Dissertation

submitted to the

Combined Faculty of Natural Sciences and Mathematics of the Ruperto Carola University Heidelberg, Germany for the degree of Doctor of Natural Sciences

> Presented by M.Sc. Tanja Marzluf born in Offenburg Oral examination: 15th November 2018

The *P. vivax* merozoite surface protein 1 (MSP-1): Development of a recombinant protein production process and structural characterisation of MSP-1

Referees: Prof. Dr. Michael Lanzer Dr. Silvia Portugal

Summary

Malaria in humans is most often caused by either *Plasmodium falciparum* or *P. vivax*. Research efforts have mostly focused on *P. falciparum*, despite *P. vivax* being the most widespread *Plasmodium* species. This parasite species differs from *P. falciparum* in some key aspects like its ability to form dormant liver forms, an early appearance of gametocytes in the blood stream and its strict tropism to reticulocytes. Although *P. vivax* and *P. falciparum* invade different developmental stages of red blood cells, both express a highly abundant protein on the surface of their merozoites, the merozoite surface protein 1 (MSP-1). Initially expressed as a precursor protein, MSP-1 undergoes proteolytic processing and is cleaved into four processing fragments called p83, p30, p38 and p42. These fragments remain non-covalently associated and are attached to the surface of the merozoite via a GPI anchor at the C-terminal fragment p42. MSP-1 is considered a promising vaccine candidate and *P. falciparum* MSP-1D is currently tested as a vaccine in a clinical trial. However, no similar vaccine was developed for *P. vivax* yet.

The establishment of a protein production process for *P. vivax* MSP-1 (PvMSP-1) was an essential part of this study. Two different methods were developed, which both use individually expressed fusion proteins to assemble the full length MSP-1 protein. One fusion protein consists of the fragments p83 and p30, the other includes p38 and p42. In the first approach, the two halves are refolded together and purified by ion exchange chromatography and size exclusion chromatography. This protocol was also applied to produce *P. berghei* MSP-1 and test whether the assembly of a hybrid protein consisting of one *P. falciparum* half and one *P. vivax* half is possible. In the second approach contaminants are first removed from the individual fusion proteins by high resolution ion exchange chromatography before reconstitution of MSP-1. Both techniques result in a pure MSP-1 preparation, but the second approach results in a higher yield.

The produced recombinant proteins were structurally characterised using a combination of bioinformatical, biophysical and biochemical methods. All recombinant proteins were cleaved by *P. falciparum* SUB1, indicating a high conservatism of the cleavage motifs. CD spectroscopy showed a mostly alpha-helical structure and only partial unfolding of the protein at high temperatures. Interestingly, this unfolding increased when MSP-1D was processed by SUB1 before performing thermal denaturation. This suggests an increased stability of the MSP-1 heterodimer compared to the complex consisting of four processing fragments. Prediction of the secondary structure revealed unstructured regions in all analysed MSP-1 proteins. The function of these regions remains unknown, but these sites could be responsible for interactions with other proteins. Cross-linking of MSP-1 paired with mass spectrometry uncovered a previously unknown interaction between the p83 and p42 processing fragment, leading to an updated model of MSP-1 with a globular instead of asymmetrical shape.

This project provides two production processes for recombinant *P. vivax* MSP-1, which is now available for further immunological studies. The structural information acquired with this and other MSP-1 proteins can be used as a base to further elucidate its function and three-dimensional structure.

Zusammenfassung

Malaria wird im Menschen meistens entweder von *P. falciparum* oder *P. vivax* verursacht. Die Forschung konzentrierte sich bisher hauptsächlich auf *P. falciparum*, obwohl *P. vivax* die am meisten verbreiteste *Plasmodium* Spezies ist. Dieser Parasit unterscheidet sich von *P. falciparum* in einigen Schlüsselaspekten, wie die Fähigkeit Dauerformen in der Leber zu bilden, dem frühen Erscheinen von Gametozyten in der Blutbahn und dem strengen Tropismus zu Retikulozyten. Obwohl *P. falciparum* und *P. vivax* in rote Blutzellen unterschiedlicher Entwicklungsstadien eindringen, exprimieren beide ein sehr häufiges Protein auf der Oberfläche ihrer Merozoiten, das "merozoite surface protein 1" (MSP-1). MSP-1 wird als Vorläuferprotein exprimiert, welches proteolytisch prozessiert und in vier Fragmente namens p83, p30, p38 und p42 geschnitten wird. Diese Fragmente bleiben nicht-kovalent verbunden und sind mittels eines GPI Ankers am C-terminalen p42 Fragment mit der Oberfläche des Merozoiten assoziiert. MSP-1 wird als vielversprechender Impfstoffkandidat betrachtet und *P. falciparum* MSP-1D wird derzeit in einer klinischen Studie als Impfstoff getestet. Dennoch existierte bis jetzt keine entsprechende Präparation von *P. vivax* MSP-1.

Die Entwicklung eines Verfahrens zu Herstellung von *P. vivax* MSP-1 war der Hauptbestandteil dieser Arbeit. Zwei verschiedene Methoden wurden entwickelt, die beide darauf beruhen zwei separat exprimierte Fusionsproteine zur Zusammensetzung des Gesamtproteins zu nutzen. Ein Fusionsprotein entspricht den Fragmenten 83 und 30, das andere beinhaltet p38 und p42. Im ersten Ansatz werden die beiden Hälften zusammen rückgefaltet und mittels Ionenaustauschchromatographie und Größenauschlusschromatographie aufgereinigt. Dieses Protokoll wurde auch zur Herstellung von *P. berghei* MSP-1 und für den Versuch ein Hybridprotein aus einer *P. falciparum* Hälfte und einer *P. vivax* Hälfte zusammenzusetzen genutzt. Bei der zweiten Variante werden Kontaminationen von den beiden Fusionsproteinen vor der Rückfaltung durch hochauflösende Ionenaustauschchromatographie abgetrennt. Beide Verfahren resultieren in sauberen MSP-1 Präparationen, mit der zweiten Methode wird allerdings eine höhere Ausbeute erreicht.

Die produzierten rekombinanten Proteine wurden mit bioinformatischen, biophysikalischen und biochemischen Methoden strukturell charakterisiert. Alle rekombinanten Proteine wurden von *P. falciparum* SUB1 geschnitten, was eine hohe Konservierung der Schnittstellen suggeriert. CD Spektroskopie ergab eine größtenteils alpha-helikale Struktur und eine nur teilweise Entfaltung des Proteins bei hohen Temperaturen. Interessanterweise entfaltete sich MSP-1D stärker, wenn es vor der thermischen Denaturierung mit PfSUB1 prozessiert wurde. Diese Beobachtung deutet darauf hin, dass das MSP-1 Heterodimer eine größere Stabilität als der Komplex aus vier Prozessierungsfragmenten aufweist. Die Vorhersage der Sekundärstruktur offenbarte unstrukturierte Bereiche in allen analysierten MSP-1 Proteinen. Die Funktion dieser Bereiche bleibt unbekannt, könnte aber in der Interaktion mit anderen Proteinen liegen. Cross-linking und massenspektrometrische Analysen enthüllten eine vorher unbekannte Interaktion zwischen den Prozessierungsfragmenten p83 und p42, was zu einem aktualisierten MSP-1 Modell mit einer globulären anstelle einer asymmetrischen Form führt.

Im Rahmen dieser Arbeit wurden zwei Herstellungsverfahren für rekombinantes *P. vivax* MSP-1 entwickelt, welches nun für weitere immunologische Studien zur Verfügung steht. Die Strukturinformationen, die mit diesem und anderen MSP-1 Proteinen erarbeitet wurden, können als Grundlage zur weiteren Untersuchung der Funktion und dreidimensionalen Struktur dienen.

Acknowledgements

First of all I would like to express my gratitude to Prof. Hermann Bujard for supervising me during this thesis. Thank you for our motivating discussions and always considering my ideas and suggestions. I'm especially grateful for all collaboration opportunities you initiated, without which parts of this project would not have been possible.

I would like to thank Prof. Michael Lanzer for giving me the opportunity to work in the parasitology department, his helpful advice during my TAC meetings, proofreading this thesis and for being my first examiner.

Special thanks go to Dr. Silvia Portugal for being the second examiner in my PhD committee.

I want to thank Dr. Faith Osier for allowing me to continue my work in her laboratory and being part of my PhD committee.

Thank you Prof. Matthias Mayer for being one of my examiners, but most importantly for allowing me to use your CD spectrometer, your help with the data analysis and your advice regarding the structural characterization of MSP-1.

I highly appreciate the help of my collaboration partners without whom many experiments would not have been possible. Thanks go to Marcelo Urbano Ferreira for sending us sera from *P. vivax* infected people in Brazil and for giving me the opportunity to attend the Science of Eradication course in 2014.

I also need to thank Prof. Mike Blackman for providing the recombinant PfSUB1.

I would also like to thank Dr. Marcel Deponte for allowing me to use his ÄKTA system while our own was in need of repair.

I'm very grateful to Dr. Görlich and Jürgen Schünemann for welcoming me at the MPI in Göttingen and your guidance during my time there. I want to especially thank Jürgen, for his effort to continue with experiments, even when I had to return to Heidelberg.

Thank you Dr. Bernd Hessling for performing the mass spectrometry experiments with cross-linked MSP-1.

I would also like to thank Dr. Christian Epp for giving me the chance to work on this project and his guidance and support.

Thanks go to Dr. Ann-Kristin Müller for her helpful comments in my TAC meetings.

During the past years, I've been lucky to work with amazing colleagues: Kristin Fürle, Anja Jäschke, Dennis Obonjo, Irene Nkumama, Fauzia Musasia and Beate Kühn. Thank you for the great atmosphere in the lab and interesting talks during breaks. Special thanks go to Kristin who not only introduced me to the ÄKTA system ("Air is the enemy!") but also for her calm presence and helpfulness. The lab would not be the same without you!

Special thanks go to Britta Nyboer for not only keeping my blood sugar level up with her delicious cupcakes but also for proofreading my thesis. Thank you for introducing me to new hobbies, our trips within and outside of Germany and simply being a fantastic friend! You will not get rid of me anytime soon!

I'm particularly thankful to Miriam Griesheimer and Sandra Niebel for taking care of everything outside of the lab. Thank you for your patience and going the extra mile to make things easier for me and other students.

I would also like to thank the people from the Lanzer lab – both current and former members – for the entertaining lunch and coffee breaks and helpful suggestions and comments throughout the years.

I'm also thankful to Kirsten Heiss, Julia Sattler and Yvonne Maier for our dinners and Christmas market visits throughout the years.

I wish to thank Carrie Anderson for her assistance and for organising beach volleyball sessions, hikes etc. Thank you for making the lab more than just a workplace.

Thank you Elizabeth Woo, for your motivating speeches, advice and help whenever a software failed me ;-).

Last but not least, I'm deeply grateful to my parents for your love, encouragement and continuous support. Thanks for everything you have done for me.

Table of contents

Sum	SummaryI		
Zus	ZusammenfassungII		
Ack	now	ledgementsIII	
Tab	le of	contentsV	
List	of a	bbreviationsIX	
List	of fi	guresXI	
List	of ta	ables XIII	
1.	Int	roduction1	
1.1.		Malaria1	
1.2.		Life cycle of Plasmodium	
1.3.		Invasion of red blood cells5	
1.4.		Plasmodium vivax7	
1.5.		Clinical manifestation and immunity10	
1.6.		Vaccine development11	
1.7.		The Merozoite surface protein 114	
1	.7.1.	Proteolytic maturation of MSP-115	
1	.7.2.	Function of MSP-1 in invasion and egress16	
1	.7.3.	MSP-1 structure	
1	.7.4.	MSP-1 as a vaccine	
1.8.		Aim of this thesis21	
2.	Ma	aterials & Methods22	
2.1.		Materials	
2	.1.1.	Laboratory equipment	
2	.1.2.	Consumables	
2	.1.3.	Chemicals and reagents	
2	.1.4.	Buffers, media and solutions25	
	2.1	1.4.1. Antibiotic stock solutions	
	2.1	1.4.2. Molecular biology	
	2.1	1.4.3. Culture and transformation of <i>E. coli</i>	
	2.1	1.4.4. Microbiological methods25	
	2.1	1.4.5. Biochemical methods	
2	.1.5.	Commercial kits	
2	.1.6.	Enzymes	
2	.1.7.	Markers	
2	.1.8.	Antibiotics	
2	.1.9.	Antibodies	

2.1.10. Prepacked chromatography columns for Äkta [™] systems	31
2.1.11. Software & Online Tools	31
2.1.12. <i>E.coli</i> strains	32
2.1.13. Amino acid sequences of MSP-1	32
2.1.14. Primers	32
2.1.15. Expression plasmids	33
2.2. Methods	34
2.2.1. Molecular biological methods	34
2.2.1.1. Obtaining the synthetic genes	34
2.2.1.2. Isolation of plasmid DNA from <i>E. coli</i>	34
2.2.1.3. Agarose gel electrophoresis of DNA	35
2.2.1.4. Purification of DNA from agarose gels	35
2.2.1.5. Determination of nucleic acid concentration	35
2.2.1.6. Primer Design	35
2.2.1.7. Polymerase chain reaction (PCR)	36
2.2.1.8. Subcloning in TOPO vector	37
2.2.1.9. DNA restriction	37
2.2.1.10. Ligation	38
2.2.1.11. Sequencing	38
2.2.2. Microbiological methods	39
2.2.2.1. Preparation of chemically competent cells	39
2.2.2.2. Transformation of chemically competent cells	39
2.2.2.3. Preparation of glycerol stocks for cryopreservation	39
2.2.2.4. Testexpression	39
2.2.2.5. Expression of recombinant proteins	40
2.2.2.6. Cell rupture	40
2.2.3. Biochemical methods	40
2.2.3.1. Pulse renaturation	40
2.2.3.2. Purification of His-tagged proteins	41
2.2.3.3. Ion-exchange chromatography using Q-Sepharose	41
2.2.3.4. Size-exclusion chromatography using a Superdex 200 column	42
2.2.3.5. Purification of inclusion bodies using Mono S/Mono Q	42
2.2.3.6. Desalting using prepacked columns	42
2.2.3.7. Analytic size-exclusion chromatography	42
2.2.3.8. Protein quantification using Bradford Assay	43
2.2.3.9. Protein quantification using Biuret Assay	44
2.2.3.10. Dialysis	44
2.2.3.11. SDS polyacrylamide gel electrophoresis (SDS-PAGE)	44
2.2.3.12. Western Blot analysis	45

	2.2.3.13. Enzyme-linked immunosorbent assay (ELISA)	45
	2.2.3.14. PfSUB1 cleavage assay	45
	2.2.3.15. Analytical ultracentrifugation	
	2.2.3.16. Circular Dichroism (CD) spectroscopy	
	2.2.3.17. Cross-linking	47
3.	Results	
3.1.	Design of synthetic MSP-1 gene	
3.2.	Construction of pZE expression plasmids	
3.	2.1. Cloning of expression plasmids for His-tagged PvMSP-1 subunits	
3.	2.2. Expression plasmids for full length PvMSP-1	
3.	2.3. Design of <i>P. berghei msp-1</i> expression plasmids	52
3.3.	Expression of recombinant <i>P. vivax</i> and <i>P. berghei</i> MSP-1	53
3.4.	Purification of <i>P. vivax</i> MSP-1	55
3.	4.1. PvMSP-1 enrichment using ion-exchange and size exclusion chromatogra	phy58
3.	4.2. Optimised PvMSP-1 purification using Mono S/Mono Q columns	61
3.	4.3. Purification of <i>P. vivax</i> MSP-1 processing subunits	66
3.5.	Purification of <i>P. berghei</i> MSP-1	70
3.6.	Generation of a MSP-1 hybrid protein	73
3.7.	Structural characterisation of Plasmodium MSP-1	76
3.	7.1. Molecular weight determination of PvMSP-1	76
3.	7.2. Construction of phylogenetic tree and sequence identity	79
3.	7.3. Processing by subtisilin-like protease 1	81
3.	7.4. Secondary structure prediction	83
3.	7.5. Circular dichroism (CD) spectroscopy	84
3.	7.6. Crosslinking of MSP-1	
3.8.	Recombinant PvMSP-1 is recognised by naturally infected donors	91
4.	Discussion	93
4.1.	Production of recombinant <i>P. vivax</i> MSP-1	93
4.2.	Production of <i>P. berghei</i> MSP-1	99
4.3.	Creation of a hybrid protein consisting of P. falciparum and P. vivax MSF	² -1 halves 100
4.4.	Structural characterisation	
4.	4.1. Molecular weight determination	
4.	4.2. Bioinformatic analysis	
4.	4.3. Processing by SUB1	
4.	4.4. Secondary structure prediction and CD spectroscopy	
4.	4.5. Cross-linking	
4.5.	Outlook	
5.	References	

6.	Appendix
6.1.	Multiple sequence alignment
6.2.	CD spectra of recombinant MSP-1 proteins124

List of abbreviations

А	adenine
ACT	artemisinin-based combination therapy
Amp	ampicillin
APS	ammonium persulfate
BCIP	5-Bromo-4-chloro-3-indolyl phosphate
BS ³	bis(sulfosuccinimidyl)suberate
BSA	bovine serum albumin
С	cytosine
CD	circular dichroism
CFSE	carboxyfluorescein succinimidyl ester
CSP	circumsporozoite protein
D	diffusion constant
Da	dalton
DARC	Duffy antigen receptor for chemokines
DC	dendritic cell
DDT	dichlorodiphenyltrichloroethane
DSSO	disuccinimidyl sulfoxide
DTT	1,4-Dithiothreitol
EDTA	ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
EtOH	ethanol
f	frictional coefficient
G	guanine
G6PD	glucose-6-phosphate dehydrogenase
GM-CSF	granulocyte macrophage colony-stimulating factor
GMP	good manufacturing practice
GSH	reduced glutathione
GSSG	oxidised glutathione
Ig	Immunoglobulin
IL(1)	interleukin-1
IPTG	isopropyl β -D-1-thiogalactopyranoside
k	Boltzmann constant
Kan	kanamycin
kDa	kilodalton
L	litre
LB	lysogeny broth

mAU	milli-absorption units
min	minutes
mS	milli-siemens
MVA	modified vaccinia virus Ankara
MWCO	molecular weight cut-off
NaCl	sodium chloride
NaOH	sodium hydroxide
nm	nanometre
NTA	nitrilotriacetic acid
Р	Plasmodium
PCR	polymerase chain reaction
Pf	P. falciparum
PfEMP1	Plasmodium falciparum erythrocyte membrane protein 1
Pv	P. vivax
PV	parasitophorous vacuole
s	seconds
\$20,w	Svedberg coefficient
SDS	sodium dodecyl sulfate
SEC	size exclusion chromatography
SUB1	subtilisin-like protease-1
Т	tyrosine
T_{abs}	absolute temperature
TAE	Tris-acetate-EDTA
TCA	trichloroacetic acid
TEMED	tetramethylethylenediamine
TfR1	transferrin receptor 1
TNF	tumor necrosis factor
Tris	Tris(hydroxymethyl)-aminomethan
WHO	World Health Organisation
η	viscosity

List of figures

Figure 1	World Map highlighting countries and territories with indigenous malaria cases	1
Figure 2	Life cycle of <i>Plasmodium vivax</i>	3
Figure 3	Invasion of red blood cells by merozoites	5
Figure 4	Distribution of <i>P. vivax</i> worldwide.	7
Figure 5	Microscopic image of a hypnozoite	8
Figure 6	MSP-1 is the most abundant protein on the surface of merozoites	14
Figure 7	Interspecies conserved and conserved blocks within the MSP-1 sequence	14
Figure 8	Cleavage sites of PfSUB1	15
Figure 9	Processing of MSP-1 by SUB1 and SUB2 takes place at specific stages of the parasite's	
	life cycle	16
Figure 10	Structure of p19 and model of the MSP-1 complex with MSP-6, MSP-7, MSPDBL-1 and MSPDBL-2	18
Figure 11	Synthetic genes of MSP-1D and MSP-1F and their correlation to field isolates MAD20	
	and K1	20
Figure 12	Alignment of PfSUB1 cleavage sites within the MSP-1 sequence from various	
	Plasmodium strains	49
Figure 13	Map of a pZE13 plasmid used for expression of P. vivax MSP-1 processing fragments	50
Figure 14	Map of a pZE23 expression plasmid used for expression of the two halves of the	
	MSP-1 heterodimer	51
Figure 15	Alignment of the N-terminal amino acid sequence of the 83/30 half of P. falciparum	
	and P. vivax	51
Figure 16	Plasmid map of pZE13-HXPb83/30	52
Figure 17	IPTG-inducible expression of P. vivax processing fragments using pZE13 plasmids	53
Figure 18	The two halves of PvMSP-1 were expressed in W3110Z1 after addition of IPTG	54
Figure 19	Recombinant protein expression of the two halves of P. berghei MSP-1	54
Figure 20	Scheme of recombinant MSP-1 production including reconstitution of the heterodimer	
	and protein enrichment using ion exchange and size exclusion chromatography	56
Figure 21	Both halves of P. vivax MSP-1 accumulate in inclusion bodies	57
Figure 22	P. vivax MSP-1 purification using Q-sepharose	59
Figure 23	Size exclusion chromatography can separate PvMSP-1 from smaller contaminants	60
Figure 24	New approach for the production of <i>P. vivax</i> MSP-1	61
Figure 25	Preparation of inclusion bodies for purification with Mono Q or Mono S	
	columns	62
Figure 26	Purification of Pv38/42 inclusion bodies using a Mono S column	63
Figure 27	Purification of PfPv83/30 inclusion bodies using Mono Q columns	64
Figure 28	Test for PvMSP-1 heterodimer formation using size exclusion chromatography	65
Figure 29	Recombinant proteins representing the P. vivax processing fragments differ in their	
	solubility	66
Figure 30	Three out of four P. vivax MSP-1 processing fragments were successfully purified	
	using Ni-NTA	67
Figure 31	Purification of Pv83 using ion exchange chromatography	69
Figure 32	Additional size exclusion chromatography to purify Pv83	70
Figure 33	Cell rupture of cultures expressing Pb83/30 and Pb38/42	71

Figure 34	P. berghei MSP-1 purification using Q-sepharose	72
Figure 35	Gel filtration of P. berghei MSP-1 could remove most contaminants	73
Figure 36	Potential hybrid protein enrichment via ion exchange chromatography	74
Figure 37	Separation of excess halves and the potential hybrid protein using SEC	75
Figure 38	Analytical ultracentrifugation run of MSP-1D	76
Figure 39	Plot of sedimentation coefficient distribution against Svedberg coefficient for MSP-1D	77
Figure 40	Analytical ultracentrifugation of P. vivax MSP-1 revealed a less homogenous protein	
	preparation	78
Figure 41	Analytical size exclusion chromatography to determine the hydrodynamic radius of	
	MSP-1 by comparison with the High Molecular Weight calibration kit	78
Figure 42	Molecular weight calculation based on hydrodynamic radius and Svedberg coefficient	79
Figure 43	Evolutionary analysis based on MSP-1	80
Figure 44	Predicted SUB1 cleavage sites in different MSP-1 proteins	81
Figure 45	PfSUB1 assay with different MSP-1 proteins derived from P. vivax, P. knowlesi and	
	P. berghei	82
Figure 46	Scheme of MSP-1 with predicted unstructured regions	83
Figure 47	CD spectra of MSP-1D at different temperatures	85
Figure 48	Fitted data from thermal denaturation of MSP-1D monitored by CD spectroscopy	85
Figure 49	CD spectra of all measured MSP-1 proteins	86
Figure 50	CD spectra of MSP-1D after processing with PfSUB1	87
Figure 51	Structure of the cross-linkers BS ³ and DSSO	88
Figure 52	Interactions within <i>P. falciparum</i> MSP-1 proteins analysed by cross-linking with BS ³	89
Figure 53	Detected interactions using the MS-cleavable cross-linker DSSO	90
Figure 54	Cross-linking of presumable hybrid proteins created from two halves of different	
	Plasmodium strains or species	91
Figure 55	Recombinant PvMSP-1 and its processing fragments are recognised by naturally	
	acquired antibodies	92
Figure 56	Old and new model of MSP-1	107
Figure 57	Scheme of the ex vivo invasion assay using P. vivax	108
Figure 58	CD spectra and melting curve of P. vivax MSP-1	124
Figure 59	CD spectra and melting curve of MSP-1F	125
Figure 60	CD spectra and melting curve of P. knowlesi MSP-1	126
Figure 61	CD spectra and melting curve of a hybrid MSP-1 protein consisting of MSP-1D 38/42	
	and MSP-F 83/30	127
Figure 62	Melting curve of MSP-1D after proteolytic processing by PfSUB1	128

List of tables

Table 1	Overview of used antibiotics	30
Table 2	Antibodies used for Western Blots or ELISA	30
Table 3	List of prepacked chromatography columns	31
Table 4	E. coli strains used for cloning and protein expression	32
Table 5	Uniprot entries for used MSP-1 amino acid sequences	32
Table 6	Oligonucleotides used in this study	32
Table 7	Overview of expression plasmids for MSP-1 halves and processing fragments	33
Table 8	Components of a preparative PCR	36
Table 9	PCR protocol using Phusion polymerase	36
Table 10	Overview of used primer sets and the respective temperature used during PCR	36
Table 11	Setup for inserting a PCR product into the TOPO vector	37
Table 12	Restriction digest for cloning and control purposes	37
Table 13	Components of a ligation reaction	38
Table 14	Proteins included in the High Molecular Weight Calibration Kit and their elution volume	
	using the Superdex 200 Increase 10/300 GL	43
Table 15	Protease inhibitors used in the PfSUB1 cleavage assay	46
Table 16	Sequence identities of MSP-1 from several Plasmodium strains according to the Prote	ein
	Blast by NCBI	81
Table 17	Exact position of unstructured regions within MSP-1 according to the secondary	
	structure prediction	84
Table 18	Melting temperatures of MSP-1 determined by CD spectroscopy	86

1.1. Malaria

In the Western world, most people only think about malaria when planning a vacation in a foreign country. However, nearly half of the world's population is at risk of infection with this vector-borne disease. Current estimations state that there were 216 million cases of malaria and 445,000 deaths caused by the disease in 2016, making malaria one of the most common and most severe infectious diseases worldwide. Pregnant women, non-immune travellers and children under the age of 5 years are particularly susceptible to malaria with 70% of deaths occurring in this age group. The vast majority of infections occurred in Africa (90%) followed by South East Asia (7%) and the Eastern Mediterranean region (2%)(Figure 1) (World Health Organization, 2017).



Figure 1. Word Map highlighting countries and territories with indigenous malaria cases. Malaria infections most often occur on the African continent, followed by South America and South East Asia (red). Countries shown in blue had no reported malaria cases for the past 3 years. Tajikistan (yellow) had no cases in 2016. Green marks countries that were certified malaria free since 2000 and includes Krygystan and Sri Lanka, which received malaria free status in 2006. Source: (World Health Organization, 2017)

Malaria cases were first documented in the Early Greek period by Hippocrates, who described a disease characterised by intermittent fever, which could be classified into febris tertian (every third day) and frebris quartana (every fourth day) (Cox, 2002). But malaria has affected human populations even before that since ancient DNA for *Plasmodium falciparum*, a parasite causing malaria in humans, was identified in mummy tissues from ~4,000 years ago (Nerlich *et al.*, 2008). The name malaria, meaning "bad air" in medieval Italian, was invented by the Romans in the middle ages, based on the belief that the illness was caused by toxic fumes arising from swamps. It was only in 1880 that Charles Alphonse Laveran discovered *Plasmodium* parasites in the blood of an infected soldier (Bruce-Chwatt, 1981). Six years later, Camillo Golgi reported that the parasite replicates asexually in red blood cells and that lysis of these cells and the release of the parasites caused the fever in infected patients (Golgi, 1886). The link between

Plasmodium parasites and mosquitoes was not discovered until 1989, when Ronald Ross found the parasite in the stomach of an *Anopheles* mosquito (Ross, 1898). Proof that female Anopheles mosquitoes could transmit malaria was published two years later by Amico Bignami and Giovanni Grassi (Grassi, 1900).

Today, more than 150 *Plasmodium* species have been discovered, but only five have been shown to be pathogenic to humans: *Plasmodium malariae*, *P. ovale*, *P. vivax*, *P. falciparum* and *P. knowlesi* (Singh and Daneshvar, 2013). Almost all malaria infections (99%) in sub-Saharan Africa are caused by *Plasmodium falciparum*, while *P. vivax* is more prevalent in the Americas (64%), South East Asia (30%) and in the Eastern Mediterranean regions (40%) (World Health Organization, 2017).

Encouraged by the availability of insecticides, such as DDT, and antimalarial medications, the World Health Organisation (WHO) initiated the malaria eradication program in 1955 with the aim to eliminate malaria worldwide. Measures to fight the disease included indoor residual spraying, insecticide-treated mosquito nets, drainage of breeding sites and treatment of infections with chloroquine. Despite these efforts, permanent success could only be achieved in Europe and North America, leading to the abandonment of the program in 1969. Failure of the campaign resulted from the emergence of parasite resistance to antimalarial medicines and mosquito resistance to insecticides, poor infrastructure, political and cultural problems and lack of consistent funding.

Resistance of *P. falciparum* to antimalarial drugs still poses a serious problem. The artemisinin-based combination therapy (ACT) is the most powerful treatment to date, but parasites resistant to artemisinin have been found in five countries in South East Asia (World Health Organization, 2017). However, the WHO recently updated the information on artemisinin resistance to clarify that "resistant" parasites only display a partial resistance as the resistance mechanisms developed by the parasite affect only the ring stage and lead to a delayed clearance of parasites in patients after ACT treatment. The development of new antimalarial drugs can be beneficial, but an effective vaccine would be the most powerful tool in malaria control. Malaria research has made significant progress by establishing a culture system for *P. falciparum* (Trager and Jensen, 1976) and sequencing the genomes of *P. falciparum* (Gardner *et al.*, 2002), *P. vivax* (Carlton, Adams, *et al.*, 2008), *P. knowlesi* (Pain *et al.*, 2008) and the mosquito *Anopheles gambiae* (Holt *et al.*, 2002). Nevertheless, the most advanced vaccine candidate RTS,S/AS01 (GlaxoSmithKline Biologicals), a circumsporozoite protein (CSP)-based subunit vaccine, was only partially effective against severe malaria in clinical phase III trials (RTS, 2015). This result was particularly discouraging as the efficacy in the high risk group of children under the age of five was below 30%. Hence, alternative vaccine candidates need to be examined to achieve better protection against malaria.

1.2. Life cycle of Plasmodium

The *Plasmodium* parasite is an eukaryotic unicellular microorganism belonging to the phylum Apicomplexa. This obligate intracellular parasite can infect a wide range of vertebrate hosts, including birds, reptiles, non-human primates and humans. Its complex life cycle includes switching from an invertebrate to a vertebrate host, replicating asexually and sexually, passing through several stages in different cell types and tissues, and subsequently requires adaption to changing environmental conditions like temperature, pH and the host's metabolism. (Figure 2).



Figure 2. Life cycle of *Plasmodium vivax.* When an infected *Anopheles* mosquito transmits sporozoites to the human host, the parasites actively move through the skin until they penetrate a blood vessel and are getting carried by the bloodstream to the liver. During the first few days the sporozoite replicates asexually in hepatocytes and forms merozoite-filled merosomes. Disintegration of these structures releases merozoites, which in turn invade red blood cells. *P. vivax* parasites have a strict preference for reticulocytes, while *P. falciparum* invades erythrocytes. The erythrocytic stage can be separated into a ring, trophozoite and schizont stage. The end result is a schizont containing up to 32 daughter merozoites, which are released upon reticulocyte rupture and invade new reticulocytes. Due to a developmental switch from asexual to sexual proliferation in the red blood cell, immature gametocytes are formed that can be taken up by a mosquito during a blood meal. In the mosquito midgut, male and female gametocytes differentiate into microgametes and macrogametes, respectively. The microgametes fertilise macrogametes to form a zygote, which develops into a motile ookinete that passes through the midgut epithelium and settles as an oocyst in the outer gut lining. Within that oocyst, sporozoites are created by asexual replication and ultimately migrate to the salivary glands to restart the life cycle. Additionally, *P. vivax* has the ability to form dormant liver forms, so-called hypnozoites, which can begin another cycle of parasite development within the human host even after all parasites were cleared from the blood circulation. Source: (Mueller *et al.*, 2009).

Infection in humans is initiated by a bite of an infected female Anopheles mosquito. Approximately 15-123 sporozoites are injected during a blood meal and deposited into the skin of the vertebrate host (Rosenberg et al., 1990; Medica and Sinnis, 2005). The sporozoites actively migrate through the skin, but only around 35% of them reach a blood vessel that passively transports them through the body and into the liver (Douglas et al., 2015). The remaining sporozoites stay in the skin or end up in lymph vessels (Graewe et al., 2012). Once in the liver, sporozoites invade hepatocytes by invagination of the host cell plasma membrane, which forms a parasitophorous vacuole (PV) and creates a physical barrier between the parasite and the host (Lingelbach and Joiner, 1998). In 5-15 days, the parasite develops into a multinucleated schizont by asexual replication. One of these schizonts can contain up to 30,000 merozoites, meaning that one sporozoite reaching the liver and completing its development is enough to initiate a malaria infection. Individual vesicles filled with merozoites are created by repeated invaginations of the host plasma membrane. These vesicles, called merosomes, bud of the host cell and are released into the blood stream (Sturm et al., 2006). Rupture of these merosomes in the lung capillaries and the release of infectious merozoites ends the clinically silent phase and starts the erythrocytic stage (Baer et al., 2007). P. vivax also generates dormant liver forms, known as hypnozoites, which can trigger a new round of parasite development within the human host independently of new transmissions (Mueller et al., 2009). Once in the bloodstream, merozoites invade red blood cells. P. vivax parasites have a strict tropism for reticulocytes, whereas P. falciparum invades erythrocytes. Infected red blood cells can be identified by a typical ring-like structured parasite within the cell. This ring stage is followed by a trophozoite and schizont stage during which the parasite replicates asexually to form more merozoites. The exact duration of this intraerythrocytic development varies among the Plasmodium species. P. knowlesi is the fastest and completes one cycle in 24 hours while it takes between 42-48 hours for P. falciparum, 48 hours for P. vivax and P. ovale, and 72 hours for P. malariae (Singh et al., 2004; Greenwood et al., 2005; Tuteja, 2007). The end of each cycle is marked by the rupture of red blood cells and the release of 16-32 merozoites into the bloodstream where they quickly invade and exploit the next red blood cell. The clinical symptoms of malaria are exclusively caused by this asexual blood stage.

In addition to asexual proliferation, some parasites differentiate into gametocytes, the sexual parasite stages. These gametocytes enter the blood flow and can be taken up by an *Anopheles* mosquito during a blood meal. In the mosquito midgut, external signals such as the shift in pH and temperature and the presence of xanthurenic acid triggers the parasite's development (Billker *et al.*, 1998). The nucleus of male gametocytes divides into 4-8 nuclei and forms microgametes, which undergo exflagellation. These microgametes fertilise mature female gametocytes (macrogametes). The resulting zygote develops into a motile ookinete that traverses the midgut wall and transforms into a spherical oocyst between the midgut epithelium and the basal lamina (Sinden and Billingsley, 2001). Asexual multiplication within the oocyst forms sporozoites, which break through the oocyst wall and enter the haemolymph before migrating to the salivary glands. Once these sporozoites have reached the salivary glands, they can be transmitted to the next vertebrate host by a mosquito bite and begin the cycle anew.

1.3. Invasion of red blood cells

The asexual blood stage relies on the parasite's ability to penetrate red blood cells. These host cells offer both a rich nutrient source and a niche where the parasite is hidden from the immune system. *Plasmodium* merozoites are the invasive form of the blood stage. Their elongated and polarised form with characteristic organelles at the apical end directs the movement direction of the parasite (**Figure 3**). Merozoites are the smallest parasite stage in the life cycle of *Plasmodium* and can be between 1.5-2.5 µm long and 1.0-2.0 µm wide. These differences in size are species-specific. For example, *P. vivax* merozoites were observed to be larger than *P. falciparum* ones. The outer surface of merozoites is covered in numerous merozoite surface proteins, often leading to the description of a fuzzy coat. The apical organelles, such as micronemes and rhoptries, are characteristic for *Apicomplexan* organisms and contain parasite proteins that are involved in the invasion mechanism (Cowman and Crabb, 2006). Dense granules and exonemes are additional reservoirs of such proteins, like the subtilisin-like protease-1 (PfSUB1) (Yeoh *et al.*, 2007). In addition to the apical organelles, the merozoites also possess typical organelles found in eukaryotes, like a nucleus, mitochondrium, endoplasmic reticulum, Golgi, microtubules and ribosomes. The apicoplast is a non-hotosynthetic plastid that is crucial for the parasite's survival. However, the exact function is still not completely elucidated.



Figure 3. Invasion of red blood cells by merozoites. (A) Morphology of a merozoite. The structure is pear-shaped and its surface is heavily covered in proteins creating a fibrillar coat. Apart from the typical eukaryotic organelles, Apicomplexan parasites have specialised organelles such as rhoptries and micronemes and dense granules, as well as exonemes that contain parasite proteins involved in the invasion process. The polar ring at the apical end functions as a microtubule organising centre. (B) Invasion of red blood cells by merozoites includes three distinct steps. First, the merozoite attaches reversibly to a potential new host cell. Surface proteins on the merozoite allow the identification of permissive red blood cells. Next, the merozoite reorientates itself and the apical end binds to the red blood cell membrane. This binding is irreversible and a tight junction is formed which actively pushes the parasite into the host cell during the final step. The fuzzy surface protein coat is shed during the parasite's entry into the red blood cell. After invasion, the parasite is surrounded by the parasitophorous vacuole, which was created by invagination of the host cell membrane. Source: (Cowman and Crabb, 2006).

The invasion process can be seen as the parasite's Achilles' heel. Plasmodium is intracellular for most of its life cycle with the merozoites being the only extracellular form in the blood stage. Hence, the time between egress and invasion of another red blood cell, poses a brief chance for the immune system to attack. If the invasion process were to be disrupted by therapeutic measures or immune mechanisms induced by vaccination, the symptom-causing asexual multiplication of *Plasmodium* could come to a standstill. The invasion of red blood cells by merozoites was first studied using P. knowlesi (Dvorak et al., 1975) and previously made observations were later confirmed with P. falciparum (Gilson and Crabb, 2009). Generally, the invasion event can be divided into three phases. (i) Initial reversible attachment of the merozoite to the red blood cell membrane. In this phase, the merozoite must distinguish between host cells suitable for invasion and other cell types. These initial contacts were shown to cause heaving perturbations of the red blood cell membrane, giving these interactions a dynamic appearance (Gilson and Crabb, 2009). Specific recognition of permissive host cells is thought to be mediated by GPI-anchored surface proteins and their associated partners (Sanders et al., 2005; Cowman and Crabb, 2006). (ii) The next step is reorientation of the merozoite and irreversible attachment of the apical pole to the red blood cell. Micronemes and rhoptries discharge proteins involved in tight junction formation, such as reticulocyte-binding protein homologues and erythrocyte binding antigens (Rayner et al., 2001; Taylor, Grainger and Holder, 2002; Duraisingh et al., 2003; Gilberger et al., 2003). (iii) Finally, the parasite enters the red blood cell by invagination of the cell membrane, creating the parasitophorous vacuole. The extracellular protein coat is shed during this step as the tight junction moves from the apical to the posterior end using its actin-myosin motor, propelling the merozoite into the host cell (Keeley and Soldati, 2004; Cowman and Crabb, 2006).

Usually, merozoites recognize a new target cell within one minute after their release from a schizont and parasite entry is completed 30 seconds after initial contact (Gilson and Crabb, 2009). However, studies on the invasive half-life of merozoites showed that an immediate invasion is not required for the parasite's survival. The *in vitro* determined invasive half-life was 8 minutes at 37°C. 20% of merozoites were still able to invade erythrocytes after 10 minutes of incubation without red blood cells. At room temperature, this observed invasive half-life even increased to 20 minutes (Boyle, Wilson and Beeson, 2013). Interestingly, incubation at 40°C reduced the invasive capacity of merozoites, indicating that fever is indeed an effective response by the immune system (Boyle, Wilson, *et al.*, 2010). Loss of invasive potential could be due to spontaneous cleavage or shedding of surface proteins (Langreth, Nguyen-Dinh and Trager, 1978; Johnson *et al.*, 1981; Blackman *et al.*, 1991), premature release of proteins from rhoptries and micronemes (Johnson *et al.*, 1980) or loss of metabolic activity.

1.4. Plasmodium vivax

Most fatal malaria cases are caused by *Plasmodium falciparum*, but another *Plasmodium* species called *P. vivax* is actually more widespread and is responsible for more than 50% of malaria cases in non-African regions. South America and South East Asia are reporting the majority (66%) of *P. vivax* infections, but the parasite was also observed in some African and Eastern Mediterranean regions (Figure 4, (World Health Organization, 2017)). Current estimations state that 2.6 billion people are at risk of an infection with *P. vivax* and 80-300 million clinical cases per year are caused by this parasite (Arévalo-Herrera, Chitnis and Herrera, 2010).



