

*-Project Title-*

**Feasibility, optimisation and scale-up studies for  
the production of *P. falciparum* putative  
vaccine candidates for a malaria  
microarray-based antigenicity screening.**

A thesis submitted to Imperial College London for the degree of  
Doctor of Philosophy  
in the Department of Life Sciences

by

**Giorgio Mazzoleni, BSc, MSc**

October 2011

Department of Life Sciences  
Division of Cell and Molecular Biology  
Imperial College London, South Kensington Campus  
London, SW7 2AZ



# Copyright

The copyright of this thesis rests with the author and no quotation or images from it or information derived from it may be published without the prior written consent of the author.

# Declaration of originality

I, Giorgio Mazzoleni, hereby declare that the work presented in this thesis is my own. All the work and data analysis of the results were performed by myself, unless otherwise specified in the text.

Giorgio Mazzoleni



# Acknowledgements

My sincere gratitude goes to Professor Andrea Crisanti and Microtest Matrices Ltd. for giving me the opportunity to work in their laboratories and for supporting me throughout the four years of this PhD. A special thanks goes to Tania Dottorini, as team leader and as a friend, for her essential support, for always having time to discuss things and for her encouragement to explore different ideas. I would also like to thank all the people involved in the FightMal project, as without them it would not be possible to progress in this monumental study. A special mention must go to Francesco Trivelli, Davide Pettinato and Carla Proietti for sharing with me the burden of hundreds of different PCRs, ligations, expressions and purifications. To Francesca Baldracchini for her help, support and friendship along all ups and downs of this long journey. I would also like to thank Walter Low and Julian Gray for their guidance and scientific contributions during the initial stages of this work. I am grateful to the PX'Therapeutics team, for their help and suggestions in setting up the purification protocols and for their warm welcome in their facility where I spent a few unforgettable months (thanks Walt for the BBQs). A special thanks goes to Lorenzo D'episcopo and all the other colleagues (past and present) at Microtest Matrices and in the Crisanti's lab for their invaluable input and time. Together we shared some forgettable and unforgettable moments and for that I thank you all. A special thanks goes to Mauro Maccari that introduced me to the fascinating world of  $\text{\LaTeX}$  and spent a few hours of his spare time (sorry Elena) helping me with some of the figures of this thesis (grazie!). I am also grateful to Clelia Supparo and Lucy Collins for their help in managing the admin side of my PhD, they have always given me their time and without them I would still be submerged in bureaucratically encrypted documents and forms. Finally, my most sincere gratitude goes to my family, my Mum and my Dad for allowing me to live this precious experience and for their complete support in everything I have ever done, grazie di cuore. Last, but not least, I would like to express my deepest thanks to Helena, for all her valuable time

spent reviewing this thesis (night and day), for her unconditional support and encouragement, and for always being there for me. Grazie per tutto quello che hai fatto per me amore, I could not have made it without you!

# ABSTRACT

Malaria is a tropical parasitic disease spread worldwide by protozoan parasites of the genus *Plasmodium*. Attempts to eradicate or control the disease have largely failed and the weapon-of-choice in combating malaria, a vaccine, still eludes us. Now that the *Plasmodium falciparum* (*Pf*) genome has been revealed, it is essential to search for new potential antigens. On that basis, my project aims to select and produce putative vaccine candidates that together with protein microarray techniques will unravel possible correlates of protection against the human malaria parasite *Pf*, responsible for the majority (90%) of deaths from malaria. My PhD project forms part of FightMal a consortium-based study that first use bioinformatic analysis to identify those proteins within the *Pf* proteome that are known or anticipated to be either anchored on the parasite/infected host cell surface or secreted. Such *in silico* analysis produced 3 ‘candidate lists’ of interest, one for each of the 3 types of protein chips planned for production: i) current vaccine candidates (n.25), ii) putative immune targets (n.260), iii) variant surface antigens, VSA (n.74). Candidates belonging to CHIP 2 were cloned and expressed (*in vivo*) using customized high-throughput platforms. To date, 394 (88%) Chip 2 DNA targets have been successfully amplified by PCR, 325 (82%) of these have been inserted into the expression plasmid and sequenced. A total of 269 constructs resulted suitable for heterologous expression. Due to the complexity of soluble expression of plasmodial proteins a number of different expression conditions were tested. Both the strain of the *E. coli* host and the composition of the expression medium resulted crucial variables for satisfactory results. Via a multi-step optimisation phase we increased the initial number of soluble candidates from 18% to 44%, which eventually led to a 6-fold increase in the number of targets available for purification. During an early evaluation study, 77% (10/13) of the recombinant proteins produced in large-scale were purified using embedded affinity tags yielding a  $\geq 90\%$  purity level. Preliminary results, obtained spotting 40 *in vitro* expressed *Pf* candidates, successfully

