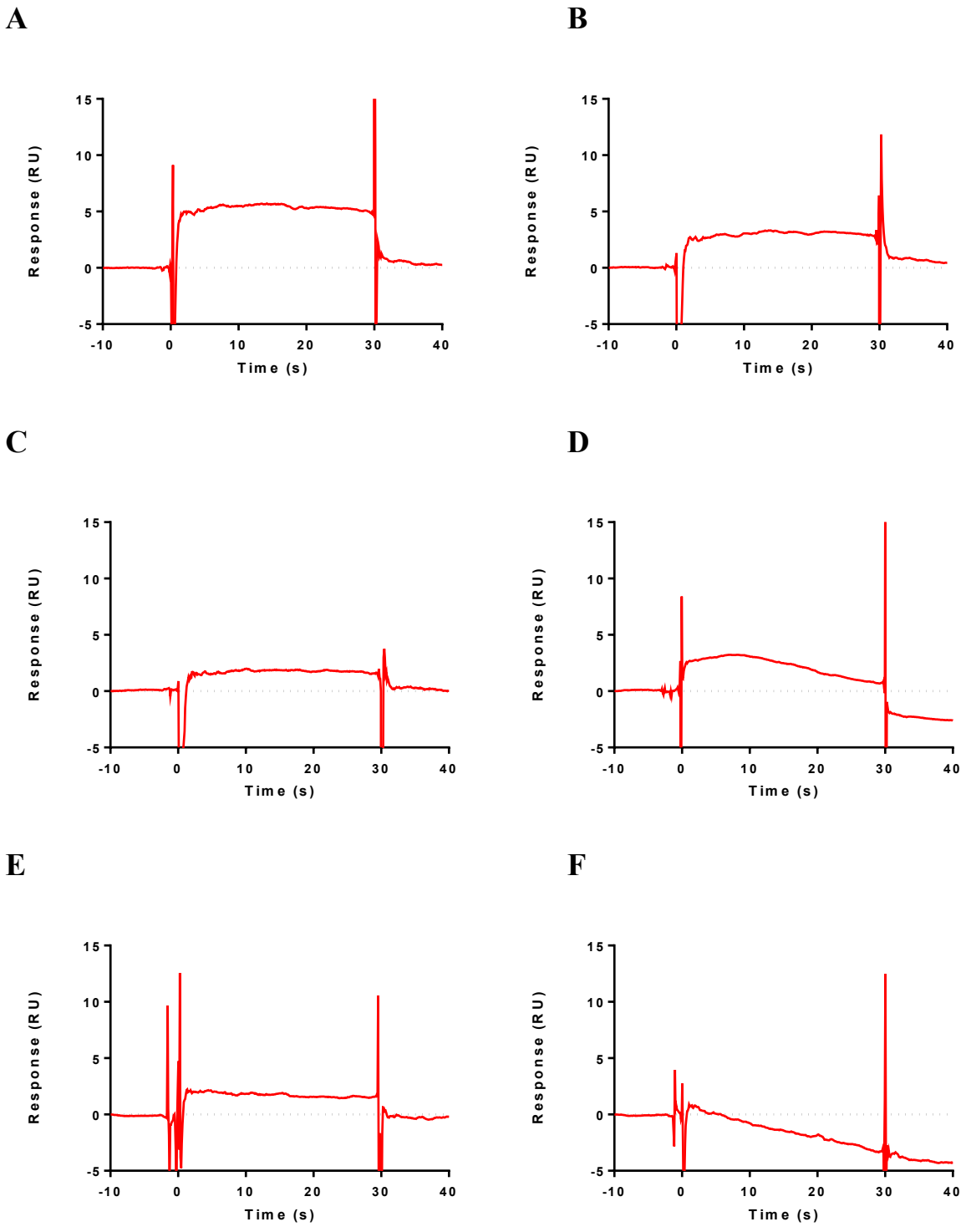


Figure S5. Binding of 2-AT **4** to immobilized AMA1 (A), SPSB2 (B), CAII (C), oxidoreductase 1 (D), kinase (E), and oxidoreductase 2 (F) by SPR. All compounds were screened at 200 µM.

^{19}F NMR Studies of 5-Trp-SPSB2

Construction of the 5-F-Trp-SPSB2 protein, and ^{19}F NMR determination of thiazole binding, were carried out as described previously.[1] ^{19}F NMR studies were carried out at 30 °C in 50 mM Phosphate, 50 mM NaCl (pH 7.4), with a 5-F-Trp-SPSB2 concentration of 100 µM, with or without 500 µM control peptide or 3 mM thiazole, as indicated.