Figure 4. Distribution of *P. vivax* worldwide. *Plasmodium vivax* is the most widespread *Plasmodium* species infecting humans. South America and South East Asia carry the majority of the *P. vivax* burden, but some African regions also report *P. vivax* infections. The data on this map represents cases reported in 2010. The colour code ranges from light blue for low endemicity to red for high endemicity. Source: The Malaria Atlas Project.

Although both *P. falciparum* and *P. vivax* can infect humans, there are some key differences between the two. *P. vivax* shows a strict tropism for CD71⁺ reticulocytes and does not invade mature erythrocytes (Russell *et al.*, 2011; Malleret *et al.*, 2014). This restriction also poses a numerical limit as reticulocytes usually only represent ~ 2.5% of all circulating red blood cells. Because of this restriction, the parasitemia observed during *P. vivax* infections is generally lower compared to *P. falciparum* infections. Sequestration is another pathological process that occurs both in *P. falciparum* and *P. vivax*, but the observed cytoadhesion levels of *P. vivax* were 10-fold lower in *in vitro* assays and the *in vivo* numbers might be even lower (Carvalho *et al.*, 2010). This finding implies that sequestration is a rather insignificant factor in *P. vivax* infections.

Another major characteristic of *P. vivax* is the ability to differentiate into hypnozoites (**Figure 5**). These dormant liver forms are undetectable by blood tests and can remain in the liver for months or even years before causing a relapse of the disease. This additional *P. vivax* reservoir within the human host can re-establish the erythrocytic stage of the parasite's life cycle without a second transmission of sporozoites.



Figure 5. Microscopic image of a hypnozoite. The hypnozoite (green dot on the right) is shown next to a liver schizont. This dormant form can remain in the liver for years before causing a relapse and is undetectable by commonly used blood tests. Source: (Vogel, 2013)

Prevention of these relapses can be achieved by radical cure treatment, which consists of co-administration of the antimalarial drugs chloroquine or artemisinin, targeting blood stages, and a hypnozoite-eliminating drug like primaquine. This compound is metabolised in the liver by an enzyme called cytochrome CYP2D6, suggesting that the actual hypnozoite killing agent is a metabolite (Baird and Hoffman, 2004). Hence, treating patients without a functional CYP2D6 metabolism could result in treatment failure (Bennett *et al.*, 2013). Also, people suffering from glucose-6-phosphate dehydrogenase (G6PD) deficiency cannot benefit from primaquine medication as this substance causes fatal haemolysis in these patients. Poor compliance is another factor that hinders successful treatment as primaquine needs to be taken for 14 days. Tafenoquine is an alternative compound that is currently under investigation for its potential use as a single-dose radical cure treatment alongside with chloroquine or artemisinin-based combination therapy. However, tafenoquine belongs to the same chemical family as primaquine and is equally contraindicated for treatment of G6PD-deficient patients and pregnant or lactating women (Watson *et al.*, 2018).

The life cycle of *P. vivax* seems to put a stronger emphasis on generating gametocytes than *P. falciparum*, as *P. vivax* gametocytes can already be detected in the blood of infected patients before they experience any symptoms. This quick occurrence of sexual stages, which can be taken up by mosquitoes to further transmit the disease and the fact that up to 71 mosquito species can carry *P. vivax* parasites might contribute to this parasite's wide distribution (Vogel, 2013).

Individuals who are permanently exposed to *P. vivax* develop milder clinical manifestations or even asymptomatic infections. This naturally acquired immunity progressively builds up during the first two decades of life (Doolan, Dobaño and Baird, 2009) but protection against *P. vivax* could also be achieved by immunising human volunteers with attenuated sporozoites (Clyde, 1975). Due to the apparently milder symptoms, *P. vivax* was used as malariotherapy between 1917 and 1940, which means patients suffering from certain diseases like neurosyphilis were infected with *P. vivax* to induce high fevers that were meant to combat the primary disease. In 1927, Julius Wagner-Jauregg received the Nobel prize for inventing this technique. However, this treatment has since been abandoned because it killed 15% of the treated patients (White, 2011). Today, *P. vivax* should no longer be considered harmless as reports of severe disease and even deaths accumulate (Baird, 2007; Anstey *et al.*, 2009; Mueller *et al.*, 2009; Lacerda *et al.*, 2012).

The central dogma of *P. vivax* infection used to be that the parasite relies on an interaction of the *P. vivax* Duffy binding protein and the Duffy antigen receptor for chemokines (DARC) on reticulocytes (Horuk *et al.*, 1993). Reduced levels of the Duffy antigen were associated with resistance to *P. vivax* infection (Zimmerman *et al.*, 1999). This resistance could have been the driving factor for the single nucleotide polymorphism in the DARC promoter, which results in the complete absence of the Duffy receptor in the homozygous state (Tournamille *et al.*, 1995). Following this theory, *P. vivax* may have been highly prevalent in west and central Africa and the parasite itself enforced the fixation of the Duffy negative allele in the human population (Carter and Mendis, 2002). However, there is a growing number of reported *P. vivax* infections in Duffy-negative individuals, suggesting that either an unidentified alternative pathway exists or that the parasite is currently in the process to evolve such a mechanism (Rosenberg, 2007; Ménard *et al.*, 2010; Zimmerman *et al.*, 2013; Russo *et al.*, 2017). It has been suggested that Duffy-negative individuals might function as a reservoir for *P. vivax* parasites in Duffy-negative regions, creating an environment in which parasites that are able to invade reticulocytes without the need for this antigen could be selected for (Ménard *et al.*, 2010).

As the Duffy antigen is shown to be less important than initially believed, another receptor-ligand interaction was demonstrated to be responsible for parasite's strict tropism for reticulocytes. The interaction between the *P. vivax* reticulocyte-binding protein 2b (PvRBP2b) and the transferrin receptor-1 (TfR1) also known as CD71, is thought to allow the exclusive selection of young reticulocytes. TfR1 is an iron transporter, which is found on the surface of reticulocytes but not on mature erythrocytes. A cryo-electron microscopy structure of the ternary complex of PvRBP2b bound to human TfR1 and transferrin was recently published by (Gruszczyk, Huang, *et al.*, 2018). The same study showed that PvRBP2b residues involved in complex formation are conserved, implying PvRBP2b as a new *P. vivax*-specific drug target.

This selectivity does not only concentrate *P. vivax* invasion on immature CD71⁺ reticulocytes, it also localises the parasite burden out of the peripheral blood as these host cells are generally restricted to the bone marrow (Malleret *et al.*, 2014). This extravascular invasion is most likely another contributing factor to lower parasitemia in the peripheral blood. TfR1-deficient erythroid cells were shown to be refractory to *P. vivax* invasion and anti-PvRBP2b monoclonal antibodies can block invasion in field isolates (Gruszczyk, Kanjee, *et al.*, 2018).

Once the *P. vivax* parasite has invaded a CD71⁺ reticulocyte, the host cell undergoes rapid remodelling and transforms from a rigid immature reticulocyte to a deformable CD71⁻ cell within 6 hours post-invasion. These remodelled cells have lost all reticular matter and several surface markers (Malleret *et al.*, 2014). The increase in deformability is particularly interesting as *P. falciparum* infected erythrocytes become more rigid, whereas *P. vivax* infection has the opposite effect on the host cell. Experiments using laminar shear flow and microfluid conditions suggested that the increased deformability represents an adaption to avoid splenic clearance of parasitised blood cells (Suwanarusk *et al.*, 2004; Handayani *et al.*, 2009). In contrast to erythrocytes, reticulocytes still own the translation machinery to express surface

proteins, such as human leukocyte antigen-1 (HLA-1), which presents antigens to CD8⁺ T cells to activate them. Recent studies suggest that *P. vivax* depletes the reticulocyte cell membrane of cholesterol and thereby increases the infected reticulocyte's susceptibility to granulysin, a substance released by cytotoxic CD8⁺ T cells that forms pores in the membrane and delivers granzyme B, triggering apoptosis in the infected cell (Junqueira *et al.*, 2018). These findings suggest that CD8⁺ T cells could contribute significantly to *P. vivax* parasite clearance and that T cell-mediated immunity should not be neglected during vaccine development targeting *P. vivax* infections.

All things considered, these recent developments strongly indicate that *P. vivax* cannot be neglected in malaria research any longer and the focus on drug discovery and vaccine development should not be restricted to *P. falciparum* alone anymore.

1.5. Clinical manifestation and immunity

The first symptoms of a malaria infection can be observed between 7 and 18 days after a mosquito bite, depending on the *Plasmodium* species. These symptoms are usually unspecific in the beginning and include tiredness, diarrhea, nausea, sweating, chills and dizziness. During the progression of the infection, patients experience more severe effects such as anaemia, respiratory distress, hypoglycaemia and in the worst cases also multi-organ failure and cerebral malaria (World Health Organization, 2017). These clinical signs of malaria are caused solely by the asexual erythrocytic replication of the *Plasmodium* parasite. Once the parasite has successfully invaded blood cells, three different mechanisms contribute to the disease pattern. (i) Rupture of infected erythrocytes leads to anaemia and impairs oxygen transport, consequently resulting in respiratory distress. (ii) Parasite metabolites, like hemozoin, and parasite antigens that are shed when the merozoites invade the next cell can stimulate the immune system to produce pro-inflammatory cytokines (Shio *et al.*, 2010). These cytokines include tumour necrosis factor alpha (TNF- α), interleukin-1 (IL-1) and interferon- γ (Clark, 1987), which increase body temperature and lead to the typical fever associated with the rupture of red blood cells. Interestingly, many merozoite surface antigens are anchored to the parasite membrane by a glycosylphosphatidylinositol (GPI) moiety that has been shown to induce TNF and IL-1 production in macrophages (Ramasamy, 1998). (iii) Infected erythrocytes are modified by the parasite and express parasite surface antigens, like the *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1), which confer binding to the endothelium of blood vessels and inner organs such as brain, liver, lung and kidneys. This process called sequestration can result in low parasite numbers in the peripheral blood and allows the parasite to evade clearance in the spleen (Howard and Gilladoga, 1989). Rosetting, the agglutination of infected and uninfected red blood cells can further increase these effects. Together, sequestration and rosetting can block the blood flow, cause oxygen deprivation, inflammation in the capillaries and in the worst case result in multi-organ failure. The most deadly course of the disease is termed cerebral malaria and is characterized by sequestration to the microvasculature of the brain. Initial symptoms of impaired vision, coordination and movement problems quickly progress to coma and

death after a short period of time. But even if a patient survives cerebral malaria, the neurological consequential damage will impact the quality of life permanently.

Malaria in pregnant women is also a significant complication. *Plasmodium* is mostly replicating in the placenta, which seems to be an immunologically naïve site (English and Newton, 2002). Sequestration in the placental capillaries can have a series of negative effects on the mother and baby, ranging from miscarriage, premature delivery, low birth weight, severe anaemia and increased mortality of the newborn and the mother (Tuteja, 2007).

However, the human body is not entirely defenceless against malaria infections. People living in malaria endemic areas can develop semi-immunity after multiple infections. This semi-immunity protects against severe consequences of recurrent malaria infections, but semi-immune individuals still function as a reservoir for the parasite (Ramasamy, 1998). Genetic polymorphisms can be advantageous in terms of malaria as well. Hemoglobinopathies, altered erythrocyte membrane proteins (such as Duffy negativity), altered red blood cell enzymes (G6PD) and modifications of immunity related genes (TNF- α) can confer resistance to *Plasmodium* infections or the accompanying symptoms (Williams, 2006).

1.6. Vaccine development

The total number of malaria cases has dropped worldwide but emerging resistances against insecticides and antimalarial drugs put this achievement at risk (World Health Organization, 2017). Hence, a safe and efficient vaccine remains a requirement for malaria eradication worldwide. The target groups for vaccine development include those who suffer most from the disease, namely infants, children and pregnant women, but also migrants who move to malaria-endemic regions, travelers and military personnel (Hollingdale and Sedegah, 2017). The US military has defined a need for at least 80% efficacy for a malaria vaccine and including not only *P. falciparum* but also *P. vivax* infections (Teneza-Mora, Lumsden and Villasante, 2015).

Anti-sporozoite vaccines aim to inhibit the invasion of hepatocytes by sporozoites and to reduce the initial parasite load. The most advanced vaccine candidate to date is RTS,S/AS01, which is a subunit vaccine consisting of part of the repeat region and the remaining C-terminus of the pre-erythrocytic circumsporozoite protein (CSP) of *P. falciparum* sporozoites fused to a viral envelope protein of the hepatitis B virus (Crompton, Pierce and Miller, 2010). However, the results of a phase III clinical trial were sobering as RTS,S/AS01 showed no efficacy over a 7 year period in African children (Olotu *et al.*, 2016). Another anti-sporozoite vaccine candidate is based on the thrombospondin-related adhesive protein (TRAP), but a phase II clinical study in adults revealed no efficacy against *P. falciparum* malaria (Mensah *et al.*, 2016).

Whole parasite vaccines are thought to mediate immunity by inducing cytotoxic T cell responses against liver stages. The used sporozoites are manipulated so that they cannot complete their development and can be safely administered to malaria-naïve individuals. There are three different modes of developmental arrest. Radiation attenuated sporozoites (RAS) were exposed to γ -radiation to induce multiple random double-strand breaks in their DNA, leading to an early arrest in liver stage development (Hoffman *et al.*, 2002). While sterile immunity was achieved by intravenous injection in humans, a high dose was required (Lyke *et al.*, 2017). Interestingly, using *P. vivax* irradiated sporozoites for vaccination was also able to induce sterile protection in study participants (Arévalo-Herrera *et al.*, 2016), suggesting that this technique is reproducible for other *Plasmodium* species.

Another approach uses gene deletion to create genetically attenuated parasites (GAP), which arrest at a defined developmental stage in the liver(Mueller *et al.*, 2005). Immunisation experiments in mice showed an improved protection in comparison to radiation attenuated sporozoites (Butler *et al.*, 2011). Alternatively, wildtype sporozoites can be injected under simultaneous chemoprophylaxis using chloroquine azithromycin or primaquine to arrest the parasite's development at a defined stage (Putrianti *et al.*, 2009; Friesen *et al.*, 2010; Mordmüller *et al.*, 2017). This method was able to induce sterile protection in humans when sporozoites were delivered by mosquito bite (Roestenberg *et al.*, 2009, 2011) or by intravenous injection of cryopreserved *P. falciparum* sporozoites (Mordmüller *et al.*, 2017).

As promising as these results might seem, a whole parasite vaccine faces three problems. (i) The laborious production and purification needs to be improved by establishing an *in vitro* culture system to produce sporozoites of sufficient quantity and quality. (ii) Intravenous injection is not an accepted mode of vaccine administration, but intramuscular or subcutaneous injections require about 23 times more sporozoites (Gómez-Pérez *et al.*, 2015). (iii) Parasites need to be stored in liquid nitrogen or dry ice, which is a logistical challenge in remote African regions, hence this kind of vaccine is probably not applicable in the area where it is most needed.

Transmission blocking vaccines follow the altruistic principle and will not be able to protect the individual but aim to reduce transmission within a population by inducing antibodies against the sexual stages of *Plasmodium* (Wu *et al.*, 2015). These antibodies are taken up by the mosquito during a blood meal and inhibit parasite development in the mosquito midgut. This vaccine type alone will most likely not be sufficient to eliminate malaria and should be combined with other vaccines targeting erythrocytic or pre-erythrocytic stages.

Blood stage vaccines are intended to reduce parasitemia or prevent clinical disease by targeting infected red blood cells and merozoites. Most vaccine candidates rely on the humoral immune response, but an optimal vaccine should also be able to elicit T cell-mediated immunity to induce multistage protection. These antigens are commonly produced as a recombinant protein and formulated with a suitable adjuvant or inserted into a viral vector for optimal immunogenicity. One vaccine candidate found on infected erythrocytes is PfEMP1. This parasitic protein enables infected red blood cells to bind to endothelial cells and is a key mediator in cytoadherence and sequestration of *P. falciparum* parasites. However, vaccine

development based on PfEMP1 is difficult as this protein is encoded by 60 var genes and consequently is extremely diverse. Other blood stage vaccine candidates focus on merozoite proteins, such as the apical membrane antigen 1 (AMA-1), erythrocyte-binding antigen 175 (EBA-175), serine repeat antigen 5 (SERA-5) or glutamate-rich protein (GLURP). For years, AMA-1 has been an attractive target due to its involvement in erythrocyte invasion, but despite a promising immunogenicity profile, it showed no efficacy in a phase IIb study (Ouattara *et al.*, 2010) and even after reformulation with another adjuvant, higher antibody titres were unable to confer protection against sporozoite challenge (Duncan *et al.*, 2011). Antibodies against EBA-175, another protein on the merozoite surface that interacts with glycophorin A on the red blood cell, caused only modest growth inhibition in *in vitro* assays and immunisation with GLURP, a protein of unknown function, resulted in no protective efficacy in a phase IIb study (Sirima *et al.*, 2016). SERA-5 has been proposed to function as a protease in merozoite egress (Beeson *et al.*, 2016) and could reduce the incidence rate of high parasitemia in volunteers in a phase Ib study (Palacpac *et al.*, 2013). The most abundant protein on merozoites, the merozoite surface protein 1 (MSP-1) is a well-studied vaccine candidate that will be discussed in more detail in 1.7.

Despite all the progress being made in recent years, a malaria vaccine that is cost-effective, targets different stages of the *Plasmodium* life cycle and is able to confer long-lived protection against more than one *Plasmodium* species is still missing.

1.7. The Merozoite surface protein 1

Merozoites are the invasive form of *Plasmodium* blood stages, hence they require surface proteins that facilitate the identification and interaction with new host cells before and during invasion. The surface of merozoites is often compared to a fuzzy layer, with one protein in particular accounting for the majority of these surface proteins, leading to the name merozoite surface protein 1 (MSP-1)(Figure 6). MSP-1 was detected in all *Plasmodium* species so far and albeit it being considered a blood stage antigen, MSP-1 was also detected in late trophozoites and schizonts developing in hepatocytes (Szarfman *et al.*, 1988; Haussig, Matuschewski and Kooij, 2011).



Figure 6. MSP-1 is the most abundant protein on the surface of merozoites. Right: Photograph of a *Plasmodium* blood culture stained with Giemsa with mixed stages including a just ruptured schizont and free merozoites. Middle: A merozoite in the process of penetrating a new host cell. M: merozoite, E: erythrocyte, V: vacuole. Source: (Miller *et al.*, 1979). Right: Model of the non-covalently associated MSP-1 complex on the surface of a merozoite.

The primary sequence of *P. falciparum* MSP-1 can be divided into 17 blocks of low, moderate and high conservation among different isolates (Tanabe *et al.*, 2007). Despite this variability, all *P. falciparum msp-1* sequences can be classified into one of two allelic isoforms, either the MAD20 or K1 allelic variants, which are field isolates from Papa New Guinea and Thailand, respectively. Cross-species comparisons also revealed areas within the *msp-1* sequence that are highly conserved among species (Figure 7). Ten blocks with >48% sequence identity among *P. falciparum*, *P. vivax* and *P. yoelii* were identified, plus three blocks of >50% sequence identity when only *P. falciparum* and *P. vivax* were considered.



Figure 7. Interspecies conserved and conserved blocks within the MSP-1 sequence. 10 interspecies conserved blocks with a sequence identity of >48% were identified in *P. vivax*, *P. falciparum* and *P. yoelii* MSP-1. The conserved blocks share a sequence identity of >50%, referring only to *P. vivax* and *P. falciparum* sequences. Source: (del Portillo *et al.*, 1991).

1.7.1. Proteolytic maturation of MSP-1

In all *Plasmodium* species, MSP-1 is encoded by a gene of approximately 5,000 bp on chromosome 9 and is initially expressed as a precursor of 185-200 kDa. A hydrophobic N-terminal signal sequence of 19 amino acids is cleaved from the protein after translocation into the endoplasmic reticulum. The full length protein is not glycosylated except for the addition of a glycosylphosphatidylinositol moiety at the C-terminus (Berhe et al., 2000). However, the precursor protein is proteolytically processed by a protease called subtilisin-like protease 1 (SUB1), which is localised in exonemes and secreted into the PV prior to egress (Blackman et al., 1998). As a result, MSP-1 is cleaved into the four fragments p83, p30, p38 and p42, which are named according to their approximate molecular weight. These fragments remain non-covalently associated to each other and the merozoite surface. Cleavage by SUB1 is not restricted to MSP-1 as several other targets were identified via a bioinformatic screen, including the serine-rich antigens SERA 4, 5, 6, MSP-6, MSP-7 and the rhoptry associated protein 1 (RAP-1) (Yeoh et al., 2007; de Monerri et al., 2011). The cleavage of P. falciparum MSP-1 into four processing fragments occurs in a distinct but variant-specific order. In the MAD20 variant cleavage starts at the 30/38 site, followed by 83/30 and finally the 38/42 site, whereas processing of the K1 variant simply begins at the N-terminus with the 83/30 site and follows the order of cleavage sites as they occur in the primary sequence up to the C-terminus (Child et al., 2010). SUB1 processing was shown to induce structural changes in the conformation of its substrates and thus changes their binding properties, which prepares the merozoites for invasion and egress (Kauth et al., 2006; Koussis et al., 2009). This proteolytic processing seems to be of vital importance for the function of MSP-1 as several alternative cleavage sites for PfSUB1 have been identified (Cooper and Bujard, 1992). The cleavage motif of SUB1 always contains an aliphatic residue at the P4 position, a small uncharged amino acid at P2, a polar residue at P1 and an acidic residue at one or more of the P1'-P5' positions (de Monerri et al., 2011; Withers-Martinez et al., 2012) (Figure 8).



Figure 8. Cleavage sites of PfSUB1. (A) The subtilisin-like protease 1 recognises its cleavage sites by an aliphatic residue at P4, an uncharged amino acid at P2 followed by a polar residue at P1. The P1' to P5' positions contain one or more acidic residue. Source: (de Monerri *et al.*, 2011) (B) Scheme of the primary sequence of MSP-1D, corresponding to the MAD20 isolate as previously published by (Kauth *et al.*, 2006). Cleavage sites for PfSUB1 and PfSUB2 are indicated with downward arrows. Conserved, dimorphic and oligomorphic regions are shown in white, grey and black, respectively. SP: signal peptide, GPI: glycosylphosphatidylinositol signal sequence.

MSP-1 undergoes a second proteolytic processing step just before or during invasion. Another subtilisin-like protease called SUB2 is discharged from micronemes onto the merozoite surface and cleaves the C-terminal p42 subunit into two fragments called p33 and p19 (Harris *et al.*, 2005). While p19 remains tethered to the merozoite membrane via the GPI anchor, the remaining MSP-1 subunits are shed from the surface of the merozoite during its entry into the host cell (Blackman *et al.*, 1991; Blackman and Holder, 1992) (Figure 9).



Figure 9. Processing of MSP-1 by SUB1 and SUB2 takes place at specific stages of the parasite's life cycle. Initially, MSP-1 is expressed as a precursor protein that is connected to the merozoite membrane via a GPI anchor. Before egress from the host cell, SUB1 is released from exonemes and cleaves the precursor protein into the four fragments p83, p30, p38 and p42. A second proteolytic processing occurs during invasion into a new host cell. SUB2, a protease released from micronemes cleaves p42 into p33 and p19. The C-terminal p19 portion is internalised into the red blood cell with the merozoite while the remaining MSP-1 subunits are shed from the merozoite surface.

1.7.2. Function of MSP-1 in invasion and egress

MSP-1 is one of the most studied *Plasmodium* proteins. However, its precise function is still under investigation while several studies indicate its involvement in both invasion and egress. That MSP-1 plays a crucial role in the parasite's development became apparent when attempts to genetically knockout the *msp-1* gene failed. Only the EGF-like domains of the C-terminal part could be functionally replaced by the same sequence from another *P. falciparum* isolate (Drew *et al.*, 2004) or by the homologue sequence from the rodent parasite *P. chabaudi* (O'Donnell *et al.*, 2000).

Current studies suggest that MSP-1 could function as an anchoring platform for other parasite proteins during initial attachment to the red blood cell (Lin *et al.*, 2016). Previous experiments using recombinant MSP-1, MSP-6 and MSP-7 proteins showed that these three proteins form a complex after being processed by SUB1 (Kauth *et al.*, 2006). Other reported interaction partners are band 3, an anion transporter in the erythrocyte membrane that interacted with the p42 and p19 portion of MSP-1, heparin-like proteoglycans, glycophorin A, spectrin and recently also MSPDBL-1 and MSPDBL-2 (Herrera *et al.*, 1993; Goel *et al.*, 2003; Boyle, Richards, *et al.*, 2010; Baldwin *et al.*, 2015). The latter are members of the MSP-3 family and harbour Duffy-binding like domains, which are thought to mediate the interaction with red blood cells (Lin *et al.*, 2014). However, the exact receptor on red blood cells used for this

interaction is unidentified. MSPDBL-1, MSPDBL-2, MSP-3, MSP-6 and MSP-7 were all shown to bind to MSP-1 independently of each other, allowing the formation of different MSP-1 dependent complexes with presumably overlapping functions. Interestingly, only MSP-1 complexes containing MSP-6, MSPDBL-1 and MSPDBL-2 were able to attach to human erythrocytes (Lin *et al.*, 2016). The same study pointed out that p83 seems to play a key role in the assembly of these complexes as their formation could be prevented by antibodies directed against that subunit. This antibody-mediated inhibition also reduced parasite growth *in vitro*. Another experiment using monoclonal antibodies directed against the EGF-like domains of the p19 MSP-1 fragment showed that invasion of merozoites could be inhibited if the secondary processing of MSP-1 was blocked (Blackman *et al.*, 1990, 1994).

At the same time, evidence is accumulating that MSP-1 is not only crucial for invasion but also for egress. When genetic knockouts of *msp-1* failed, conditional mutagenesis experiments were used as an alternative approach to study whether the parasite could develop normally in absence of MSP-1. The lack of MSP-1 caused an abnormal development in the liver stage, where merozoite formation was noticeably disturbed and the infected hepatocytes showed internal budding and misshaped daughter cells (Combe *et al.*, 2009). Additionally, MSP-1 must not only be present, it also has to be processed by SUB1 to fulfil its function as inhibition of SUB1 discharge or the expression of MSP-1 mutants that are refractory to SUB1 processing causes a severe egress defect (Yeoh *et al.*, 2007; Das *et al.*, 2015). In contrast to MSP-1's role as an anchoring platform during invasion, MSP-1 itself is thought to be interacting with the cytoskeleton of the host cell before egress. Proteolytic processing by SUB1 enables MSP-1 to bind to heparin and spectrin. Following the rupture of the parasitophorous vacuole membrane, intracellular merozoites can bind to spectrin in the plasma membrane of the host cell. This interaction could cause internal shear forces that disrupt the structural integrity of the cytoskeleton and finally, lead to the rupture of the red blood cell membrane and release of merozoites (Das *et al.*, 2015).

1.7.3. MSP-1 structure

Structural information can often complement data from functional assays to elucidate the properties of a protein and its potential as a drug target or vaccine candidate. So far, little is known about the three-dimensional structure of MSP-1. Due to its size and high flexibility of the N-terminal part, all attempts to obtain a crystal structure have failed so far. Only the structure of the C-terminal p19 fragment could be solved by NMR spectroscopy for *P. falciparum* and *P. vivax* (Morgan *et al.*, 1999; Babon *et al.*, 2007), and by x-ray analysis for *P. falciparum*, *P. cynomolgi* and *P. knowlesi* (Chitarra *et al.*, 1999; Garman *et al.*, 2003; Pizarro *et al.*, 2003). All these studies reveal two EGF-like domains which contain three disulphide bonds each, indicating structural and functional conservation of the p19 fragment among *Plasmodium* species. The entire p19 fragment is compact and flat, and both EGF-domains are folded back on one another in a side-by-side arrangement. Attempts to generate individual point mutations in this MSP-1 subunit proved to be difficult and but the EGF-like domains could be functionally replaced with the corresponding domains from *P. chabaudi* MSP-1 and *P. berghei* MSP-8 (O'Donnell *et al.*, 2000; Drew *et al.*, 2004).



Figure 10. Structure of p19 and model of the MSP-1 complex with MSP-6, MSP-7, MSPDBL-1 and MSPDBL-2. (A) Structure of the *P. knowlesi* p19 fragment solved by x-ray crystallography. The EGF-domains (D1, D2) are arranged side by side. Beta-strands are shown as ribbons and numbered in each domain. Disulphide bonds are indicated in yellow. Source: (Garman *et al.*, 2003). (B) Model of MSP-1 in its PfSUB-processed form and known interaction partners. MSP-6 forms a tetramer after processing and binds to the p38 fragment. MSP-7 also requires processing by SUB1 to be able to bind to p83, p30 and p38 (Kauth *et al.*, 2006). MSPDBL-1 and MSPDBL-2 form an oligomer and interact with p83, p38 and p42, but not with p30 (Lin *et al.*, 2016).

Intramolecular interactions were analysed using recombinant MSP-1 processing fragments that were incubated together followed by affinity chromatography using an affinity resin that binds to only one of the MSP-1 fragments. In these experiments, p30 was shown to interact with p83, p38 and p42. The C-terminal p42 and p38 proteins also interacted with each other, whereas p83 had no second interaction partner within MSP-1 (Kauth *et al.*, 2003). Interactions with other proteins such as MSP-6 and MSP-7 were demonstrated as well, but only the processed form of MSP-6, namely MSP-6₃₆, could bind to MSP-1. MSP-6₃₆ tetramerises and binds solely to p38, while MSP-7 interacts with p38, p30, and p83 (Kauth *et al.*, 2006). MSPDBL-1 and MSPDBL-2 have also been identified as interaction partners of MSP-1 and were shown to bind to p42, p38 and p83 (Lin *et al.*, 2014). Another study focused on the C-terminal half of *P. falciparum* MSP-1 only by using a fusion protein comprising p38 and p42 to identify interaction partners. MSP-3, MSP-6, MSP-7 and MSP-9 were all found to interact with this fusion protein, but not with the p19 fragment, suggesting that only sequences upstream of the p19 C-terminal region are involved in molecular interactions with other MSP proteins (Paul *et al.*, 2018).

1.7.4. MSP-1 as a vaccine

MSP-1 is a promising vaccine candidate because of its abundance on the merozoite surface and crucial role in the parasite's life cycle. Several epidemiological studies in malaria endemic regions demonstrated a positive correlation between naturally acquired antibodies directed against different parts of *P. falciparum* and *P. vivax* MSP-1 and acquired clinical immunity. For instance, antibodies against the N-terminal p83 fragment of *P. falciparum* MSP-1 were associated with protection against malaria in adolescents in Mali (Tolle *et al.*, 1993), while antibodies targeting the C-terminal p42 and p19 fragments were correlated with clinical immunity in children in West Africa and Papa New Guinea (al-Yaman *et al.*,

1996; Egan *et al.*, 1996). Interestingly, antibodies against the oligomorphic block 2 (Figure 11) strongly correlated with protection against *P. falciparum* malaria in children in Gambia and Ghana (Conway *et al.*, 2000; Cavanagh *et al.*, 2004), suggesting that both the conserved and the oligomorphic regions contribute to significant immune reactions.

An immune response based on MSP-1 vaccination could attack the *Plasmodium* parasite using a variety of mechanisms. The main contributor to an effective immune response is thought to be the humoral immune response. Antibodies directed against either the C-terminal p19, N-terminal p83 fragment or the full length MSP-1 protein were able to directly inhibit the growth of blood stages *in vitro* (Blackman *et al.*, 1990; Woehlbier *et al.*, 2006). Additionally, MSP-1 induced antibodies can interact with complement factors as shown for anti-p19 antibodies *in vitro* (Boyle *et al.*, 2015). Recruitment of other immune cells, such as neutrophils and monocytes, pose other mechanisms to eliminate the parasite in the host. The p19 fragment was also shown to activate polymorphonuclear neutrophils and trigger the release of reactive oxygen species in a reaction called antibody-dependent respiratory burst (Joos *et al.*, 2010). Another indirect growth inhibition mechanism relying on the recruitment of monocytes was demonstrated with antibodies against the N-terminal block 2 of *P. falciparum* MSP-1 (Galamo *et al.*, 2009).

But the blood stage is not the only part of the parasite's life cycle that can be targeted by an MSP-1 induced immune response. Vaccination of mice with p42 using viral vectors AdHu5 and MVA induced CD4⁺ and CD8⁺ T cells and resulted in an enhanced survival rate and reduced parasite burden presumably due to increased IFN- γ levels (Draper *et al.*, 2009). Several CD8⁺ T cell epitopes were identified within the MSP-1 sequence, suggesting that MSP-1 has the potential to induce cellular immune responses in humans and could lead to multistage protection (Carralot *et al.*, 2008). Such a stage-overlapping immunity is of particular importance for *P. vivax* infections, as an effective immune response in the liver could eliminate hypnozoites and hence, reduce or eliminate the risk of recurring relapses.

In malaria research, immunisation experiments regularly use mice and monkeys as animal models to validate the efficacy of a vaccine candidate or antimalarial drug. Such experiments were performed using both native MSP-1 isolated from parasite cultures and recombinant protein fragments. Purified MSP-1 from P. yoelii parasite cultures (Holder and Freeman, 1981) and recombinant p19 and p42 produced in E. coli (Daly and Long, 1995; Tian et al., 1997) or S. cerevisiae (Hirunpetcharat et al., 1997) induced protection against P. yoelii infection in mice. Similar experiments using native P. falciparum MSP-1 or recombinant MSP-1 fragments showed that whereas the complete protein could induce protection against lethal infections in Saimiri and Aotus monkeys (Perrin et al., 1984; Herrera et al., 1990; Etlinger et al., 1991; Cavanagh et al., 2014), the recombinant fragments only caused partial protection (Herrera et al., 1990; Etlinger et al., 1991; Cavanagh et al., 2014). Immunisation trials using only the C-terminal p42 fragment in adenoviral vectors or as a recombinant protein combined with adjuvant were unsuccessful in inducing significant protection in humans (Ogutu et al., 2009; Sheehy et al., 2012). However, other studies have pointed out the importance of the dimorphic and oligomorphic regions within the MSP-1 sequence. A longitudinal study in Mali revealed that antibodies detected in the population were predominantly directed against dimorphic regions (Früh et al., 1991) and as mentioned previously, antibodies directed against the block 2 sequence in P. falciparum were able to activate other immune cells (Galamo et al.,
Introduction

2009). These findings suggest that instead of focusing on specific parts of MSP-1, the full length protein should be used as a vaccine.

Many vaccination experiments focused on MSP-1 fragments instead of the full length protein as obtaining it as a recombinant protein in sufficient quantity and quality is technically challenging due to the high AT content (~80%) of the *Plasmodium falciparum* sequence, but also because of the sheer size of the molecule. Optimising the codon usage to human codon frequency reduced the AT content to 55% and allowed the synthesis of two MSP-1 proteins, namely MSP-1D and MSP-1F based on the MAD20 and K1 isolates (Pan *et al.*, 1999) (Figure 11). The MSP-1D variant is identical to the MAD20 isolate, except for block 2 and 4, which were replaced with the K1 sequence. Block 2 has also been replaced by the sequence of the MAD20 isolate in the MSP-1F protein, which largely resembles the K1 isolate.

A vaccine formulation composed of MSP-1D and the IDRI-GLA-SE adjuvant was developed in the research group of Prof. Hermann Bujard and has recently been tested in a phase Ia clinical trial at the University Clinic Heidelberg.



Figure 11. Synthetic genes of MSP-1D and MSP-1F and their correlation to field isolates MAD20 and K1. The codon optimised sequence of MSP-1D is mostly based on the MAD20 isolate from Papa New Guinea. The dimorphic block 2 and 4 were replaced with the sequence from the K1 isolate. In reverse, the MSP-1F protein is based on the K1 isolate from Thailand and only has block 2 replaced with the MAD20 variant. Arrows indicate processing sites for SUB1 and SUB2 leading to fragments p83, p30, p38, p42 after SUB1 processing and p33 and p19 resulting from p42 cleavage by SUB2. The numbering of the sequence blocks is based on (Tanabe *et al.*, 1987). White: conserved region, black: MAD20 sequence, grey: K1 sequence.

Introduction

1.8. Aim of this thesis

A recombinant protein vaccine based on *P. falciparum* MSP-1 has recently been tested in a phase I clinical trial. However, a similar formulation for *P. vivax* is missing, despite rising numbers of infections and the parasite's proven ability to cause severe disease in humans (Baird, 2007; Anstey *et al.*, 2009; Mueller *et al.*, 2009; Lacerda *et al.*, 2012). The core part of this project was the establishment of a protein production procedure for recombinant *P. vivax* MSP-1 (PvMSP-1) that can be used for structural and immunological analysis.

This protein production process should meet the following criteria: (i) The structural conformation of PvMSP-1 has to be identical or at least as close as possible to the native confirmation to ensure the inclusion of structural epitopes. (ii) The process itself has to be robust and result in a high yield at low costs. Most countries affected by *P. vivax* malaria are newly industrialised countries and although this project is the very preliminary stage of vaccine development, cost-efficiency should be considered. (iii) Upscaling of the protocol to pharmaceutical manufacturing facilities following GMP guidelines has to be possible.

Structural information about MSP-1 is still limited and no crystal structure of the full length protein is available. An efficient recombinant protein production protocol can help overcoming limitations due to low quantity or quality of a protein and facilitate in depth structural analysis. In this study, structural information was gathered by CD spectroscopy of *P. vivax* MSP-1 and other recombinant MSP-1 proteins to control proper folding of the protein. *In vitro* processing using *P. falciparum* SUB1 was used to check for conserved cleavage sites among MSP-1 proteins from different *Plasmodium* species. Interactions within MSP-1 were identified by combining cross-linking techniques with mass spectrometry, leading to an updated model of MSP-1 and the arrangement of its processing subunits.

Immunological experiments proved to be difficult due to the lack of a long-term *P. vivax* cell culture or animal model. In recent years, other laboratories in proximity to *P. vivax* endemic areas managed to establish short-term cultures or have collected blood samples from different cohorts to be used in *P. vivax* research. During the course of this study, preparations were made to collaborate with other research facilities to analyse the immunological potential of PvMSP-1. Recombinant proteins corresponding to individual PvMSP-1 processing fragments were purified and can be used to analyse fragment-specific immune responses. However, cohort studies and *in vitro* experiments have their limitations compared to animal models. As neither *P. falciparum* nor *P. vivax* MSP-1 can be analysed in non-humanised mouse models, *P. berghei* MSP-1 was produced in the same manner as *P. vivax* MSP-1. This protein based on the rodent parasite *P. berghei* ANKA is ready to be used in immunisation studies to analyse the potential protective effect of MSP-1 against cerebral malaria.