demonstrated the potentials of the protein microarray technology as an efficient serum-based multiplex assay capable of identifying new antigens. Using an optimised protocol, over 7000 antigen-antibody interactions were evaluated. 92% of the *Pf* protein panel was recognised by both total and cytophilic IgGs, with 6 novel candidates resulting as antigenic as some of the antigens evaluated in vaccine clinical trials for malaria. It is envisaged that my PhD will deliver multi-level technical know-how, which could improve the production and evaluation of novel plasmodial vaccine candidates. The final *Pf* microarray will be employed to screen clinical samples collected from protected and non-protected children enrolled in a longitudinal, case-control study in an area of Uganda where malaria is endemic. It is expected that the FightMal project overall will eventually generate novel immunological data and information pertaining to both the complexity of the plasmodial proteome and variability of the naturally-induced immune response against the malaria parasite.

# Contents

Copyright	2
Declaration of originality	3
Acknowledgements	4
Abstract	6
<b>ABBREVIATIONS</b>	<b>16</b>
Common abbreviations . . . . .	16
Amino acid abbreviations and properties . . . . .	18
Power prefix abbreviations . . . . .	19
<b>1 INTRODUCTION</b>	<b>20</b>
1.1 Malaria: the global situation and figures . . . . .	20
1.2 Biology of malaria parasite . . . . .	23
1.2.1 The <i>Plasmodium falciparum</i> . . . . .	25
<i>The Pf genome</i> . . . . .	25
<i>The life cycle of Pf</i> . . . . .	27
<i>Immune response to Pf</i> . . . . .	30
1.3 Malaria control and treatments . . . . .	36
1.3.1 Prevention of malaria transmission: vector control . . . . .	37
1.3.2 Diagnosis of malaria . . . . .	39
1.3.3 Current malaria treatments: antimalarial drugs . . . . .	41
1.3.4 Rational for a malaria vaccine . . . . .	45
1.4 Protein microarrays . . . . .	54
1.4.1 Protein microarrays: sub-classes . . . . .	56
1.4.2 Protein microarrays: use in serum profiling . . . . .	57
1.5 Aim of my PhD project . . . . .	60

<b>2</b>	<b>MATERIALS AND METHODS</b>	<b>62</b>
2.1	Overview . . . . .	62
2.2	Bioinformatic selection of <i>Pf</i> antigens . . . . .	62
2.3	Isolation and cloning of <i>Pf</i> candidates . . . . .	66
2.3.1	Polymerase Chain Reaction (PCR) . . . . .	67
2.3.2	Ligation into the pET160/GW/D-TOPO <sup>®</sup> vector . . . . .	69
2.3.3	Transformation of bacteria . . . . .	70
2.3.4	Colony PCR . . . . .	70
2.3.5	Miniprep of plasmid DNA . . . . .	71
2.3.6	Sequencing analysis . . . . .	71
2.3.7	Plasmid storage . . . . .	72
2.4	<i>In vivo</i> expression of the <i>Pf</i> candidates in <i>E. coli</i> . . . . .	72
2.4.1	Transformation of Rosetta-Gami <sup>TM</sup> B(DE3) . . . . .	72
2.4.2	Medium preparations for <i>Pf</i> protein expression . . . . .	73
2.4.3	<i>In vivo</i> expression of <i>Pf</i> candidates . . . . .	75
2.4.4	Bacterial lysis . . . . .	76
2.5	<i>In vitro</i> expression of the <i>Pf</i> candidates . . . . .	77
2.5.1	Expressway <sup>TM</sup> Plus <i>E. coli</i> system . . . . .	77
2.5.2	RTS wheat germ & rabbit reticulocyte system . . . . .	77
	<i>pIVEX</i> plasmid optimization and subcloning . . . . .	78
	<i>Expression using the two eukaryotic systems</i> . . . . .	79
2.6	Protein purification procedures . . . . .	80
2.6.1	Immunoaffinity chromatography: Magnetic beads . . . . .	81
2.6.2	Immobilised Metal Affinity Chromatography . . . . .	81
2.6.3	Gravity column chromatography . . . . .	83
2.6.4	Protein Liquid Chromatography . . . . .	84
2.7	Protein characterization procedures . . . . .	85
2.7.1	SDS-PAGE protein analysis . . . . .	85
2.7.2	Immunoblot analysis . . . . .	85
2.8	Protein microarray development . . . . .	86
2.8.1	Construction of protein microarray chips . . . . .	86
2.8.2	Processing of protein microarray chips . . . . .	87
	<i>Preparation of human serum samples</i> . . . . .	87
	<i>Processing of the printed slides</i> . . . . .	88
	<i>Quantitative analysis of image data</i> . . . . .	89