2-ATs PAINS subclasses

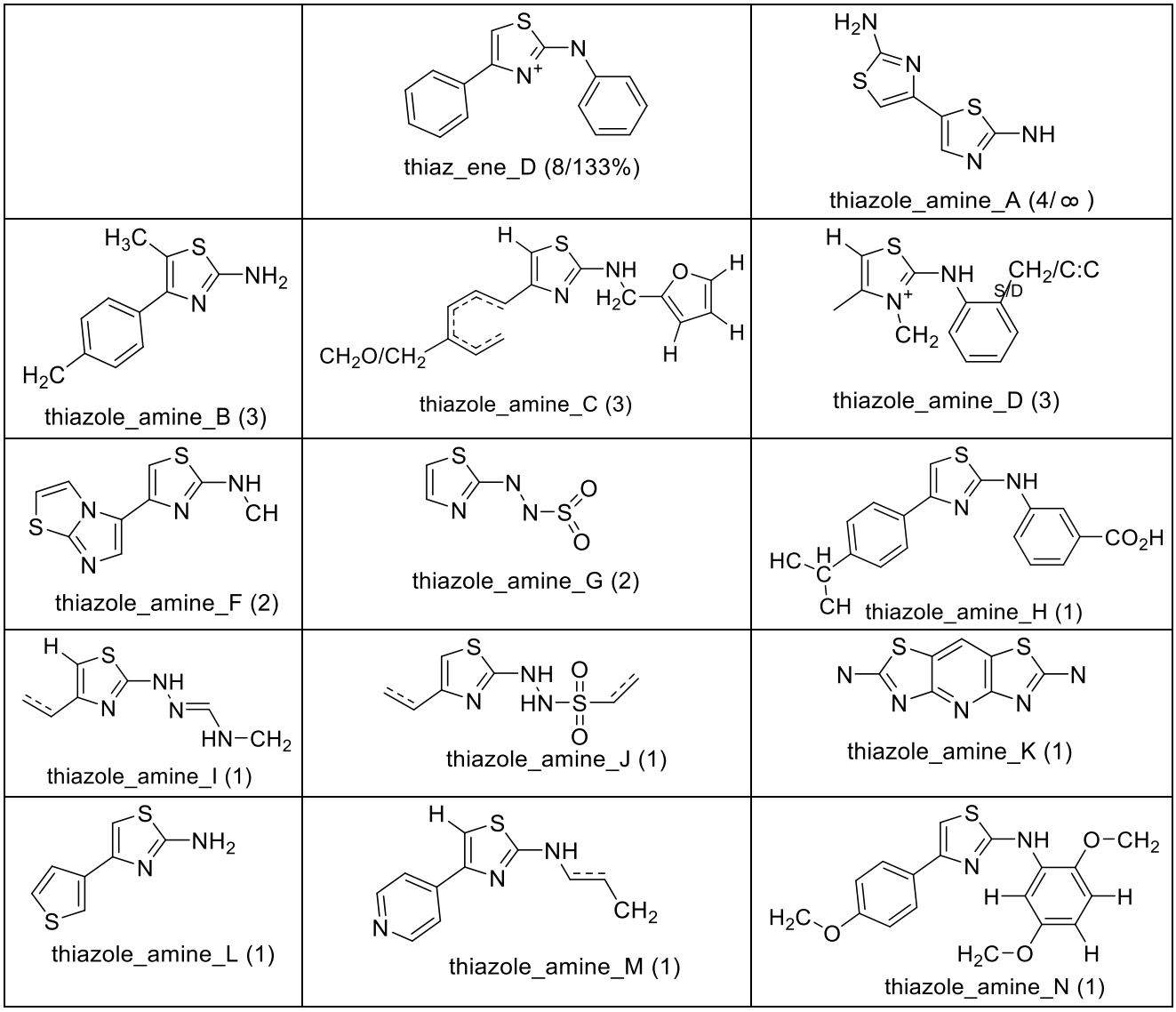


Figure S6. Fourteen PAINS 2-AT subclasses with number of analogues in brackets. For thiaz_ene D and thiazole_amine A, enrichment factors are also given that indicate relative promiscuity. For benign compounds this is between 10-20%. Here, it varies from a very high 133%, to infinity, the latter implying that none of the four analogues was clean. The integer in parentheses represents the number of analogues belonging to that PAINS substructure and the percentage where listed is the enrichment factor as previously defined.[2]

Frequent-hitter incidence

a) b) c)

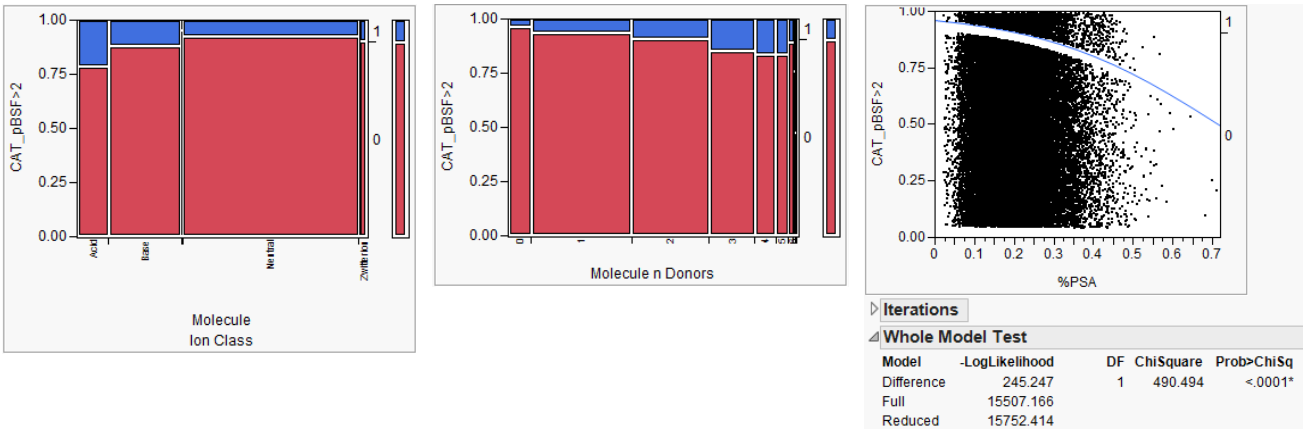


Figure S7. Trends of frequent-hitter incidence with a) ion class b) donor count c) %PSA

References

(1) Leung, E. W. W.; Yagi, H.; Harjani, J. R.; Mulcair, M. D.; Scanlon, M. J.; Baell, J. B.; Norton, R. S. (19) F NMR as a Probe of Ligand Interactions with the iNOS Binding Site of SPRY Domain-Containing SOCS Box Protein 2. *Chem. Biol. Drug Des.* **2014**, in press. doi: 10.1111/cbdd.12355.

(2) Baell, J. B.; Holloway, G. A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) From Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem.* **2010**, *53*, 2719–2740.