2.1. Materials

2.1.1. Laboratory equipment

ABI 7500 Real-Time PCR System Analytical ultracentrifuge Autoclave 5075 ELV BioPhotometer BioTek Cytation Hybrid multi-mode reader CD spectrometer J-715 Centrifuges Heraeus Megafuge 16 Heraeus Pico 17 Heraeus Megafuge 1.0R Chromatography system, ÄKTA Purifier 100 Dialysis clips Digital timer Electrophoresis Power Supply EV243 ELISA reader Multiskan FC Freezers -20°C -80°C Fridges

Gel Electrophoresis System EasyPhor GeneTouch Thermal Cycler Heating block Ice machine Incubator shaker, Innova 4000/4300 Incubators Magnetic stirrer C-MAG HS7 Microwave oven Nalgene[™] Polysulfone Filter Holder pH meter InoLab® pH7110 Pipetting aid Pipetus® Precision balances Protein Electrophoresis system Mighty Small II Roll mixer RS-TR05 Single channel pipettes HTL Discovery comfort Sonicator Sonopuls HD 2070 Standard Analog Shaker

Life technologies, CA, USA Beckman Coulter, CA, USA Systec, Wettenberg Eppendorf, Hamburg BioTek Instruments, Bad Friedrichshall Jasco, Pfungstadt Thermo Fisher Scientific, MA, USA Thermo Fisher Scientific, MA, USA Heraeus Instruments, Hanau GE Healthcare, Little Chalfont, UK WeLoc, Smålandsstenar, Sweden Carl Roth, Karlsruhe Sigma-Aldrich, MO, USA Thermo Fisher Scientific, MA, USA Liebherr International GmbH, Biberrach an der Riß Thermo Fisher Scientific, MA, USA Liebherr International GmbH, Biberrach an der Riß Bosch, Stuttgart Biozym, Hessisch Oldendorf Biozym, Hessisch Oldendorf VWR International GmbH, Darmstadt Ziegra Eismaschinen, Isernhagen New Brunswick Scientific Co. Inc., Enfield, CT, USA Heraeus Instruments, Hanau IKA, Staufen Zanussi, Nuremberg Thermo Fisher Scientific, MA, USA WTW, Weilheim Hirschmann Laborgeräte, Eberstadt Kern EG 2200-2NM, Balingen Hoefer, Inc., Holliston, MA, USA Phoenix Instruments, Garbsen Kinesis GmbH, Langenfeld Bandelin electronics, Berlin VWR International GmbH, Darmstadt

Vortex

Water bath Precitherm PFV Water purifcation system Purelab Plus Wet blot system XCell SureLockTM VWR International GmbH, Darmstadt Labora, Mannheim ELGA Labwater, Celle Thermo Fisher Scientific, MA, USA

2.1.2. Consumables

Amicon Ultra Centrifugal filters Cell strainer Cuvettes Falcons (15 mL, 50 mL) Gloves Immobilon-P PVDF Membrane (0,45 µm) Immobilon-P Transfer Membrane, PVDF Microcentrifuge tubes 1.5 mL 2 mL5 mL -Parafilm M® PCR tubes PD10 Desalting columns Pipette tips Diamond (10 µL, 200 µL, 1000 µL) Plastic pipettes (1 ml, 2 ml, 5 ml, 10 ml, 25 ml) Polystyrene columns Spectra/Por Dialysis Membrane Sterile filters, Rotilabo® (0.2µm; 0.45µm) Whatman Paper

2.1.3. Chemicals and reagents

2-Mercaptoethanol 4-Nitrophenyl phosphate disodium salt hexahydrate tablets Agar Bacteriology grade Ammonium persulfate BCIP tablets Bissulfosuccinimidyl suberate Brilliant Blue G250 Calcium chloride dihydrate cOmplete Tablets mini EDTA-free EASY*pack* Diethanolamine Dipotassium phosphate Disodium hydrogen phosphate Merck Milipore, Darmstadt Thermo Fisher Scientific, MA, USA Sarstedt AG & Co., Nümbrecht Greiner Bio-One, Frickenhausen Ansell, Yarra City, Australia Millipore, Corp. Billerica, MA, USA Merck Milipore, Darmstadt

Sarstedt AG & Co., Nümbrecht Sarstedt AG & Co., Nümbrecht Eppendorf, Hamburg Bemis Flexible Packaging, WI, USA Sarstedt AG & Co., Nümbrecht GE Healthcare, Little Chalfont, UK Gilson International, Bad Camberg Greiner Bio-One, Frickenhausen Thermo Fisher Scientific, MA, USA Spectrum Laboratiries Inc., CA, USA Carl Roth, Karlsruhe GE Healthcare, Little Chalfont, UK

Sigma-Aldrich, MO, USA Sigma-Aldrich, MO, USA AppliChem, Darmstadt Grüssing, Filsum Sigma-Aldrich, MO, USA Thermo Fisher Scientific, MA, USA Carl Roth, Karlsruhe AppliChem, Darmstadt Roche, Basel, Switzerland Carl Roth, Karlsruhe Grüssing, Filsum Disuccinimidyl sulfoxide DTT E64 EDTA 0.1 mol/L EDTA disodium salt dihydate Ethanol Ethidiumbromide Glutathione oxidised Glycerol Glycine Guanidine hydrochloride Hydrochloric acid Imidazole IPTG L-Arginine L-Glutathione reduced LB broth Leupeptin Lysozyme Magnesium chloride Midori Green Ni-NTA agarose Superflow Pefabloc Potassium acetate Potassium chloride Potassium dihydrogen phosphate Powdered milk Protein Assay Dye Reagent Concentrate Rotiphorese Gel 30 (30% acrylamide) SDS Ultra pure Sodium bicarbonate Sodium carbonate Sodium chloride Sodium dihydrogen phosphate Sodium hydroxide TCA TEMED Tris Tween 20 UltraPureTM Agarose

Thermo Fisher Scientific, MA, USA Carl Roth, Kalrsruhe Sigma-Aldrich, MO, USA J.T.Baker, PA, USA AppliChem, Darmstadt Zentralbereich Neuenheimer Feld, Heidelberg Carl Roth, Karlsruhe Carl Roth, Karlsruhe **VWR** Chemicals Sigma-Aldrich, MO, USA Sigma-Aldrich, MO, USA Honeywell Fluka, NJ, USA Merck & Co., Inc, NJ, USA Carl Roth, Karlsruhe Carl Roth, Karlsruhe Carl Roth, Karlsruhe Sigma-Aldrich, MO, USA Sigma-Aldrich, MO, USA Carl Roth, Karlsruhe Merck, Darmstadt Biozym, Hessisch Oldendorf Qiagen, Hilden Sigma-Aldrich, MO, USA Sigma-Aldrich, MO, USA AppliChem, Darmstadt Carl Roth, Karlsruhe Carl Roth, Karlsruhe Bio-Rad, CA, USA Carl Roth, Karlsruhe Carl Roth, Karlsruhe Sigma-Aldrich, MO, USA Sigma-Aldrich, MO, USA Sigma-Aldrich, MO, USA Grüssing, Filsum Honeywell Fluka, NJ, USA Sigma-Aldrich, MO, USA Carl Roth, Karlsruhe Carl Roth, Karlsruhe Carl Roth, Karlsruhe Thermo Fisher Scientific, MA, USA

2.1.4. Buffers, media and solutions

2.1.4.1. Antibiotic stock solutions

Ampicillin	50 mg/mL in ddH ₂ O
Kanamycin	25 mg/mL in ddH ₂ O
Chloramphenicol	34 mg/mL in Ethanol

2.1.4.2. Molecular biology

TAE buffer (50x)	2 M Tris
	50 mM EDTA, pH 8.0
	1 M acetic acid
	ad 1L with ddH_2O

2.1.4.3. Culture and transformation of E. coli

Transformation buffer	70 mM CaCl ₂
	35 mM MgCl ₂
LB medium	10 g/L tryptone
	5 g/L yeast extract
	5 g/L NaCl

2.1.4.4. Microbiological methods

Lysis buffer	20 U/mL DNAseI
	10 μg/mL lysozyme
	1 protease inhibitor tablet (Roche)
	in PBS, pH 7.4
Solubilisation buffer	6 M guanidine hydrochloride
	2 mM EDTA
	50 mM NaH ₂ PO ₄
	50 mM DTT
	pH 8.0
Freezing solution	20% PBS
	80% glycerol
Solution A	50 mM CaCl
	10 mM potassium acetate, pH 6.2

Solution B	50 mM CaCl
	20% glycerol
	10 mM potassium acetate, pH 6.2

2.1.4.5. Biochemical methods

Blocking buffer	TBST + 1% skim milk powder
Coomassie staining solution	250 mL ethanol
	50 mL acetic acid
	200 mL ddH ₂ O
	1.25 g coomassie blue
D.1	<u>For Pv83:</u>
	18.9 mM Na ₂ HPO ₄
	1.1 mM NaH ₂ PO ₄
	20 mM NaCl
	pH 8.5
	For PvMSP-1:
	18.9 mM Na ₂ HPO ₄
	1.1 mM NaH ₂ PO ₄
	50 mM NaCl
	pH 8.0
D.2	18.9 mM Na ₂ HPO ₄
	1.1 mM NaH ₂ PO ₄
	300 mM NaCl
	pH 8.0
D.3	18.9 mM Na ₂ HPO ₄
	1.1 mM NaH ₂ PO ₄
	1 M NaCl
	pH 8.0
Destaining solution	100 mL acetic acid
-	200 mL ethanol 96%
	ad 1 L H ₂ O

ELISA coating buffer	34 mM Na2CO3 16 mM NaHCO3 pH 10.6
ELISA substrate buffer	99 mM Diethanolamine 1 mM MgCl ₂ pH 9.5
Lower Tris	1.5 M Tris 0.4% SDS pH 8.8
Mono Q buffer A	8 M Urea 20 mM Tris 4 mM DTT pH 8.0
Mono Q buffer B	8 M Urea 20 mM Tris 500 mM NaCl 4 mM DTT pH 8.0
Mono S buffer A	8 M Urea 20 mM NaPO4 4 mM DTT pH 6.5
Mono S buffer B	8 M Urea 20 mM NaPO4 500 mM NaCl 4 mM DTT pH 6.5
Ni-NTA elution buffer	50 mM NaH2PO4 300 mM NaCl 250 mM imidazole pH 8.0
Ni-NTA wash buffer	50 mM NaH2PO4 300 mM NaCl 20 mM imidazole pH 8.0

Phosphate buffer

Quenching buffer (crosslinking)

Renaturation buffer

SDS gel (10%)

10 mM NaH₂PO₄ pH 7

1 M Tris-HCl pH 7.5

For Pv42 and PbMSP-1: 500 mM arginine 50 mM NaH₂PO₄ 1 mM EDTA 10 mM reduced glutathione 1 mM oxidised glutathione pH 8

<u>For Pv83:</u> 500 mM arginine 50 mM NaH2PO4 1 mM EDTA 5 mM DTT pH 8.5

<u>For PvMSP-1:</u> 1 M arginine 100 mM Tris 1 mM EDTA 10 mM reduced glutathione 1 mM oxidised glutathione pH 8.5

For one separating gel:

1.85 mL 30% acrylamide
 1.25 mL lower Tris buffer
 2.1 mL H₂O
 10 μL TEMED
 50 μL 10% APS

	For one stacking gel:
	330 μL 30% acrylamide
	625 µL upper Tris buffer
	1.55 mL H ₂ O
	5 μL TEMED
	25 µL 10% APS
SDS-PAGE running buffer (10x)	10 g SDS
	30,3 g Tris
	144,1 g glycine
SDS-PAGE sample buffer (4x)	8% SDS
	50% upper Tris buffer
	40% glycine
	0.08% bromphenol blue
	100 mM DTT
SUB1 buffer	50 mM Tris-HCl
	15 mM CaCl
	рН 7.6
TBST	150 mM NaCl
	10 mM Tris
	0.05% Tween 20
	pH 8.0
Transfer buffer	25 mM Tris
	192 mM glycine
	0.01% SDS
	20% EtOH
Upper Tris	0.5 M Tris
	0.4% SDS
	pH 6.8
2.1.5. Commercial kits	
High Molecular Weight Calibration Kit	GE Healthcare, Little Chafont, UK

High Molecular Weight Calibration Kit	GE Healthcare, Little Chafont, UK
NucleoSpin Gel and PCR Clean-up	Macherey-Nagel
NucleoSpin Plasmid	Macherey-Nagel
Protein Kit (Biuret Method)	Merck
Zero Blunt® TOPO PCR Cloning Kit	Thermo Fisher Scientific, MA, USA

2.1.6. Enzymes

BamHI-HF	New England Biolabs, MA, USA
BglII	New England Biolabs, MA, USA
ClaI	New England Biolabs, MA, USA
DNAse I	Sigma-Aldrich, MO, USA
Phusion® High-Fidelity DNA polymerase	New England Biolabs, MA, USA
PstI	New England Biolabs, MA, USA
PvuI-HF	New England Biolabs, MA, USA
SacI-HF	New England Biolabs, MA, USA
SphI-HF	New England Biolabs, MA, USA
T4 ligase	New England Biolabs, MA, USA

2.1.7. Markers

2 log DNA ladder	New England Biolabs, MA, USA
Color Plus Protein Marker	New England Biolabs, MA, USA
Protein ladder 10-250 kDa	New England Biolabs, MA, USA
Unstained protein standard Broad range 10-200 kDa	New England Biolabs, MA, USA

2.1.8. Antibiotics

All antibiotics were prepared as a 1,000x stock solution, sterile filtered and stored at -20°C.

Table 1. Overview of used antibiotics.

	stock concentration	manufacturer
Ampicillin	50 mg/mL	Carl Roth, Karlsruhe
Chloramphenicol	34 mg/mL	Carl Roth, Karlsruhe
Kanamycin	25 mg/mL	Carl Roth, Karlsruhe

2.1.9. Antibodies

Table 2. Antibodies used for Western Blots or ELISA.

antibody	manufacturer
Anti-Human IgG-alkaline phosphatase	Sigma-Aldrich, MO, USA
Anti-Mouse IgG-alkaline phosphatase	Sigma-Aldrich, MO, USA
Anti-Penta-His monoclonal antibody	Thermo Fisher Scientific, MA, USA
Anti-Rabbit IgG-alkaline phosphatase	Sigma-Aldrich, MO, USA

2.1.10. Prepacked chromatography columns for Äkta[™] systems

Table 3. List of prepacked chromatography columns

model name	manufacturer
HiLoad TM 26/600 Superdex TM 200 prep grade	GE Healthcare, Little Chalfont, UK
HiPrep 26/10 Desalting	GE Healtchare, Little Chalfont, UK
Mono Q 10/10	GE Healthcare, Little Chalfont, UK
Mono S 10/10	GE Healthcare, Little Chalfont, UK
Superdex TM 200 Increase 10/300 GL	GE Healthcare, Little Chalfont, UK

2.1.11. Software & Online Tools

Adobe Illustrator	Adobe Systems Software, Dublin, Republic of Ireland
Adobe Photoshop	Adobe Systems Software, Dublin, Republic of Ireland
Ape Version 2.0.47	M. Wayne Davis
Enzyme X 3	Mekentosj, Dordrecht, Netherlands
Microsoft Excel	Microsoft Corporation, Redmond, WA, USA
Microsoft Word	Microsoft Corporation, Redmond, WA, USA
Prism 6	GraphPad Software Inc., CA, USA
Serial Cloner Version 2.6.1	Franck Perez, SerialBasics
Snapgene Viewer	GSL Biotech LLC, IL, USA

Online Tools

Clustal Omega	Embl-Ebi, Cambridgeshire, UK
ExPASy Bioinformatics Resource Portal	Swiss Institute of Bioinformatics, Switzerland
GeneArt TM Synthesis	Thermo Fisher Scientific, MA, USA
Graphical Codon Usage Analyser	Dr. Thomas Schödl, University of Regensburg, Germany
JPred4	University of Dundee, Dundee, UK
mFold web server	The RNA Institute, State University of New York at Albany
NCBI BLAST	National Center for Biotechnology Information, MD, USA
NEB Tm Calculator	New England Biolabs, MA, USA
NEBioCalculator	New England Biolabs, MA, USA
PoPS	Dr. Sarah Boyd, Monash University, Melbourne, Australia
WatCut	University of Waterloo, Canada

2.1.12. E.coli strains

Table 4. E. coli strains used for cloning and protein expression.

Strain	Genotype					Source
Dh5alphaZ1	Δ(lac)U169,	end	A1,	gyrA46,	hsdR17(rK-mK+),	Dr. Rolf Lutz, ZMBH
	phi80,4(lacZ)M15, recA1,relA1, supE44, thi-1, laciq, tetR,SPr					
W3110Z1	ATCC Nr. 2732	25, laciq,	tetR, SP	r		Dr. Rolf Lutz, ZMBH

2.1.13. Amino acid sequences of MSP-1

Table 5. Uniprot entries for used MSP-1 amino acid sequences.

Plasmodium strain	Uniprot entry
Berghei ANKA	A0A077XCK6
Falciparum 3D7	Q8I0U8
Falciparum FCB-1	P04933
Knowlesi	A7BG19
Vivax Belem	Q8I1M7
<i>Vivax</i> Salvador I	A5K724

2.1.14. Primers

Table 6. Oligonucleotides used in this study.

Oligonucleotide	Sequences 5' to 3'	T _M (°C)
pMA-Pv42seq	CAGTGAGCGGAAGGC	53.2
Pv30Claup	CCATCGATAGCGAGACAGGCACCACC	59.3
Pv30Pstdown	TTCTGCAGTTAGTTGCCGCCCACTCTTCTG	59.0
Pv3842BamHup	CGGGATCCATGAGCGAGGAAAAGCCC	52.1
Pv38Claup	CCATCGATAGCGAGGAAAAGCCCGAG	57.8
Pv38Pstdown	TTCTGCAGTTACTCGCCGGTGGTCACTTGATC	60.3
Pv42Claup	CCATCGATGCCGAGAGCGAGGCCCCTGAAATTC	65.5
Pv42Pstdown	TTCTGCAGTCAGCTGGAGCTGC	51.5
Pv8330BamHup	CGGGATCCATGGAAACCGAGAGCTACAAGCAG	55.7
Pv8330Cladown	CCATCGATTTAGTTGCCGCCCACTCTTCTG	59.0
Pv83Claup	CCATCGATGAAACCGAGAGCTACAAGCAG	55.7
Pv83Pstdown	AACTGCAGTTAGGCGCCTCTCAGAGAC	55.3
Pv83Seq1	GCGACCTGATCATTGCCG	56.7
Pv83Seq2	GACATTCGGCGACCTC	53.0
T1Avr	CCTAGGTCTAGGGCGGCGGATTTGTCC	65.7
Pb8330ClaXaup	CCATCGATATCGAAGGTCGTGAAACCATCGAGGTGTACAACGACATCATC	60.5

Pb8330Pstdown	TTCTGCAGTTAGGACTCGGCTCTGGTGGTG	60.7
Pb3842BamHup	CGGGATCCATGGAGGAAGATATCCCCGCC	54.7
Pb3842Cladown	CCATCGATTTAGCTGCTGGAGGAGCAGAACAC	60.6
pZE23 N-terminal	GATTCAATTGTGAGCGG	50.0
pZE23 C-terminal	CTCAGGAGAGCGTTCACCGAC	59.7
Pb8330Seq1	GCACCACCGGAACA	46.7
Pb8330Seq2	CAGAACATCCTGGT	46.1
PfPv8330	CGGGATCCATGGTGACTCACGAATCGTATCAAGAATTGGTGAAGAAGCTCGAA	60.6
	GCTTTAGAGGACGCCGTATTGGGCTACGAGCTGTTCCACAAGAAGAA	

2.1.15. Expression plasmids

 Table 7. Overview of expression plasmids for MSP-1 halves and processing fragments.

vector	resistance marker
pZE13-HXPv30	Amp
pZE13-HXPv38	Amp
pZE13-HxPv42	Amp
pZE13-HxPv83	Amp
pZE13-HxPv83/30	Amp
pZE13-XaPb83/30	Amp
pZE23-Pb38/42	Kan
pZE23-PfPv83/30	Kan
pZE23-Pv38/42	Kan

2.2. Methods

2.2.1. Molecular biological methods

2.2.1.1. Obtaining the synthetic genes

All expression plasmids of PvMSP-1 and its processing subunits were cloned using a synthetic gene of PvMSP-1. The sequence of this gene was based on the MSP-1 nucleotide sequence found in the Salvador I *P. vivax* strain (GeneID: PVX_099980, PlasmoDB.org). The nucleotide sequence was optimised for two reasons: i) The DNA of *Plasmodium* is usually very AT-rich, which can cause problems during cloning. ii) Optimising the sequence for humans allows the usage of the synthetic gene in adenoviruses for vaccination challenges.

The nucleotide sequence was optimised according to human codon usage via the GeneArt Synthesis online tool. The three restriction sites used for cloning (BamHI, ClaI, PstI) were excluded from this procedure to avoid these restriction sites in the synthetic gene. For cloning purposes, a start codon (ATG) was added at the 5' end of the sequence and a stop codon plus ClaI restriction site at the 3' end.

The synthetic gene was split into three fragments to lower the costs of gene synthesis. For this purpose, the nucleotide sequence was scanned for possible silent mutations using the WatCut online tool by the University of Waterloo, Canada. The usage of normal restriction sites would have caused a frameshift or insertion of additional amino acids into the sequence, while silent mutations do not alter the amino acid sequence. A BgIII restriction site was added to split the gene into fragment one and two, a SphI restriction site was included to allow ligation of fragment two and three. The fragments were later ligated using these restriction sites and PvuI, which cuts once within the GeneArt vector, to obtain the full synthetic gene.

Expression plasmids for PbMSP-1 were created following the same principle. The coding sequence of *P. berghei* ANKA MSP-1 (GeneID: PBANKA_0831000, PlasmoDB.org) was optimised according to codon frequency in mice to enable insertion into adenoviral vectors for immunisation purposes. The restriction sites used for cloning into pZE expression plasmids were excluded from this optimisation. Instead of ordering the synthetic gene in three fragments, the entire coding sequence was ordered as one gene and the nucleotide sequence coding for each of the MSP-1 halves was amplified via PCR before ligation into the expression plasmids.

2.2.1.2. Isolation of plasmid DNA from E. coli

Gene fragments and expression plasmids were amplified in the *E. coli* strain Dh5alphaZ1 and purified by alkaline lysis using the NucleoSpin Plasmid Kit (Macherey-Nagel). Three mL overnight culture were pelleted (11,000 g, 1 min) and resuspended in buffer A1 (250 μ L). The addition of a buffer containing sodium hydroxide (buffer A2, 250 μ L) alkalised the mixture and broke the hydrogen bonds between the complementary DNA strands of both genomic and plasmid DNA. When a neutralisation buffer was added (buffer A3, 300 μ L), the plasmid DNA refolds, while the genomic DNA remained unfolded. In the following steps, the refolded plasmid DNA was purified via its interaction with a silica membrane. First, the lysate was spun at 11,000 g for 10 minutes to separate cell debris from the solution containing the DNA. The supernatant was loaded onto a NucleoSpin Plasmid column (11,000 g, 1 min), which was then washed with buffer A4 (600 μ L, 11,000 g, 1 min). The silica membrane was dried by another centrifugation step (11,000 g, 2 min). Elution was achieved by adding the elution buffer (50 μ L) to the column, incubating for 1 minute at room temperature and centrifugation (9,000 g, 2 min) after placing the column in a new microcentrifuge tube.

2.2.1.3. Agarose gel electrophoresis of DNA

DNA electrophoresis was frequently used to check purified plasmid DNA and PCR products during the cloning procedure. The agarose gels used in this study contained 1% agarose in TAE buffer. Ethidium bromide was added directly to the gel before polymerisation to visualise the DNA (3 μ L in 50 mL agarose). Gels were run at 80-120 V and analysed under UV light.

2.2.1.4. Purification of DNA from agarose gels

Restriction digested plasmids and PCR products were separated on a 1% agarose gel running at 80 V. The desired bands were cut from the gel and used for purification with the NucleoSpin® Gel and PCR Cleanup Kit. Usually, 1 mL of NTI buffer was added to the gel samples before they were heated at 50 °C. When all agarose was melted, the solution was transferred into the column and centrifuged at 11,000 g for 30 s. The silica membrane was washed with 700 µL NT3 buffer (11,000 g, 30 s), followed by a drying step (11,000 g, 1 min) to remove all buffers completely. For the elution, 15-30 µL of elution buffer were pipetted on the silica membrane and incubated for 1 min before centrifugation (11,000 g, 1 min).

2.2.1.5. Determination of nucleic acid concentration

There are several different techniques to measure DNA concentration. In this study, two methods to determine the concentration of DNA were used. At the beginning of the experiments, the DNA concentration was determined by running a specific volume of the sample on a 1% agarose gel and comparing it to the 2-log DNA ladder bands or by direct measurement with UV/Vis absorbance. The manual of the DNA ladder states the amount of DNA in nanogram per band, allowing direct comparison of the bands to estimate the concentration. Later on, the DNA concentration was measured using the BioTek Cytation Hybrid multi-mode reader and the Take3 Micro-volume plate. This setup allows the measurement of the nucleic acid concentration via UV/Vis absorbance using only 2.5 µL of sample.

2.2.1.6. Primer Design

The construction of expression vectors for all individual recombinant proteins required specific primers for each plasmid. These primers were designed to contain a homologous region, which binds to the template DNA, a restriction site for cloning and a start or stop codon if such a codon was not included in the amplified region of the template. The length of the homologous region varied between 20-30 nucleotides and was adjusted to include cytosine or guanine at the end to ensure a strong binding to the template. The length was also adjusted depending on the GC content and melting temperature of the primer. If the PCR product was intended to be inserted into pZE23 expression plasmids, a BamHI restriction site was included in the forward primer and a ClaI restriction site in the reverse primer. Ligation into the pZE13 expression plasmids required a ClaI cleavage site in the forward primer and a PstI restriction site in the reverse primer. The aimed GC content was 40-60% and the melting temperature was designed to be approximately 60°C. The primers were ordered at Thermo Fisher as custom oligos.

2.2.1.7. Polymerase chain reaction (PCR)

The polymerase chain reaction is a standard technique to amplify specific nucleotide sequences for cloning and control purposes. The coding sequences for the two halves of MSP-1 and its processing fragments were amplified using primers that contained the start and stop codon for protein expression and restriction sites for insertion into pZE13 or pZE23 expression plasmids. The less error-prone Phusion polymerase was used to lower the risk of amino acid substitutions or truncations of the recombinant protein. The exact setup and thermocycling conditions for a preparative PCR are given in **Table 8**and **Table 9**. All primers were originally prepared as a 100 µM stock and diluted 1:10 before being used in a PCR.

Component	Volume	Final concentration
5x Phusion HF buffer	10 µL	1x
10 mM dNTPs	1 μL	200 µM
10 µM forward primer	2.5 μL	0.5 µM
10 µM reverse primer	2.5 μL	0.5 µM
Template DNA	1 µL	< 250 ng
Phusion DNA polymerase	0.5 μL	
Water	Ad 50 μL	

Table 8. Components of a preparative PCR.

Table 9. PCR protocol using Phusion polymerase.

Step	Temperature	Time
Initial denaturation	98°C	30 s
30 cycles	98°C	5 s
	T _m	30 s
	72°C	15-30 s per kb
Final extension	72°C	10 min

The melting temperature used in the PCR was calculated for each primer pair using the T_m calculator provided on the website of New England Biolabs. This calculation takes the used polymerase, buffer and primer length and concentration into consideration. All used melting temperatures for each primer set are listed in Table 10.

Table 10	Overview of	used primer	sets and the	respective t	emperature u	sed during PCR.
----------	--------------------	-------------	--------------	--------------	--------------	-----------------

Forward primer	Reverse primer	Calculated melting temperature
Pb8330ClaXaup	Pb8330Cladown	68°C
Pb3842BamHup	Pb3842Cladown	63°C
Pv8330BamHup	Pv8330Cladown	59°C
Pv3842BamHup	Pv3842Cladown	59°C
Pv38Claup	Pv38Pstdown	67°C
Pv30Claup	Pv30Pstdown	59°C

Pv42Claup	Pv42Pstdown	55°C
Pv83Claup	Pv83Pstdown	60°C
PfPv8330	Pv8330Cladown	68°C

2.2.1.8. Subcloning in TOPO vector

Some PCR products were subcloned into the commercially available TOPO vector to amplify the PCR product in *E. coli* before moving on to restriction digestion and ligation into the intended expression plasmid. The PCR product was mixed with the salt solution included in the Zero Blunt[®] TOPO PCR Cloning Kit and water before the vector, which is fused to the TOPO isomerase, was added (see Table 11). The mixture was incubated at room temperature for 5-30 minutes before transformation into the *E. coli* strain Dh5alphaZ1.

Table 11. Setup for inserting a PCR product into the TOPO vector.

Reagent	Volume
PCR product	2 μL
Salt solution	1 μL
Water	Ad 5 μL
pCR TM II-Blunt-TOPO	1 μL
Final volume	6 μL

2.2.1.9. DNA restriction

DNA restriction is a standard technique to insert a coding nucleotide sequence at a specific site in an expression plasmid or to check vectors for correct insertion. In this study two types of expression plasmids were used. The pZE13 plasmid contains an upstream ClaI restriction site immediately downstream of the His-tag, and a PstI cleavage site downstream of the insert. For the expression of full-length MSP-1, the pZE23 vectors were usually used, which contain a BamHI restriction site at the N-terminal end of the insert and a ClaI cleavage site at the C-terminal end. The ligation of these expression plasmids with a PCR product required a double digest, using two enzymes cleaving the DNA sequences with sticky ends. If the cleaved DNA was to be used in further cloning steps, a preparative digestion was performed. If the DNA restriction was done as a control for correct insertion, a smaller volume of the reaction mixture was prepared. See Table 12 for details.

Table 12. Restriction digest for cloning and control purposes.

Reagent	Standard protocol	Preparative digestion
Cutsmart buffer	1 μL	5 μL
Restriction enzyme	0.3 μL each	1 μL each
DNA	100 – 200 ng	$\sim 1~\mu g$ or all of the PCR product
Water	ad 10 µL	ad 50 μL

The mixture was incubated at 37 °C for one hour (standard protocol) or 2 hours (preparative digestion). The resulting digested DNA was run on a 1% agarose gel for analysis or subsequent purification steps.

2.2.1.10. Ligation

Ligation of restriction digested plasmids and inserts was performed using T4 DNA ligase. The specific volumes of DNA were calculated using the NEBioCalculator based on the length of the nucleotide sequence (see formula below). Usually, a ratio of 3:1 and 1:1 was used unless otherwise stated. The ligation reaction was performed at 16°C overnight.

Table 13. Components of a ligation reaction.

Component	Volume/amount
10x T4 Ligase buffer	2 µL
Expression plasmid	50 ng
Insert	Calculated with NEBioCalculator
T4 DNA ligase	1 μL
Water	ad 20 µL

Formula for calculating ligation mixture:

 $mass insert (g) = desired \frac{insert}{vector} molar ratio \times mass of vector (g) \times ratio of insert to vector lengths$

2.2.1.11. Sequencing

To confirm the correctness of expression plasmids, all cloned vectors were sent to GATC Biotech for sequencing using the SUPREMERun protocol. If possible, primers offered by the GATC Biotech company were used, otherwise custom primers were sent with the DNA samples. The sample requirements were $30-100 \text{ ng/}\mu\text{L}$ for the DNA to be sequenced and 10 pmol/ μL for the primers. Both DNA and oligonucleotides were sent in a total volume of 20 μL .

2.2.2. Microbiological methods

2.2.2.1. Preparation of chemically competent cells

Bacterial strains used for cloning and protein expression are usually made either chemically or electrocompentent to increase the uptake of foreign DNA. The cells used in this study were washed in two different CaCl₂ containing buffers to render them chemically competent. First, an overnight culture of the strain was grown in 5 mL LB medium with the appropriate antibiotic at 37°C. The next day, 500 mL LB medium with antibiotic were inoculated with 2 mL of overnight culture and grown to an OD₆₀₀ of 0.3-0.4. The culture was cooled on ice for 15 min, centrifuged at 4,000 rpm for 7 min at 4°C and resuspended in 250 mL cold solution A. After another incubation step on ice for 30 min, the cell suspension was centrifuged again and resuspended in 25 mL of cold solution B. The bacteria suspension was distributed into aliquots of 300-500 µL and immediately frozen. The chemically competent cells were stored at -80°C.

2.2.2.2. Transformation of chemically competent cells

Cloning of DNA and production of recombinant proteins in *E. coli* exploits the ability of bacteria to take up foreign DNA, called transformation. The *E. coli* strains used in this study were treated with CaCl₂ to enhance their competence in DNA uptake. Constructs were amplified using the Dh5alphaZ1 strain, while the W3110Z1 strain was used for protein expression.

Transformation of both strains was performed using the same protocol: $1-5 \ \mu$ L of purified plasmid DNA or ligated DNA were pipetted to 100 μ L of chemically competent *E. coli* bacteria and 100 μ L of transformation buffer. The mixture was incubated on ice for 30 minutes, followed by a heat shock at 42°C in a water bath. After another incubation step of 15 minutes on ice, 1 mL of LB medium was added and the bacteria were grown at 37°C, 220 rpm for 1 hour. The cells were pelleted at 6,000 g for 3 minutes. Most of the supernatant was removed, leaving 200-300 μ L in the tube, in which the cell pellet was resuspended. The bacteria were plated on LB agar plates containing the appropriate antibiotic.

2.2.2.3. Preparation of glycerol stocks for cryopreservation

Once an *E. coli* strain was successfully transformed with the desired expression plasmid, frozen stocks were created as a backup and for the inoculation of overnight pre-cultures. 1 mL of LB medium containing the appropriate antibiotic was inoculated with 30 μ L overnight culture. The culture was grown at 37°C for 2-3 hours. 675 μ L of this culture was mixed with 375 μ L freezing solution and stored at -80°C.

2.2.2.4. Testexpression

A small scale expression test was performed to identify W3110Z1 clones expressing the recombinant protein. An overnight culture of 5 mL with the appropriate antibiotic (kanamycin in case of pZE23, ampicillin in case of pZE13 expression plasmids) was prepared. Next day, 3 mL LB medium with antibiotic were inoculated with 60 μ L of overnight culture. Once the culture has reached an OD₆₀₀ of 0.4-0.5 the culture was split into two tubes. One was induced with 1 mM IPTG, the other was grown without the addition of IPTG. Both cultures were incubated for further 3 hours at 37°C before centrifuging 200 μ L of each culture at 6,000 g for 5 minutes at room temperature. The pellet was resuspended in 40 μ L of SDS sample buffer with DTT. Analysis of protein expression was done via SDS-PAGE comparing the noninduced and the induced sample.

2.2.2.5. Expression of recombinant proteins

500 mL LB medium were autoclaved in a 2 L Erlenmeyer flask the day before protein expression was performed. Either ampicillin or kanamycin were added to the medium directly before use (see **Table 7** for resistance markers of used plasmids). The medium was inoculated with 10 mL of an overnight pre-culture of the *E. coli* strain W3110Z1 (see Table 4) containing the expression plasmid of the protein of interest. The culture was grown to an OD₆₀₀ of 0.4-0.5 at 37°C, 220 rpm. Protein expression was induced by addition of 1 mM IPTG after a sample of the noninduced culture was taken. The induced culture was incubated for further 3 hours at 37°C, 220 rpm. A sample of the induced culture was taken at the end of the incubation step to compare the noninduced and induced culture via SDS-PAGE. The bacterial cells were harvested at 4,000 rpm, 4°C for 25 minutes. The bacterial pellet was transferred into a 50 mL falcon and was frozen at -80°C until cell rupture was performed.

2.2.2.6. Cell rupture

Some recombinant proteins are expressed in soluble form, others accumulate in aggregates of insoluble proteins, known as inclusion bodies. Therefore, the first step of every protein purification protocol is to separate the soluble from the insoluble fraction. The bacterial pellets were thawed on ice. 2 mL of lysis buffer per gram of cell pellet were used to resuspend the pellet. The suspension was transferred into a beaker and stirred at 4°C for 20 minutes. This mixture was poured into a small centrifuge tube and sonicated 6 times on ice for 30 seconds at 50% power. The soluble and insoluble parts were separated by centrifugation at 12,000 rpm, 4 °C for 20 minutes. The pellet was resuspended a second time with the same amount of lysis buffer, followed by sonication and centrifugation. The remaining pellet was resuspended using solubilisation buffer to solubilise inclusion bodies. The suspension was sonicated as described before, then stirred for 2 hours at room temperature before centrifugation. The supernatants of each step were transferred into a 15 mL Falcon and stored at -80°C or at 4°C until further usage. Additionally, a sample for SDS-PAGE was collected after every centrifugation step.

2.2.3. Biochemical methods

2.2.3.1. Pulse renaturation

Recombinant proteins of high molecular weight often accumulate in inclusion bodies and have to be refolded to gain a functional protein. One method to refold insoluble proteins is the pulse-renaturation technique (Rudolph and Lilie, 1996). Briefly, solubilised inclusion bodies are quickly diluted by adding them directly to renaturation buffer. This step was repeated 8 times every hour. Correctly folded proteins are removed from the reaction and cannot aggregate anymore. The aimed final concentration was approximately 1 mg/mL for optimal yield without protein precipitation. Refolded proteins were dialysed against the buffer used during following protein enrichment steps. The formula below was used to calculate the exact volume of protein added per pulse.

 $volume of \ each \ pulse = \frac{0.125 \ mg/mL \ x \ volume \ of \ renaturation \ buffer}{protein \ concentration}$

The inclusion bodies had to be dialysed against solubilisation buffer without DTT for optimal renaturation and the renaturation buffer had to be adjusted for every protein as well, depending on the presence of disulphide bonds in the refolded protein and its isoelectric point. The pH value of the renaturation buffer should be at least one pH value above the isoelectric point of the protein. Arginine is always included in the buffer, varying from 0.5–1 M and acts

as a mild denaturing agent. Cysteines in the protein cause the formation of interchain disulphide bonds, which reduce solubility in the absence of reducing agents. The addition of reduced and oxidised glutathione paired with a mildly alkaline milieu (pH 8-8.5) allows a thiol-disulphide exchange, which leads to the reduction of these interchain disulphide bonds and the formation of correct and energetically favourable disulphide bonds. If there are no cysteines in the protein, a redox system does not have to be included in the renaturation buffer.

The renaturation of PvMSP-1 and PbMSP-1 required an estimation of the protein concentration of the two halves. The solubilised inclusion bodies of both halves were run on a SDS gel at 3-4 dilutions (1:25, 1:50, 1:100, 1:200) to determine the ratio of 83/30 to 38/42. The exact protein concentrations were measured via Biuret assay.

2.2.3.2. Purification of His-tagged proteins

Recombinant proteins that are fused to an isolation tag can be purified by affinity chromatography. The MSP-1 processing fragments were all expressed with an N-terminal 6xHis-tag, which generally allows enrichment using Ni-NTA. All steps of this protocol were performed at cool temperatures to limit protein degradation. 1 mL of 50% Ni-NTA slurry was used for 4 mL lysate or refolded protein. The appropriate amount of Ni-NTA was equilibrated with lysis buffer and 10 mM imidazole. The mixture was gently inverted and centrifuged at 1200 rpm for 3 minutes to remove the buffer. Imidazole was added to the lysate to a final concentration of 10 mM to prevent binding of contaminants to the Ni-NTA. The mixture was gently shaken on a rotary shaker for 60 minutes at 4°C. Unbound proteins were removed by a centrifugation step at 1,200 rpm for 3 minutes at 4°C. A washing step with wash buffer (same volume as volume of used lysate) containing 20 mM imidazole was used to remove loosely bound proteins. The mixture was poured into an empty column with a frit at the bottom. The Ni-NTA is retained in the column, while the buffer flows through. The proteins were eluted in two steps. First, 4 mL of elution buffer containing 50 mM imidazole was added, followed by a stronger elution with 250 mM imidazole. The eluted proteins were stored at 4°C or -80°C.

If the Ni-NTA was intended to be reused for the same protein, it was washed with 30 CVs 0.5 M sodium hydroxide and 10 CVs 30% ethanol directly after the purification procedure and stored in the fridge.

2.2.3.3. Ion-exchange chromatography using Q-Sepharose

The first purification step after refolding recombinant MSP-1 from its two halves is Ion-exchange chromatography, using Q-Sepharose as an anion exchanger. Therefore, the renatured protein was dialysed against buffer D.1 to remove arginine from the renaturation buffer and charge the protein negatively. The negative charge is a result of the buffer's pH being at least one unit above the isoelectric point of the protein, causing deprotonation of the protein.

The volume of the Q-sepharose matrix was adjusted to the measured protein concentration in the load. 1 mL of Q-sepharose has a binding capacity of 10 mg. The matrix was initially washed with ddH₂O, equilibrated with buffer D.1 and then activated with 5 CV of buffer D.3 (1 M NaCl). All buffers used for the ÄKTA system were filtered and degassed. The column was equilibrated with buffer D.1 again before loading the protein. Loading of the protein was performed using a 150 mL loop and a flow rate of 0.8 mL/min. The protein was eluted directly with 75 mM, 100 mM and 150 mM NaCl by mixing buffer D.1 and D.2 appropriately. Remaining proteins were eluted from the column with buffer D.3. After each purification, the column was cleaned with H₂O. If another protein was purified next, the column was regenerated with 1 M NaOH for 1 hour before rinsing the column with H₂O again. All ÄKTA columns were stored in 20% ethanol.

2.2.3.4. Size-exclusion chromatography using a Superdex 200 column

Size-exclusion chromatography (SEC) was used as an additional purification step after ion-exchange chromatography. With this technique, individual proteins can be separated based on their hydrodynamic radius. Briefly, large molecules will pass the matrix faster than smaller ones and therefore elute earlier. All size-exclusion chromatographies were performed in filtered, degassed PBS with a constant flow rate of 1.4 mL/min. The protein sample was loaded using a 6 mL sample loop, which was directly filled using a 10 mL syringe. Samples were collected when the UV rises on the spectrum. The eluted protein samples were analysed via SDS-PAGE before concentrating the desired fractions using Amicon Ultra centrifugal filters.

2.2.3.5. Purification of inclusion bodies using Mono S/Mono Q

The production workflow of recombinant *Plasmodium vivax* MSP-1 was optimised using Mono S and Mono Q columns to increase the final yield. In this case, the inclusion bodies were solubilised in 8 M Urea and purified further before the MSP-1 heterodimer was assembled.

The buffer of the solubilised inclusion bodies was exchanged for the Mono Q buffer A or Mono S buffer A for PfPv83/30 and Pv38/42, respectively. Approximately 20 mL solubilised inclusion bodies were loaded onto the respective column. For PfPv83/30 the Mono Q 10/10 column was used and the Mono S 10/10 column for Pv38/42. A gradient elution was performed with Mono Q buffer B or Mono S buffer B, with both containing 500 mM NaCl. The purified inclusion bodies of both halves did not require further purification and were used for direct buffer exchange into the refolding buffer to reconstitute the MSP-1 heterodimer.