<b>3</b>	<b>RESULTS and DISCUSSION</b>	<b>90</b>
3.1	Overview . . . . .	90
3.2	Bioinformatic selection of the <i>Pf</i> antigens . . . . .	91
3.2.1	Bioinformatic selection: overview . . . . .	91
3.2.2	CHIP 1 candidate list . . . . .	93
3.2.3	CHIP 2 candidate list . . . . .	94
3.2.4	CHIP 3 candidate list . . . . .	102
3.2.5	Discussion . . . . .	105
3.3	Cloning of <i>Pf</i> CHIP 2 candidates . . . . .	110
3.3.1	Selection of the cloning platform . . . . .	110
3.3.2	Selection of N- and C- terminal tags . . . . .	114
3.3.3	Selection of the optimal amplification procedure . . . . .	116
3.3.4	High-throughput cloning of 260 gene targets . . . . .	119
3.3.5	Discussion . . . . .	122
3.4	<i>In vivo</i> expression of <i>Pf</i> candidates in <i>E. coli</i> . . . . .	128
3.4.1	Selection of the optimal <i>in vivo</i> expression platform . . . . .	128
	<i>Evaluation of the most suitable E. coli strain</i> . . . . .	128
	<i>Selection of the optimal expression medium</i> . . . . .	130
	<i>Characterisation of the TM expression medium</i> . . . . .	138
3.4.2	Mini-scale solubility analysis of 189 <i>Pf</i> candidates ex- pressed in TM medium . . . . .	145
3.4.3	Large-scale solubility analysis of 189 <i>Pf</i> candidates ex- pressed in TM medium . . . . .	145
3.4.4	Discussion . . . . .	147
3.5	<i>In vitro</i> expression of <i>Pf</i> candidates . . . . .	154
3.5.1	Prokaryotic cell-free system: <i>E. coli</i> . . . . .	154
3.5.2	Eukaryotic cell-free system: wheat germ . . . . .	159
3.5.3	Discussion . . . . .	162
3.6	Purification of <i>Pf</i> recombinant proteins . . . . .	166
3.6.1	Discussion . . . . .	178
3.7	Protein microarray feasibility study . . . . .	183
3.7.1	Production of protein microarray chips . . . . .	183
3.7.2	Processing of protein microarray chips . . . . .	185
3.7.3	Discussion . . . . .	192
<b>4</b>	<b>CONCLUSIONS</b>	<b>194</b>
4.1	The <i>Pf</i> genome and proteome analysis . . . . .	194
4.2	Expression of selected <i>Pf</i> candidates in heterologous hosts . . . . .	195

4.3	The <i>Pf</i> microarray platform as a platform to identify malaria vaccine candidates and as an <i>in vitro</i> diagnostic tool . . . . .	197
4.4	FightMal project: future work . . . . .	199
<b>REFERENCES</b>		<b>201</b>
<b>APPENDIX</b>		<b>223</b>



# List of Figures

1.1	Malaria, the Global situation . . . . .	22
1.2	Electron microscopy images of <i>Pf</i> and <i>Anopheles gambiae</i> . . .	24
1.3	<i>Pf</i> life cycle . . . . .	29
1.4	Morphology of <i>Pf</i> stages . . . . .	30
1.5	Merozoite and iRBC, targets of NAI . . . . .	33
1.6	Fluctuation of <i>Pf</i> parasitemia in an infected individual . . . . .	34
1.7	Malaria, the controls and treatments . . . . .	36
1.8	Variations in the malaria indicators from 1998 to 2008 in relation to the introduction or increase of control measures . . . . .	44
1.9	The different stages of the life cycle of the parasite that can be targeted by a malaria vaccine . . . . .	47
1.10	Protein microarray applications . . . . .	55
1.11	Protein microarray classes . . . . .	57
2.1	Methodology overview . . . . .	63
2.2	Nested PCR strategy and primers . . . . .	68
2.3	pIVEX 1.4 WG vector, functional elements . . . . .	78
2.4	pIVEXpET160 Pfcassette MAL7p1.177 . . . . .	79
2.5	Immobilised metal ion affinity chromatography (IMAC) . . . . .	82
2.6	Structure comparison, imidazole versus histidine . . . . .	82
2.7	IMAC purification procedure . . . . .	83
3.1	Time-line of the my PhD activities . . . . .	91
3.2	Structure of the bioinformatic analysis for CHIP 1, 2, 3 . . . . .	92
3.3	Example of redundancy in the preliminary antigen selection . . .	97
3.4	Summary of the bioinformatic analysis for CHIP 2 . . . . .	99
3.5	Summary of the gene splitting criteria . . . . .	100
3.6	Results of the gene splitting activity . . . . .	101
3.7	Summary of the bioinformatic selection for CHIP 1,2,3 . . . . .	106
3.8	Gateway <sup>®</sup> Technology . . . . .	110

3.9	List of vectors and features available with Gateway <sup>®</sup> Technology	112
3.10	pET160/GW/D-TOPO <sup>®</sup> Cloning site . . . . .	113
3.11	Polymerase enzyme evaluation tests . . . . .	117
3.12	Standard PCR Vs nested PCR . . . . .	118
3.13	Overall results of the cloning activity . . . . .	121
3.14	Biomass Comparison with 2xYT Vs TM . . . . .	134
3.15	Influence of temperature and length of incubation on biomass and solubility, 2xYT vs TM . . . . .	137
3.16	Analysis of pH and biomass, 2xYT Vs TM . . . . .	138
3.17	Amino acids usage, <i>E. coli</i> Vs <i>Pf</i> . . . . .	139
3.18	Amino acid composition of <i>Pf</i> challenging proteins (insoluble in 2xYT and soluble in TM) . . . . .	141
3.19	Influence of amino acid compositions on bacterial biomass and protein solubility . . . . .	144
3.20	Amino acid influence on solubility, WB . . . . .	144
3.21	Mini-scale studies, overall results: AIM vs 2xYT vs TM . . . . .	146
3.22	Expression level plotted as a function of antigen size . . . . .	157
3.23	Expression level plotted as a function of antigen size . . . . .	157
3.24	Wheat germ <i>in vitro</i> expression test . . . . .	162
3.25	Immunoaffinity purification via magnetic beads . . . . .	167
3.26	Gravity flow IMAC, PF14_0010 . . . . .	168
3.27	Gravity flow IMAC, 5 additional candidates . . . . .	169
3.28	Initial IMAC-FPLC purification test . . . . .	171
3.29	IMAC-FPLC purification: results . . . . .	172
3.30	Second step FPLC purification, gel filtration . . . . .	173
3.31	IMAC-FPLC followed by SEC . . . . .	174
3.32	IMAC-LPLC purification study at MtM . . . . .	176
3.33	Purification feasibility studies . . . . .	177
3.34	Microarray printing scheme and features . . . . .	184
3.35	Microarray printing scheme and features . . . . .	185
3.36	Reactivity of serum samples against <i>Pf</i> antigens . . . . .	186
3.37	Serum antibody reactivities of sera from heavily exposed Gam- bian adults . . . . .	190
3.38	Immune reactivity, IgG1 vs IgG3 subtypes . . . . .	191