2.2.3.6. Desalting using prepacked columns

Desalting of a protein solutions can be necessary when salts, amino acids or additives in the buffer are interfering with the next step in protein purification or with a specific assay. Small volumes up to 2.5 mL were desalted using commercial PD10 columns, whereas two HiPrep 26/10 columns were connected in series to perform a buffer exchange with larger volumes. Both column types are filled with a porous sepharose resin, which allows separation of proteins with a larger hydrodynamic radius from those with a smaller one, identical to a size exclusion chromatography but with lower resolution. As a result, the proteins are passing through the column faster than any salt or additive in the buffer. In this project, this principle was frequently used to remove arginine from lyophilised protein samples or after the refolding step.

Before desalting, the columns were equilibrated with the desired buffer. In case of the PD10 columns, at least 25 mL of buffer has to be used to remove any UV-absorbing stabilisers, which are included in the storage solution. Next, the protein solution was loaded onto the column, followed by 3.5 mL of buffer to elute the protein in fractions à 500μ L. The collected fractions were measured at 280 nm using the BioTek reader to determine samples containing the highest protein concentrations. The HiPrep26/10 column was used with the ÄKTA system, allowing a more precise collection of eluted protein according to the UV spectrum.

2.2.3.7. Analytic size-exclusion chromatography

Analytic size-exclusion chromatography was used to analyse the homogeneity of the sample and to determine the hydrodynamic radius, which was used in combination with the sedimentation coefficient to calculate the molecular weight of the recombinant protein. As proteins vary in size and shape, analytical size-exclusion chromatography should not be used to directly determine the molecular weight.

The column volume of the used Superdex 200 Increase 10/300 GL column was 24 mL. The dead volume of 8 mL was identified using blue dextran. The High Molecular Weight Calibration Kit (see Table 14) was used to gain a spectrum matching elution volumes to specific known hydrodynamic radiuses. A flow rate of 0.6 mL/min was used for the calibration kit and the analyte. PBS was used as a standard buffer for these experiments. The hydrodynamic radius of MSP-1 was determined by matching the elution volume to those of the marker proteins in the calibration kit.

Table 14	Proteins	included in	the H	High	Molecular	Weight	Calibration	Kit	and the	ir elution	volume	using	the
Superdex	200 Incre	ease 10/300	GL.										

Protein	Molecular weight	Stoke's radius	Elution volume
Thyroglobulin	669 kDa	85.0	8.8 mL
Ferritin	440 kDa	61.0	10.1 mL
Aldolase	158 kDa	48.1	12.4 mL
Conalbumin	75 kDa	NA	13.8 mL
Ovalbumin	44 kDa	30.5	14.9 mL

2.2.3.8. Protein quantification using Bradford Assay

The Bradford assay is a commonly used colorimetric test to quantify protein in solution. This method is based on the binding of Coomassie brilliant blue G-250 to protein, which causes an absorbance shift in acidic conditions. The unbound red dye is stabilised by binding to protein and turns blue. This increase in absorbance is highest at 595 nm and is proportional to the protein concentration. High concentrations of commonly used detergents, such as SDS or guanidine hydrochloride interfere with this reaction. Thus, if proteins in guanidine hydrochloride buffer have to be measured, the Biuret assay is a more accurate alternative. The Bradford assay is linear over a range of 0 µg/mL to 2 mg/mL. Higher concentrated samples have to be diluted for precise measurements.

The standard protocol for such a measurement is to dilute the Bradford reagent 1:5 in water. For each protein, four different concentrations are prepared containing either 2.5 μ L, 5 μ L, 10 μ L or 20 μ L of the original sample in Milli-Q water to a total volume of 20 μ L. 1 mL of the diluted dye is added to each of these dilutions and incubated for 20 minutes. The OD₅₉₅ is measured using a photometer and the diluted dye as blank. The exact protein concentrations are calculated using a standard curve created with BSA dilutions. The standard curve used for this assay resulted in the following equation for direct determination of protein concentrations:

concentration
$$\left(\frac{mg}{mL}\right) = \frac{\frac{OD_{595}}{0.0583}}{sample \ volume} \times optional \ predilution$$

2.2.3.9. Protein quantification using Biuret Assay

The Biuret assay is a method of protein quantification, which is commonly used when the protein solution contains substances that cause incorrect results in the Bradford assay. This chemical test is based on the formation of violet-coloured complexes in the presence of peptide bonds in an alkaline solution. 10, 25 and 50 μ L of the protein solution were mixed with 1 mL of H₂O and precipitated with 0.5 mL 50% TCA, followed by a 10 min centrifugation step at 10,000 g. The precipitate is resolved in 2 mL Biuret reagent, which contains NaOH, hydrated copper(II) sulphate and potassium sodium tartrate. The nitrogen atoms of the peptide bonds react with the cupric ions from the Biuret reagent in alkaline conditions and form a chelate complex, which cause the violet colour of the solution. The colour intensity is concentration-dependent and was measured at 546 nm.

The protein concentration was calculated for each individual measurement using the following equation. For further experiments, the average of the three obtained values was considered as the correct protein concentration.

$$concentration \left(\frac{mg}{mL}\right) = \frac{OD_{546} \times 19 \times 2}{sample \ volume \ \times 5}$$

2.2.3.10. Dialysis

Dialysis was used to transfer a protein into a different buffer to remove a reducing agent, such as DTT. The protein solution was transferred into a dialysis membrane with a molecular weight cut-off (MWCO) lower than the molecular weight of the protein. The MWCO represents the molecular weight of the protein that is 90% retained in the membrane. In case of PvMSP-1 and its subunits, a dialysis membrane with a MWCO of 12-14 kDa was used. The dialysis membrane was filled with the protein solution, closed with dialysis clips and placed into new buffer for at least 4 hours or overnight. The remaining concentration of the substance that is to be removed can be calculated by multiplying the initial concentration with the sample volume, divided by the volume of dialysis buffer.

A faster way to exchange the buffer is dialysis via centrifugation using Amicon Ultra Centrifugal filters. Usually a MWCO of 30,000 Da was used. Centrifugation was performed at 4°C, with a maximum speed of 4,000 g when using a swinging bucket rotor or 5,000 g when using a fixed angle rotor. The samples were immediately removed from the centrifugal filters after centrifugation for optimal protein recovery.

2.2.3.11. SDS polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was frequently used to analyse protein samples during and after protein expression. In general, self-poured 10% gels were used unless otherwise stated. Samples were prepared using the SDS-PAGE sample buffer (4x) with or without DTT and boiled for 10 minutes at 80°C. Usually, 7 μ L of protein ladder and 10 μ L of protein sample were loaded into the pockets of the gel. Initially, the gel was run at 120 V, then at 200 V when the samples had reached the separating gel. After the run, the gels were stained with Coomassie for at least 25 minutes, followed by shaking in destaining solution.

2.2.3.12. Western Blot analysis

Protein samples were separated on a 10% SDS gel before transferring the proteins onto a PVDF membrane via wet transfer (125 mA, 2 hours). The membrane was blocked with 3% milk powder in TBST for 1 hour at room temperature. Both the primary and secondary antibody were diluted in TBST with 3% milk powder. Usually, an anti-His antibody was used at a dilution of 1:2,000. Secondary antibodies are often more sensitive and can be used at a higher dilution of 1:20,000. Incubation with the primary antibody was done overnight in the cold room. Next, the membrane was washed three times with TBST for 10 minutes on a shaker. The membrane was incubated with the secondary antibody for 1 hour at room temperature. The used secondary antibody was conjugated with alkaline phosphatase to allow development of the Western Blot without a machine after adding BCIP as a substrate. One BCIP tablet was dissolved in 10 mL H₂O to obtain a substrate solution. After washing the membrane as described previously, the substrate was added to the membrane. The reaction was observed and stopped before the bands were overexposed or at the latest after 30 minutes by washing the membrane with water.

2.2.3.13. Enzyme-linked immunosorbent assay (ELISA)

Recombinant proteins were diluted in coating buffer at a final concentration of 100 nM. A 96 well plate was coated using 100 μ L of the diluted protein, covered with parafilm and incubated at 4°C overnight. The plate was washed twice with 200 μ L TBST per well before adding 200 μ L of blocking buffer. After incubation for 1 hour at room temperature, the plate was washed again with 200 μ L TBST per well. 100 μ L of blocking buffer were added into the wells. The sera from Brazilian volunteers were diluted in blocking buffer (1:100 if the plate was coated with PvMSP-1, Pv83 or Pv42 and 1:25 if the plate was coated with Pv30 or Pv38). 100 μ L of the prediluted sera were added to the wells and mixed with the previously added blocking buffer. All sera were measured in duplicates. The sera were incubated for 2 hours at room temperature before washing the plate 4 times with TBST. The secondary antibody, anti-human IgG coupled to alkaline phosphatase, was diluted 1:30,000 in blocking buffer and 100 μ L were added to each well followed by incubation for 1 hour at room temperature. The plate was washed twice with TBST and twice with substrate buffer. 4-Nitrophenyl phosphate disodium salt hexahydrate tablets were used a substrate. One tablet was dissolved in 20 mL of substrate buffer. 100 μ L of substrate solution were added to each well and incubated for another hour at room temperature in the dark. The reaction was stopped by addition of 0.2 M NaOH (100 μ L/well) before measuring the absorption at 405 nm.

2.2.3.14. PfSUB1 cleavage assay

The following procedure is based on a publication by (Koussis *et al.*, 2009) and was used to verify whether MSP-1 proteins from other *Plasmodium* strains can be processed by the subtilisin-like protease 1 (SUB1) from *P. falciparum* in a similar fashion than PfMSP-1.

The protein to be analysed was dialysed into SUB1 buffer to create optimal conditions for the protease. The used enzyme PfSUB1 was kindly provided by Michael J. Blackman. After dialysis, the sample was centrifuged to remove any precipitates. The protein concentration was measured and adjusted to 0.25 mg/mL with SUB1 buffer. The total volume required for the assay was 350 µL. The protease inhibitors listed in **Table 15** were added to make sure that the protein will not be cleaved by any other enzymes than PfSUB1.

Table 15. Protease inhibitors used in the PfSUB1	cleavage assay.
--	-----------------

Inhibitor	concentration	function
E64	10 µM	irreversible cysteine protease inhibitor
Leupeptin	10 μg/mL	cysteine and serine protease inhibitor
Pefabloc	100 μg/mL	irreversible serine protease inhibitor

1.5 units of PfSUB1 were added per 100 μ L of sample volume. The mixture was divided into 6 aliquots of 50 μ L each. All samples were incubated at 16 °C, but varying incubation times were applied. Two controls were incubated overnight, one without PfSUB1 and one without protease inhibitors. The remaining six samples were incubated for 0 min, 30 min, 1 hour, 3 hours, 6 hours and overnight. After each time point, the samples were centrifuged for 2 min at 10,000 g. The supernatant was transferred into a new Eppendorf tube before stopping the reaction by boiling for 10 min at 95 °C in 1x SDS sample buffer and 4.5 μ L DTT. As a final step, all samples were analysed via SDS-PAGE.

2.2.3.15. Analytical ultracentrifugation

A sedimentation velocity run can be used to determine the Svedberg coefficient (s) and homogeneity of a sample. The Svedberg constant can then be used in combination with the hydrodynamic radius determined by size exclusion chromatography to calculate the molecular weight.

The protein solution was centrifuged at 14,000 rpm for 10 min to remove most of the aggregates. The sample cell was filled with 500 μ L PBS (left side) and 450 μ L MSP-1 (right side). Multiple scans were performed at 42,000 rpm, 20°C. The absorbance was measured at 280 nm.

2.2.3.16. Circular Dichroism (CD) spectroscopy

Circular dichroism is the difference in absorbance of left- and right-handed circularly polarised light by an optically active molecule. 19 out of 20 amino acids are chirally active and as a consequence, all proteins are as well and can therefore be studied using this technique. The method used in this study involves thermal degradation, meaning CD spectra were recorded at different temperatures (10-85 °C) to study the structure and unfolding temperature of MSP-1.

The analyte was dialysed in 10 mM phosphate buffer, pH 7, to remove any light-absorbing substances, like salts from PBS or arginine phosphate, which had been added during the lyophilisation process in case of MSP-1D. The spectrometer was flooded with nitrogen 15 min before the measurement. The quartz cuvette was rinsed with H₂O, then with phosphate buffer. 150 µL of protein solution were filled into the cuvette and the spectrum was recorded from 190- 250 nm. A temperature/wavelength scan was performed to follow the unfolding of the protein at 10°C, 20°C, 30°C, 37°C, 40°C, 42°C, 50 °C, 60°C, 70°C and 85°C. After the measurement, the cuvette was cleaned using 10% SDS and concentrated nitric acid.

Individual CD spectra at each temperature were created with Prism. The melting temperature was determined by fitting the data using Prism.

2.2.3.17. Cross-linking

Cross-linking combined with mass spectrometry analysis was used to identify interactions within MSP-1. Initial experiments were performed using the bissulfosuccinimidyl suberate (BS³) cross-linker, whereas follow-up experiments used disuccinimidyl sulfoxide (DSSO) to facilitate data analysis. The optimal cross-linker concentration was titrated using MSP-1D, which contains 211 lysine residues. Due to the high arginine content in the lyophilised preparation, 150 µg MSP-1D was first dissolved in 150 µL Milli-Q water, followed by a buffer exchange into PBS to remove arginine before adding the cross-linker. Stock solutions of each cross-linker were freshly prepared according to the manufacturer's instructions. The suitable amount of cross-linker was determined by analysing cross-linked MSP-1D via SDS-PAGE. In the end, a 2.5-fold molar excess of BS³ and a 20-fold molar excess of DSSO were used to cross-link MSP-1 proteins. When MSP-1 preparations were already in PBS, a buffer exchange was not necessary for the cross-linking reaction. MSP-1 was incubated with BS³ for 30 minutes and with DSSO for 60 minutes at room temperature before stopping the reaction with quenching buffer. The trypsin digest and mass spectrometry analysis was performed by the mass spectrometry core facilities in the ZMBH and DKFZ.

3. Results

3.1. Design of synthetic MSP-1 gene

The prerequisite for structural and immunological analysis of *P. vivax* MSP-1 was obtaining sufficient amounts of protein. Direct extraction from the *Plasmodium* parasite was not an option, as the process itself only resulted in low yields using *P. falciparum*, but even more so as *P. vivax* is difficult to culture long-term (Thomson-Luque *et al.*, 2017). Alternatively, recombinant protein production techniques can be used to acquire the needed material. In case of *Plasmodium* MSP-1, *E. coli* was chosen as a suitable expression system as native MSP-1 is not glycosylated (Berhe *et al.*, 2000), which would otherwise have required a mammalian expression system. Protein expression in *E. coli* usually offers a high protein yield combined with low costs and is less labour-intensive than other expression systems.

P. vivax research so far focused mostly on the Salvador I strain, resulting in its genome being sequenced (Carlton, Adams, *et al.*, 2008) and the establishment of functional assays using this strain, such as a sporozoite challenge in *Aotus* monkeys for testing antimalarial vaccines (Jordán-Villegas *et al.*, 2005). The MSP-1 sequence of the Salvador I strain was also chosen for the development of a protein production process in this project (see Uniprot entry A5K724). The sequence identity with the second most common strain *P. vivax* Belem is 84% (NCBI protein blast, (Altschul *et al.*, 1997)) which leads to the assumption that a recombinant protein based on the Salvador I sequence might resemble the Belem variant very closely.

The AT-rich nucleotide sequences found in *Plasmodium* have often caused severe difficulties for cloning techniques (Gardner *et al.*, 2002). To overcome this and also enable the use of *P. vivax msp-1* in adenoviral vectors as a vaccine, the *msp-1* gene was codon optimised for humans before ordering it as a synthetic gene. While the AT-content in *Plasmodium vivax msp-1* is considerably lower than in *P. falciparum*, the AT content could still be reduced from 57% to 46%. A further modification of the native *msp-1* gene was the removal of the N-terminal signal peptide and the C-terminal glycosylphosphatidylinositol (GPI) anchor signal sequence as both were not required for the planned analysis.

Since the production process for the *P. falciparum* protein was already established, the approach for the *P. vivax* variant was based on the corresponding patent by Prof. Hermann Bujard (Bujard *et al.*, 2011) and uses the pZE expression plasmids previously developed by (Lutz and Bujard, 1997). It is important to note that this approach uses known processing sites to split the MSP-1 protein into two halves, which are expressed separately and refolded together at a later step to reconstitute the full length protein. The base for this method is the processing of MSP-1 by SUB1 and the corresponding cleavage sites can be seen in **Figure 12**. While sequences matching the SUB1 cleavage motif can be found in all MSP-1 variants, the exact positions are sometimes shifted, resulting in processing fragments of varying length. However, in

this thesis, the nomenclature of the *P. falciparum* processing fragments is used for all MSP-1 subunits and the *Plasmodium* species is indicated (e.g. Pv83 for the *P. vivax* 83 fragment) to simplify the comparison of structural and immunological data.



Figure 12. Alignment of PfSUB1 cleavage sites within the MSP-1 sequence from various *Plasmodium* strains. Predictions were made using the PfSUB1 model and the PoPs online tool. Cleavage sites resulting in four processing fragments were found in all analysed sequences (highlighted in yellow). However, the exact positions of these cleavage sites are shifted by up to ~230 amino acids (e.g. p38/42 cleavage sites), resulting in processing fragments of varying length. All protein sequences were originally retrieved from the Uniprot database.

3.2. Construction of pZE expression plasmids

3.2.1. Cloning of expression plasmids for His-tagged PvMSP-1 subunits

The focus of this project is on the structural analysis of the full length protein. Nevertheless, some experiments ask for recombinant proteins representing only a specific MSP-1 processing fragment. The protein sequence of *P. vivax* MSP-1 was analysed using the *P. falciparum* SUB1 model (de Monerri *et al.*, 2011) and the Prediction of Protease specificity (PoPs) online tool (Boyd *et al.*, 2005). The nucleotide sequences corresponding to predicted processing fragments were amplified by PCR using primers containing ClaI and PstI restriction sites. All obtained PCR products were subcloned into TOPO vectors for easier amplification and digestion with restriction enzymes before ligation with the pZE13 expression plasmid backbone. These plasmids contain an ampicillin resistance marker, an N-terminal 6xHis-tag to facilitate protein purification and use the IPTG inducible $P_{A1lacO-1}$ promoter (Lutz and Bujard, 1997) (Figure 13).



Figure 13. Map of a pZE13 plasmid used for expression of *P. vivax* **MSP-1 processing fragments.** The shown map represents the plasmid used for the *P. vivax* 83 processing fragment and comprises an N-terminal 6xHis-tag and an ampicillin resistance marker. The insert was ligated with the plasmid backbone following a classical cloning strategy using the restriction enzymes Clal and Pstl. PA1lacO-1: LacR-regulated promoter (Lutz and Bujard, 1997).

3.2.2. Expression plasmids for full length PvMSP-1

Obtaining the full length *P. vivax* MSP-1 protein was of paramount importance for this study. This protein was not only to be used for acquiring structural data, but could also be used as an antimalarial vaccine targeting the so-far neglected *P. vivax* parasite. A requirement for this was the production of a recombinant protein without any purification tags, such as a His-tag, as demanded by regulatory agencies. This process is based on the previously patented protein production protocol (Bujard *et al.*, 2011) and includes splitting the MSP-1 protein sequence into an N-terminal and C-terminal half according to the predicted cleavage site between the Pv30 and Pv38 fragment. The resulting sequences are named Pv83/30 and Pv38/42. Both were amplified via PCR using primers with BamHI and ClaI restriction sites to facilitate insertion into pZE23 expression plasmids. The pZE23 plasmids differ from their pZE13 counterparts by including a kanamycin resistance marker instead of the ampicillin one and the lack of a purification tag (Figure 14).



Figure 14. Map of a pZE23 expression plasmid used for expression of the two halves of the MSP-1 heterodimer. The plasmid carries a kanamycin resistance cassette and can be used for expression of recombinant proteins without a purification tag. The gene of interest is inserted using the restriction sites BamHI and Clal. Shown here is the expression plasmid for Pv38/42. The expression plasmid for Pv83/30 is identical except for the coding sequence (blue) which is replaced accordingly.

The pZE23 expression plasmid achieved high expression levels of Pv38/42, but Pv83/30 showed similar difficulties as previously observed with *P. falciparum* 83/30 in our laboratory. In case of *P. falciparum* 83/30, these issues were overcome by targeted mutation of the N-terminal nucleotide sequence to remove hairpin structures obstructing ribosomal binding. Analysis of the Pv83/30 mRNA structure revealed some hairpin formation which could be overcome by addition of an N-terminal tag or replacement with an already optimised nucleotide sequence. The first 22 amino acids of *P. falciparum* and *P. vivax* MSP-1 have a sequence identity of 65% (Figure 15), leading to the assumption that a replacement would not alter the *P. vivax* protein in a severe manner. Consequently, the pZE23-PfPv83/30 plasmid was generated by replacement of 66 nucleotides at the N-terminal end.



Figure 15. Alignment of the N-terminal amino acid sequence of the 83/30 half of *P. falciparum* and *P. vivax*. The depicted sequences correspond to the first 22 amino acids of MSP-1 in both *Plasmodium* species. The signal peptide found in the wildtype protein has been excluded. The direct alignment only reveals 6 direct matches, but the overall sequence identity was determined as 65% by protein blast.

3.2.3. Design of P. berghei msp-1 expression plasmids

The goal of this project was to obtain precise structural information of MSP-1. In addition to the *P. vivax* variant, MSP-1 from other parasite species such as *P. knowlesi* and *P. falciparum* were available for analysis. However, including a *Plasmodium* species that solely infects rodents but is well studied could improve our understanding of consistent features within MSP-1. For this purpose, a protein production protocol was developed for *P. berghei* ANKA MSP-1 (PbMSP-1) based on the previously mentioned method. Here, the full length protein was considered sufficient and expression of PbMSP-1 processing fragments was omitted.

P. berghei MSP-1 was analysed with the PoPs tool (Figure 12) and coding sequences corresponding to the N-terminal Pb83/30 and C-terminal Pb38/42 halves were cloned into pZE23 expression plasmids as described in 3.2.2. As a future use of PbMSP-1 in vaccination experiments using mice was not excluded, the synthetic gene was codon optimised for mice to allow usage in adenoviral vectors. This optimisation also decreased the AT content of the original nucleotide sequence from 75% to 45%.

Initially, the coding sequences of the PbMSP-1 halves were cloned into pZE23 expression plasmids as described for the *P. vivax* protein in 3.2.2. Whereas good expression levels were observed for Pb38/42, the pZE23-Pb83/30 expression plasmid showed similar issues as previously noticed for *P. vivax* and *P. falciparum* 83/30. As the *P. berghei* MSP-1 was never intended to be used as a human vaccine, the addition of an N-terminal His-tag was chosen as the easiest solution to overcome these expression difficulties. Therefore, a pZE13 expression plasmid was used for Pb83/30 instead and a factor Xa cleavage site was added via PCR to allow His-tag removal if required (Figure 16).



Figure 16. Plasmid map of pZE13-HXPb83/30. As expression of the N-terminal half of *P. berghei* MSP-1 without a purification tag was difficult but still desirable, a pZE13 plasmid was designed which includes a removable His-tag. The plasmid's backbone is identical to previously shown pZE13 plasmids, except for the addition of a factor Xa cleavage site between the His-tag and the protein sequence.

3.3. Expression of recombinant P. vivax and P. berghei MSP-1

All previously described pZE plasmids were cloned using the *E. coli* strain Dh5alphaZ1, but as this strain has a low growth rate, the faster growing W3110Z1 strain was more suitable for large scale production of recombinant MSP-1. All constructs and halves of MSP-1 were expressed separately in standard LB medium at 37°C. Expression of His-tagged processing fragments using pZE13 plasmids was successfully induced after addition of IPTG (Figure 17) followed by further incubation at 37°C for 3 hours.



Figure 17. IPTG-inducible expression of *P. vivax* **processing fragments using pZE13 plasmids.** All four processing fragments were successfully produced after addition of IPTG. First lane: before IPTG addition, second lane: after IPTG addition. The calculated molecular weights are 83 kDa (Pv83), 31 kDa (Pv30), 41 kDa (Pv38) and 44 kDa (Pv42). *P. vivax* MSP-1 has a high proline content, which influences the running behaviour on SDS gels. This is especially noticeable with the Pv30 protein, which runs close to the 60 kDa marker instead of the 30 kDa one.

Obtaining full length MSP-1 revealed to be more difficult than the processing fragments. The 83/30 proteins from *P. vivax* and *P. berghei* could not be expressed using the standard pZE23 plasmid (data not shown). While both could be produced after addition of an N-terminal His-tag, this approach was not ideal as proteins with an isolation tag cannot be used as a vaccine for humans. Fortunately, replacement of the first 22 amino acids as described in 3.2.2 was sufficient to achieve protein expression in W3110Z1 under standard conditions (Figure 18).

The PvMSP-1 processing fragments Pv30, Pv38 and the PfPv83/30 half migrate on an SDS gel as if they have a higher molecular weight. This is probably due to the very high proline content of *P. vivax* MSP-1. As prolines cause kinks and structural rigidity, the electrophoretic mobility of proteins with a high proline content is decreased (Hames, 1998). This altered migration behaviour is most noticeable for Pv30, a protein consisting of 287 amino acids out of which 23 are prolines (8%). The full length *P. vivax* MSP-1 is made of 1,751 amino acids, including 73 prolines (4.2%).



Figure 18. The two halves of PvMSP-1 were expressed in W3110Z1 after addition of IPTG. Expression of the N-terminal half was achieved by replacing the first 22 amino acids with the already optimised nucleotide sequence for *P. falciparum* 83/30, resulting in the PfPv83/30 protein. Pv38/42 shows good expression without modification of the original nucleotide sequence. The expected molecular weights are 110 kDa (PfPv83/30) and 82 kDa (Pv38/42).

P. berghei msp-1 showed a similar protein expression behaviour as observed for its *P. vivax* and *P. falciparum* counterparts. The Pb38/42 half can be produced without an isolation tag or any modification using a pZE23 expression plasmid. However, the Pb83/30 half showed no expression using the same plasmid backbone. Inserting the coding sequence into a pZE13 plasmid enabled production of Pb83/30 with an N-terminal His-tag. Nevertheless, in comparison to the noninduced sample, the induced one contains additional bands at approximately 140 kDa, 120 kDa and 90 kDa. This leads to the conclusion that although Pb83/30 can be expressed, it is prone to degradation or premature termination of translation, which severely limits the obtainable yield.



Figure 19. Recombinant protein expression of the two halves of *P. berghei* **MSP-1.** Pb83/30 was expressed with an N-terminal 6xHis-tag and a factor Xa cleavage site using a pZE13 expression plasmid. The C-terminal half Pb38/42 was produced using a pZE23 expression plasmid with no additional tags. Molecular weights were predicted as 118 kDa for Pb83/30 and 79 kDa for Pb38/42. The induced sample of the Pb38/42 culture contains one strong band. But Pb83/30 seems to be prone to degradation as the induced sample contains additional bands at ~140 kDa, 120 kDa and 90 kDa compared to the noninduced sample.

3.4. Purification of P. vivax MSP-1

The primary goal of this project was to elucidate the structure of MSP-1 using recombinant proteins. Since our research group's focus has been on vaccine development, a secondary goal of this project was to develop a production protocol for *P. vivax* MSP-1, which not only supplies enough material for biophysical and biochemical experiments, but can also be applied for large scale production meeting Good manufacturing practice (GMP) standards.

The initial approach for *P. vivax* MSP-1 purification was based on the patented purification protocol for *P. falciparum* MSP-1 (Bujard *et al.*, 2011). Due to the fact that these two proteins don't share the same amino acid sequence, they also behave slightly differently during protein purification. Therefore, adjustments were made to certain buffer compositions to improve the final yield of *P. vivax* MSP-1. Technical progress since the development of the patented procedure also allowed the replacement of a few cost- and labour-intensive steps with faster or more efficient alternatives.

High expression levels of heterologous proteins in *E. coli* often lead to accumulation of insoluble proteins in form of inclusion bodies (Singh *et al.*, 2015). These cytoplasmic aggregates can be easily separated from soluble proteins but need to be refolded *in vitro* to obtain recombinant proteins similar or ideally identical to the native form. The PvMSP-1 processing fragments Pv30 and Pv38 were expressed as soluble proteins whereas Pv83, Pv42 and the two halves of the full length protein were insoluble. What is usually an obstacle in protein production was utilised to reconstitute the complete MSP-1 heterodimer using two denatured halves and refolding them together. The scheme in **Figure 20** shows the basic steps of the PvMSP-1 protecin is refolded and purified using a combination of ion exchange and size exclusion chromatography.


Figure 20. Scheme of recombinant MSP-1 production including reconstitution of the heterodimer and protein enrichment using ion exchange and size exclusion chromatography. The two halves of MSP-1 are expressed separately and inclusion bodies are solubilised in buffer containing guanidine hydrochloride and DTT. The heterodimer is renatured in refolding buffer containing arginine and a glutathione redox system. Refolded MSP-1 is purified by ion-exchange chromatography using Q-sepharose after arginine has been removed by dialysis. Remaining contaminants or degradation products are separated from MSP-1 during size exclusion chromatography (SEC). The protein solution can be concentrated between purification procedures and/or after the final chromatography step.

After protein expression, bacterial cells were lysed by sonication using pre-cooled PBS-based lysis buffer containing lysozyme, DNase and protease inhibitors. Lysozyme breaks down bacterial cell walls in addition to the mechanical sonication and DNase is used to degrade nucleic acids to improve the viscosity of the lysate and prevent interference with chromatography steps. Handling of bacterial lysates was performed at low temperatures and in presence of protease inhibitors to decrease protein degradation by endogenous proteases from *E. coli*. Soluble and insoluble proteins were separated by centrifugation before a second round of sonication was performed. Finally, inclusion bodies were solubilised in 6 M guanidine hydrochloride buffer with dithiothreitol (DTT), pH 8. Usage of the chaotropic agent guanidine hydrochloride in high concentrations results in the complete denaturation of proteins, while DTT reduces disulphide bonds. The goal of this step is to unfold the insoluble proteins entirely as a base for renaturation without any folding intermediates found in inclusion bodies. A control of this cell rupture and preparation of solubilised inclusion bodies can be seen in **Figure 21** for the two halves of *P. vivax* MSP-1. Both proteins are insoluble and can only be seen in the inclusion bodies samples.



Figure 21. Both halves of *P. vivax* **MSP-1 accumulate in inclusion bodies.** Bacterial cells were lysed twice by sonication to separate soluble from insoluble proteins. In the last step, insoluble proteins that had aggregated in inclusion bodies were solubilised. The inclusion bodies samples sometimes show an altered running behaviour on SDS gels due to the high content of guanidine hydrochloride. Samples were analysed in dilutions of 1:10 and 1:20 or 1:20 and 1:50. The molecular weights are 110 kDa (PfPv83/30) and 82 kDa (Pv38/42). SN: supernatant containing soluble proteins, IB: inclusion bodies.

Renaturation of recombinant MSP-1 follows the so-called "pulse renaturation" principle (Rudolph and Lilie, 1996). Here, denatured proteins are rapidly diluted in a buffer containing arginine and reduced and oxidised glutathione to reduce precipitation of folding intermediates. Arginine is a commonly used additive in renaturation procedures as it prevents aggregation of proteins. Glutathione functions as a redox system, which improves the formation of disulphide bonds during the refolding process. MSP-1 contains 6 disulphide bonds at the C-terminal end, so using a redox system was crucial for renaturing the full length protein.

The solubilised inclusion bodies were dialysed against 6 M guanidine hydrochloride, 2 mM EDTA, 50 mM NaPO₄, pH 3 to remove DTT as this reducing agent would disturb the refolding process. The low pH prevents formation of intra- and intermolecular disulphide bonds before the denatured proteins are added to refolding buffer. The exact composition of the refolding buffer was adjusted for each recombinant protein. *P. vivax* MSP-1 required 1 M arginine, 100 mM Tris, 1 mM EDTA, 10 mM reduced glutathione, 1 mM oxidised glutathione and a pH of 8.5 for optimal refolding. As MSP-1 is reconstituted from two separate proteins during the renaturation step, the protein concentration of the two halves was determined beforehand by Biuret assay and SDS-PAGE to make sure both halves are added in equal amounts. The name "pulse renaturation" refers to the way the solubilised inclusion bodies are added to a 100-fold higher volume of renaturation buffer every hour. In case of MSP-1, a maximum protein concentration of 0.12 mg/mL per step was used. After 10 steps, a concentration of 1.2 mg/mL was reached and the mixture was incubated for 12-16 hours at 4°C before precipitated protein was removed by centrifugation.

3.4.1. PvMSP-1 enrichment using ion-exchange and size exclusion chromatography

The previously refolded P. vivax MSP-1 was purified using a combination of the most commonly used purification methods for tagless proteins, ion exchange and size exclusion chromatography. The principle of ion exchange chromatography is that ionisable groups bind to a stationary phase and can be eluted by displacement with mobile ions of higher affinity. Thus, the success of this purification technique relies on the pH of the used buffer as the net charge of a protein is influenced by the pH value. A pH too close to the isoelectric point of a protein results in no net charge and only weak binding to the matrix. Apart from choosing the wrong pH for ion exchange chromatography, the presence of arginine can also hinder binding to the column. Accordingly, P. vivax MSP-1 was dialysed against buffer D.1 (18.9 mM Na₂HPO₄, 1.1 mM NaH₂PO₄, 50 mM NaCl, pH 8) to remove arginine and to change the net charge of the protein. The isoelectric point of *P. vivax* MSP-1 is 6.6, so it should be negatively charged in a buffer of pH 8 and be able to bind to an anion exchanger. Q-sepharose is a strong anion exchanger consisting of cross-linked 6% agarose beads, with quaternary ammonium (Q) strong anion exchange groups. The matrix was packed into a column suitable for the ÄKTA system and activated with buffer D.3, which contains 1 M NaCl before equilibrating it with buffer D.1. PvMSP-1 was loaded onto the column with a 150 mL sample loop and a flow rate of 0.5 mL/min. Initially, a gradient elution was performed to find the optimal settings for washing and elution. This method uses the ÄKTA system to gradually mix buffer D.1 and buffer D.2, which contains 300 mM NaCl, to slowly increase the salt concentration. But as PvMSP-1 already eluted at 75 mM NaCl, a washing step was omitted and a stepwise elution protocol was established using 75, 100 and 150 mM NaCl. Fractions were collected as soon as the UV spectrum rose significantly and samples were analysed on SDS gel. Buffer D.3 was used as a final washing step to remove all remaining proteins from the column.



Figure 22. *P. vivax* **MSP-1 purification using Q-sepharose.** A) The OD280 and conductivity spectrum was recorded during the purification. The numbers above individual peaks correspond to the samples in the SDS gel. B) SDS gel of all collected fractions to identify fractions containing PvMSP-1. Refolded PvMSP-1 in D.1 buffer containing 50 mM NaCl bound to the anion exchanger. However, it was already eluting at 75 mM NaCl (sample 1), suggesting only weak binding. Removal of excess halves, which had not refolded to form the full length MSP-1 was achieved in two steps. First, excess Pv8330 did not bind to the column and can be seen in the flow through sample. Second, Pv3842 bound strongly to the column and eluted at higher concentrations than the reconstituted heterodimer. The two halves can be seen as two separate bands on SDS gels. The Pv8330 half runs at ~140 kDa and the Pv3842 half at ~90 kDa. L: load, FT: flow through, numbers refer to fractions collected.

Full length MSP-1 separates again into the two halves on denaturing SDS-gels resulting in two bands at ~140 kDa and ~90 kDa. Suitable elutions were pooled and dialysed against PBS for storage or a second purification step if needed.

Here, the elution with 75 mM NaCl was the purest and required no further purification. However, some preparations contained more contaminants of lower molecular weight that can be eliminated by size exclusion chromatography. This technique is also known as gel filtration and is one of the mildest and simplest chromatography methods. Proteins are separated based on their hydrodynamic radius with larger molecules eluting earlier than smaller molecules. There is no binding to the matrix and buffer composition does not directly affect the resolution. To simplify the process, all size exclusion chromatographies were performed with PBS because it is a versatile buffer for storage and following experiments. PvMSP-1 preparations from ion exchange purifications were concentrated to a maximum concentration of 1.2 mg/mL. The solution was filled into a 6 mL sample loop via direct injection and loaded onto a Superdex 200 prep grade 26/600 column, which consists of a composite matrix of cross-linked agarose and dextran. The run was performed with a constant flow rate of 1.4 mL/min. Fractions were collected according to the OD280 curve and analysed on SDS gel.



Figure 23. Size exclusion chromatography can separate PvMSP-1 from smaller contaminants. A) Selected elutions from the Q-sepharose purification were concentrated and loaded onto a Superdex 200 prep grade column. The UV spectrum shows two significant peaks. Numbers indicate collected fractions, which were further analysed by SDS-PAGE. B) Analysis on SDS gel reveals that PvMSP-1 can be found in the first peak, whereas other proteins of lower molecular weight can be found in the second peak. The Pv8330 half runs at ~140 kDa and the Pv3842 half at ~90 kDa. L: load.

Typically, the UV spectrum showed two peaks (Figure 23). According to SDS-PAGE, PvMSP-1 elutes in the first peak and other contaminating proteins elute in the second peak. While this technique works well to separate PvMSP-1 from considerably smaller degradation products or *E. coli* proteins, contaminating proteins too close to the hydrodynamic radius of PvMSP-1 cannot be removed. Protein samples are diluted during size exclusion chromatography, so PvMSP-1 containing fractions were concentrated to ~1 mg/mL before storage.

3.4.2. Optimised PvMSP-1 purification using Mono S/Mono Q columns

The previously described purification procedure for *P. vivax* MSP-1 was suitable to produce protein preparations of high purity. However, the rather laborious protocol resulted in a low yield, making it both time- and material-consuming to obtain sufficient material for further analysis. The cause of the low yield is not found in low expression levels, but in protein precipitation during refolding and dialysis, and also during the chromatography steps as PvMSP-1 cannot be completely recovered from the column. Therefore, a new approach was developed together with our collaboration partners Dr. Dirk Görlich and Jürgen Schünemann from the Max Plank Institute in Göttingen.



Figure 24. New approach for the production of *P. vivax* **MSP-1.** Inclusion bodies of both halves are solubilised in 8 M urea and further purified using Mono Q or Mono S columns. These matrixes are strong ion exchangers and offer a high resolution. After these individual purification steps, both halves are mixed in a ratio of 1:1. The buffer is replaced with refolding buffer using a desalting column and the solution is incubated overnight at 4°C. Next, the buffer is exchanged with PBS or any other buffer required for downstream experiments. Size exclusion chromatography (SEC) can be performed if further purification is required.

The major difference between the new and the previously established protocol is the purification of solubilised inclusion bodies individually before reconstitution of the heterodimer (**Figure 24**). The goal is to use highly pure inclusion bodies so no degradation products or contaminants are present during the refolding step. Apart from creating optimal conditions for refolding, this technique would also make any further purification steps after the renaturation obsolete. This is especially beneficial as most protein had precipitated or bound too strongly to the chromatography resins after the refolding step in the previous protocol.

In comparison to the Q-sepharose fast flow matrix, Mono Q and Mono S columns provide an even higher resolution. These chromatography resins consist of monodispersed 10 µm porous beads substituted with quaternary ammonium (Q) or methyl sulfonate (S) groups, making them strong ion exchangers.

As a preparation for purification, inclusion bodies were solubilised in 8 M urea instead of 6 M guanidine hydrochloride as the latter would behave similarly to NaCl at the pH values used for ion exchange chromatography, resulting in reduced binding or premature elution of the protein.



Figure 25. Preparation of inclusion bodies for purification with Mono Q or Mono S columns. In contrast to previous cell rupture procedures, inclusion bodies were solubilised using 8 M urea instead of guanidine hydrochloride. PfPv83/30 runs at ~130 kDa, while Pv38/42 can be seen at ~90 kDa. - : noninduced culture, + : induced culture, BL: bacterial lysate, SN: supernatant containing soluble proteins, IB: inclusion bodies.

As before, the halves PfPv83/30 and Pv38/42 were expressed separately and accumulated in inclusion bodies. These were solubilised in 8 M urea, 20 mM NaPO₄, 5 mM DTT, pH 6.5 for cation exchange chromatography using a Mono S column (Figure 25). Although both proteins display similar expression levels, the solubilised inclusion bodies of Pv83/30 contained less contaminating proteins than those of Pv38/42. The isoelectric points of the individual halves are 6.5 and 6.6 for Pv38/42 and PfPv83/30, respectively. In theory, both proteins should have only minimal to no net charge at the used standard buffer for cation exchange with Mono S. Nevertheless, Pv38/42 did bind to the column while the main contaminant did not. 25 mL of solubilised inclusion bodies were loaded onto a Mono S 10/10 column and eluted using a salt gradient with a final concentration of 500 mM NaCl. Fractions were collected automatically using the ÄKTA system at increased OD280 values. Similar to observations with PvMSP-1 and the Q-sepharose resin, Pv38/42 starts to elute at >60 mM NaCl. Fortunately, most contaminating proteins did not bind to the Mono S column and the majority of those that had bound eluted at a higher salt concentration (Figure 26).