# List of Tables

1.1	Worldwide estimation of malaria cases and deaths from 2000 to 2009 . . . . .	21
1.2	<i>Pf</i> genome summary and comparison to other organisms . . . . .	26
1.3	<i>Pf</i> proteome summary . . . . .	27
1.4	Scientific basis for NAI . . . . .	32
1.5	Antimalarial drug resistances . . . . .	42
1.6	Vaccines against malaria: development status . . . . .	51
1.7	<i>Pf</i> antigens utilised in the malaria vaccine development . . . . .	53
1.8	List of <i>in vitro</i> expression-based proteome microarrays . . . . .	58
2.1	Exportome and/or parasite surface attributes, list of algorithms	65
2.2	Candidate structure analysis, list of algorithms . . . . .	66
2.3	Priority selection score for CHIP 2 . . . . .	67
2.4	Sequences of the nested-PCR primers . . . . .	69
2.5	Rosetta-Gami™ B (DE3) <i>E. coli</i> Key Feature . . . . .	73
2.6	Amino acid mixtures composition . . . . .	75
3.1	List of candidates for CHIP 1 . . . . .	93
3.2	CHIP 2 antigen screening, initial pool of candidates . . . . .	95
3.3	CHIP 2 antigen screening, preliminary selection . . . . .	96
3.4	CHIP 2 antigen screening, snapshot of the internal scoring function	98
3.5	CHIP 2 antigen screening, the final results . . . . .	98
3.6	Final list of the CHIP 3 antigens . . . . .	104
3.7	List of possible C-terminal tags . . . . .	114
3.8	Attributes of Flag, cMYC and Strep-tag II . . . . .	115
3.9	Panel of polymerase enzymes: performance and costs . . . . .	116
3.10	Panel of <i>E. coli</i> strains evaluated for the heterologous production of <i>Pf</i> proteins . . . . .	129
3.11	<i>E. coli</i> strain evaluation studies . . . . .	130
3.12	Expression medium PX' study: LB vs TB vs AIM . . . . .	131

---

3.13	Expression medium evaluation studies . . . . .	132
3.14	Expression medium MtM study: final feasibility study 2xYT Vs TM Vs AIM . . . . .	133
3.15	Expression medium MtM study, a retrospective evaluation of TM Vs AIM . . . . .	135
3.16	Amino acids usage, <i>E. coli</i> Vs <i>Pf</i> . . . . .	140
3.17	TM amino acid supplements, Invitrogen vs internal . . . . .	143
3.18	Large-scale expression of 25 candidates in TM . . . . .	147
3.19	TM formulation . . . . .	153
3.20	List of <i>Pf</i> genes utilised in the <i>E. coli in vitro</i> expression . . . . .	155
3.21	<i>In vitro</i> expression levels . . . . .	158
3.22	Comparison of cell-free expression systems . . . . .	159
3.23	Wheat germ <i>in vitro</i> expression test . . . . .	162
3.24	Serum sample reactivity data . . . . .	187
3.25	Antigen reactivity per unit of expression . . . . .	188
3.26	Antigen expression and reactivity levels . . . . .	189

# Abbreviations

## Common abbreviations

---

Abbreviation	Name
°C	Celsius
6xHis	Hexa-histidine (HHHHHH)
AA	Amino acid
ACT	Artemisinin-based combination therapy
AIDS	Acquired Immune Deficiency Syndrome
AMA	Apical membrane antigen
b	Base
bp	Base pairs
BSA	Bovine serum albumin
cDNA	Complementary DNA
CSP	Circumsporozoite protein
Da	Dalton
ddH <sub>2</sub> O	Double-deionised water
DNA	Deoxyribonucleic acid
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbant assay
EV	Erythrocytic stage vaccine
EXP	Exported antigen
<i>g</i>	Centrifugal force
g	Gram
GLURP	Glutamine-rich protein
GSOR	Gene Selected Outside Region
h	hour
HIV	Human Immunodeficiency Virus
HRP	Horseradish peroxidase
IgG	Immunoglobulin G
iRBC	Red blood cell
ITN	Insecticides treated nets
L	Litre
LSA	Liver-stage antigen

---

Abbreviation	Name
m	Meter
M	Molar
MA	Microarray
mAb	Monoclonal antibody
min	Minute
mol	Mole
MS	Mass Spectrometry
MSP	Merozoite surface protein
MtM	Microtest Matrices Ltd.
MVI	Malaria Vaccine Initiative
NAI	Natural acquired immunity
NCBI	National Center for Biotechnology Information
ORF	Open reading frame
<i>P. falciparum</i>	<i>Plasmodium falciparum</i>
PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction
PEV	Pre-Erythrocytic Vaccine
<i>Pf</i>	<i>Plasmodium falciparum</i>
Pfu	<i>Pyrococcus furiosus</i>
PIESP	Parasite-infected erythrocyte surface proteins
PX'	PX'Therapeutics
RBC	Red blood cell
RBS	Ribosomal binding site
RESA	Ring-infected erythrocyte surface protein
RNA	Ribonucleic acid
rpm	Revolutions per minute
RT	Room temperature
S/N	Signal to noise Ratio
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SDS-PAGE	SDS-Polyacrylamide gel electrophoresis
sec	Second
SEL	Short External Loops
SIT	Sterile Insect Technique
SSV	Sexual stage vaccine
TAE	Tris-acetate-EDTA
Taq	<i>Thermus aquaticus</i>
TBST	Tris-buffered saline Tween 20
TE	Tris-EDTA
tRNA	RNA transfer
UV	Ultraviolet
V	Volt

Abbreviation	Name
v	Volume
w	Weight
WB	Western blot assay
WHO	World Health Organization

## Amino acid abbreviations

Amino acid	Three-letter	One-letter	Properties
Alanine	Ala	A	Non Polar
Arginine	Arg	R	Positive R
Asparagine	Asn	N	Polar
Aspartic acid	Asp	D	Negative R
Cysteine	Cys	C	Polar
Glutamic acid	Glu	E	Negative R
Glutamine	Gln	Q	Polar
Glycine	Gly	G	Polar
Histidine	His	H	Positive R
Isoleucine	Ile	I	Non Polar
Leucine	Leu	L	Non Polar
Lysine	Lys	K	Positive R
Methionine	Met	M	Non Polar
Phenylalanine	Phe	F	Aromatic R
Proline	Pro	P	Polar
Serine	Ser	S	Polar
Threonine	Thr	T	Polar
Tryptophan	Trp	W	Aromatic R
Tyrosine	Tyr	Y	Aromatic R
Valine	Val	V	Non Polar

## Power prefix abbreviations

Prefix	Symbol	Factor	Decimal
mega	M	$10^6$	1 000 000
kilo	k	$10^3$	1 000
hecto	h	$10^2$	100
deca	da	$10^1$	10
-	-	$10^0$	1
deci	d	$10^{-1}$	0.1
centi	c	$10^{-2}$	0.01
milli	m	$10^{-3}$	0.001
micro	$\mu$	$10^{-6}$	0.000 001
nano	$\eta$	$10^{-9}$	0.000 000 001
pico	$\rho$	$10^{-12}$	0.000 000 000 001
femto	$f$	$10^{-15}$	0.000 000 000 000 001
atto	$a$	$10^{-18}$	0.000 000 000 000 000 001
zepto	$z$	$10^{-21}$	0.000 000 000 000 000 000 001



# Chapter 1

## INTRODUCTION

### 1.1 Malaria: the global situation and figures

Malaria is considered by far to be the most important tropical parasitic disease. Malaria together with pneumonia, tuberculosis, diarrhoeal diseases, measles and HIV/AIDS account for 90% of all infectious diseases deaths and for half of the premature deaths worldwide (World Health Report, WHO). Mortality figures due to malaria range from 0.6 to 1 million deaths per year (WHO, 2010a). The disease primarily affects children, pregnant women and people with HIV/AIDS due to their weakened immune systems that make them more vulnerable to develop malaria. Although only a small subset of infected children will die as a result of the disease, the ones that are strong enough to survive will suffer serious malaria-related illnesses such as anaemia, periodic fever, brain damage and blindness (Carter et al., 2005; Fernando et al., 2003). Contracting malaria during pregnancy can be particularly serious as it can lead to severe anaemia, miscarriage (up to 60%) and maternal death (10-50%) (WHO, 2010a; Guyatt and Snow, 2001). International travellers and immigrants from endemic areas living in non-endemic areas and returning home are also at high risk of malaria.