Figure 26. Purification of Pv38/42 inclusion bodies using a Mono S column. A) OD280 and conductivity spectrum as recorded during the chromatography run. 25 mL of inclusion bodies were loaded. Bound proteins were eluted using a NaCl gradient with a maximum concentration of 500 mM (%B). Significant peaks were observed after 138 and 144 mL matching samples 4-8 on the SDS gel. B) The SDS gel of collected fractions shows a good separation of Pv38/42 (runs at ~90 kDa) and other proteins. The main contaminant did not bind to the Mono S column and is found in the flow through. The majority of Pv38/42 elutes in samples 3-8. L: load, FT: flow through, numbers refer to collected fractions.

Cation exchange using Mono S did not work for PfPv83/30 as the protein did not bind the column. An alternative with identical resolution is using the anion exchanger Mono Q. The inclusion bodies were solubilised in 8 M urea, 20 mM Tris, 5 mM DTT, pH 8. As the pH value is higher than PfPv83/30's isoelectric point, the protein's net charge should be negative allowing it to bind to the positively charged Mono Q resin. 25 mL of solubilised inclusion bodies were loaded onto a Mono Q 10/10 column followed by a gradient elution with 500 mM NaCl. All proteins bound to the resin, so the binding alone had no purification effect as seen for Pv38/42. However, the majority of contaminating proteins eluted at different salt concentrations than PfPv83/30 and could be removed by strict fractionization (**Figure 27**). Fractions 4–9 contain mostly PfPv83/30 and were used to reconstitute the *P. vivax* MSP-1 heterodimer.



Figure 27. Purification of PfPv83/30 inclusion bodies using Mono Q columns. A) OD280 and conductivity were monitored during the purification. 25 mL of Inclusion bodies in 8 M urea buffer were loaded onto the column. Elution was performed using a salt gradient from 0 – 500 mM NaCl. The OD280 spectrum shows several peaks, with the highest peak occurring at 72 mL. B) Collected fractions were analysed on an SDS gel. Almost all proteins bound to the column. According to the OD280 curve and the SDS gel analysis, PfPv83/30 started to elute from the column at a NaCl concentration of 170 mM (see peak at 72 mL and samples 4-8 on the SDS gel). Most contaminating proteins eluted at other salt concentrations. PfPv83/30 runs at 120 kDa. L: load, FT: flow through, numbers refer to collected fractions.

Both purified inclusion bodies had a concentration of 1.2 mg/mL and were mixed in a ratio of 1:1 for refolding. At this point, the mixed halves added up to a volume of 94 mL, which exceeded the capacity of the pulse renaturation technique. Instead, the urea buffer was directly replaced with renaturation buffer (1 M arginine, 100 mM Tris, 1 mM EDTA, 10 mM GSH and 1 mM GSSG, pH 8.5) using a HiPrep 26/10 desalting column. The solution had a concentration of 1.1 mg/mL and was incubated at 4°C overnight. Precipitated protein was removed via centrifugation.

A sample of the refolded protein was analysed on a Superdex 200 Increase 10/300 column to check if both halves elute together. This column is generally used for analytical purposes and provides a much higher resolution than the preparative size exclusion columns. 300 μ L of refolded protein were injected onto the column and analysed at a constant flow rate. Samples were collected in 500 μ L fractions and loaded on an SDS-gel. As seen in **Figure 28**, both halves elute together despite minor differences in their individual concentrations. This suggests that not all halves have bound to their respective partner, but the main proportion is expected to have reconstituted the *P. vivax* MSP-1 heterodimer.



Figure 28. Test for PvMSP-1 heterodimer formation using size exclusion chromatography. The refolded protein was analysed using a Superdex 200 Increase 10/300 column which offers a high resolution. The aim of this test is to see if the two halves elute together or separately. If the two halves had not bound to each other, PfPv83/30 would elute earlier than Pv38/42. While there are some minor differences when comparing the PfPv83/30 and Pv38/42 bands within each sample, no half completely runs without the other, suggesting that the main proportion is present as a heterodimer. L: load.

This PvMSP-1 preparation was lyophilised in arginine buffer without any additional buffer exchange because tests with lower arginine concentrations or PBS did not show an improved lyophilisation behaviour. Following experiments also require different buffers, so storing the protein in the refolding buffer offers a higher flexibility. The lyophilised PvMSP-1 is stored at -80°C.

The final yield using this protocol was 90 mg of *P. vivax* MSP-1 from 31 g *E. coli* wet paste. This corresponds to 2.9 mg per 1 g wet paste. In comparison, the previous approach yielded only 1 mg from 6 g wet paste, thus 0.17 mg per 1 g wet paste. The wet paste values always refer to the sum of both halves and not each half individually.

3.4.3. Purification of P. vivax MSP-1 processing subunits

Purification tags offer the possibility of a quick and simple protein enrichment using affinity chromatography. All *P. vivax* MSP-1 processing fragments were expressed with an N-terminal 6x His-tag, which is one of the most common isolation tags and allows binding of recombinant proteins to nickel-charged affinity resins. Pv30 and Pv38 were produced as soluble proteins whereas Pv83 and Pv42 were insoluble and required refolding prior to purification (Figure 29).



Figure 29. Recombinant proteins representing the *P. vivax* processing fragments differ in their solubility. Cells were lysed by sonication and soluble and insoluble proteins were separated by centrifugation. Insoluble proteins were solubilised in 6 M guanidine hydrochloride. The proteins of lower molecular weight, namely Pv30 and Pv38 are found as soluble and insoluble proteins (see B and C), whereas those of higher molecular weight, Pv42 and Pv83, accumulate in inclusion bodies (A and D). Pv83 runs at 83 kDa, Pv30 at ~55 kDa, Pv38 at 50 kDa and Pv42 at 41 kDa. Discrepancies to the theoretical molecular weight can be explained by high proline content found in these proteins. SN: supernatant containing soluble proteins, IB: inclusion bodies.Pv30 and Pv38 were purified from the first supernatant

obtained after cell rupture. A total concentration of 10 mM imidazole was added to the bacterial lysate to reduce unspecific binding to the affinity resin Ni-NTA. This material consists of nitrilotriacetic acid (NTA), a chelating ligand, coupled to a cross-linked agarose resin and is precharged with nickel. 1 mL of Ni-NTA slurry was used per 4 mL of bacterial lysate using the batch method. This means that Ni-NTA and lysate or buffers were incubated in a falcon tube and separated by centrifugation instead of using a column for gravity flow chromatography. The Ni-NTA and lysate were mixed in a 50 mL falcon tube, rolled for one hour to bind the recombinant protein to the matrix followed by centrifugation. Unspecifically bound proteins were removed using a wash buffer (50 mM NaH₂PO₄, 300 mM NaCl, pH 8) with 20 mM imidazole. Finally, His-tagged proteins were eluted with 250 mM imidazole. All steps were performed at 4°C to limit protein degradation.

Pv42 could be purified following the same protocol after the protein had been refolded using the "pulse renaturation" technique as described previously. Samples of each purification step were analysed on SDS gel (Figure 30). Attempts to purify Pv42 in a denatured state using Ni-NTA were unsuccessful (data not shown).



Figure 30. Three out of four *P. vivax* MSP-1 processing fragments were successfully purified using Ni-NTA. The 6xHis-tagged PvMSP-1 fragments Pv30, Pv38 and Pv42 bound to the Ni-NTA matrix. However, most of Pv30 remained in the flow through, which can be caused by exceeding the binding capacity of the applied Ni-NTA. This effect can also be observed for Pv38 and Pv42 to a lesser extent. The Ni-NTA matrix was washed with 20 mM imidazole to remove loosely bound proteins and impurities. Finally, proteins of interest were eluted using 250 mM imidazole. L: load, FT: flow through, W: wash, E: eluate.

Three out of four PvMSP-1 processing fragments were successfully enriched using Ni-NTA. However, not all of the His-tagged proteins bound to the Ni-NTA and were found in the flow through, as can be seen in **Figure 30**. This could mean that the binding capacity of Ni-NTA was exceeded in these cases. Unfortunately, some contaminants remain in all eluates, likely caused by histidine rich proteins found in the lysate or degradation products. The high concentration of imidazole was removed by dialysis against PBS before storing the proteins at -80°C.

The final yield of Pv30 and Pv38 was 0.75 mg per gram *E. coli* wet paste, although this value could be increased by additional purification using the second bacterial lysate and refolding the protein that accumulated in inclusion bodies as these proteins were expressed as both soluble and insoluble proteins (**Figure 29**). The final yield of Pv42 after refolding the protein and enrichment by Ni-NTA affinity chromatography was 1.6 mg per gram *E. coli* wet paste.

Pv83 was the most challenging to purify of all *P. vivax* MSP-1 processing fragments. Similar to Pv42, Pv83 was refolded in 500 mM arginine buffer with 5 mM DTT, pH 8.5 following the previously described "pulse renaturation" technique. The majority of the refolded protein did not bind to the Ni-NTA and was found in the flow through despite using generous amounts of Ni-NTA resin. The remaining Pv83 that had bound to the Ni-NTA eluted equally during the washing and elution steps and the final eluate was also highly contaminated with other proteins of lower molecular weight (data not shown). While these contaminants could represent degradation products, obtaining a purer preparation was desirable. Additional washing steps with lauroyl-sarcosine did not lead to the improved purity, so instead of pursuing the affinity chromatography approach further, a combination of ion exchange chromatography and size exclusion chromatography was tested for Pv83.

Refolded Pv83 was dialysed against buffer D.1 (18.9 mM Na₂HPO₄, 1.1 mM NaH₂PO₄, 20 mM NaCl, pH 8.5) to prepare the protein for ion exchange chromatography. The isoelectric point of Pv83 is 7.3, meaning a buffer with a pH of 8.5 changes the net charge of Pv83 so that it becomes negatively charged and can bind to a positively charged medium, such as Q-sepharose. A low amount of NaCl was added to the buffer to reduce unspecific binding to the matrix. Ion Exchange chromatography was performed using the ÄKTA system with a low flow rate (0.5 mL/min) during the binding step. Afterwards, the column was washed with different NaCl concentrations starting from 50 mM to 1 M NaCl to elute the protein. Fractions were collected when a peak was observed in the UV spectrum and samples were analysed on SDS gel (Figure 31). Most of Pv83 bound to the column. The little amount visible in the flow through can be further reduced by using more Q-sepharose matrix. The elution with 75 mM NaCl is the purest elution in this purification. Using 100–150 mM NaCl elutes remaining Pv83 but also other proteins of 40 kDa from the column. The 1 M NaCl elution removes most still bound proteins from the column but is usually not included in further purification steps as it contains a lot of contaminants besides Pv83.



Figure 31. Purification of Pv83 using ion exchange chromatography. A) Purification of refolded Pv83 using Q-sepharose and a NaCl gradient to elute the protein from the matrix. Conductivity and OD280 were monitored during the purification. Numbers indicate collected fractions, which were analysed via SDS-PAGE. B) Most of the protein bound to the Q-sepharose resin as only a little amount can be seen in the flow through sample. But not all of Pv83 could be eluted with up to 150 mM NaCl (samples 1-4) as the final wash with 1 M NaCl (sample 5) still contains Pv83. The calculated molecular weight for Pv83 is 83 kDa. L: load, FT: flow through.

The collected fractions ranging from 75 mM to 150 mM NaCl were pooled and purified further using size exclusion chromatography with a Superdex 200 column (total column volume = 320 mL). The protein solution was concentrated from 0.1 mg/mL to 1.3 mg/mL in a total volume of 6 mL before directly injecting it into a sample loop. The entire run was performed using PBS and a flow rate of 1.4 mL/min. The 280 nm UV spectrum was observed during the run and revealed two separate peaks. Fractions were collected and analysed on SDS gel (**Figure 32**). The first sample contains the majority of Pv83 and only low amounts of contaminants. Most of the proteins around 40 kDa were found in the second sample and were therefore successfully separated from Pv83. The final yield of Pv83 was 3 mg per gram *E. coli* wet paste.



Figure 32. Additional size exclusion chromatography to purify Pv83. Left: Selected elutions from the ion exchange chromatography were pooled, concentrated and loaded onto a Superdex 200 column to remove contaminants of lower molecular weight by size exclusion. PBS was used for equilibration and performing the run. The 280 nm spectrum shows one small peak at around 100 mL and two bigger peaks ranging from 150 to 250 mL. Samples were collected from the two main peaks and analysed on SDS gel. Right: SDS-PAGE analysis of collected samples. Pv83 was mainly found in the first sample, referring to the first peak in the UV spectrum. The main portion of contaminants was in the second sample and could be separated from Pv83. L: load.

3.5. Purification of P. berghei MSP-1

The production of recombinant *P. berghei* MSP-1 followed the same protocol as previously explained for *P. vivax* MSP-1 (Figure 20) and includes protein refolding prior to ion exchange and size exclusion chromatography purifications. After expression of both halves, bacterial cultures were lysed as described in 3.4. Again, both halves accumulated in inclusion bodies and were solubilised in 6 M guanidine hydrochloride buffer with dithiothreitol (DTT), pH 8 (Figure 33). The theoretical molecular weights of the two halves are 118 kDa (Pb83/30) and 79 kDa (Pb38/42). However, Pb83/30 contains 75 prolines, which corresponds to 7% of its amino acids and influences its running behaviour on SDS gels in a similar manner as observed for *P. vivax* MSP-1 proteins.

Another difficulty of Pb83/30 purification becomes visible in SDS gels as seen in Figure 19 and Figure 33 as the samples do not only contain one strong band matching the calculated molecular weight or a protein of apparent higher molecular weight. Instead, the induced samples contain bands running at approximately 140 kDa, 120 kDa and 90 kDa, which were not found in the noninduced sample. This observation leads to the assumption that Pb83/30 is prone to degradation during protein expression.



Figure 33. Cell rupture of cultures expressing Pb83/30 and Pb38/42. *E. coli* cells were lysed by sonication. Soluble and insoluble proteins were separated by centrifugation. After washing the bacterial pellet twice with lysis buffer, inclusion bodies were solubilised using 6 M guanidine hydrochloride. Both halves of *P. berghei* MSP-1 are expressed as insoluble proteins. The inclusion bodies of Pb83/30 contain various bands that could represent full length Pb83/30 or degradation products. Samples were analysed in dilutions of 1:10 and 1:20 or 1:20 and 1:50. The molecular weights were calculated as 118 kDa (Pb83/30) and 79 kDa (Pb38/42). SN: supernatant containing soluble proteins, IB: inclusion bodies.

The protein concentrations were estimated by SDS-PAGE and Biuret assay measurements, but the high amount of degraded protein makes it difficult to determine the exact concentration of intact Pb83/30 for refolding with Pb38/42. Refolding of *P. berghei* MSP-1 was performed using 500 mM arginine, 50 mM NaH₂PO₄, 1 mM EDTA, 10 mM reduced glutathione and 1 mM oxidised glutathione, pH 8 according to the pulse renaturation method. A ratio of 1:1 was aimed for to reconstitute the PbMSP-1 heterodimer, however due to the high amount of presumable degradation products, it cannot be excluded that some Pb38/42 bound to partially degraded Pb83/30. Attempts to enrich intact Pb83/30 using the N-terminal His-tag and Ni-NTA did not improve the purity of the inclusion bodies as the contaminants had bound to the column as well and eluted at the same imidazole concentration.

The theoretical isoelectric point of PbMSP-1 is 5.8, so dialysis against D.1 buffer (18.9 mM Na₂HPO₄, 1.1 mM NaH₂PO₄, 50 mM NaCl, pH 8) should result in a negatively charged protein that can be purified using anion exchange chromatography. Additionally, arginine from the refolding step was removed and precipitated protein was separated from the soluble protein by centrifugation.

The protein was loaded onto a self-packed Q-sepharose ÄKTA column using a 150 mL sample loop and a flow rate of 0.8 mL/min. Elution was performed using a salt gradient from 50 mM to 300 mM NaCl to achieve a better separation compared to stepwise elution. PbMSP-1 started to elute at a salt concentration of 160 mM. Finally, the column was washed with 1 M NaCl to remove any remaining bound proteins. Samples were analysed on SDS-gel to select suitable PbMSP-1 containing fractions for further purification (Figure 34). The load contains the four bands described as Pb83/30 and presumable degradation products

thereof. However, eluted fractions seem to contain only one band matching Pb83/30 at approximately 140 kDa and the Pb38/42 band at 80 kDa in addition to contaminating proteins visible at lower molecular weights (labelled 3-8 in Figure 34).



Figure 34. *P. berghei* **MSP-1 purification using Q-sepharose.** A) This spectrum including OD280 and conductivity was recorded during ion exchange chromatography and reveals one pronounced peak after increasing the NaCl concentration. B) According to the SDS gel, all of the loaded protein bound to the Q-sepharose resin. Increasing the NaCl concentration gradually eluted PbMSP-1 with contaminating proteins of approximately 50 kDa. However, only one band matching Pb83/30 can be found in the eluates, meaning other degradation products still present in the load were eliminated. Pb83/30 runs at 140 kDa, Pb3842 at 80 kDa. L: load, FT: flow through, numbers refer to collected fractions.

Fractions 3 to 8 were pooled and concentrated to 1.4 mg/mL for size exclusion chromatography. The aim of this purification step was the removal of contaminating proteins found at approximately 50 kDa and below (Figure 34). 6 mL of concentrated protein solution were loaded on a Superdex 200 prep grade 26/600 column and separated at a constant flow rate of 2 mL/min using PBS. Fractions were collected according to the OD280 curve and analysed by SDS-PAGE. The OD280 curve shows three distinct peaks. The first peak does not seem to contain the protein of interest in meaningful concentrations as no bands

are visible in the SDS-gel. Bands matching the expectations for recombinant *P. berghei* MSP-1 are seen in sample 5 and 6 and also very weakly in sample 4. The highest peak in the UV spectrum was observed later and probably contains proteins of smaller hydrodynamic radius and lower molecular weight. Fractions 5 and 6 were concentrated to a final concentration of ~1 mg/mL and stored at -80°C. The final yield of *P. berghei* MSP-1 using this approach was 0.25 mg per gram *E. coli* wet paste.



Figure 35. Gel filtration of *P. berghei* **MSP-1 could remove most contaminants.** Eluates from the ion exchange chromatography were concentrated and loaded onto the size exclusion column. A) The OD280 spectrum includes three peaks. Several samples were collected mostly for the first two peaks and one for the third peak. B) Bands matching the expected molecular weight for *P. berghei* MSP-1 were found in sample 5 and 6. The load contains numerous contaminants, which are apparently not present in the fractions 5 and 6. However, running a protein solution over a size exclusion column always results in dilution of the sample. Hence, proteins whose bands were already weak in the load sample might simply not be detected by Coomassie staining in later samples anymore. The main contaminating proteins around 35-50 kDa were prominent in the load sample and cannot be seen in sample 5 and 6, suggesting successful removal of these contaminants. Pb83/30 runs at 140 kDa and Pb38/42 at 80 kDa. L: load, numbers refer to collected fractions.

3.6. Generation of a MSP-1 hybrid protein

Generally, refolding the MSP-1 heterodimer using two halves works well when both halves are derived from the same *Plasmodium* species. As halves of *P. falciparum* MSP-1 and *P. vivax* MSP-1 were available, it was also interesting to attempt creating a hybrid MSP-1 protein. If successful, this could be used as a vaccine to induce immune reactions against both parasite species instead of using both full length MSP-1 proteins in one formulation.

Here, solubilised inclusion bodies of PfPv83/30 and Pf38/42 were refolded in the same buffer as used for *P. vivax* MSP-1 (1 M arginine, 100 mM Tris, 1 mM EDTA, 10 mM reduced glutathione, 1 mM oxidised glutathione, pH 8.5) using the pulse renaturation method. After removing precipitated protein, the buffer was replaced with D.1 buffer using a HiPrep 26/10 desalting column. The benefit of this column is the efficient and controllable removal of arginine, which would hinder binding to the Q-sepharose resin.

50 mL of protein solution were purified using the anion exchange approach and a salt gradient for elution with stringent fractionation (Figure 36). The spectrum shows various peaks, suggesting a heterogenous mixture. Analysis via SDS-PAGE showed that first, PfPv83/30 elutes alone (samples 1-4). Then, there are samples which contain both halves, but also an excess of Pf38/42 (samples 6-9). The latter were pooled and further purified by gel filtration.



Figure 36. Potential hybrid protein enrichment via ion exchange chromatography. A) The hybrid protein was refolded using the PfPv83/30 half and the Pf38/42 half, followed by ion exchange chromatography. The OD280 curve rose as soon as the salt concentration was increased, suggesting only weak binding of some proteins. B) The SDS gel shows that a portion of PfPv83/30 elutes early, whereas Pf38/42 and the presumable hybrid complex elute at higher NaCl concentrations. In general, all collected samples contain only few contaminating proteins of lower molecular weight. Numbers refer to collected fractions.

In theory, size exclusion chromatography should be able to separate a potential hybrid complex and the excess of both individual halves if the resolution of the used column is high enough. The pooled fractions were concentrated to a volume of 5 mL and loaded onto a Superdex 200 prep grade 26/600 column using a constant flow rate of 3 mL/min and PBS as running buffer. The absorbance spectrum shows one peak with a plateau after 200 mL (Figure 37). The SDS-PAGE analysis revealed two bands representing both halves in equal intensity in the first 3 samples. Afterwards, more Pf38/42 is present than PfPv83/30 (sample 4). The retention volume of the first samples matches the values for other full length MSP-1 complexes, suggesting that these fractions might indeed contain a hybrid complex. However, additional analytical techniques were required to test proper complex formation.



Figure 37. Separation of excess halves and the potential hybrid protein using SEC. A) After ion exchange chromatography, suitable fractions were pooled and purified by gel filtration to remove protein halves that are not part of the MSP-1 heterodimer. The OD280 curve reveals one peak with a plateau after 200 mL. B) The first fraction collected during the purification was not analysed on SDS gel due to its low protein concentration. Samples 2 and 3, which eluted at a similar retention volume than other MSP-1 complexes, seem to contain PfPv83/30 and Pf38/42 in the same amount. Sample 5-7 include a more pronounced band of Pf38/42 and were therefore discarded. L: load, numbers refer to collected fractions.

3.7. Structural characterisation of Plasmodium MSP-1

3.7.1. Molecular weight determination of PvMSP-1

A combination of sedimentation velocity and size exclusion chromatography was used to determine if the previously described production protocol resulted in correctly assembled MSP-1 heterodimers.

This method of molecular weight calculation requires two experimentally determined variables, the Svedberg coefficient and the hydrodynamic radius. The Svedberg coefficient can be determined using a sedimentation velocity run, an analytical ultracentrifugation method that measures the rate at which molecules move in response to centrifugal force. This sedimentation rate is influenced by the mass and shape of the molecules. The hydrodynamic radius can be defined by analytical size exclusion using proteins with a known hydrodynamic radius as reference.

Sedimentation velocity runs were performed in collaboration with Dr. Norbert Mücke, from the DKFZ in Heidelberg. This process was first established using MSP-1D (GMP material). Lyophilised MSP-1D was solved in PBS and additionally dialysed against PBS to remove arginine that had been added during the lyophilisation process. The analytical ultracentrifugation was performed at 20°C and 42,000 rpm. Comparison of OD230 measurements at 3,000 rpm and 42,000 rpm revealed the presence of protein aggregates in the sample (Figure 38).



Figure 38. Analytical ultracentrifugation run of MSP-1D. The OD230 signal dropped when the speed was increased from 3,000 rpm to 42,000 rpm, suggesting that the sample contains protein aggregates. R is referring to the radius from the centre of the rotor to the sample; A in OD is the measured absorbance at 230 nm.



Figure 39. Plot of the sedimentation coefficient distribution against the Svedberg coefficient for MSP-1D. Three species were detected, with the species 1 (purple) representing 10% of the sample, species 3 (turquoise) represents ~15% of the sample and the main proportion being species 2 (green). The determined Svedberg coefficient is 7.36 S.

The plot of the measured sedimentation coefficient distribution against the Svedberg coefficient includes three species (**Figure 39**). One of them has a Svedberg coefficient of 3.9 S and contributes to 10% of the protein sample (purple peak). Another 15% of the sample contain another species with a sedimentation coefficient of 11.10 (turquoise). The green peak represents the majority of the sample. This species has a Svedberg coefficient of 7.36 S and was subsequently used for the molecular weight calculation.

The same method was used to analyse *P. vivax* MSP-1 (Figure 40). However, the sample was less homogenous as the previously analysed GMP material of MSP-1D. As with the previous run, three species were identified, but the proportions of species one was remarkably increased. Still, the majority of the PvMSP-1 sample consisted of a species with a Svedberg coefficient similar to the one of MSP-1D. As no major contaminations were found in the SDS gel, species one could represent free MSP-1 halves that have not bound their respective partner to form the heterodimer.



Figure 40. Analytical ultracentrifugation of *P. vivax* **MSP-1 revealed a less homogenous protein preparation**. Three species were detected, with the first and second contributing to the majority of the protein sample, whereas species 3 was detected in lower amounts. The main peak (green) corresponds to a similar Svedberg coefficient than determined for MSP-1D while species one could represent unbound MSP-1 halves.

The hydrodynamic radius was determined using an analytical Superdex 200 Increase 10/300 GL column and the High Molecular Weight Calibration Kit. The proteins of the kit and the MSP-1 samples were analysed in separate runs. The recorded spectra were overlaid with the Unicorn software (Figure 41). Comparison of the peaks from the calibration kit run and the recombinant proteins led to a hydrodynamic radius of 6.17 nm and 6.5 nm for MSP-1D and PvMSP-1, respectively.



Figure 41. Analytical size exclusion chromatography to determine the hydrodynamic radius of MSP-1 by comparison with the High Molecular Weight calibration kit. The proteins were analysed in PBS and an analytical Superdex 200 Increase column. MSP-1D is represented by the green curve, *P. vivax* MSP-1 by the blue one. The hydrodynamic radiuses of MSP-1D and PvMSP-1 were determined as 6.17 nm and 6.5 nm, respectively.

The molecular weight calculation based on the Svedberg coefficient and hydrodynamic radius are shown in **Figure 42**. The hydrodynamic radius is first used to define the diffusion constant (D) before calculating the molecular weight by dividing the Svedberg coefficient by the diffusion constant. In case of MSP-1D, the result was 189 kDa, which matches the values found in literature. For PvMSP-1, the same Svedberg coefficient was used as no significant difference was observed in the sedimentation velocity run for species 2 and no purer sample was available for additional measurements at the time. The result was a molecular weight of 199 kDa, which also corresponds to published values.



Figure 42. Molecular weight calculation based on hydrodynamic radius and Svedberg coefficient. The diffusion coefficient (D) is calculated by inserting the hydrodynamic radius as determined by analytical size exclusion chromatography into the Stokes-Einstein relation. The Svedberg coefficient is divided by the diffusion constant to calculate the molecular weight. This technique led to a molecular weight of 189 kDa for MSP-1D and 199 kDa for PvMSP-1.

3.7.2. Construction of phylogenetic tree and sequence identity

MSP-1 can be found in all *Plasmodium* species and some similarities but also differences already became apparent when the production process for recombinant MSP-1 proteins was adjusted in our laboratory. Bioinformatical analysis can complement data obtained from biophysical and biochemical assays to attain structural information. Here, a combination of multiple sequence alignment, protein blast and phylogenetic tree generation was used to highlight common characteristics of MSP-1.

Amino acid sequences of MSP-1 from *Plasmodium falciparum*, *P. vivax*, *P. berghei* and *P. knowlesi* were aligned using the CLUSTAL OMEGA multiple sequence alignment online tool (version 1.2.4). All sequences include an N-terminal signal peptide and a C-terminal GPI anchor signal sequence. Both of these were not present in the recombinant proteins used for subsequent experiments. The complete alignment can be found in the appendix (6.1). All six aligned sequences were used to create a phylogenetic tree by inferring evolutionary history using the neighbour-joining method (Saitou and Nei, 1987). The

evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. The method applied for the creation of this phylogenetic tree included partial deletion, which means all positions with less than 95% site coverage were excluded. The final dataset includes 1528 positions. The robustness of this phylogenetic tree was tested using the Bootstrap method with 500 replicates and showed a high reliability.

The created phylogenetic tree suggests an evolutionary split between the *Plasmodium* species infecting humans and *P. berghei*, the only parasite species not infecting humans included in this analysis (Figure 43). *P. knowlesi* MSP-1 shows a higher similarity with both *P. vivax* strains than with *P. falciparum*.



Figure 43. Evolutionary analysis based on MSP-1. The phylogenetic tree was created using the amino acid sequence of MSP-1 and the MEGA7 software (Kumar, Stecher and Tamura, 2016). Evolutionary history was modelled using the neighbour-joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch length = 1.77042031 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The Poisson correction model was used to calculate evolutionary distances (Zuckerkandl and Pauling, 1965), which are in the units of the number of amino acid substitutions per site. The analysis involved 6 amino acid sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 1528 positions in the final dataset.

0.10

Another indicator of molecular evolution is sequence identity. The same amino acid sequences as used for the phylogenetic tree were compared to each other by Protein Blast (NCBI protein blast, version 2.8.0, (Altschul *et al.*, 1997)). The computed sequence identities range from 36% to 84% (**Table 16**). The highest similarity was found between MSP-1 from *P. vivax* Salvador I and *P. vivax* Belem. MSP-1 from the two *Plasmodium* falciparum strains 3D7 and FCB-1 only show a sequence identity of approximately 40% when compared with *P. vivax* MSP-1. The amino acid sequence of *P. knowlesi* MSP-1 is closer to *P. vivax* than *P. falciparum* MSP-1 with ~63% vs. ~40% sequence identity. A comparison of *P. berghei* MSP-1 with other *Plasmodium* strains revealed a sequence identity value of ~36% for all of them.

Table 16. Sequence identities of MSP-1 from several *Plasmodium* strains according to Protein Blast by NCBI. The full amino acid sequences were analysed including the GPI anchor signal sequence and signal peptide. PvSal: *P. vivax* Salvador I, PvBelem: *P. vivax* Belem, PbANKA: *P. berghei* ANKA, Pf3D7: *P. falciparum* 3D7, PfFCB-1: *P. falciparum* FCB-1, Pk: P. knowlesi.

	PvSal1	PvBelem	PbANKA	Pf3D7	PfFCB-1	Pk
PvSal1		84%	36%	39%	40%	62%
PvBelem	84%		36%	38%	40%	64%
PbANKA	36%	36%		36%	34%	38%
Pf3D7	39%	38%	36%		59%	40%
PfFCB-1	40%	40%	34%	59%		39%
Pk	62%	64%	38%	40%	39%	

3.7.3. Processing by subtisilin-like protease 1

The proteolytic processing of MSP-1 by the subtilisin-like protease 1 (SUB1) is crucial for egress and invasion of red blood cells. It is assumed that all *Plasmodium* parasites express MSP-1 as a precursor, which is cleaved by a protease into four fragments. The amino acid sequences from six *Plasmodium* strains were analysed using the *P. falciparum* SUB1 model (de Monerri *et al.*, 2011) and the PoPs online tool (Boyd *et al.*, 2005). All analysed MSP-1 sequences contain at least three predicted cleavage sites. The exact locations of these motifs are not the same for all MSP-1 proteins, leading to fragments of varying length. The alignment in **Figure 44** includes the predicted cleavage sites with the highest scores. Other suggested sequences were regarded as alternative cleavage sites.

FCB-1	KTEGQ SDNS	701	EVSAN	DDTS	914	VVTGE	AVTP	1268
3D7	P <mark>LVAA SET</mark> T	726	QITGT	SSTS	914	VVTGE	AISV	1331
berghei_ANKA	V <mark>VTGE SEE</mark> T	738	TTRAE	SEED	1090	TILGQ	SEEY	1593
knowlesi	S <mark>LRGA SET</mark> A	699	TVQAS	SEEE	1067	PLFGN	DEDD	1431
vivax_Sall	S <mark>LRGA SET</mark> G	730	RVGGN	SEEK	1009	VTTGE	AESE	1359
vivax_Belem	L <mark>LRGS SEA</mark> A	713	RVGGN	SEEK	973	VTTGE	AESE	1335

Figure 44. Predicted SUB1 cleavage sites in different MSP-1 proteins. MSP-1 amino acid sequences of several *Plasmodium* strains were analysed using the PfSUB1 model and the PoPs online tool (Boyd *et al.*, 2005; de Monerri *et al.*, 2011). Predicted cleavage sites with the highest score were aligned and highlighted.

Recombinant proteins based on *P. vivax* Salvador I, *P. knowlesi* and *P. berghei* ANKA MSP-1 were incubated with PfSUB-1 to test if the proteolytic processing is indeed conserved. Samples were taken after 30 min, 1 hour, 3 hours, 6 hours and 24 hours and analysed by SDS-PAGE (Figure 45). In all cases, the two bands representing full length MSP-1 are gone in the overnight sample and only bands representing proteins of lower molecular weight remain. Bands presumably representing processing fragments are labelled in relation to their *P. falciparum* counterpart, irrespective of their actual molecular weight. In case of *P. knowlesi*, the calculated molecular weights of the predicted MSP-1 fragments are 79 kDa, 43 kDa, 40 kDa and 45 kDa for Pk83, Pk30, Pk38 and Pk42, respectively. For *P. berghei* values of 82 kDa (Pb83), 35 kDa (Pb30), 57 kDa (Pb38) and 23 kDa (Pb42) were calculated.



Figure 45. PfSUB1 assay with different MSP-1 proteins derived from *P. vivax*, *P. knowlesi* and *P. berghei*. A) Recombinant proteins were incubated with PfSUB1. Samples were taken at the start of the assay, after 30 min, 1 hour, 3 hours, 6 hours and 24 hours. The two bands representing full length MSP-1 disappear over time and cannot be found in the overnight sample anymore. Instead, bands of lower molecular weight appear, presumably representing the processing fragments (indicated with arrows). B) *P. knowlesi* MSP-1 processing fragments have a molecular weight of 79 kDa (Pk83), 43 kDa (Pk30), 40 kDa (Pk38) and 45 kDa (Pk42). Pk30 contains 43 prolines, leading to an altered running behaviour on SDS gels as observed for some *P. vivax* fragments. C) Processing of *P. berghei* MSP-1 should result in proteins of 82 kDa (Pb30), 35 kDa (Pb30), 57 kDa (Pb38) and 23 kDa (Pb42).

3.7.4. Secondary structure prediction

Most biophysical assays can only be used to obtain overall structural information and do not allow the identification of specific structures or lack thereof at precise locations. Bioinformatical tools can partially fill that gap and deliver location specific predictions based on a protein or nucleotide sequence. The JPred4 software (Drozdetskiy et al., 2015) uses a combination of the Inet algorithm (Cuff and Barton, 2000) and the LUPAS method to predict beta-sheets, alpha-helical and coil regions based on amino acid sequences (Lupas, 1997). The protein sequences of MSP-1 from P. falciparum 3D7, P. vivax Salvador I, P. berghei ANKA and P. knowlesi were analysed using default settings. On average, approximately 40% of the protein sequences were classified as alpha-helical and only a low percentage of 4-5% matched the criteria for a beta-sheet. Surprisingly, all proteins include large areas of unpredicted and therefore unstructured regions. Figure 46 shows an overview of MSP-1 with known conserved, dimorphic and oligomorphic regions of MSP-1D in white, grey and black, respectively. Cleavage sites used for proteolytic processing of the precursor protein are indicated with arrows and their position is given below. Unstructured regions were defined as stretches of >40 amino acids with no predicted secondary structure and are shown in the scheme in green. Notably, these unstructured regions can be found in similar positions in all Plasmodium species included in this comparison. The p30 processing fragment contains two, sometimes even three unstructured regions. In case of P. knowlesi, these regions are separated only by short coil regions. The precise position of these unstructured regions are given in Table 17.



Figure 46. Scheme of MSP-1 with predicted unstructured regions. The secondary structure of MSP-1 was computed using JPred4, a software that recognises alpha-helical, beta-sheet and coiled structures. The top part shows a previously published overview of MSP-1D with conserved, dimorphic and oligomorphic regions in white, grey and black (Kauth *et al.*, 2006). Arrows indicate cleavage sites used for proteolytic processing of the precursor protein, resulting in four fragments p83, p30, p38 and p42. Unpredicted regions are shown in green. SP: signal peptide, blue: glycosylphosphatidylinositol signal sequence (also labelled GPI).

Plasmodium strain	unstructured region no.	position in the amino acid sequence
P. falciparum 3D7	1	53 -138
	2	713 – 791
	3	904 – 994
	4	1247 – 1287
P. vivax Salvador I	1	207 – 287
	2	716 – 821
	3	944 – 982
	4	1255 – 1316
P. berghei	1	277 - 376
	2	680 - 833
	3	942 – 1121
	4	1340 – 1392
P. knowlesi	1	204 - 258
	2	691 - 857
	3	873 – 931
	4	946 - 1042
	5	1320 - 1403

Table 17. Exact position of unstructured regions within MSP-1 according to the secondary structure prediction.

3.7.5. Circular dichroism (CD) spectroscopy

Circular dichroism (CD) spectroscopy is a frequently used tool to check whether recombinant proteins are folded correctly. This method is based on circular dichroism, which is the difference in absorption of left- and right-handed circularly polarised light, meaning it can be used to study chirally active molecules, such as proteins and macromolecules. These measurements usually result in a characteristic spectrum, representing the secondary structure that is most common in the molecule. An a-helix causes negative bands at 208 and 222 nm and a positive band at 193 nm, while an antiparallel b-sheet can be recognized by negative bands at 218 nm and positive bands at 195 nm, making these two structures easy to distinguish.

Here, recombinant MSP-1 proteins were dialysed against 10 mM NaH₂PO₄ buffer (pH 7) to remove any absorbing salts or arginine from the lyophilisation process. Measurements were performed using a protein concentration of ~0.3 mg/mL over a wavelength range of 190 to 250 nm. Changes in these spectra were monitored while performing thermal denaturation by increasing the temperature from 10°C to 85°C.

The CD spectrum of MSP-1D represents a mostly alpha-helical secondary structure with negative bands at 208 nm and 222 nm and a positive band at 193 nm (Figure 47). The peak heights decrease at higher temperatures, as the protein begins to unfold. However, even at 85°C MSP-1D is not completely unfolded.



Figure 47. CD spectra of MSP-1D at different temperatures. The spectrum matches the typical CD spectrum for alpha-helical proteins with two negative bands at 208 and 222 nm and a positive band at 193 nm. The intensity of this spectrum decreases with rising temperatures, but even at very high temperatures, such as 85°C, the protein is not completely unfolded.

The melting temperature of MSP-1D was determined by measuring the CD signal at 222 nm during thermal denaturation. Scans were performed in 0.1°C intervals while the temperature was increased from 10°C to 85°C. The measured CD signal was plotted against the temperature and the melting temperature was calculated by applying a nonlinear fit in Prism (**Figure 48**). For MSP-1D, the melting temperature was determined as 45.6°C \pm 0.09 with an R² value of 0.99.



Figure 48. Fitted data from thermal denaturation of MSP-1D monitored by CD spectroscopy. Scans were performed at 222 nm in 0.1°C intervals while the temperature increased from 10°C to 85°C. The melting point was determined as 45.6°C \pm 0.09 by using a nonlinear fit in Prism.

Recombinant MSP-1 proteins derived from *P. vivax*, *P. knowlesi*, *P. falciparum* FCB-1 and a hybrid protein consisting of one 3D7 half and one FCB-1 half (hybrid D/F) were prepared and measured the same way as MSP-1D. All CD spectra show the typical shape of a mostly alpha-helical secondary structure (**Figure 49**). Different peak heights are caused by slight variations in protein concentration but also by different extinction coefficients of the MSP-1 proteins. Individual melting temperatures are shown in **Table 18**. The average melting temperature is 44°C. CD spectroscopy results for each measured protein can be found in the appendix (6.2).



Figure 49. CD spectra of all measured MSP-1 proteins. The CD measurements were performed individually for each recombinant MSP-1 protein at 20°C in phosphate buffer. Differences in relative peak heights are caused by slight variations in protein concentration. All CD spectra correspond to the typical spectrum of an alpha-helical protein.

Table 18. Melting temperatures of MSP-1 determined by CD spectroscopy.

Recombinant MSP-1 based on	melting temperature (in °C)	Standard error
3D7 (MSP-1D)	45.6	0.09
- processed MSP-1D	44.7	0.05
FCB-1	44.0	0.1
vivax	42.5	0.15
knowlesi	44.9	0.07
hybrid 3D7/FCB-1	42.9	0.1

All recombinant proteins used in this study were assembled from two halves. However, MSP-1 only occurs as a precursor protein or as a non-covalently associated complex composed of four fragments. To allow a comparison between such a complex and the other recombinant proteins, MSP-1D was incubated with PfSUB1 overnight and subsequently measured by CD spectroscopy. Successful proteolytic processing was confirmed by SDS-PAGE. The salts present in the SUB1 cleavage buffer would interfere with CD spectroscopy, so the buffer was exchanged with 10 mM NaH₂PO₄ buffer (pH 7) using a PD10 desalting column. The CD spectrum measured at low temperatures matches the one for unprocessed MSP-1D (Figure 50). However, the processed protein seems to be completely unfolded at temperatures higher than 42°C. The melting point was calculated as 44.7 °C.