The World Health Organisation (WHO) shows, in their annual reports and updates, how this old disease is assuming the semblance of a persistent "plague", with nearly 10 millions of deaths and over 2 billions of cases in the last 10 years (see Table 1.1). About half of the worldwide population is at risk from contracting malaria, and morbidity figures show that each year we are facing up to 300 millions estimated new infections (WHO, 2010a). While these figures show that the impact of malaria can be considered on a global scale, its distribution is not. The incidence of malaria is mostly confined to

CASES (in thousands)	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Uncertainty bounds	
											Lower	Upper
<b>African</b>	173,000	178,000	181,000	185,000	187,000	188,000	187,000	186,000	181,000	176,000	117,000	241,000
<b>Americas</b>	2,800	2,300	2,200	2,100	1,900	1,900	1,700	1,500	1,100	1,100	1,000	1,300
<b>Eastern Mediterranean</b>	15,000	16,000	17,000	16,000	14,000	12,000	12,000	12,000	13,000	12,000	10,000	15,000
<b>European</b>	47	34	27	22	13	7	4	2	1	1	1	1
<b>South-East Asia</b>	38,000	38,000	35,000	35,000	37,000	39,000	34,000	32,000	34,000	34,000	28,000	41,000
<b>Western Pacific</b>	2,800	2,500	2,200	2,500	2,800	2,300	2,500	2,100	1,900	2,300	2,000	2,500
<b>World</b>	233,000	236,000	237,000	241,000	243,000	244,000	238,000	233,000	231,000	225,000		
<i>lower bound</i>	181,000	181,000	182,000	184,000	185,000	185,000	179,000	175,000	171,000	169,000		
<i>upper bound</i>	302,000	304,000	308,000	313,000	314,000	317,000	310,000	304,000	298,000	294,000		