Figure 50. CD spectra of MSP-1D after processing with PfSUB1. A) MSP-1D was incubated with PfSUB1 overnight. Successful proteolytic processing was confirmed by SDS-PAGE. B) After a buffer exchange to remove light-absorbing salts, CD measurements were performed as previously described. The recorded CD spectrum shows the positive band at 193 nm and two negative bands at 208 and 222 nm, which is characteristic for an alpha-helical structure. The CD spectra measured at 50°C or higher are flatter without any significant peaks, suggesting that the protein is mostly unfolded at these temperatures. The melting temperature was determined as 44.7°C by repeated CD signal measurements at 222 nm during thermal denaturation.

3.7.6. Crosslinking of MSP-1

Interactions within the MSP-1 complex were analysed by cross-linking combined with mass spectrometry analysis. Initial experiments were performed using the cross-linker BS³, which is a water-soluble homobifunctional *N*-hydroxysuccinimide ester (NHS) with a defined spacer arm. These esters form stable amide bonds with primary amino groups (-NH₂), which are present in the side chain of lysine residues and at the N-terminus of each polypeptide. The water-insoluble cross-linker disuccinimidyl sulfoxide (DSSO) follows the same principle, but can be cleaved in the gas phase during tandem mass spectrometry by collision-induced dissociation (Kao *et al.*, 2011). This cleavage facilitates data acquisition and analysis as the fragmentation patterns are unique to specific cross-linking types, allowing the identification of interlinks, intralinks or dead ends. The spacer arms of BS³ and DSSO are 11.4 and 10.3 Angström, respectively (Figure 51).



Figure 51. Structure of the cross-linkers BS³ **and DSSO**. A) The chemical structure of BS³ contains amine-reactive N-hydroxysuccinimide (NHS) esters at each side of a spacer of 11.4 Angström. The reactive groups (highlighted in grey) form stable bonds with primary amino groups present in side chains of lysine residues. B) DSSO includes of two symmetric collision-induced dissociation (CID)-cleavable sites that facilitate identification of crosslinked peptides based on a distinct fragmentation profile. The spacer length of DSSO is 10.3 Angström.

Cross-linking reactions with BS³ were performed using MSP-1D and MSP-1F and analysed at the mass spectrometry core facility in the ZMBH. Both *P. falciparum* derived MSP-1 proteins included a mixture of cross-links within processing fragments but also between them (**Figure 52**). Such interactions between processing fragments were found for p83 with p30, p38 and p42. P30 interacted only with p83 in these experiments, p38 only with p42 and p30, while p42 interacted with p38 and p83. Interestingly, no interactions involving the p19 portion of MSP-1 were found. All shown cross-links were manually checked in addition to software-based analysis of valid interactions.



Figure 52. Interactions within *P. falciparum* **MSP-1** proteins analysed by cross-linking with **BS**³. Recombinant MSP-1D and MSP-1F were incubated with a 2.5 fold molar excess of BS³ for 30 minutes. The cross-linking reaction was stopped with quenching buffer and samples were digested with trypsin before mass spectrometry analysis. According to this data, p83 interacts with all processing fragments of MSP-1. P30 interacts with p83 only, p38 binds to p83 and p42. The C-terminal p42 binds to p38 and p83 but not to p30. No interactions involving the very C-terminal end which comprises the p19 fragment were found in any samples.

Due to technical difficulties with the BS³ cross-linker, DSSO was used as an alternative. This cross-linker has the additional advantage that analysis is more straightforward, especially when intermolecular interactions are analysed. The experiments shown here only look for interactions within MSP-1, but the same cross-linker can be used for further studies in the future. DSSO was used at a 20-fold molar excess, leading to more cross-links being detected compared to BS³. *P. falciparum* MSP-1D, *P. vivax* and *P. knowlesi* MSP-1 were analysed under the same conditions. The results are similar to the ones detected using BS³ but there are also some differences. In the MSP-1D sample, p83 does not interact with p30, but with p38 and p42 (**Figure 53**). In general, p30 shows only a low amount of interactions with other processing fragments in this analysis. The *P. vivax* and *P. knowlesi* MSP-1 proteins shows a similar interaction pattern among their processing fragments, but here p83 interacts with all other processing fragments, including p30. However, in all three samples p30 does not interact with p42. P38 appears to bind to all other processing fragments and the C-terminal p42 interacts with p38 and p83 in all analysed MSP-1 heterodimers. In agreement with previous findings using BS³ cross-linked samples, no interactions involving the p19 region were detected.



Figure 53. Detected interactions using the MS-cleavable cross-linker DSSO. All shown links have reached a score of >80. The cross-linker was added to the recombinant proteins at a 20-fold molar excess. In all three proteins the p83 processing fragment interacts with p42 and p38, while an additional interaction with p30 was found in *P. vivax* and *P. knowlesi* MSP-1. P38 interacts with all other processing fragments, whereas p30 seems to only bind to p83 and p38. Interestingly, no interactions were found for the very C-terminal part of MSP-1.

In addition to strain-specific MSP-1 proteins, recombinant proteins created from halves of two different *Plasmodium* strains or species were cross-linked and analysed (Figure 54). One hybrid protein was assembled using the N-terminal 83/30 half of MSP-1F and the C-terminal 38/42 half of MSP-1D. While interactions between both fragments were detected after cross-linking with DSSO, the majority of interactions occur within the N-terminal fragment itself. This might be explained by an excess of the 83/30 half in the sample, leading to more detected cross-links within that fragment only. However, the interaction pattern is similar to previously tested MSP-1 proteins and includes interactions between p83 and p42, as well as between p30, p83 and p38. An interaction between p30 and p42 was not detected, which is identical to the findings for MSP-1D, *P. vivax* and *P. knowlesi* MSP-1.

Cross-linking and mass spectrometry analysis of the presumable hybrid protein consisting of *P. vivax* 83/30 and *P. falciparum* 38/42 revealed a very low number of interactions and all of them were limited

to interactions within the two halves. As no links between the two fragments were found, despite including results with a very low score of ~20, it has to be assumed that the creation of a hybrid protein derived from two different species was unsuccessful.



Figure 54. Cross-linking of presumable hybrid proteins created from two halves of different *Plasmodium* strains or species. The hybrid protein created using the 83/30 half of MSP-F and the 38/42 half of MSP-1D shows a similar interaction pattern than the other recombinant MSP-1 proteins. However, most interactions were detected within the N-terminal fragment, suggesting that there might be an excess of 83/30 in the sample. Successful generation of a hybrid protein can be assumed as cross-links between both halves were found. Unfortunately, this was not the case for the second protein. Here, the creation of a hybrid protein based on one half of *P. vivax* MSP-1 and one half of *P. falciparum* was attempted. In this case also cross-links of a score as low as 20 are shown. The lack of interactions between both fragments indicate that the assembly of a hybrid protein consisting of MSP-1 halves from two different *Plasmodium* species is not possible with the applied protocol.

3.8. Recombinant PvMSP-1 is recognised by naturally infected donors

The most crucial criterium for a vaccine based on a recombinant protein is achieving the identical or at least highly similar structure compared to the native protein. Sera from volunteers living in a malaria-endemic region in Brazil were used to measure the interaction between recombinant PvMSP-1 and naturally acquired antibodies. Two types of samples were available from each volunteer. The T1 sample, which was taken at the beginning of a cohort study in an agricultural settlement (Bastos *et al.*, 2007) and the T2 sample, which was collected during an acute infection with *P. vivax*.

A 96 well plate was coated with 100 nM PvMSP-1 or one of its processing fragments in coating buffer. After washing with TBST, the plate was incubated with blocking buffer, followed by another washing step with TBST before sera from the Brazilian cohort were added. Predilutions of the sera were determined with an initial test using only 5 sera with the most volume. PvMSP-1, Pv83 and Pv42 are apparently the most immunogenic proteins as test sera could be prediluted at a ratio of 1:100. ELISAs using Pv30 and
Results

Pv38 required a lower predilution of 1:25. All sera were prediluted in blocking buffer and incubated on the plate for 2 hours at room temperature. Successful binding was detected using anti-human IgG coupled with alkaline phosphatase and 4-Nitrophenyl phosphate disodium salt hexahydrate as substrate. All sera were tested in duplicates, but a quantitative analysis was not possible because of insufficient sera volumes. N-terminal Pv83, C-terminal Pv42 and the full length PvMSP-1 were well recognised by naturally acquired antibodies. Analysis of Pv30 or Pv38 resulted in a lower signal with all sera. Statistical significance between T1 and T2 samples was calculated using the Wilcoxon signed-rank test. In theory, an ongoing *P. vivax* infection should boost the immune response. However, the increased signal between T1 and T2 samples was only significant for Pv83 and Pv30. In some cases the signal measured for the T2 sample was even lower than the one for the T1 sample (**Figure 55**).



Figure 55. Recombinant PvMSP-1 and its processing fragments are recognised by naturally acquired antibodies. 96 well plates were coated with 100 nM of protein and incubated with sera from Brazilian donors collected at the beginning of a cohort study (T1 sample) and when they suffered from an acute *P. vivax* infection (T2 sample). PvMSP-1 and its processing fragments Pv83 and Pv42 showed the highest recognition and sera could be used at a predilution of 1:100. Less naturally acquired antibodies recognised Pv30 or Pv38 and sera had to be used at a lower predilution of 1:25. Statistical significance between T1 and T2 samples was calculated with the Wilcoxon signed-rank test. An expected boost effect was not observed for many samples and the absorbance increase between the T1 and T2 samples was only significant for Pv83 and Pv30.

The merozoite surface protein 1 is one of the most researched proteins on the surface of *Plasmodium* parasites. However, most studies use only fragments of the full length MSP-1 protein, such as the C-terminal p42 or p19 fragments. While the information gathered from such experiments can be useful and has created the base for most immunological studies regarding this protein, using the full length protein has several advantages. (i) The overall structure of a protein of this size will most likely differ from those of the individual fragments, (ii) binding interactions and binding sites could only be revealed when the entire protein sequence is present and (iii) crucial immunological studies is a supply of sufficient amounts of MSP-1. The development of recombinant MSP-1 as a vaccine is a main objective in our laboratory and a production protocol for *P. falciparum* MSP-1 was successfully developed and patented in 2011 (Bujard *et al.*, 2011).

However, another *Plasmodium* parasite, *P. vivax* has received more attention over the years as reports of severe malaria have accumulated and the numbers of *P. vivax* infections have increased in areas where *P. falciparum* infections have been reduced. But research efforts focusing on *P. vivax* face more difficulties than *P. falciparum* studies as a long-term cell culture system is missing and freshly isolated or thawed parasites have to be used for experiments, which are not readily available all around the world.

The aim of this project was to establish a protein production process for *P. vivax* and *P. berghei* MSP-1, which can be used as reliable sources for structural and immunological assays. The comparison of structural information obtained from recombinant MSP-1 proteins of several *Plasmodium* species could provide insights into the protein's function and deliver information about which elements could be exploited by a vaccine. Recombinant *P. berghei* MSP-1 can be used for vaccination studies in mice to analyse if immunisation with MSP-1 has a beneficial influence on experimental cerebral malaria development in mice. While immunological studies using *P. vivax* MSP-1 (PvMSP-1) are challenging to execute in our laboratory, a robust protein production protocol puts us into the position to collaborate with other research groups that have access to cohorts in *P. vivax* endemic areas.

4.1. Production of recombinant P. vivax MSP-1

The development of a malaria vaccine is the primary goal in our laboratory. As a consequence, the production process developed in this thesis had to fulfil several requirements in addition to providing enough material for experiments. The protocol has to be planned with the option to upscale the process for mass production, which eliminates a number of expression systems, such as mammalian cell lines. *E. coli* was chosen as a suitable expression host as this system is cost-effective, commonly used and can be adapted to an upscaled format. The lack of posttranslational modifications in *E. coli* can be a problem for the production of other pharmaceutical or therapeutic proteins, which is why other expression systems such as mammalian or insect cells are chosen for the manufacture of some protein based products. This

shortcoming of the *E. coli* system is an advantage for MSP-1 expression as it was shown to be not glycosylated despite the presence of potential glycosylation sites (Berhe *et al.*, 2000). Using other expression systems could result in glycosylation of MSP-1 and therefore potentially compromise epitopes that would otherwise lead to immune responses, rendering the expression host unsuitable for vaccine production.

One risk of the *E. coli* host is the potential presence of endotoxins in the final product. These lipopolysaccharides are pyrogenic components of the cell membrane, which are released during cell lysis of gram-negative bacteria and can cause severe reactions in humans, such as fever, leukopenia, leucocytosis, induction of thrombosis and septic shock (Sandle, 2016). While less important for functional or structural assays in the laboratory, this can become crucial if the production process is upscaled for the manufacture of a human vaccine. The amount of endotoxins was quantified for the *P. falciparum* protocol, which is highly similar to the *P. vivax* production process, using the *Limulus* amebocyte lysate (LAL) testing method. This assay is based on the reaction of lipopolysaccharides with a clottable protein isolated from amoebocyte cells from the blood of *Limulus polyphemus*. The purified *P. falciparum* MSP-1 protein contained below 50 endotoxin units (EU) per mg protein. The typical vaccine formulation used in ongoing clinical trials includes only 150 µg of MSP-1 per dose, resulting in an estimated endotoxin level of 7.5 EU/dose. For comparison, the commercially available vaccines against typhus and cholera contain 100,000 EU/mL and 1,000,000 EU/mL, with an average volume of 0.5 mL per dose (Brito and Singh, 2011). As the purification principle of *P. vivax* MSP-1 is very similar to *P. falciparum* MSP-1, comparable endotoxin levels are expected although they were not quantified during this project.

The *P. vivax* MSP-1 sequence chosen for protein expression is based on the Salvador I strain, which is the best analysed P. vivax strain with its genome being sequenced in 2008 (Carlton, Adams, et al., 2008). Many molecular cloning approaches in *Plasmodium* face the problem of AT-richness. In comparison to P. falciparum, whose genome has an AT composition of 80% (Gardner et al., 2002), the genome of P. vivax is already less AT-rich (58%)(Carlton, Adams, et al., 2008). However, the AT percentage of the P. vivax MSP-1 gene could be further reduced from 57% to 46%, a reduction by 11%, by optimising a synthetic gene for human codon usage. This adaptation also has the advantage that the synthetic gene can be used for integration into adenoviruses if a protein-based vaccine is not successful. Additional modifications of the original MSP-1 sequence include the removal of the signal peptide and GPI anchor signal sequence as those are not required for the production of a soluble non-membrane bound protein. Expression of the full length MSP-1 protein as one recombinant protein was problematic as it is very prone to degradation. A tandem purification using a GST-tag at the 5' end and a His-tag at the 3' end resulted in a low yield and was subsequently abandoned as a protein expression technique (Kauth, 2006). Splitting the protein into two halves and assembling MSP-1 from two individually expressed proteins was a more successful alternative. Accordingly, the coding sequence of PvMSP-1 was analysed using a model for the subtilisin-like protease-1 (SUB1) (de Monerri et al., 2011) and the Prediction of Protease specificity online tool (Boyd et al., 2005) to determine the cleavage sites of each MSP-1 processing fragment. Splitting the protein at a naturally occurring cleavage site was considered superior to cutting the sequence into two

halves of equal length. As MSP-1 is proteolytically processed at the site separating the 83/30 and the 38/42 fragments, two recombinant proteins who are separated at this site have a higher probability of folding into the same conformation as the native protein. In that regard, assembling the MSP-1 protein from four processing fragments might result in an even higher structural similarity to the native protein than creating a heterodimer, but such a protocol would be even more difficult to combine with GMP guidelines. The sequence analysis revealed several cleavage sites, but only those with the highest score values were considered as potential cleavage sites in *P. vivax* MSP-1. In addition to defining the two halves of PvMSP-1, this analysis was also used to identify the four processing fragments which are created during SUB1 processing before egress. Nonetheless, it should be remembered that the chosen cleavage sites might not represent all possible MSP-1 fragments as alternative cleavage sites, which were found for *P. falciparum* (Das *et al.*, 2015) most likely also exist in the *P. vivax* MSP-1 sequence.

All recombinant MSP-1 halves and the individual processing fragments could be expressed under standard conditions (LB medium, 37°C) using pZE expression plasmids and the E. coli strain W3110-Z1. This system offers a tight regulation and no recombinant protein expression was detected before addition of IPTG. The His-tagged processing fragments and the Pv38/42 half showed very high expression levels after induction. The N-terminal Pv83/30 protein was not significantly expressed without the addition of an N-terminal 6xHis-tag or the replacement of the first 22 amino acids with the already optimised sequence from another P. falciparum 83/30 expression plasmid. The reasons for these difficulties were analysed in more detail when the expression plasmids for P. falciparum MSP-1 were created and suggested that the ribosomal binding site is not accessible in the unaltered sequence. The mRNA structure of Pv83/30 was analysed using the mFold web server and revealed some unfavourable structures, such as partially closed or hairpin structures, that could hinder binding of the ribosome during translation. Another contributing factor is the instability of the 83/30 fragment, which is prone to degradation by proteases. Susceptibility to proteases can be enhanced by flexible regions or destabilising amino acids at the N-terminus, a phenomenon known as the N-end rule (Varshavsky, 1997). But the 83/30 sequence of all MSP-1 proteins used in this study have either glutamic acid or valine at their N-terminal position, which are stabilising amino acids in E. coli. As these recombinant protein expression issues were overcome by adding an Nterminal 6xHis-tag or changing 66 base pairs in the N-terminal coding sequence, access to the ribosomal binding site seems to have a stronger negative impact on protein expression than degradation.

The majority of recombinant proteins based on MSP-1 accumulate in inclusion bodies, which consist of unfolded or partially folded proteins. The only exception are the Pv30 and Pv38 processing fragments, which are present as soluble and insoluble proteins and were purified using the bacterial lysate. The Pv83 and Pv42 processing fragments and both halves of PvMSP-1, comprised of Pv83/30 and Pv38/42 are expressed in insoluble form. This result is not surprising given the large molecular weight of these proteins and the presence of intramolecular disulphide bonds at the C-terminus of MSP-1. *E. coli* has a reducing milieu in the cytosol that can be problematic for the expression of soluble proteins with disulphide bonds. However, this outcome was beneficial for the reconstitution of the full length MSP-1 protein as both

halves can be refolded together from solubilised inclusion bodies. Another advantage of the accumulation in inclusion bodies is the low amount of *E. coli* host proteins found in inclusion bodies and the protection from endogenous proteolytic enzymes of the host cell. This can be particularly beneficial for the Pv83/30 half as all recombinant MSP-1 proteins produced in our laboratory showed a decreased stability for this half in particular.

Reconstitution of full length MSP-1 was achieved by using the "pulse renaturation" technique during which solubilised inclusion bodies are quickly diluted into the refolding buffer. This method has the advantage that only a low amount of folding intermediates are present during each "pulse" and correctly folded proteins are removed from the equilibrium and are not available for aggregation when the next load of solubilised inclusion bodies is added. An alternative approach is refolding by dialysis, which results in a slow removal of denaturing agents, allowing the formation of folding intermediates. These can be beneficial for the refolding process but some proteins also tend to aggregate more when dialysis is used as a refolding method. The key components of the refolding buffer include arginine, a glutathione redox system and a pH of 8.5. The exact mechanisms how arginine contributes to protein refolding is not known. One theory states that the guanidinium group of arginine interacts with tryptophan in the protein to suppress aggregation. This might be a contributing factor for P. vivax MSP-1, which includes tryptophan at three positions, but P. falciparum MSP-1 lacks this amino acid entirely. Another proposed mechanism is the interaction of arginine with amino acid side chains and peptide bonds, which is a more generally applicable idea (Arakawa et al., 2007). The mildly reducing conditions in the refolding buffer paired with a ratio of reduced glutathione to oxidised glutathione of 10 mM to 1 mM aid in protein refolding and disulphide bonds formation by repeatedly breaking energetically unfavourable conformations until the native conformation is reached. The latter are of particular importance for the C-terminal half of MSP-1. For optimal reconstitution of MSP-1 with the correct stoichiometry an equal amount of 83/30 and 38/42 were added. Initial attempts to refold P. vivax MSP-1 with the standard arginine concentration of 500 mM were not successful and resulted in both halves running separately on an SDS gel when initial purification attempts were tested (data not shown). Increasing the arginine concentration to 1 M could reduce these issues although not all MSP-1 halves successfully bound to each other and excess proteins had to be removed during the purification process.

Whereas the usage of isolation tags was legitimate for the enrichment of recombinant proteins meant for laboratory experiments only, the production protocol for full length PvMSP-1 had to be transferable for human vaccine manufacturing. Regulatory agencies do not allow the usage of any purification tags, such as 6xHis-tags, due to their immunogenic potential in humans. So in contrast to the MSP-1 processing fragments, no isolation tags were allowed for the expression of both PvMSP-1 halves. The pZE23 expression plasmids can be used for such protein expression demands, however, as mentioned above, Pv83/30 could not be expressed unless the N-terminal sequence was altered. The addition of an N-terminal His-tag was fine for initial structural experiments, but the replacement of the first 22 amino acids is a preferable alternative for vaccine applications. The overall sequence identity of these 22 amino acids of *P. falciparum* MSP-1 and *P. vivax* is 65% and includes 6 direct matches. The coding sequence for *P. falciparum* 83/30 was previously optimised to overcome similar protein expression difficulties and

while the same optimisation process can theoretically be done for the original *P. vivax* sequence, it was considered too time-consuming for this project.

The P. vivax MSP-1 processing fragments were expressed with an N-terminal His-tag and showed very high expression levels using the pZE13 expression plasmids (Figure 17). All induced samples contain a very pronounced band representing the protein of interest and some weaker bands of lower molecular weight that could result from protein degradation. Pv30 and Pv38 were purified from the bacterial lysate as soluble proteins and required no further refolding. Pv42 was expressed in insoluble form, which is most likely due to the six disulphide bonds not being formed in the reducing conditions found in the cytosol during protein expression. In theory, Ni-NTA can also be used to purify unfolded His-tagged proteins. In case of Pv42 this was unsuccessful. The causes for this were not further analysed as refolding the protein allowed protein enrichment in the same manner as for Pv30 and Pv38. Pv42 was refolded using the standard refolding buffer with 500 mM arginine, a glutathione redox system and a pH of 8.0. The final eluates of the Ni-NTA purifications contain mostly the desired protein, but also low amounts of contaminating proteins of both higher and lower molecular weight. These proteins could be degradation products, which still have an intact N-terminal His-tag, or histidine-rich E. coli proteins. The final yield of Pv42 was 1.6 mg/g E. coli wet paste, whereas Pv30 and Pv38 were obtained at a lower yield of 0.75 mg/g E. coli wet paste. An explanation for this is that both Pv30 and Pv38 were expressed as soluble and insoluble proteins in almost equal amounts, whereas Pv42 was almost exclusively in inclusion bodies. As refolding of recombinant proteins requires additional steps and Pv30 and Pv38 were only required in low amounts for this project, the inclusion bodies containing these proteins were discarded and not considered for the final yield calculations.

Pv83 was expressed as an insoluble protein as well, but as there are no disulphide bonds in this protein, the glutathione redox system was replaced with 5 mM DTT to reduce the formation of intermolecular disulphide bonds between the 3 cysteines found in Pv83. In contrast to Pv42, Pv83 could not be purified by His-tag affinity chromatography despite the presence of the His-tag as proven by Western Blot. Possible reasons could be based on the three-dimensional structure of the refolded protein. If the N-terminal His-tag is in the protein core, it is not accessible for binding to the affinity resin. Instead, Pv83 was purified in a similar manner as the full length protein by combining two chromatography techniques that do not rely on the presence of purification tags. First, Pv83 was enriched by anion exchange chromatography before contaminants of lower molecular weight were removed by size exclusion chromatography. Pv83 began to elute from the Q-sepharose column at a salt concentration of 75 mM, a concentration that is commonly used to wash the column and remove any unwanted contaminants. However, most of Pv83 remained bound to the column and could only be removed at a NaCl concentration of 1 M. This large range of binding strength and required salt concentration to elute the protein indicated that anion exchange chromatography is accompanied with a high loss of Pv83 protein. The final yield after additional size exclusion chromatography was 3 mg per gram *E. coli* wet paste.

The purification of *P. vivax* MSP-1 after refolding using both halves faced similar challenges than the purification of Pv83 alone. Protein enrichment without relying on affinity tags limits the possible purification procedures and required the combination of anion exchange chromatography and size exclusion chromatography with multiple buffer exchange steps in between. Technical progress since the development of the patent for *P. falciparum* MSP-1 production allowed the replacement of some labour-intensive steps, such as using desalting columns for buffer exchange instead of overnight dialysis. Nevertheless, some disadvantages of the original protocol remained after adaption to *P. vivax* MSP-1.

(i) All recombinant MSP-1 proteins produced in our laboratory can only be handled at a maximum concentration of 1.2 mg/mL without precipitation. This is not a problem for final storage of the protein, but it complicates chromatography runs as the volume limits for columns and sample loops are reached quicker than when a protein solution of higher concentration can be used. Also, protein concentration steps in between can result in protein loss if the maximum concentration is exceeded by accident. (ii) On the one hand, MSP-1 binds very strongly to the Q-sepharose matrix used for anion chromatography, but also begins to elute at low NaCl concentrations. This hinders efficient washing steps to remove contaminating proteins that might still be present after the refolding step. Additionally, protein is lost during this purification step since a large proportion of it binds very strongly to the column and cannot be eluted unless 1 M of NaCl is applied. This eluate, often termed "final wash" is highly contaminated with other proteins and including it into a following size exclusion chromatography would forfeit the purpose of the anion exchange chromatography step. (iii) Despite the efforts to add both P. vivax MSP-1 halves in equal amounts, not all of them formed a heterodimer and excess PfPv83/30 and Pv38/42 have to be removed during purification steps. In theory, both halves should have the same negative net charge at a buffer at pH 8 since their pIs are close with 6.6 and 6.5 for PfPv83/30 and Pv38/42, respectively. However, both halves display a distinct binding affinity to Q-sepharose with the C-terminal Pv38/42 half binding stronger to the matrix as PfPv83/30, enabling the separation of full length PvMSP-1 from excess PfPv83/30 during anion exchange chromatography, as the latter remains in the flow through. Remaining unbound halves and smaller degradation products that were co-eluted with PvMSP-1 can be removed during size exclusion chromatography. Such contaminating proteins of around 40 kDa were observed for both PvMSP-1 and Pv83, suggesting that these are either degradation products derived from the N-terminal portion of the protein or host proteins that are co-enriched by ion exchange chromatography. While the purification protocol resulted in protein preparations with only a small amount of remaining contaminants, the final yield of this laborious technique was rather low with 0.17 mg per g wet paste.

During this project I could successfully demonstrate the efficiency of an alternative purification process for full length PvMSP-1. In this approach, the individually expressed halves are purified in their denatured state using strong ion exchange resins before refolding the protein. This not only allows working with higher concentrations during protein purification, it also significantly reduces protein precipitation that occurs after the refolding step in the previously described protocol. Additionally, all dialysis steps are replaced by a less time-consuming buffer exchange using desalting columns. The different binding properties of the PfPv83/30 and the Pv38/42 could also be observed in this approach. Denatured Pv38/42

could be purified using a Mono S column, whereas PfPv83/30 did not bind to this cation exchanger and required the anion exchanger Mono Q instead. A major contaminant in the solubilised inclusion bodies of Pv38/42 did not bind to the Mono S resin, suggesting that this contaminant is not a degradation product, as these would have a similar net charge and hence similar binding behaviour. Applying different NaCl concentrations efficiently removed most contaminating proteins during the purification of PfPv83/30. The purified PvMSP-1 halves were refolded by direct buffer exchange into refolding buffer while not exceeding a maximum concentration of 1.2 mg/mL for this final step. Complex formation was tested by size exclusion chromatography and showed no significant separation of the two halves, indicating successful reconstitution of MSP-1.

The advantages of this approach are (i) a higher yield of 2.9 mg per g wet paste, (ii) a faster purification process and (iii) a reduction of MSP-1 complexes containing contaminating proteins or incomplete halves as the individual halves undergo rigorous purification before refolding. This alternative protocol uses more expensive chromatography materials, but time-saving qualities and higher yield could contribute to a more cost-effective manufacturing process if PvMSP-1 production were upscaled for vaccine production under GMP conditions. In that case, size exclusion chromatography should be added as mandatory step after refolding to remove low amounts of unbound MSP-1 halves.

4.2. Production of P. berghei MSP-1

The production of recombinant *P. berghei* MSP-1 (PbMSP-1) and the challenges associated with its expression and purification are very similar to other MSP-1 proteins. As with *P. vivax* MSP-1, the base for cloning the expression plasmids was a synthetic gene. The AT content of the original coding sequence could be reduced from 75% to 45% by optimising the codon usage for mice. PbMSP-1 is not intended for human use but the codon optimised gene can be inserted into adenoviral vectors for usage in vaccination studies in mice.

Initially, only the pZE23 expression plasmids without a His-tag were used for recombinant protein expression in the *E. coli* host W3110Z1. The Pb38/42 half was expressed after addition of IPTG, but the Pb83/30 half demonstrated expression problems as observed for *P. vivax* 83/30. Instead of replacing the coding sequence with the nucleotide sequence of another *Plasmodium* species, an N-terminal His-tag followed by a factor Xa cleavage site was added to enable protein expression and include the option to remove the His-tag by enzymatic cleavage. As PbMSP-1 will not be used as a human vaccine, regulatory restrictions regarding isolation tags do not apply to this protein.

Another difficulty regarding Pb83/30 is its instability. All SDS gels of induced *E. coli* cultures expressing this protein show bands at 140 kDa, 120 kDa and 90 kDa that are not found in the noninduced sample. This suggests that *P. berghei* 83/30 is even more prone to degradation or premature termination during translation as *P. vivax* MSP-1. This severely limits the obtainable yield and complicates reconstitution of the MSP-1 heterodimer. Not only is it more difficult to precisely determine the concentration of intact Pb83/30 that has to be added during the pulse renaturation, the chance of Pb38/42 binding to degradation products is much higher. Interestingly, degradation of Pb83/30 seems to begin at the C-terminal end as

the attempt to enrich intact Pb83/30 using Ni-NTA enriched the degradation products as well (data not shown). Refolding of PbMSP-1 was performed using the same technique as for P. vivax MSP-1, although an arginine concentration of 500 mM instead of 1 M was sufficient. Interestingly, anion exchange chromatography seemingly eliminated the presumable Pb83/30 degradation products and the final eluate after size exclusion chromatography contained only two bands matching the Pb83/30 and Pb38/42 halves (Figure 35). This effect could be caused by an unequal charge distribution or the degradation products lacking an interaction site, which is involved in binding to Pb38/42, resulting in them themselves not being able to bind to the anion exchanger. However, freezing and thawing purified PbMSP-1 apparently caused further degradation as the previously eliminated proteins of 120 and 90 kDa can be seen again in a later assay (Figure 45). Because of the instability of the Pb83/30 half, P. berghei MSP-1 was excluded from most experiments which had the purpose to elucidate the structure of MSP-1. The full length protein might still be suitable for immunisation studies in mice, but such experiments were not performed during this project due to the lack of a suitable adjuvant. The final yield of PbMSP-1 using this protocol was 0.25 mg per gram E. coli wet paste, which is similar to other yields obtained with this protocol. The alternative production process as described for P. vivax MSP-1 was not tested with other MSP-1 proteins but could be beneficial for P. berghei MSP-1 in regard to final yield but also in purity as degraded Pb83/30 halves could potentially be removed before assembling the full length protein. However, the instability problem of Pb83/30 as observed after freezing and thawing most likely remains.

4.3. Creation of a hybrid protein consisting of P. falciparum and P. vivax MSP-1 halves

Malaria endemic areas are often known for either having mostly P. falciparum or P. vivax infections despite both parasite species being found in the same area. The ideal vaccine against malaria induces cross-reactive immune responses targeting both of these parasite species. The sequence identity of the full length MSP-1 protein of P. falciparum and P. vivax is 39%, with the N-terminal 83/30 half sharing a sequence identity of 36% and the C-terminal 38/42 half achieving a value of 44% due to the highly conserved sequence of p19. The availability of MSP-1 halves from both of these species led to the idea to create an MSP-1 heterodimer consisting of one P. falciparum and one P. vivax half. For this experiment, the N-terminal 83/30 half of P. vivax was refolded with the C-terminal 38/42 half of P. falciparum followed by purification via ion exchange chromatography and size exclusion chromatography. Anion exchange chromatography could eliminate excess 83/30 that did not bind to the resin, but free Pf38/42, and the potential hybrid MSP-1 complex eluted at the same salt concentrations. Size exclusion chromatography can be used to separate individual proteins or in this case unbound MSP-1 halves from a complex or from each other if the used column offers sufficient resolution. The preparative column used here could not provide the resolution required for separating excess halves and the potential complex into two distinct peaks in the UV spectrum. According to the SDS gel, the final eluate contains both halves in equal concentrations (see sample 3 and 4 in Figure 37). However, unbound 83/30 fragment elutes first from a size exclusion chromatography column due to its higher hydrodynamic radius compared to the 38/42 processing fragment. If the resolution of the used column is not high enough, both proteins could

have eluted separately and the beginning elution of 38/42 overlapped with the elution of 83/30, resulting in samples containing both halves in an unbound state. Hence, size exclusion chromatography was unsuitable to test whether the two halves had bound together and additional cross-linking experiments were performed (**Figure 54**). In these experiments, no interactions between the two processing fragments were found, indicating that the binding affinity between the *P. falciparum* and *P. vivax* MSP-1 halves is too low to form a hybrid complex.

4.4. Structural characterisation

4.4.1. Molecular weight determination

Heterologous expression of P. vivax MSP-1 and its processing fragments was successful as described above. But especially the PfPv83/30, Pv30 and Pv38 protein migrate in SDS gels as if they had a higher molecular weight than calculated. P. vivax MSP-1 consists of 1751 amino acids and contains 73 prolines (4.2%), which are known as "helix breakers". The cyclic side group of this amino acid restricts the structural flexibility and introduces kinks into the polypeptide backbone. As a consequence, the electrophoretic mobility of proteins consisting of many prolines is decreased (Hames, 1998). The most noticeable shift can be seen for Pv30. This MSP-1 fragment has an expected molecular weight of 31 kDa, but according to the SDS gel it is closer to the 60 kDa marker. As this is double the expected molecular weight, dimerization of Pv30 might also be a possible explanation. However, this is unlikely as such a dimer should have been disrupted by DTT in the sample buffer or guanidine hydrochloride, which is used to solubilise inclusion bodies (Figure 29). The Pv30 processing fragment consists of 287 amino acids and 23 of those are prolines (8%), which are probably the cause for the decreased mobility in SDS gels. The P. falciparum MSP-1 protein as found in the 3D7 strain consists of 1720 amino acids and includes 67 prolines. But these prolines are equally distributed throughout the protein and the percentage of proline is in the range of 3.8 to 4.3% in each of the processing fragments. This difference probably explains why P. falciparum MSP-1 based recombinant proteins do not show such an altered migration behaviour in SDS gels.

As SDS-PAGE was obviously not a reliable method to determine the precise molecular weight of *P. vivax* MSP-1, a combination of size exclusion chromatography and sedimentation velocity analysis was used to calculate the molecular weight and check if the recombinant protein meets the expectations. The GMP product MSP-1D was used as a positive control for this method. Some companies sell molecular weight marker kits to analyse the molecular weight via size exclusion chromatography alone. However, results can vary a lot since the running behaviour during gel filtration is influenced by the hydrodynamic radius of the molecule and not the molecular weight directly. For *P. vivax* MSP-1, the result from the size exclusion chromatography alone would suggest a molecular weight of 400 kDa, indicating the formation of a heterotetramer instead of a heterodimer. Alternative ways to determine the molecular weight include analytical ultracentrifugation. A sedimentation equilibrium analysis can measure the molecular weight directly, but requires a lot of material. Combining size exclusion chromatography to determine the

hydrodynamic radius and a sedimentation velocity run to define the Svedberg coefficient was a more feasible approach.

The Svedberg coefficient is defined as the ratio of a particle's sedimentation velocity in response to centrifugal force and it is influenced by the molecular mass and geometrical shape of the molecule. The purity of MSP-1D was demonstrated in several individual tests as preparation for the clinical trial, which justified using it as a comparison for PvMSP-1. The sedimentation velocity run revealed three distinct populations in the PvMSP-1 sample of similar forms than those observed for MSP-1D. The first population most likely contains excess Pv38/42, which was not successfully removed during the protein purification following the original protocol. The second and main population contains the MSP-1 complex in heterodimeric form and the third population probably consists of aggregated protein. All samples were centrifuged before analysis to remove aggregates, but residual or newly aggregated protein can still be detected during the sedimentation velocity run. However, the Svedberg coefficient for *P. vivax* MSP-1 is very close to the value for MSP-1, indicating correct complex formation.

The molecular weight of *P. vivax* MSP-1 of 199 kDa was calculated using the determined values for the hydrodynamic radius (6.5 nm) and a Svedberg coefficient of 7.36 S. According to literature, PvMSP-1 has a molecular weight of 200 kDa. A value of 199 kDa for the recombinant protein matches this value since the GPI anchor signal sequence and signal peptide were removed. In addition to precisely determining the molecular weight, this analysis also shows that the recombinant *P. vivax* MSP-1 produced in this thesis forms a heterodimer and not a heterotetramer.

4.4.2. Bioinformatic analysis

Finding similarities between different *Plasmodium* species and strains was considered useful to evaluate whether the P. falciparum protein alone could be sufficient to induce cross-reactive immune responses or if it would be advisable to formulate a vaccine including MSP-1 from P. vivax or P. knowlesi as well. Three bioinformatic tools were combined to analyse MSP-1 sequences and to gather information biophysical and biochemical experiments could not provide at the time. A multiple sequence alignment combined with the predicted and successfully tested SUB1 processing sites highlighted the different lengths of MSP-1 processing fragments as described above. A protein blast using MSP-1 sequences of P. falciparum, P. vivax, P. knowlesi and P. berghei and the generation of a phylogenetic tree based on MSP-1 alone revealed how similar some *Plasmodium* species are to each other. The phylogenetic tree was created using the neighbour-joining method and the Poisson method to calculate evolutionary distances (Zuckerkandl and Pauling, 1965; Saitou and Nei, 1987). The generated phylogenetic tree is in agreement with previously reviewed evolutionary data based on partial mitochondrial genomes (Carlton, Escalante, et al., 2008). P. knowlesi MSP-1 shows a higher similarity with P. vivax MSP-1 than with P. falciparum MSP-1, which can be seen as an indication that P. knowlesi and P. vivax MSP-1 have a higher chance of inducing cross-reactive antibodies than P. vivax and P. falciparum. This data also indicates an early split between parasite species infecting humans or monkeys and rodents. However, the overall sequence identity of P. berghei MSP-1 and the other Plasmodium species included in this analysis is between

34-38%, which is almost identical to the values estimated for *P. vivax* and *P. falciparum* (38-40%). *P. vivax* and *P. knowlesi* share a higher sequence similarity of 62-64%, meaning the chance of these two MSP-1 proteins sharing immunogenic epitopes is higher. Interestingly, the two *P. falciparum* strains 3D7 and FCB-1 also only share a sequence identity of 59%, whereas both *P. vivax* strains Belem and Salvador I reached a value of 84%. Effective immune responses against multiple *Plasmodium* strains or species have to target epitopes found in more than one strain and chances of achieving this are higher if more than one MSP-1 protein is included in the vaccine. To elucidate this possible strategy, hybrid proteins consisting of *P. falciparum* and *P. vivax* as well as both *P. falciparum* strains were produced as recombinant proteins and were included in some structural studies in this project. Nevertheless, the sequence identity gives no information about structural epitopes and should never be considered as the only indicator of cross-reactive potential.

4.4.3. Processing by SUB1

The maturation of MSP-1 by SUB1 mediated cleavage is thought to induce conformational changes within the MSP-1 molecule leading to the exposure of interaction sites that are required to bind to other proteins. One example is the increased binding affinity of processed MSP-1 to spectrin, an interaction that leads to the disruption of the red blood cell membrane and egress of merozoites (Das et al., 2015). A sequence alignment and analysis of MSP-1 sequences from P. falciparum, P. vivax, P. knowlesi and P. berghei using the Prediction of Protease Specificity online tool (Boyd et al., 2005) and the published SUB1 model (de Monerri et al., 2011) revealed cleavage motifs in all of the analysed sequences. The exact position of these motifs within the amino acid sequence is not conserved though, resulting in processing fragments of varying length. This variation is the most striking between P. falciparum and P. berghei. The p30 fragment of P. berghei consists of 350 amino acids and is almost double as long as the one of P. falciparum MSP-1, which comprises 189 amino acids. Instead, the p42 fragment of P. falciparum is the longer one with 373 amino acids compared to 185 amino acids of the P. berghei p42 protein. As a result of these varying lengths, some interaction sites responsible for MSP-1 complex formation or interaction with binding partners might be part of different processing fragments in each Plasmodium species. P. vivax and P. falciparum show less variation, but P. knowlesi also has an increased length of the p30 processing fragment (368 amino acids) without a simultaneously shortened p42 processing fragment.

The recombinant MSP-1 proteins derived from *P. vivax*, *P. knowlesi* and *P. berghei* were incubated with recombinant *P. falciparum* SUB1 to test if the proteolytic processing is not only theoretically conserved and if the cleavage motifs are universal enough to be recognised by the *P. falciparum* protease. Samples were taken at specific time points to follow the processing in time. After 24 hours, all recombinant MSP-1 proteins were cleaved into four fragments as shown in Figure 45. Obviously, this experiment does not represent the timeline of proteolytic processing in the parasite as the quantities of MSP-1 protein and PfSUB1 were not adjusted to the amount in the parasite. In case of *P. knowlesi* and *P. vivax* MSP-1, the C-terminal 38/42 half seems to be processed quicker than the 83/30 half, suggesting an increased

accessibility of the cleavage motif in that fragment. The *P. berghei* MSP-1 sample shows significant degradation of the 83/30 fragment after freezing and thawing the purified protein, making a similar observation difficult in this assay. However, the *P. falciparum* SUB1 protease was able to cleave all tested MSP-1 proteins, indicating a high conservatism of these motifs and the process itself.

4.4.4. Secondary structure prediction and CD spectroscopy

Whenever recombinant proteins are used for functional characterisation, the structural integrity has to be tested. Bioinformatic tools can predict the secondary structure and sometimes even model the three-dimensional structure if certain criteria, like commonly found motifs and a high level of structured regions, are met. The amino acid sequence of MSP-1 from the *Plasmodium* species *P. falciparum*, *P. vivax*, P. knowlesi and P. berghei were analysed using the JPred software, which predicts alpha-helical structures, beta-sheets and coil regions. 40% of the protein sequence were alpha-helical, but surprisingly, all analysed sequences contain large stretches of unpredicted sequences with at least 40 amino acids in length. Such disordered regions are characterised by low sequence complexity, low proportions of bulky hydrophobic amino acids (phenylalanine, tyrosine and tryptophan) and high proportions of charged and hydrophilic amino acids (arginine, histidine, lysine, aspartic acid and glutamic acid) (Wright and Dyson, 2015). About a third of the amino acid composition of MSP-1 from all analysed Plasmodium species consists of charged and hydrophilic amino acids. Lysine is the most common charged and hydrophilic amino acid in MSP-1 and usually comprises about 10% of all amino acids found in the protein sequence. The locations of these presumable unstructured regions are similar albeit not identical when different Plasmodium species are compared. For instance, one unstructured region is found in the N-terminal p83 region, two and, in case of P. knowlesi, even three of these regions are in the p30 fragment and another one occurs in p38 in all species. This pattern suggests a non-coincidental occurrence. The presence of unstructured domains challenges the central dogma of biology that function is closely linked to the three-dimensional structure of a protein. But these unstructured regions possess an enhanced flexibility and could form loops that include interaction sites for other proteins. For MSP-1 these sequences might function as docking sites for MSP-6, MSP-7 and MSPDBL-1 and MSPDBL-2. These proteins were previously shown to form a complex with MSP-1 (Kauth et al., 2006; Lin et al., 2014), but the precise interaction sites remain unknown.

Protruding flexible loops could also harbour cleavage sites for proteases as these are typically found on the surface and have to be sterically accessible. The SUB1 cleavage site separating p83 and p30 and the one separating p30 and p38 overlaps with unpredicted domains in *P. berghei* MSP-1 and the SUB1 cleavage motif causing the split between p30 and p38 in *P. falciparum* MSP-1 is also found in such an unstructured region. However, other cleavage sites are found in structured regions, suggesting that in general, these sites do not have to be exposed in form of an unstructured loop for productive proteolytic processing.

The disordered domains could also change their conformation upon binding to an interaction partner. This mechanism is observed for some intrinsically disordered proteins involved in cell cycle control and

some disordered regions have been linked to biological functions, such as transcriptional activation domains (Wright and Dyson, 1999). Such an intrinsic plasticity could enable MSP-1 to recognise several biological targets without sacrificing specificity and be the base for the formation of various MSP-1 complexes with different interaction partners as previously proposed (Lin *et al.*, 2016). Induced folding events can also be a control mechanism that ensures the incorporation of individual components of a multiprotein complex in the correct order (Wright and Dyson, 1999).

The amount of unstructured regions found in MSP-1 complicates the construction of a reliable three-dimensional model and could also pose problems for crystallography and cryo-electron microscopy. Unstructured or only partially structured proteins have an increased susceptibility to protease degradation (Wright and Dyson, 2015), which might be a contributing factor for the decreased stability of the N-terminal 83/30 MSP-1 half, as this part of the protein contains more and longer stretches of disordered regions.

Circular dichroism (CD) spectroscopy is a commonly used biophysical method to analyse the secondary structure of a protein. In this project recombinant MSP-1 proteins were analysed by thermal denaturation, which includes recording a CD spectrum in certain intervals while the temperature increases from 10°C to 85°C. All tested MSP-1 proteins display the CD spectrum typical for a mostly alpha-helical secondary structure and melting temperatures range from 42.5 to 45.6°C with an average of 44°C. Interestingly, a hybrid protein consisting of the *P. falciparum* strains 3D7 and FCB-1 achieved a similar stability profile with a melting temperature of 42.9°C. This observed alpha-helical structure corresponds to the predictions made by the JPred software. However, this method cannot provide residue-specific information so no further analysis in regard to disordered domains is possible with CD spectroscopy. Recombinant MSP-1 proteins that were not processed by SUB1 were still partially folded at 85°C. MSP-1D, which was incubated with SUB1 and processed into four fragments was less stable and was mostly unfolded at temperatures >42°C. This suggests that the heterodimeric composition of recombinant MSP-1 stabilises the protein in addition to the disulphide bonds in the EGF-like domains at the C-terminus. Accordingly, heterologous production of MSP-1 based on two halves instead of four fragments is beneficial for stability and possibly also storage of the final product.

In addition to CD spectroscopy, correct conformation of recombinant *P. vivax* MSP-1 was tested with sera from study participants in Brazil who were infected with *P. vivax* parasites. These sera were used in ELISA experiments to test if naturally acquired antibodies recognise the recombinant *P. vivax* MSP-1 protein and its processing fragments. In this assay, previous observations made for *P. falciparum* processing fragments (Epp, 2003; Woehlbier *et al.*, 2006) were confirmed for *P. vivax*: While all recombinant PvMSP-1 proteins were recognised, Pv30 and Pv38 are apparently less immunogenic as the sera had to be used at a higher concentration in ELISA experiments than when Pv83, Pv42 and the full length protein were tested. The available sera corresponded to two different time points. T1 was collected at the beginning of the study, T2 during an ongoing *P. vivax* infection. However, the T1 sample does not correspond to a naïve patient. Still, the T2 sample was expected to show an increased reaction to *P. vivax* MSP-1 proteins, but this was not the case in all samples. Some study participants had good antibody

responses at baseline with a decline during the infection. One possible explanation for this is that this subgroup has defective B cell responses, for example due to atypical B cells, during acute-phase infections. The absence of the expected boosting response might also be due to antigenic diversity if the recombinant antigen does not match the variant expressed by the infecting parasites. However, this effect would be more pronounced with the most polymorphic domains of PvMSP-1 and less evident for the whole antigen, which includes both conserved and polymorphous domains. Nevertheless, all recombinant *P. vivax* proteins were recognised by naturally acquired antibodies, indicating that the heterologous production process established in this thesis can be used to manufacture PvMSP-1 in its natural conformation.

4.4.5. Cross-linking

Cross-linking paired with mass spectrometry can provide peptide-specific information about intra- and intermolecular interactions. Binding interactions within the P. falciparum MSP-1 heterodimer were previously analysed by co-expression, co-renaturation or incubation of separately refolded processing fragments followed by affinity chromatography to determine which MSP-1 processing fragments are binding to each other. In these experiments, p30 was the only processing fragment which interacted with all other fragments. While this seems surprising at first since p30 is dimorphic in P. falciparum MSP-1 and a protein forming the core of a complex is usually expected to be highly conserved, a closer look at the sequence reveals a middle grade of conservatism of block 9, which comprises 2/3 of p30 (Tanabe et al., 1987). P83 was shown to exclusively interact with p30 in these experiments and p38 and p42 interact with each other in addition to binding to p30. A close proximity of p38 and p42 was already suggested by another study, which identified a monoclonal antibody recognising a structural epitope formed by these two processing fragments (Lyon et al., 1987). These co-purification experiments can only provide a limited amount of information regarding interactions and binding of the processing fragments needs to be strong enough to allow co-purification using an isolation tag that is only present in one of the proteins. Hence, weak interactions were most likely missed and interactions which mask isolation tags could lead to a false-negative result. Cross-linking combined with mass spectrometry is a more precise and sensitive method to further elucidate and localise interactions among the MSP-1 processing fragments. Due to technical difficulties, the initially used BS³ cross-linker was substituted with DSSO, which is a cleavable cross-linker and facilitates data analysis due to its specific fractionation pattern. Most striking was the detection of an interaction between p83 and p42 in all MSP-1 heterodimers as this interaction was not found in previous experiments. Another change involved the p30 fragment, which seems to interact only with p83 and p38 but not with p42. Previous experiments already showed that p30 has a weaker affinity to p42 than p42 to p38 and p30 to p83 (Kauth et al., 2003). As the cross-linking experiments performed in this study used the full length MSP-1 protein, interactions of lower affinity might not be detected when binding sites of higher affinity are available. If these indications are taken into account, the MSP-1 model would change from the previously proposed asymmetric shape (Kauth et al., 2003) and form a more globular molecule (Figure 56).



Figure 56. Old and new model of MSP-1. Left: The previously published model with p30 forming the core of the MSP-1 complex and p83 only binding to p30. This model was based on affinity chromatography-based experiments and suggested a more elongated, asymmetric form of the protein complex. Right: According to the cross-linking data, the MSP-1 model was changed to include the interactions of p83 with other processing fragments and p30 only binding to p83 and p38.

The discrepancies between previous data and the cross-linking results might also be explained by the low sensitivity of the co-purification technique. Especially the p83 fragment of *P. falciparum* was difficult to use in the previous experiments as this protein was highly instable during co-expression in *E. coli*. As an alternative approach, p83 and a fusion protein of p38/42 were refolded separately before incubating them and continuing with the affinity chromatography step (Kauth *et al.*, 2003), which could lead to different structural conformations as when both proteins are refolded together. The crosslinking experiments performed here use the MSP-1 heterodimer reconstituted from two halves, which can also result in some variation compared to experiments using a complex assembled from all four processing fragments. One major advantage of this cross-linking technique is the sequence specific information that allows the localisation of found interactions. The previously mentioned unstructured regions within MSP-1 could be included to the protein of the sequence of the sequence specific information that allows the localisation of found interactions. The previously mentioned unstructured regions within MSP-1 could be included to the protein of the protein of

involved in interactions within the MSP-1 protein or with other binding partners, such as MSP-6, MSP-7, MSP-3 or MSPDBL1 and MSPDBL2. A closer look at the found cross-links revealed that most interactions within the MSP-1 protein do not occur between unstructured regions, suggesting that these unstructured domains could be reserved for other interactions, while the structured regions are responsible for interactions within the protein and potentially stabilise its three-dimensional structure. The p30 processing fragment contains most unstructured regions in all analysed MSP-1 proteins and according to previous experiments binds to MSP-7 (Kauth *et al.*, 2006). However, further cross-linking experiments using suggested binding partners for MSP-1 are required to test this hypothesis. The usage of DSSO as a cross-linker will be beneficial for such studies as the fractionation pattern allows an easier distinction between interactions found within MSP-1 and MSP-1 and another protein.

4.5. Outlook

The availability of recombinant full length *P. vivax* MSP-1 facilitates further immunological research on *P. vivax*, as long as access to *P. vivax* parasites or cohorts in *P. vivax* endemic areas is possible. In that sense, a collaboration with Dr. Sócrates Herrera, who has access to such cohorts in Columbia and offered to perform immunological assays, is planned.

A long-term cell culture of *P. vivax* parasites is still not available, but some laboratories have made remarkable progress in short-term cultures and functional assays using *P. vivax* parasites. One of these assays is an *ex vivo* invasion inhibition assay (Russell *et al.*, 2011; Cho *et al.*, 2015) which could be used to test *P. vivax* MSP-1 antibodies for their invasion-blocking potential. Briefly, enriched cord blood reticulocytes are mixed with matured trypsin-treated *P. vivax* schizonts, which were isolated and concentrated from clinical isolates. Target reticulocytes are labelled with carboxyfluorescein succinimidyl ester (CFSE) to distinguish newly invaded reticulocytes from those of the initial inoculum. After 24 hours of incubation, the numbers of invaded reticulocytes are quantified by flow cytometry (**Figure 57**). Polyclonal antibodies against PvMSP-1 and its processing fragments were already generated in rabbits (data not shown) and IgGs were purified, ready to be used in this or similar assays.



Figure 57. Scheme of the *ex vivo* **invasion assay using** *P. vivax. P. vivax* infected reticulocytes are enriched from clinical isolates and mixed with CFSE labelled target cells. After 24 hours of incubation, invaded CFSE⁺ reticulocytes are quantified by flow cytometry. Antibodies could be tested for their invasion-blocking potential by adding them when the inoculum and target cells are mixed. The graphic above is based on the scheme published by Cho *et al.*, 2015.

The cross-linking data presented here already includes new insights compared to previous immunoaffinity-based experiments, but only looked at interactions within the MSP-1 protein so far. The next logical step is to generate complexes of MSP-1 and its known binding partners MSP-6, MSP-7, MSPDBL-1 and MSPDBL-2, and cross-link them with DSSO. The advantage of this cross-linker compared to BS³ is the distinct fragmentation pattern for inter- and intramolecular cross-links (Kao *et al.*, 2011), which allows to easily identify interactions between MSP-1 and a binding partner. The peptide-specific information could be used to further optimise the MSP-1 model and localise the exact interaction sites. This information could be used to design drugs or vaccines that inhibit MSP-1 complex formation with MSP-6 or MSPDBL proteins and prevent MSP-1 complex-mediated attachment of the parasite to red blood cells.

The best way to obtain structural information about MSP-1 would be to obtain a three-dimensional structure. Attempts to crystallise MSP-1 have failed in our laboratory, most likely because of the presence of unstructured regions in the protein and the associated flexibility. Cryo-electron microscopy (Cryo-EM) is another promising technology to obtain three-dimensional structures and a collaboration with Dr. Kudryashev at the Max Planck Institute in Frankfurt is currently ongoing. Current efforts concentrate on MSP-1D as this protein is available as a GMP product, offering the highest purity. Nonetheless, using the protein on its own was less successful and preliminary data revealed a preferential orientation, which makes the rendering of a three-dimensional structure difficult. Recent efforts include the formation of a complex of MSP1-D and a Fab fragment of the monoclonal antibody 5.2, which binds to a conformational epitope in one of the EGF-like domains at the C-terminus. For one, this further increases the size of the molecule and also facilitates analysing the data as previously published structural information of the p19 fragment (Pizarro *et al.*, 2003) can be used as a starting point. However, the large stretches of unstructured regions could limit the resolution of the cryo-EM structure. Another approach would be to express MSP-1 with the GPI anchor signal sequence and signal peptide in a mammalian cell line to study membrane-bound MSP-1.

Overall, this study provides new structural information about MSP-1 from various *Plasmodium* species and the established protein production processes can be used to obtain sufficient amounts of *P. vivax* MSP-1 for further functional and immunological studies.

al-Yaman, F. *et al.* (1996) 'Assessment of the role of naturally acquired antibody levels to Plasmodium falciparum merozoite surface protein-1 in protecting Papua New Guinean children from malaria morbidity.', *American Journal of Tropical Medicine and Hygiene*, 54(5), pp. 443–448.

Altschul, S. F. *et al.* (1997) 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.', *Nucleic acids research*, 25(17), pp. 3389–3402.

Anstey, N. M. et al. (2009) 'The pathophysiology of vivax malaria.', Trends in parasitology, 25(5), pp. 220-227.

Arakawa, T. *et al.* (2007) 'Suppression of protein interactions by arginine: a proposed mechanism of the arginine effects.', *Biophysical chemistry*, 127(1–2), pp. 1–8.

Arévalo-Herrera, M. *et al.* (2016) 'Protective Efficacy of Plasmodium vivax Radiation-Attenuated Sporozoites in Colombian Volunteers: A Randomized Controlled Trial.', *PLoS neglected tropical diseases*, 10(10), p. e0005070.

Arévalo-Herrera, M., Chitnis, C. and Herrera, S. (2010) 'Current status of Plasmodium vivax vaccine.', *Human vaccines*, 6(1), pp. 124–132.

Babon, J. J. et al. (2007) 'Structural studies on Plasmodium vivax merozoite surface protein-1.', Molecular & Biochemical Parasitology, 153(1), pp. 31-40.

Baer, K. *et al.* (2007) 'Release of hepatic Plasmodium yoelii merozoites into the pulmonary microvasculature.', *PLoS pathogens*, 3(11), p. e171.

Baird, J. K. (2007) 'Neglect of Plasmodium vivax malaria.', Trends in parasitology, 23(11), pp. 533–539.

Baird, J. K. and Hoffman, S. L. (2004) 'Primaquine therapy for malaria.', *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 39(9), pp. 1336–1345.

Baldwin, M. R. *et al.* (2015) 'Merozoite surface protein 1 recognition of host glycophorin A mediates malaria parasite invasion of red blood cells.', *Blood*, 125(17), pp. 2704–2711.

Bastos, M. S. *et al.* (2007) 'Antigenic polymorphism and naturally acquired antibodies to Plasmodium vivax merozoite surface protein 1 in rural Amazonians.', *Clinical and Vaccine Immunology*, 14(10), pp. 1249–1259.

Beeson, J. G. et al. (2016) 'Merozoite surface proteins in red blood cell invasion, immunity and vaccines against malaria.', FEMS microbiology reviews, p. fuw001.

Bennett, J. W. et al. (2013) 'Primaquine failure and cytochrome P-450 2D6 in Plasmodium vivax malaria.', New England Journal of Medicine, 369(14), pp. 1381–1382.

Berhe, S. *et al.* (2000) 'Plasmodium falciparum: merozoite surface proteins 1 and 2 are not posttranslationally modified by classical N- or O-glycans.', *Experimental parasitology*, 94(3), pp. 194–197.

Billker, O. et al. (1998) 'Identification of xanthurenic acid as the putative inducer of malaria development in the mosquito.', Nature, 392(6673), pp. 289–292.

Blackman, M. J. *et al.* (1990) 'A single fragment of a malaria merozoite surface protein remains on the parasite during red cell invasion and is the target of invasion-inhibiting antibodies.', *The Journal of experimental medicine*, 172(1), pp. 379–382.

Blackman, M. J. *et al.* (1991) 'Proteolytic processing of the Plasmodium falciparum merozoite surface protein-1 produces a membrane-bound fragment containing two epidermal growth factor-like domains.', *Molecular & Biochemical Parasitology*, 49(1), pp. 29–33.

Blackman, M. J. et al. (1994) 'Antibodies inhibit the protease-mediated processing of a malaria merozoite surface protein.', *The Journal of experimental medicine*, 180(1), pp. 389–393.

Blackman, M. J. et al. (1998) 'A subtilisin-like protein in secretory organelles of Plasmodium falciparum merozoites.', *The Journal of biological chemistry*, 273(36), pp. 23398–23409.

Blackman, M. J. and Holder, A. A. (1992) 'Secondary processing of the Plasmodium falciparum merozoite surface protein-1 (MSP1) by a calcium-dependent membrane-bound serine protease: shedding of MSP133 as a noncovalently associated complex with other fragments of the MSP1.', *Molecular & Biochemical Parasitology*, 50(2), pp. 307–315.

Boyd, S. E. et al. (2005) 'PoPS: a computational tool for modeling and predicting protease specificity.', Journal of bioinformatics and computational biology, 3(3), pp. 551–585.

Boyle, M. J., Richards, J. S., *et al.* (2010) 'Interactions with heparin-like molecules during erythrocyte invasion by Plasmodium falciparum merozoites.', *Blood*, 115(22), pp. 4559–4568.

Boyle, M. J., Wilson, D. W., *et al.* (2010) 'Isolation of viable Plasmodium falciparum merozoites to define erythrocyte invasion events and advance vaccine and drug development.', *Proceedings of the National Academy of Sciences of the United States of America*, 107(32), pp. 14378–14383.

Boyle, M. J. et al. (2015) 'Human antibodies fix complement to inhibit Plasmodium falciparum invasion of erythrocytes and are associated with protection against malaria.', *Immunity*, 42(3), pp. 580–590.

Boyle, M. J., Wilson, D. W. and Beeson, J. G. (2013) 'New approaches to studying Plasmodium falciparum merozoite invasion and insights into invasion biology.', *International journal for parasitology*, 43(1), pp. 1–10.

Brito, L. A. and Singh, M. (2011) 'Acceptable levels of endotoxin in vaccine formulations during preclinical research.', *Journal of pharmaceutical sciences*, 100(1), pp. 34–37.

Bruce-Chwatt, L. J. (1981) 'Alphonse Laveran's discovery 100 years ago and today's global fight against malaria.', *Journal of the Royal Society of Medicine*, 74(7), pp. 531–536.

Bujard, H. et al. (2011) 'MSP-1 protein preparations from Plasmodium'.

Butler, N. S. *et al.* (2011) 'Superior antimalarial immunity after vaccination with late liver stage-arresting genetically attenuated parasites.', *Cell host & microbe*, 9(6), pp. 451–462.

Carlton, J. M., Escalante, A. A., *et al.* (2008) 'Comparative evolutionary genomics of human malaria parasites.', *Trends in parasitology*, 24(12), pp. 545–550.

Carlton, J. M., Adams, J. H., et al. (2008) 'Comparative genomics of the neglected human malaria parasite Plasmodium vivax', Nature, 455(7214), pp. 757–763.

Carralot, J.-P. et al. (2008) 'Mass spectrometric identification of an HLA-A*0201 epitope from Plasmodium falciparum MSP-1.', International immunology, 20(11), pp. 1451–1456.

Carter, R. and Mendis, K. N. (2002) 'Evolutionary and historical aspects of the burden of malaria.', *Clinical microbiology reviews*, 15(4), pp. 564–594.

Carvalho, B. O. *et al.* (2010) 'On the cytoadhesion of Plasmodium vivax-infected erythrocytes.', *Journal of Infectious Diseases*, 202(4), pp. 638–647.

Cavanagh, D. R. *et al.* (2004) 'Antibodies to the N-terminal block 2 of Plasmodium falciparum merozoite surface protein 1 are associated with protection against clinical malaria.', *Infection and Immunity*, 72(11), pp. 6492–6502.

Cavanagh, D. R. *et al.* (2014) 'Antibody responses to a novel Plasmodium falciparum merozoite surface protein vaccine correlate with protection against experimental malaria infection in Aotus monkeys.', *PloS one*, 9(1), p. e83704.

Child, M. A. *et al.* (2010) 'Regulated maturation of malaria merozoite surface protein-1 is essential for parasite growth.', *Molecular microbiology*, 78(1), pp. 187–202.

Chitarra, V. *et al.* (1999) 'The crystal structure of C-terminal merozoite surface protein 1 at 1.8 A resolution, a highly protective malaria vaccine candidate.', *Molecular cell*, 3(4), pp. 457–464.

Cho, J. S. et al. (2015) 'Unambiguous determination of Plasmodium vivax reticulocyte invasion by flow cytometry.', *International journal for parasitology*.

Clark, I. A. (1987) 'Monokines and lymphokines in malarial pathology.', Annals of tropical medicine and parasitology, 81(5), pp. 577-585.

Clyde, D. F. (1975) 'Immunization of man against falciparum and vivax malaria by use of attenuated sporozoites.', *The American journal of tropical medicine and hygiene*, 24(3), pp. 397–401.

Combe, A. et al. (2009) 'Clonal conditional mutagenesis in malaria parasites.', Cell host & microbe, 5(4), pp. 386-396.

Conway, D. J. et al. (2000) 'A principal target of human immunity to malaria identified by molecular population genetic and immunological analyses.', *Nature medicine*, 6(6), pp. 689–692.

Cooper, J. A. and Bujard, H. (1992) 'Membrane-associated proteases process Plasmodium falciparum merozoite surface antigen-1 (MSA1) to fragment gp41.', *Molecular & Biochemical Parasitology*, 56(1), pp. 151–160.

Cowman, A. F. and Crabb, B. S. (2006) 'Invasion of red blood cells by malaria parasites.', Cell, 124(4), pp. 755-766.

Cox, F. E. G. (2002) 'History of human parasitology.', Clinical microbiology reviews, 15(4), pp. 595-612.

Crompton, P. D., Pierce, S. K. and Miller, L. H. (2010) 'Advances and challenges in malaria vaccine development.', *The Journal of clinical investigation*, 120(12), pp. 4168–4178.

Cuff, J. A. and Barton, G. J. (2000) 'Application of multiple sequence alignment profiles to improve protein secondary structure prediction.', *Proteins*, 40(3), pp. 502–511.

Daly, T. M. and Long, C. A. (1995) 'Humoral response to a carboxyl-terminal region of the merozoite surface protein-1 plays a predominant role in controlling blood-stage infection in rodent malaria.', *Journal of immunology* (*Baltimore, Md.*: 1950), 155(1), pp. 236–243.

Das, S. *et al.* (2015) 'Processing of Plasmodium falciparum Merozoite Surface Protein MSP1 Activates a Spectrin-Binding Function Enabling Parasite Egress from RBCs.', *Cell host & microbe*, 18(4), pp. 433–444.

Doolan, D. L., Dobaño, C. and Baird, J. K. (2009) 'Acquired immunity to malaria.', *Clinical microbiology reviews*, 22(1), p. 13--36-- Table of Contents.

Douglas, R. G. *et al.* (2015) 'Active migration and passive transport of malaria parasites.', *Trends in parasitology*, 31(8), pp. 357–362.

Draper, S. J. *et al.* (2009) 'Recombinant viral vaccines expressing merozoite surface protein-1 induce antibody- and T cell-mediated multistage protection against malaria.', *Cell host & microbe*, 5(1), pp. 95–105.

Drew, D. R. *et al.* (2004) 'A common cross-species function for the double epidermal growth factor-like modules of the highly divergent plasmodium surface proteins MSP-1 and MSP-8.', *The Journal of biological chemistry*, 279(19), pp. 20147–20153.

Drozdetskiy, A. et al. (2015) 'JPred4: a protein secondary structure prediction server.', Nucleic acids research, 43(W1), pp. W389--94.

Duncan, C. J. A. et al. (2011) 'Impact on malaria parasite multiplication rates in infected volunteers of the proteinin-adjuvant vaccine AMA1-C1/Alhydrogel+CPG 7909.', PloS one, 6(7), p. e22271.

Duraisingh, M. T. et al. (2003) 'Erythrocyte-binding antigen 175 mediates invasion in Plasmodium falciparum utilizing sialic acid-dependent and -independent pathways.', Proceedings of the National Academy of Sciences, 100(8), pp. 4796–4801.

Dvorak, J. A. et al. (1975) 'Invasion of erythrocytes by malaria merozoites.', Science (New York, N.Y.), 187(4178), pp. 748–750.

Egan, A. F. et al. (1996) 'Clinical immunity to Plasmodium falciparum malaria is associated with serum antibodies

to the 19-kDa C-terminal fragment of the merozoite surface antigen, PfMSP-1.', Journal of Infectious Diseases, 173(3), pp. 765-769.

English, M. and Newton, C. R. J. C. (2002) 'Malaria: Pathogenicity and Disease', in *Malaria Immunology*. Basel: KARGER, pp. 50-69.

Epp, C. (2003) Das Merozoitenoberflächenprotein (MSP)-1 aus Plasmodium falciparum, ein Impfstoffkandidat gegen Malaria: Untersuchungen zur Struktur, Immunogenität und schützenden Wirkung in vitro.

Etlinger, H. M. et al. (1991) 'Ability of recombinant or native proteins to protect monkeys against heterologous challenge with Plasmodium falciparum.', *Infection and Immunity*, 59(10), pp. 3498–3503.

Felsenstein, J. (1985) 'CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP.', *Evolution*, 39(4), pp. 783–791.

Friesen, J. et al. (2010) 'Natural immunization against malaria: causal prophylaxis with antibiotics.', Science translational medicine, 2(40), p. 40ra49--40ra49.

Früh, K. *et al.* (1991) 'Human antibody response to the major merozoite surface antigen of Plasmodium falciparum is strain specific and short-lived.', *Infection and Immunity*, 59(4), pp. 1319–1324.

Galamo, C. D. *et al.* (2009) 'Anti-MSP1 block 2 antibodies are effective at parasite killing in an allele-specific manner by monocyte-mediated antibody-dependent cellular inhibition.', *Journal of Infectious Diseases*, 199(8), pp. 1151–1154.

Gardner, M. J. et al. (2002) 'Genome sequence of the human malaria parasite Plasmodium falciparum', Nature, 419(6906), pp. 498-511.

Garman, S. C. *et al.* (2003) 'Structure of the C-terminal domains of merozoite surface protein-1 from Plasmodium knowlesi reveals a novel histidine binding site.', *The Journal of biological chemistry*, 278(9), pp. 7264–7269.

Gilberger, T.-W. *et al.* (2003) 'The cytoplasmic domain of the Plasmodium falciparum ligand EBA-175 is essential for invasion but not protein trafficking.', *The Journal of cell biology*, 162(2), pp. 317–327.

Gilson, P. R. and Crabb, B. S. (2009) 'Morphology and kinetics of the three distinct phases of red blood cell invasion by Plasmodium falciparum merozoites.', *International journal for parasitology*, 39(1), pp. 91–96.

Goel, V. K. *et al.* (2003) 'Band 3 is a host receptor binding merozoite surface protein 1 during the Plasmodium falciparum invasion of erythrocytes.', *Proceedings of the National Academy of Sciences*, 100(9), pp. 5164–5169.

Golgi, C. (1886) 'Sul' infezione malarica', Archivio per le scienze mediche Torino, (10), pp. 109-135.

Gómez-Pérez, G. P. *et al.* (2015) 'Controlled human malaria infection by intramuscular and direct venous inoculation of cryopreserved Plasmodium falciparum sporozoites in malaria-na{\"\i}ve volunteers: effect of injection volume and dose on infectivity rates.', *Malaria Journal*, 14(1), p. 306.

Graewe, S. *et al.* (2012) 'Chronicle of a death foretold: Plasmodium liver stage parasites decide on the fate of the host cell.', *FEMS microbiology reviews*, 36(1), pp. 111–130.

Grassi, G. B. (1900) Studi di uno Zoologo sulla Malaria. Rome.

Greenwood, B. M. et al. (2005) 'Malaria.', Lancet (London, England), 365(9469), pp. 1487-1498.

Gruszczyk, J., Huang, R. K., *et al.* (2018) 'Cryo-EM structure of an essential Plasmodium vivax invasion complex.', *Nature*, 559(7712), pp. 135–139.

Gruszczyk, J., Kanjee, U., et al. (2018) 'Transferrin receptor 1 is a reticulocyte-specific receptor for Plasmodium vivax.', Science (New York, N.Y.), 359(6371), pp. 48–55.

Hames, B. D. (1998) Gel Electrophoresis of Proteins: A Practical Approach . 3rd editio. Oxford University Press.

Handayani, S. et al. (2009) 'High deformability of Plasmodium vivax-infected red blood cells under microfluidic

conditions.', Journal of Infectious Diseases, 199(3), pp. 445-450.

Harris, P. K. *et al.* (2005) 'Molecular identification of a malaria merozoite surface sheddase.', *PLoS pathogens*, 1(3), pp. 241–251.

Haussig, J. M., Matuschewski, K. and Kooij, T. W. A. (2011) 'Inactivation of a Plasmodium apicoplast protein attenuates formation of liver merozoites.', *Molecular microbiology*, 81(6), pp. 1511–1525.

Herrera, S. et al. (1990) 'Immunization of Aotus monkeys with Plasmodium falciparum blood-stage recombinant proteins.', *Proceedings of the National Academy of Sciences of the United States of America*, 87(10), pp. 4017–4021.

Herrera, S. *et al.* (1993) 'A conserved region of the MSP-1 surface protein of Plasmodium falciparum contains a recognition sequence for erythrocyte spectrin.', *The EMBO journal*, 12(4), pp. 1607–1614.

Hirunpetcharat, C. *et al.* (1997) 'Complete protective immunity induced in mice by immunization with the 19-kilodalton carboxyl-terminal fragment of the merozoite surface protein-1 (MSP1[19]) of Plasmodium yoelii expressed in Saccharomyces cerevisiae: correlation of protection with antigen', *Journal of immunology (Baltimore, Md.: 1950)*, 159(7), pp. 3400–3411.

Hoffman, S. L. et al. (2002) 'Protection of humans against malaria by immunization with radiation-attenuated Plasmodium falciparum sporozoites.', Journal of Infectious Diseases, 185(8), pp. 1155–1164.

Holder, A. A. and Freeman, R. R. (1981) 'Immunization against blood-stage rodent malaria using purified parasite antigens.', *Nature*, 294(5839), pp. 361–364.

Hollingdale, M. R. and Sedegah, M. (2017) 'Development of whole sporozoite malaria vaccines.', *Expert review of vaccines*, 16(1), pp. 45-54.

Holt, R. A. et al. (2002) 'The genome sequence of the malaria mosquito Anopheles gambiae.', Science (New York, N.Y.), 298(5591), pp. 129–149.

Horuk, R. *et al.* (1993) 'A receptor for the malarial parasite Plasmodium vivax: the erythrocyte chemokine receptor.', *Science (New York, N.Y.)*, 261(5125), pp. 1182–1184.

Howard, R. J. and Gilladoga, A. D. (1989) 'Molecular studies related to the pathogenesis of cerebral malaria.', *Blood*, 74(8), pp. 2603–2618.

Johnson, J. G. *et al.* (1980) 'Factors affecting the ability of isolated Plasmodium knowlesi merozoites to attach to and invade erythrocytes.', *Parasitology*, 80(3), pp. 539–550.

Johnson, J. G. et al. (1981) 'Identification of surface proteins on viable Plasmodium knowlesi merozoites.', The Journal of protozoology, 28(2), pp. 160–164.

Joos, C. *et al.* (2010) 'Clinical protection from falciparum malaria correlates with neutrophil respiratory bursts induced by merozoites opsonized with human serum antibodies.', *PloS one*, 5(3), p. e9871.

Jordán-Villegas, A. *et al.* (2005) 'Aotus lemurinus griseimembra monkeys: a suitable model for Plasmodium vivax sporozoite infection.', *The American journal of tropical medicine and hygiene*, 73(5 Suppl), pp. 10–15.

Junqueira, C. et al. (2018) 'Cytotoxic CD8+ T cells recognize and kill Plasmodium vivax-infected reticulocytes.', Nature medicine, 80, p. 151.

Kao, A. *et al.* (2011) 'Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes.', *Molecular & cellular proteomics : MCP*, 10(1), p. M110.002212.

Kauth, C. W. *et al.* (2003) 'The Merozoite Surface Protein 1 Complex of Human Malaria Parasite Plasmodium falciparum: INTERACTIONS AND ARRANGEMENTS OF SUBUNITS', *Journal of Biological Chemistry*, 278(25), pp. 22257–22264.

Kauth, C. W. (2006) Entwicklung eines experimentellen auf MSP-1 von Plasmodium falciparum basierenden Malaria Impfstoffes und dessen strukturelle und immunologische Charakterisierung. Kauth, C. W. et al. (2006) 'Interactions between merozoite surface proteins 1, 6, and 7 of the malaria parasite Plasmodium falciparum.', *The Journal of biological chemistry*, 281(42), pp. 31517–31527.

Keeley, A. and Soldati, D. (2004) 'The glideosome: a molecular machine powering motility and host-cell invasion by Apicomplexa.', *Trends in cell biology*, 14(10), pp. 528–532.

Koussis, K. *et al.* (2009) 'A multifunctional serine protease primes the malaria parasite for red blood cell invasion.', *The EMBO journal*, 28(6), pp. 725–735.

Kumar, S., Stecher, G. and Tamura, K. (2016) 'MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.', *Molecular biology and evolution*, 33(7), pp. 1870–1874.

Lacerda, M. V. G. *et al.* (2012) 'Postmortem characterization of patients with clinical diagnosis of Plasmodium vivax malaria: to what extent does this parasite kill?', *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 55(8), pp. e67--74.

Langreth, S. G., Nguyen-Dinh, P. and Trager, W. (1978) 'Plasmodium falciparum: merozoite invasion in vitro in the presence of chloroquine.', *Experimental parasitology*, 46(2), pp. 235–238.

Lin, C. S. *et al.* (2014) 'The merozoite surface protein 1 complex is a platform for binding to human erythrocytes by Plasmodium falciparum.', *Journal of Biological Chemistry*, 289(37), pp. 25655–25669.

Lin, C. S. *et al.* (2016) 'Multiple Plasmodium falciparum Merozoite Surface Protein 1 Complexes Mediate Merozoite Binding to Human Erythrocytes.', *Journal of Biological Chemistry*, 291(14), pp. 7703–7715.

Lingelbach, K. and Joiner, K. A. (1998) 'The parasitophorous vacuole membrane surrounding Plasmodium and Toxoplasma: an unusual compartment in infected cells.', *Journal of cell science*, 111 (Pt 1, pp. 1467–1475.

Lupas, A. (1997) 'Predicting coiled-coil regions in proteins.', *Current opinion in structural biology*, 7(3), pp. 388-393.

Lutz, R. and Bujard, H. (1997) 'Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements.', *Nucleic acids research*, 25(6), pp. 1203–1210.

Lyke, K. E. *et al.* (2017) 'Attenuated PfSPZ Vaccine induces strain-transcending T cells and durable protection against heterologous controlled human malaria infection.', *Proceedings of the National Academy of Sciences of the United States of America*, 114(10), pp. 2711–2716.

Lyon, J. A. *et al.* (1987) 'Monoclonal antibody characterization of the 195-kilodalton major surface glycoprotein of Plasmodium falciparum malaria schizonts and merozoites: identification of additional processed products and a serotype-restricted repetitive epitope.', *Journal of immunology (Baltimore, Md. : 1950)*, 138(3), pp. 895–901.

Malleret, B. et al. (2014) 'Plasmodium vivax: restricted tropism and rapid remodelling of CD71 positive reticulocytes.', Blood.

Medica, D. L. and Sinnis, P. (2005) 'Quantitative dynamics of Plasmodium yoelii sporozoite transmission by infected anopheline mosquitoes.', *Infection and Immunity*, 73(7), pp. 4363–4369.

Ménard, D. et al. (2010) 'Plasmodium vivax clinical malaria is commonly observed in Duffy-negative Malagasy people.', Proceedings of the National Academy of Sciences of the United States of America, 107(13), pp. 5967–5971.

Mensah, V. A. *et al.* (2016) 'Safety, Immunogenicity and Efficacy of Prime-Boost Vaccination with ChAd63 and MVA Encoding ME-TRAP against Plasmodium falciparum Infection in Adults in Senegal.', *PloS one*, 11(12), p. e0167951.

Miller, L. H. *et al.* (1979) 'Interaction between cytochalasin B-treated malarial parasites and erythrocytes. Attachment and junction formation.', *The Journal of experimental medicine*, 149(1), pp. 172–184.

de Monerri, N. C. *et al.* (2011) 'Global identification of multiple substrates for Plasmodium falciparum SUB1, an essential malarial processing protease.', *Infection and Immunity*, 79(3), pp. 1086–1097.

Mordmüller, B. et al. (2017) 'Sterile protection against human malaria by chemoattenuated PfSPZ vaccine.', Nature.

Morgan, W. D. *et al.* (1999) 'Solution structure of an EGF module pair from the Plasmodium falciparum merozoite surface protein 1.', *Journal of molecular biology*, 289(1), pp. 113–122.

Mueller, A.-K. *et al.* (2005) 'Plasmodium liver stage developmental arrest by depletion of a protein at the parasite-host interface.', *Proceedings of the National Academy of Sciences*, 102(8), pp. 3022–3027.

Mueller, I. *et al.* (2009) 'Key gaps in the knowledge of Plasmodium vivax, a neglected human malaria parasite.', *The Lancet. Infectious diseases*, 9(9), pp. 555–566.

Nerlich, A. G. et al. (2008) 'Plasmodium falciparum in ancient Egypt.', Emerging Infectious Diseases, 14(8), pp. 1317–1319.

O'Donnell, R. A. *et al.* (2000) 'Functional conservation of the malaria vaccine antigen MSP-119across distantly related Plasmodium species.', *Nature medicine*, 6(1), pp. 91–95.

Ogutu, B. R. *et al.* (2009) 'Blood stage malaria vaccine eliciting high antigen-specific antibody concentrations confers no protection to young children in Western Kenya.', *PloS one*, 4(3), p. e4708.

Olotu, A. et al. (2016) 'Seven-Year Efficacy of RTS,S/AS01 Malaria Vaccine among Young African Children.', New England Journal of Medicine, 374(26), pp. 2519–2529.

Ouattara, A. et al. (2010) 'Lack of allele-specific efficacy of a bivalent AMA1 malaria vaccine.', Malaria Journal, 9(1), p. 175.

Pain, A. et al. (2008) 'The genome of the simian and human malaria parasite Plasmodium knowlesi.', Nature, 455(7214), pp. 799-803.

Palacpac, N. M. Q. *et al.* (2013) 'Phase 1b randomized trial and follow-up study in Uganda of the blood-stage malaria vaccine candidate BK-SE36.', *PloS one*, 8(5), p. e64073.

Pan, W. *et al.* (1999) 'Vaccine candidate MSP-1 from Plasmodium falciparum: a redesigned 4917 bp polynucleotide enables synthesis and isolation of full-length protein from Escherichia coli and mammalian cells.', *Nucleic acids research*, 27(4), pp. 1094–1103.

Paul, G. *et al.* (2018) 'Protein-protein interaction studies reveal the Plasmodium falciparum-Merozoite surface protein-1 region involved in complex formation that binds to human erythrocytes.', *The Biochemical journal*, p. BCJ20180017.

Perrin, L. H. et al. (1984) 'Antimalarial immunity in Saimiri monkeys. Immunization with surface components of asexual blood stages.', *The Journal of experimental medicine*, 160(2), pp. 441–451.

Pizarro, J. C. *et al.* (2003) 'Crystal structure of a Fab complex formed with PfMSP1-19, the C-terminal fragment of merozoite surface protein 1 from Plasmodium falciparum: a malaria vaccine candidate.', *Journal of molecular biology*, 328(5), pp. 1091–1103.

del Portillo, H. A. *et al.* (1991) 'Primary structure of the merozoite surface antigen 1 of Plasmodium vivax reveals sequences conserved between different Plasmodium species.', *Proceedings of the National Academy of Sciences of the United States of America*, 88(9), pp. 4030–4034.

Putrianti, E. D. et al. (2009) 'Vaccine-like immunity against malaria by repeated causal-prophylactic treatment of liver-stage Plasmodium parasites.', Journal of Infectious Diseases, 199(6), pp. 899–903.

Ramasamy, R. (1998) 'Molecular basis for evasion of host immunity and pathogenesis in malaria.', *Biochimica et biophysica acta*, 1406(1), pp. 10–27.

Rayner, J. C. *et al.* (2001) 'A Plasmodium falciparum homologue of Plasmodium vivax reticulocyte binding protein (PvRBP1) defines a trypsin-resistant erythrocyte invasion pathway.', *The Journal of experimental medicine*, 194(11), pp. 1571–1581.

Roestenberg, M. et al. (2009) 'Protection against a malaria challenge by sporozoite inoculation.', New England Journal of Medicine, 361(5), pp. 468-477.

Roestenberg, M. et al. (2011) 'Long-term protection against malaria after experimental sporozoite inoculation: an open-label follow-up study.', *Lancet (London, England)*, 377(9779), pp. 1770–1776.

Rosenberg, R. et al. (1990) 'An estimation of the number of malaria sporozoites ejected by a feeding mosquito.', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 84(2), pp. 209–212.

Rosenberg, R. (2007) 'Plasmodium vivax in Africa: hidden in plain sight?', *Trends in parasitology*, 23(5), pp. 193–196.

Ross, R. (1898) 'The role of the mosquito in the evolution of the malaria parasite', The Lancet, (ii), p. 489.

RTS, S. C. T. P. (2015) 'Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial', *The Lancet*, 386(9988), pp. 31–45.

Rudolph, R. and Lilie, H. (1996) 'In vitro folding of inclusion body proteins.', FASEB journal: official publication of the Federation of American Societies for Experimental Biology, 10(1), pp. 49–56.

Russell, B. *et al.* (2011) 'A reliable ex vivo invasion assay of human reticulocytes by Plasmodium vivax.', *Blood*, 118(13), pp. e74--81.

Russo, G. *et al.* (2017) 'Molecular evidence of Plasmodium vivax infection in Duffy negative symptomatic individuals from Dschang, West Cameroon.', *Malaria Journal*, 16(1), p. 74.

Saitou, N. and Nei, M. (1987) 'The neighbor-joining method: a new method for reconstructing phylogenetic trees.', *Molecular biology and evolution*, 4(4), pp. 406–425.

Sanders, P. R. *et al.* (2005) 'Distinct protein classes including novel merozoite surface antigens in Raft-like membranes of Plasmodium falciparum.', *The Journal of biological chemistry*, 280(48), pp. 40169–40176.

Sandle, T. (2016) '11 - Endotoxin and pyrogen testing', in Sandle, T. (ed.) *Pharmaceutical Microbiology*. Oxford: Woodhead Publishing, pp. 131–145.

Sheehy, S. H. *et al.* (2012) 'ChAd63-MVA-vectored blood-stage malaria vaccines targeting MSP1 and AMA1: assessment of efficacy against mosquito bite challenge in humans.', *Molecular therapy: the journal of the American Society of Gene Therapy*, 20(12), pp. 2355–2368.

Shio, M. T. *et al.* (2010) 'Innate inflammatory response to the malarial pigment hemozoin.', *Microbes and infection*, 12(12–13), pp. 889–899.

Sinden, R. E. and Billingsley, P. F. (2001) 'Plasmodium invasion of mosquito cells: hawk or dove?', *Trends in parasitology*, 17(5), pp. 209-212.

Singh, A. *et al.* (2015) 'Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process.', *Microbial cell factories*, 14(1), p. 41.

Singh, B. *et al.* (2004) 'A large focus of naturally acquired Plasmodium knowlesi infections in human beings.', *Lancet* (*London, England*), 363(9414), pp. 1017–1024.

Singh, B. and Daneshvar, C. (2013) 'Human infections and detection of Plasmodium knowlesi.', *Clinical microbiology reviews*, 26(2), pp. 165–184.

Sirima, S. B. *et al.* (2016) 'A phase 2b randomized, controlled trial of the efficacy of the GMZ2 malaria vaccine in African children.', *Vaccine*, 34(38), pp. 4536–4542.

Sturm, A. *et al.* (2006) 'Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids.', *Science (New York, N.Y.)*, 313(5791), pp. 1287–1290.

Suwanarusk, R. et al. (2004) 'The deformability of red blood cells parasitized by Plasmodium falciparum and P. vivax.', *Journal of Infectious Diseases*, 189(2), pp. 190–194.

Szarfman, A. et al. (1988) 'Allelic forms of gp195, a major blood-stage antigen of Plasmodium falciparum, are

expressed in liver stages.', The Journal of experimental medicine, 167(1), pp. 231-236.

Tanabe, K. et al. (1987) 'Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum.', Journal of molecular biology, 195(2), pp. 273–287.

Tanabe, K. et al. (2007) 'Allelic dimorphism-associated restriction of recombination in Plasmodium falciparum msp1.', Gene, 397(1-2), pp. 153–160.

Taylor, H. M., Grainger, M. and Holder, A. A. (2002) 'Variation in the expression of a Plasmodium falciparum protein family implicated in erythrocyte invasion.', *Infection and Immunity*, 70(10), pp. 5779–5789.

Teneza-Mora, N., Lumsden, J. and Villasante, E. (2015) 'A malaria vaccine for travelers and military personnel: Requirements and top candidates.', *Vaccine*, 33(52), pp. 7551–7558.

Thomson-Luque, R. *et al.* (2017) 'A Continuous, Long-Term Plasmodium vivax In Vitro Blood-Stage Culture: What Are We Missing?', *Trends in parasitology*, 33(12), pp. 921–924.

Tian, J. H. et al. (1997) 'Comparison of protection induced by immunization with recombinant proteins from different regions of merozoite surface protein 1 of Plasmodium yoelii.', *Infection and Immunity*, 65(8), pp. 3032–3036.

Tolle, R. *et al.* (1993) 'A prospective study of the association between the human humoral immune response to Plasmodium falciparum blood stage antigen gp190 and control of malarial infections.', *Infection and Immunity*, 61(1), pp. 40–47.

Tournamille, C. et al. (1995) 'Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals.', *Nature genetics*, 10(2), pp. 224–228.

Trager, W. and Jensen, J. B. (1976) 'Human malaria parasites in continuous culture.', *Science (New York, N.Y.)*, 193(4254), pp. 673–675.

Tuteja, R. (2007) 'Malaria - an overview.', The FEBS journal, 274(18), pp. 4670-4679.

Varshavsky, A. (1997) 'The N-end rule pathway of protein degradation.', Genes to cells: devoted to molecular & cellular mechanisms, 2(1), pp. 13–28.

Vogel, G. (2013) 'The forgotten malaria.', Science (New York, N.Y.), 342(6159), pp. 684-687.

Watson, J. et al. (2018) 'Implications of current therapeutic restrictions for primaquine and tafenoquine in the radical cure of vivax malaria.', *PLoS neglected tropical diseases*, 12(4), p. e0006440.

White, N. J. (2011) 'Determinants of relapse periodicity in Plasmodium vivax malaria.', *Malaria Journal*, 10(1), p. 297.

Williams, T. N. (2006) 'Human red blood cell polymorphisms and malaria.', *Current opinion in microbiology*, 9(4), pp. 388–394.

Withers-Martinez, C. *et al.* (2012) 'Plasmodium subtilisin-like protease 1 (SUB1): insights into the active-site structure, specificity and function of a pan-malaria drug target.', *International journal for parasitology*, 42(6), pp. 597–612.

Woehlbier, U. et al. (2006) 'Analysis of Antibodies Directed against Merozoite Surface Protein 1 of the Human Malaria Parasite Plasmodium falciparum', Infection and Immunity, 74(2), pp. 1313–1322.

World Health Organization (2017) World malaria report 2017.

Wright, P. E. and Dyson, H. J. (1999) 'Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm.', *Journal of molecular biology*, 293(2), pp. 321–331.

Wright, P. E. and Dyson, H. J. (2015) 'Intrinsically disordered proteins in cellular signalling and regulation.', *Nature reviews. Molecular cell biology*, 16(1), pp. 18–29.

Wu, Y. et al. (2015) 'Development of malaria transmission-blocking vaccines: from concept to product.', Advances in parasitology, 89, pp. 109–152.

Yeoh, S. *et al.* (2007) 'Subcellular discharge of a serine protease mediates release of invasive malaria parasites from host erythrocytes.', *Cell*, 131(6), pp. 1072–1083.

Zimmerman, P. A. et al. (1999) 'Emergence of FY*A(null) in a Plasmodium vivax-endemic region of Papua New Guinea.', *Proceedings of the National Academy of Sciences*, 96(24), pp. 13973–13977.

Zimmerman, P. A. *et al.* (2013) 'Red blood cell polymorphism and susceptibility to Plasmodium vivax.', *Advances in parasitology*, 81, pp. 27–76.

Zuckerkandl, E. and Pauling, L. (1965) 'Evolutionary Divergence and Convergence in Proteins', in Bryson, V. and Vogel, H. J. (eds) *Evolving Genes and Proteins*. Academic Press, pp. 97–166.

6. Appendix

6.1. Multiple sequence alignment

CLUSTAL O(1.2.4) multiple sequence alignment

<pre>falciparum_FCB-1 falciparum_3D7 berghei_ANKA knowlesi vivax_Sal1 vivax_Belem falciparum_FCB-1 falciparum_3D7 berghei_ANKA knowlesi vivax_Sal1 vivax_Belem</pre>	<pre>signal peptide MKIIFFLCSFLFFIINTQCVTHESYQELVKKLEALEDAVLTGYSLFQKEKMVLNEGTSGT MKIIFFLCSFLFFIINTQCVTHESYQELVKKLEALEDAVLTGYSLFQKEKMVLNEEEITT MKVIGLLFSFVFFAIKCKSETIEVYNDIIQKLEKLESLSVEGLELFQKSQVIINASPP MKALLFLFSLIFFVTKCQCET-EDYKQLLVKLDKLEGLVVDGYELFHKNKISLDNIDAVQ MKALLFLFSFIFFVTKCQCET-ESYKQLVAKLDKLEALVVDGYELFHKKKLGENDIKVET MKALLFLFSFIFFVTKCQCET-ESYKQLVAKLDKLEALVVDGYELFHKKKLGENDIKVDA ** : :* *::** :: :* *::: **: ** : * .**:*:: AVTTSTPGSKG-SVASGGSGGSVASGGSVASGGSVASGGSVASGGSG</pre>	60 60 58 59 59 59 106 120 59 59 63 61
falciparum_FCB-1 falciparum_3D7 berghei_ANKA knowlesi vivax_Sal1 vivax_Belem	-NSRRTNPSDNSSDSDAKSYADLKHRVRNYLLTIKELKYPQLFDLTNHMLTLCDNIHGFK NTSSGASPPADASDSDAKSYADLKHRVRNYLFTIKELKYPELFDLTNHMLTLCDNIHGFK ETINPFSDNTFAPKLQGFITKFEELGFTEQTELVNLIKTLVPNKYGLK NIDGNNVNALAYKIRDIVGKYLELQIPGHGNLLHMIRELALDANGLK NNNNNNQVSVLTSKIRNFLSKFLELQIPGHTDLLHLIRELAFEPNGIK : : : ::::: : ** :: * : * : *:	165 180 107 106 111 109
falciparum_FCB-1 falciparum_3D7 berghei_ANKA knowlesi vivax_Sal1 vivax_Belem	YLIDGYEEINELLYKLNFYFDLLRAKLNDVCANDYCQIPFNLKIRANELDVLKKLVFGYR YLIDGYEEINELLYKLNFYFDLLRAKLNDVCANDYCQIPFNLKIRANELDVLKKLVFGYR YLIESKEEFNGLMHAINFYYDVFRDKLNDMCANNYCEIPEHLKISEEETEMLKKVILGYR YLVENYEEFNQLMHVINFNYDLLRAKLNDMCAHEYCKIPEHLKISAKELDMLKKVVLGYR YLVESYEEFNQLMHVINFHYDLLRAKLHDMCAHDYCKIPEHLKISDKELDMLKKVVLGYR YLVESYEEFNQLMHVINFHYDLLRAKLHDMCAHDYCKIPEHLKISDKELDMLKKVVLGYR **::. **:* *:: :** :*::* **:**:**:**	225 240 167 166 171 169
falciparum_FCB-1 falciparum_3D7 berghei_ANKA knowlesi vivax_Sal1 vivax_Belem	KPLDNIKDNVGKMEDYIKKNKKTIENINELIEESKKTIDKNK KPLDNIKDNVGKMEDYIKKNKTTIANINELIEGSKKTIDQNK KPIENIQDDIEKLEIYITKNKETVTALNTLIAEETKKITPENEAD	267 282 212 223 231 226
falciparum_FCB-1 falciparum_3D7 berghei_ANKA knowlesi vivax_Sal1 vivax_Belem	NATKEEEKK NATKEEEKK NADNEEGKK NTGTPVAAAAGAAAAAVPGAIASPSPVESSTPENYDQKK ASSETTQISGSSNSGSSSTGSSNSGSSSTGSSGTGSTGTGQSPPAAADASSTNANYEAKK GSTGNTETGTRSSASSNTLSGGDGTTVVGTSSPAPAAPSSTNEDYDEKK **	276 291 227 262 291 275
falciparum_FCB-1 falciparum_3D7 berghei_ANKA knowlesi vivax_Sal1 vivax_Belem	KLYQAQYDLSIYNKQLEEAHNLISVLEKRIDTLKKNENIKELLDKINEIKNPPPANSGNT KLYQAQYDLSIYNKQLEEAHNLISVLEKRIDTLKKNENIKKLLDKINEIKNPPPANSGNT IIYQAMYNVIFYKKQLAEIQKVIETLEKRVSALKKNDVIKPLLQQIEDIKAAPVTTEGQI VIFQAIYNFIFYTNQLEEAQKLMQVLEKRVKLLKEHKSIKALLEQIATEKNNLTTNNA IIYQAVYNTIFYTNQLQEAQKLIAVLEKRVKVLKEHKDIKVLLEQVAKEKEKLPSDYP KIYQAMYNGIFYTSQLEEAQKLIEVLEKRVKVLKEHKGIKALLEQVEAEKKKLPKDNT ::** *: :*** * :::: .****: **:: **	336 351 287 320 349 333
falciparum_FCB-1 falciparum_3D7 berghei_ANKA knowlesi vivax_Sal1 vivax_Belem	TTSGQSSTEPASTGTPSSGEVSTGTSTGGASAGVTNTGAATTGTTGTGAATTGTTGAEAV	336 351 347 320 349 333

falciparum ECB-1		370		
		205		
alciparum_3D7PNTLLDKNKKIEEHEEKIKEIAKTIKFNIDSLFT				
berghei ANKA	TTGNTGAEAATTGNTNTEVTQVQTVPTLTPEEKKKKMDGLYAQIKEIAKTIKFNLDGIFV	407		
knowlesi		359		
		200		
vivax_Sall	EARTEALARTONFDLDGLFT	388		
vivax_Belem	SQLVANAKTVNFDLDGLFT	373		
_	:.*: :* ***::*::*.			
falciparum FCB-1	DPLELEYYLREKNKNIDISAKVETKESTEPNEYPNGVTYPLSYNDINNAL-	420		
falciparum 3D7		130		
		439		
berghei_ANKA	NPIELEYFKKEKKKESCNLSTSSCKKNKTSETIIPLNVRYPNGIGYPLPENDVYNKIA	465		
knowlesi	NVEELEYYLREKAKMAGTLIGPESSQSTGTPGKAVPTLKETYPYGITYALPERTIYELIE	419		
vivax Sall	DAEELEYYI.REKAKMAGTI.TIPESTKSAGTPCKTVPTI.KETYPHCISYAI.AENSIYEI.TE	448		
vivan_baii		122		
vivax_Belen	DALLLYILKEKAKMAGILIIPESIKSAGIPGKIVPILKETIPHGISIALAENSIILLIE	433		
	· ***· · ** * · · · · · · · · · · · · ·			
falcıparum_FCB-1	NELNSFGDLINPFDYTKEPSKNIYTDNERKKFINEIKEKIKIEKKKIESDKKSYEDRS	478		
falciparum 3D7	ADNDKNSYGDLMNPHTKEKINEKIITDNKERKIFINNIKKKIDLEEKNINHTKEQNKKLL	499		
herchei ANKA	NNAAETTYGDI. TNPDNTPI – TEDI. ATNEOARKNI, IKA IKKKI EAEEKNI. ESI. KTNYDNKI.	524		
pergner_mun		170		
knowlesi	KFGSEESFGDLQNPDNGKQPNKGIIINETKKKTLVDKIMSKIKLEEEKLPKLKKEIDEKM	4/9		
vivax Sall	KIGSDETFGDLQNPDDGKQPKKGILINETKRKELLEKIMNKIKIEEDKLPNLKKEYEEKY	508		
vivax Belem	KIGSDETFGDLONPDDGKOPKKGILINETKRKELLEKIMNKIKIEEDKLPNLKKEYEEKY	493		
	··*** ** · · · * ** · · · · · · · · · ·			
folginorum ECP-1		530		
	KS LINDI I KE I EKILINE I I DSKE NNN I DLI NE EKEMIGKKI SI KVEKLI HINI I FASI ENSKR	550		
Ialciparum_3D7	EDYEKSKKDYEELLEKFYEMKFNNNFNKDVVDKIFSARYTYNVEKQRYNNKFSSSNNSVY	559		
berghei ANKA	ASFNOOKAPFKEAAKLFYESKFGNKLTSDIFEKFKTORTEYMNKKTELENCLYGNTKO	582		
knowlosi		537		
KIIOWIESI		557		
vivax_Sall	KVYEAKVNEFKPAFNHFYEARLDNTLVENKFDDFKKKREAYMEEKKKLESCSYEQNSN	566		
vivax Belem	KVYEAKVNEFKPAFNHFYEARLDNTLVENKFDEFKTKREAYMEEKKKLESCSYEONTN	551		
—	: : :*: ::.* : ::: * * :* :. : :.			
falciparum FCB-1	NLEKLTKALKYMEDYSLENIVVEKELKYYKNLISKIENEIETLVENIKKDEEOLFEKKIT	598		
falainanum 207		C1 0		
Tarciparum_SD/	NVQALAAALSILEDISLAAGISEADINHIIILAIGLEADIAALIEEIASSENAILEANEA	019		
berghei ANKA	LISKLNKQLNYLQDYSLRKDIISDEIEYFSNKKKGLQYNINRLAEAVQAKQNILVA	638		
knowlesi	LINKLKKOLTYLVDYTLKKDVTEDEINYFSDLEWKLKNEIYELAKEVRKNENKLIM	593		
wiway Call		622		
VIVAX_SAII		022		
vivax_Belem	LINKLKKQLTYLEDYVLRKDIADDEIKHFSFMEWKLKSEIYDLAQEIRKNENKLTV	607		
	· · * · * · · · · · · · · · · · · · · ·			
falciparum_FCB-1	KDENKPDEKILEVSDIVKVQVQKVLLMNKIDELKKTQLILKNVELKHNIHVPNSYKQENK	658		
falciparum 3D7	G-LTHSANGSLEVSDIVKLOVOKVLLIKKIEDLRKIELFLKNAOLKDSIHVPNIYKPONK	678		
horahoj ANKA		691		
Dergiler_ANIXA		0.91		
knowlesi	ENKFDFSGVLELQIHKVLMIKKIGALKNVQNLLKNAKLKDKLYIPKVYKTGQK	646		
vivax Sall	ENKFDFSGVVELQVQKVLIIKKIEALKNVQNLLKNAKVKDDLYIPKVYKTSEK	675		
vivay Belem		660		
VIVAX_DETEN		000		
	· · · · · · · · · · · · · · · · · · ·			
	03/30			
falginarum EGD 1		700		
Ialcipalum_FCB-I	QEFILLIVLAREIDALAVIMERVESLINEERANIK <mark>1</mark> <mark>EGQ5DN</mark> SEFSIEGEI	109		
talciparum_3D7	PEPYYLIVLKKEVDKLKEFIPKVKDMLKKEQAVLSSITQP <mark>LVAASET</mark> TEDGGH	731		
berghei ANKA	PEPYYLIAVKREVDRLAOFIPKIENMIAKEKEKTEOVPVVTGESEETSSVST	743		
knowlogi	ρεργγιτιζί μνετρικί μρετρικτές τη	704		
KIIOWIESI		704		
vivax_Sall	PEPYYLMVLKREIDKLKDFIPKIESMIATEKNKPTVAAADIVAKGQS <mark>LRGASET</mark> GTTGNT	735		
vivax Belem	PEPYYLMVLKREIDKLKDFIPKIESMIATEKAKPAASAPVTSGQL <mark>LRGSSEA</mark> ATEVTT	718		
—	*****:.:*:*:* *:*:*: *: *: *: *: *:			
falciparum FCB-1	AGSALEG-DSVQAQA	734		
falciparum 3D7		759		
harabai ANUA		700		
Dergner_ANKA	evVTPVTEEAQ-SSQNAP	183		
knowlesi	TATESGSTTSASTAVQQPTQQAAQAAQAASPVTVTQPTETVTQTPAPATETAGEAAQETS	764		
vivax Sall	VNAQTAVVQPQHQVVNAVTV-QPGTTGHOAOGGEAETOTN	774		
vivax Belem	QQQQQQQQQQQQQQQQQQ	744		
_	:			
falciparum_FCB-1	QEQKQAQPPVPV	746		
falciparum 3D7	TVTITLPPTOPS	771		
herahei ANKA		8 N /		
linerilect		004		
VIIOMTERT	IV SI TAFAV V SEAGIEGEEGIEV VAQFEAASGEIQIPTPGAV DASPAAPVPAGTPGTTD	024		
vivax_Sall	svqaTqtisqa	801		
vivax_Belem	QQQQTQPTSQS	771		

121

falciparum_FCB-1 falciparum_3D7 berghei_ANKA knowlesi vivax_Sal1 vivax_Belem	PVPEAKAQVPTPPAPVNNKTENVSKLDYLEKLYEFLNTSYICHKYILVSHST PPKEVKVVENSIEHKSNDNSQALTKTVYLKKLDEFLTKSYICHKYILVSNSS TTPEAATTSTTTSTTTSTTTSTTTPVMTKLYYLEKLQKFLVFSYSCHKYLLLQNST AAPEASVPAPAGSALPATTAPAAAAPAAPAMSKLEYLEKLLEFLKSSYACHKHIFVTNST PAPTQASPEPAPAAPPSTPAAAVAPAPTMSKLEYLEKLLDFLKSAYACHKHIFVTNST AAPGVSATPAPTPAAAAAPAPAMSKLEYLEKLLDFLKSAYACHKHIFVTNST : ::* **:** .** :* ***:::: :*:	798 823 860 884 859 823
falciparum_FCB-1 falciparum_3D7 berghei_ANKA knowlesi vivax_Sal1 vivax_Belem	MNEKILKQYKITKEEESKLSSCDPLDLLFNIQNNIPVMYSMFDSLNNSLSQLFMEIYE MDQKLLEVYNLTPEEENELKSCDPLDLLFNIQNNIPAMYSLYDSMNNDLQHLFFELYQ INKDALSKYALTTEEDKIRTLKRCSELDILLAIQNNMPTMYSLYESIVDGLQNIYAELYE MNPELLKQYALTTDEEKKIKESACDELDLLFNVQNNLPSMYSIYDTMINDLQNLYIELYQ MKKELLDQYKLNADEQNKINETKCDELDLLFNVQNNLPAMYSIYDSMSNELQNLYIELYQ MDKKLLKEYELNADEKTKINQNKCDELDLLFNVQNNLPAMYSIYDSMSNELQNLYIELYQ :*.*:::::::::::::::::::::::::::::::	856 881 920 944 919 883
falciparum_FCB-1 falciparum_3D7 berghei_ANKA knowlesi vivax_Sal1 vivax_Belem	KEMVCNLYKLKDNDK-IKNLLEEAKKVSTSVKTLSSSSMQPLSL KEMIYYLHKLKEENH-IKKLLEEQKQ <mark>ITGTSST</mark> SSPGNTTVNTA KEMMYHIYKLKDENPSIKSILVKAGVIDPEPVASPPTPPTPPTPPTPPTPPTLPTTPT KEMVYNIYKNKDTDTKIKAFLETLKSNAASVTPAVVPAAAPVVTPAPAEPVVTPA KEMVYNIYKNKDTDKKIKAFLETLKSKAAAPAQSAAKP- KEMVYNIYKNKDTDKKIKAFLETSNNKAAAPAQSAAKP- ***: ::* *: : ** :*	899 924 978 999 957 921
falciparum_FCB-1 falciparum_3D7 berghei_ANKA knowlesi vivax_Sal1 vivax_Belem	TDKPE <mark>V</mark> -SANDD QSATHSNSQNQQSNASSTNTQNGVAVSSGPAVVEESHD PVTPAAPS-EQTTTPEAATAASNPGASASETPASNPGASASSTPSTPAAASNPGASASST PAP-GQA-APAAAPTTTNPSTTPSGTTTNAVS S-GQA-GTTP-VTTTAPVTTAPVTT	912 962 1037 1029 974 938
falciparum_FCB-1 falciparum_3D7 berghei_ANKA knowlesi vivax_Sal1 vivax_Belem	30/38 TSHSTNLNNSLKLFENILSLGKNKNIYQ-ELIGQK PLTVLSISNDL	946 997 1094 1072 1014 978
<pre>falciparum_FCB-1 falciparum_3D7 berghei_ANKA knowlesi vivax_Sal1 vivax_Belem falciparum_FCB-1 falciparum_3D7</pre>	SSENFYEKILKDSDTFYNESFTNFVKSKADDINSLNDESKRKKLEEDINKLKKTLQLSFD EMEKFYENILKNNDTYFNDDIKQFVKSNSKVITGLTE-TQKNALNDEIKKLKDTLQLSFD ELDNLYKSYLQRIDG-NNTEFINFIKSKKELIKALTP-EKVNQLYLEIAHLKELSEHYYD NVEKIYEKHLSQMDK-YNDYFIKFLESQKEKITSMTE-EQANALGAEIEALKKKVQVSLD QVEKFYEKHLSQIDK-YNDYFQKFLESQKDEITKMDE-TKWKALGAEIEELKKKLQVSLD QVEKFYEKHLSQIDK-YNDYFKKFLESKKEEIIKMDD-TKWNALGKEIEELKKKLQVSLD . :::*:. *. * * : :*::*: * : : : * :* **. : * LYNKYKLKLERLFDKKKTVGKYKMQIKKLTLLKEQLESKLNSLNNPKHVLQNFSVFFNKK	1006 1056 1152 1130 1072 1036 1066 1116
berghei_ANKA knowlesi vivax_Sal1 vivax_Belem	RYYKYKLKLERLYEKHEQIQVSNRQIRELSILKARLLKRKQNINGIFYILSGYVNFFNKR HYGKYKLKLERFLEKKNKISNSKEHIKKLTSLKNKLERKLNFLNNPTSVLKNYIIFFNKK HYGKYKLKLERLLKKKNKISNSKDQIKKLTSLKNKLERRQNLLNNPTSVLKNYTAFFNKK * ******:*: .*:: : ::::::::::::::::::::	1212 1190 1132 1096
falciparum_FCB-1 falciparum_3D7 berghei_ANKA knowlesi vivax_Sal1 vivax_Belem	KEAEIAETENTLENTKILLKHYKGLVKYYNGESSPLKTLSEESIQTEDNYASLENFKVLS KEAEIAETENTLENTKILLKHYKGLVKYYNGESSPLKTLSEVSIQTEDNYANLEKFRVLS READKQYVDNALKNNDMLLKYYKARIKYFTSEAVPLKTLTKASLDRETNYLKIEKFRAYS KEAEKKEVENTLKNTEILLKYYKARAKYYIGEPFPLKTLSEESLQKEDNYLNLEKFRVLS RETEKKEVENTLKNTEILLKYYKARAKYYIGEPFPLKTLSEESMQKEDNYLNLEKFRVLS :*:: .:*:*:*:*:*:*: **: **: **: **: **:	1126 1176 1272 1250 1192 1156
falciparum_FCB-1 falciparum_3D7 berghei_ANKA knowlesi vivax_Sal1 vivax_Belem	KLEGKLKDNLNLEKKKLSYLSSGLHHLIAELKEVIKNKNYTGNSPSENNTDVNNALESYK KIDGKLNDNLHLGKKKLSFLSSGLHHLITELKEVIKNKNYTGNSPSENNKKVNEALKSYE RLELRFKKNINLGKEKISYVSGGLYHVFEEFKELLKNKNYTGKTNPDTVPEVTNAFEQYK RMEGRLGNNINLEKENISYLSSGLHHVFTELKEIIKNKKYTGNDHAKNTTAVKEALQAYE RLEGRLGKNIELEKENISYLSSGLHHVLTELKEIIKNKKYSGNDHTKNIAAVKEALQAYQ RLEGRLGKNIELEKENISYLSSGLLHVLTELKEIINDKKYSGKDHAKNIAEVKKALQAYQ ::: :: .*:.* *:::*::*:*: *:: *::*::*::*::*:	1186 1236 1332 1310 1252 1216

	Appendix	
falciparum_FCB-1	KFLPEGTDVATVVSESGSDTLEQ	1209
falciparum_3D7	NFLPEAKVTTVVTPPQPDVTPSPLSVRVSGSSGSTK	1272
berghei ANKA	ELLPKGVTASASPAAATTPT-SADAATQRATPESRSGSGSG	1372
knowlesi	ELLPKVATOTASLPPVAPPAVVPPVVPEAEAEAEAEAEAEPATSTOPATADTAAP	1365
vivax Sall	ELIPKVTTOEGASTTAATLPVTVPSAVPGGLPGAGVPGAAAGLTPP	1298
wiway Bolom		1271
VIVAX_BETEW	LLIPKVISQESISVAVIVPGAVVPGVPIAAAAGSGASGAVPPAAAAGSGASGAVP	12/1
	:::*: :.	
falciparum_FCB-1	S-QPKORTITTSQNVDDEVDDVIIVPIFG	1247
falciparum 3D7	E-ETQIPTSGSLLTELQQVVQLQNYDEEDDSLVVLPIFG	1310
berghei ANKA	RSGSGENAVVSG	1397
knowlesi	TOTPAA PTOT PAE PATATATTGETAAA PAA PAA PAVEVONAE VKAOE YGED YDKVITI. P <mark>LFG</mark>	1425
vivax Sall		1341
wiway Bolom		1217
VIVAX_DETEN	IN GGISIINIGGVVI GVVESNENVINNVAVDINEDIDKVINDIDKG	1317
	38/42	
falciparum FCB-1	FSFEDY DDI COV <mark>VICE AVIT</mark> OSUI DNII SKIFNEVEVI VI KDI ACUVRSI KKOI FNNVM	1305
falainanum 207		1203
	ESEDNDEILDQVVIGEAISV-IMDNILSGFENEIDVIILFLAGVIKSLAKQIEKNIF	1307
bergnei_ANKA	SSVDDNDDDDDDDDQIASGQSENA-QEKNILEAFKNESEYLYARSLGITYKSLKKHMIREFS	1456
knowlesi	<mark>NDED</mark> DVEDQEEKQIITGEAENA-QPENIVPEGINEYEVVYIKPLAGMYKSIKKQLENHVA	1484
vivax Sall	NNDDDGEEDQV <mark>TTGEAES</mark> E-APEILVPAGISDYDVVYLKPLAGMYKTIKKQLENHVN	1397
vivax Belem	NNDDDGEEDOV <mark>TTGEAES</mark> E-APEILVPAGISDYDVVYLKPLAGMYKTIKKOLENHVN	1373
—	······································	
falciparum FCB-1	TFNVNVKDILNSRFNKRENFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSY	1365
falciparum 3D7	TENLIN NUTLINGEL KKRKYFLOVI. ESDI MOFKHISSNEYI TEDSEKI LINGEOKNTI LKSY	1427
horaboi ANKA		1516
Derglier_ANKA		1510
knowlesi	AFNTNITDMLDSRLKKRNYFLDVLDSELNPFKYSSSGEYIIKDPYKLLDLEQKKKLLGSY	1544
vivax_Sall	AFNTNITDMLDSRLKKRNYFLEVLNSDLNPFKYSSSGEYIIKDPYKLLDLEKKKKLIGSY	1457
vivax Belem	AFNTNITDMLDSRLKKRNYFLEVLNSDLNPFKYSPSGEYIIKDPYKLLDLEKKKKLLGSY	1433
—	···· ···· *·····**· * ·**· ·* ···* ···	
falciparum FCB-1	NYIKDSIDTDINFANDVLGYYKILSEKYKSDLDSIKKYIND	1406
falciparum 3D7	KYIKESVENDIKFAOEGISYYEKVLAKYKDDLESIKKVIKEEKEKEPSSPP	1478
herchei ANKA		1563
In a set		1500
knowlesi	QIIGASVDKDLIIAKDGMEIINNMGELIKQHLEAVNAQIKEIEASVPGEQ	1594
vivax_Sall	KYIGASIDMDLATANDGVTYYNKMGELYKTHLDGVKTEIKKVEDDIKKQDEELKKLGNVN	151/
vivax_Belem	KYIGASIDKDLATANDGVTYYNKMGELYKTHLTAVNEEVKKVEADIKAEDDKIKKIGSDS	1493
	:* .:: $*:$: : : :: : $**$. $*$.:: :.	
falciparum_FCB-1	KQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKVLNYTYEKSNVEV	1455
falciparum 3D7	-TTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNKIDDYLINLKAKINDCNVEKDEAHV	1537
berghei ANKA	DKDEKKKYVPIFEDLKGLYET <mark>ILGOSEE</mark> YIEVLONRLDSYKNEKTEFEI	1612
knowlesi	SOLNAOKEELKKYLPFLNSIOKEYESLVNMAHTYKENLKKFINNCOIEKKETEI	1648
wiway gall		1577
vivax_baii	SUBSTITUTE TARGENERAL FERNEL OF SUBSTITUTE DATA AND A DEVELOPMENT	1552
VIVAX_BETEI	INTERIOSMARRAELERIUFFENSIORETSIVSKVNTITUDURKVINNCULERREAET	1000
		1 - 1 -
falciparum_FCB-1	KIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGMVFENLAKTVLSN	1515
falciparum_3D7	KITKLSDLKAIDDKIDLFKNPYDFEAIKKLINDDTKKDMLGKLLSTGLV-QNFPNTIISK	1596
berghei_ANKA	LTKNLETYIKIDEKLENFVENAENNKHIASIALNNLNKSGLVGEGESKKILAK	1665
knowlesi	IVKKLEDYTKIDENLEIYKKSKKESDVRSSGLLEKLKNSKLINEEESKKVLSQ	1701
vivax Sall	TVKKLODYNKMDEKLEEYKKSEKKNEVKSSGLLEKLMKSKLIKENESKEILSO	1630
vivax Belem	TVKKLODYNKMDEKLEEYKKSEKKNEVKSSGLLEKLMKSKLIKENESKEILSO	1606
VIVAX_Detem	· · · · · · · · · · · · · · · · · · ·	1000
falciparum FCB-1	LLDGNLOGMLNIS-OHOCVKK-OCPONSGCFRHLDEREECKCLLNYKOEGDKCVENPNPT	1573
falciparum 3D7	LIEGKEODMINIS-OHOCVKK-OCPENSCCERHIDEREECKCLINYKOEGDKCVENPNPT	1654
horahoi ANKA		1705
		1723
knowlesi	LLNVQT-QMLNMSSAHKCIDT-NVPENAACYRYLDGTEEWRCLLGFKEVGGKCVPAS-IT	1/58
vivax_Sall	LLNVQT-QLLTMSSEHTCIDT-NVPDNAACYRYLDGTEEWRCLLTFKEEGGKCVPASNVT	1688
vivax_Belem	LLNVQT-QLLTMSSEHTCIDT-NVPDNAACYRYLDGTEEWRCLLTFKEEGGKCVPASNVT	1664
	···· · · · · · · · · · · · · · · · · ·	
folginger ECD 1	GPI anchor signal seque	ence
rarciparum_FCB-1	CNENNGGUDADAKUTEED5G5NGKKITUEUTKPD5YPLFDG1FCSSSNFLG1SFLL1LML	1033
talciparum_3D7	CNENNGGCDADATCTEEDSGSSRKKITCECTKPDSYPLFDGIFCSSSNFLGISFLLILML	1714
berghei_ANKA	CGNNNGGCDPTAGCQTAENRENSKKIICTCKEPTPNAYYDGVFCSSSSFMGLSILLIITL	1785
knowlesi	CEENNGGCAPEAECTMDDKKEVECKCTKEGSEPLFEGVFCSSSSFLSLSFLLLILI	1814
vivax Sall	CKDNNGGCAPEAECKMTDSNKIVCKCTKEGSEPLFEGVFCSSSSFLSLSFLLLMLL	1744
vivax Belem	CKDNNGGCAPEAECKMTDSNKIVCKCTKEGSEPLFEGVFCSSSSFLSLSFLLLMLL	1720
		. = 5

* :**** * * . ::: * *.: ::*:********

falciparum FCB-1	ILYSFI*-	1639
falciparum 3D7	ILYSFI*-	1720
berghei ANKA	IVFNIF*-	1791
knowlesi	FFLSMEL*	1821
vivax Sall	FLLCMEL-	1751
vivax_Belem	FLLCMEL*	1727
—	:. :	

6.2. CD spectra of recombinant MSP-1 proteins



Figure 58. CD spectra and melting curve of *P. vivax* **MSP-1.** A) CD spectra were recorded at temperatures ranging from 10°C to 85°C. The individual CD spectra match the typical spectrum for a mostly alpha-helical protein. As seen with MSP-1D, PvMSP-1 remains partially folded even at very high temperatures. B) The calculated melting temperature based on CD222 values is 42.45°C.



Figure 59. CD spectra and melting curve of MSP-1F. A) The CD spectra recorded at temperatures ranging from 10°C to 85°C all correspond to an alpha-helical structure. B) The melting temperature of 43.96°C was calculated using a linear fit in Prism.



Figure 60. CD spectra and melting curve of *P. knowlesi* **MSP-1.** A) The obtained CD spectra of *P. knowlesi* MSP-1 also correspond to a protein with a mostly alpha-helical secondary structure. B) The melting temperature of 44.89°C was determined using a nonlinear fit in Prism using the 222 nm data from the thermal denaturation.



Figure 61. CD spectra and melting curve of a hybrid MSP-1 protein consisting of MSP-1D 38/42 and MSP-1F 83/30. A) The hybrid protein based on two *P. falciparum* strains displays an alpha-helical structure is still partially folded at very high temperatures. B) The melting temperature is similar to the ones determined for MSP-D and MSP-F. The calculated value based on thermal denaturation data is 42.92°C.
Appendix



Figure 62. Melting curve of MSP-1D after proteolytic processing by PfSUB1. MSP-D was proteolytically processed by PfSUB1 overnight before performing thermal denaturation. The CD222 data was plotted against temperature and the melting temperature was determined using a nonlinear fit in Prism. The result of this fit is a melting temperature of 44.67°C, which is one degree lower than the calculated melting temperature for the unprocessed protein.