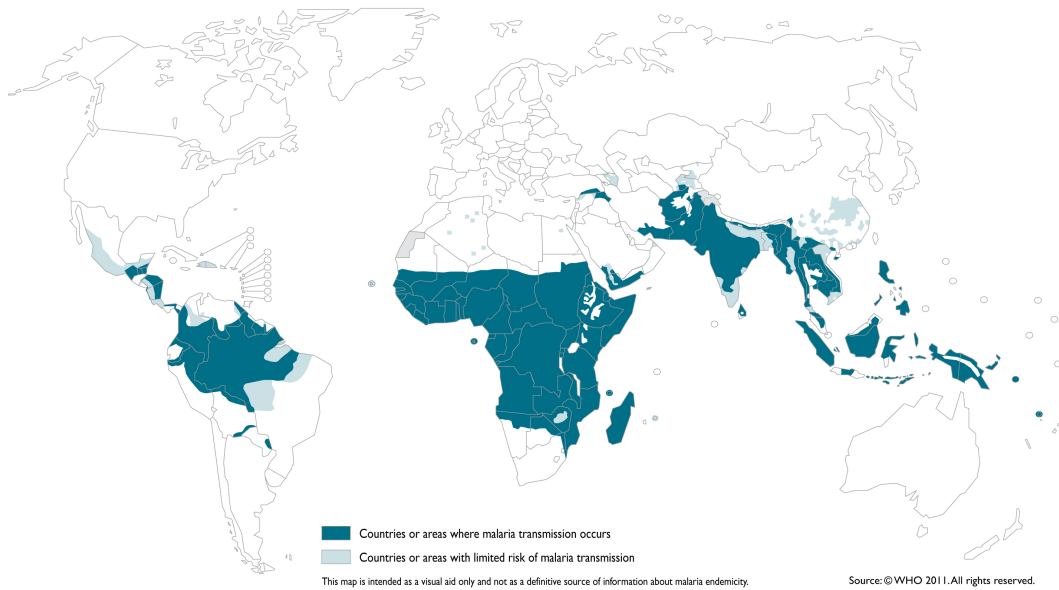
  

DEATHS	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Uncertainty bounds	
											Lower	Upper
<b>African</b>	900,000	893,000	885,000	880,000	870,000	853,000	832,000	802,000	756,000	709,000	554,000	892,000
<b>Americas</b>	2,400	2,300	1,400	1,400	1,500	1,600	1,600	1,400	1,100	1,300	900	1,700
<b>Eastern Mediterranean</b>	18,000	18,000	21,000	19,000	17,000	17,000	16,000	15,000	16,000	16,000	12,000	26,000
<b>European</b>	0	0	0	0	0	0	0	0	0	0	0	1
<b>South-East Asia</b>	58,000	55,000	51,000	50,000	52,000	50,000	48,000	43,000	48,000	49,000	37,000	63,000
<b>Western Pacific</b>	6,800	5,800	5,200	5,900	6,500	4,900	5,400	4,700	4,200	5,300	3,400	7,300
<b>World</b>	985,000	974,000	963,000	957,000	947,000	927,000	904,000	867,000	826,000	781,000		
<i>lower bound</i>	797,000	785,000	775,000	769,000	765,000	744,000	725,000	694,000	662,000	628,000		
<i>upper bound</i>	1,228,000	1,214,000	1,199,000	1,191,000	1,174,000	1,153,000	1,120,000	1,075,000	1,024,000	968,000		

**Table 1.1: Worldwide estimation of malaria cases and deaths from 2000 to 2009.** Table adapted from the World Malaria Report 2010 (WHO, 2010a).

developing countries. Around 78% of new cases and 90% of malaria deaths occur in Africa where a child under the age of five dies every 45 seconds from malaria or malaria-related causes (source: WHO Media centre, April 2010). Thanks to the first global eradication programs carried out in the late 50'-60's by the WHO, malaria was successfully eliminated from many of the affected areas such as North America (1970, year of registration into the WHO official register of areas where malaria eradication has been achieved), Europe (Spain 1964, Poland and Romania 1967, Italy and Netherlands 1970) parts of the Asia (Singapore 1982) and Australia (1981). However, if on one hand the disease was successfully eradicated in these prosperous nations, on the other hand these programs failed to eliminate the disease in other regions. Indeed, malaria still remains endemic in over hundred countries around the world. Africa (78%), South-East Asia (15%) and the Eastern Mediterranean (5%) regions are still under the burden of this disease of the poor (see Figure 1.1, WHO (2010a)). Until recently these countries were still facing a deteriorating effectiveness of front line treatments. In these affected regions resistance to anti-malarial compounds and insecticides, health care problems and a continuing lack of a vaccination tool meant that malaria was posing a greater problem than ever since the first international attempts to eradicate it.

The numbers of individuals affected or dying as a consequence of malaria are just the tip of the iceberg as they do not take into consideration the eco-



**Figure 1.1: Malaria, the Global situation**

During the 1960's malaria was successfully eradicated from developed countries within North America, Europe, parts of Asia and Australia. Unfortunately eradication programs failed to eliminate the disease in other regions. These regions now represent an appealing reservoir for the parasite. Indeed, malaria still remains endemic in 106 countries worldwide: Africa, South-East Asia, the Central/South America and Eastern Mediterranean are still heavily affected by this disease (WHO, 2010a).

epidemiological and socio-economical (e.g. agriculture productivity, fertility, educational attainment, tourism etc., Sabot et al. (2010)) consequences in the affected countries. The disease has been shown to severely impact upon economic growth, thus slowing the already difficult path towards development (Chima et al., 2003). It has been calculated that malaria has induced a reduction in economic growth of up to 1.3% annually in highly endemic regions (Sachs and Malaney, 2002) with the cost of the disease that in Africa reaches US\$12 billion per year (Greenwood et al., 2005).

In the late 90's, with a systemic failure of control strategies and growing evidence that both disease morbidity and mortality were worsening, malaria once again reached the attention of the international community (Greenwood et al., 2005; Sachs, 2002). Such renewed interest and motivation has now brought a fresh boost in both financial and political support for fighting malaria. As a result many new initiatives have started to emerge in the last two decades: the Roll Back Malaria Partnership, RBM ([www.rbm.who.int](http://www.rbm.who.int)), the Malaria Vaccine Initiative (MVI) ([www.malariavaccine.org](http://www.malariavaccine.org)), The Global Fund against AIDS, tuberculosis and malaria ([www.theglobalfund.org](http://www.theglobalfund.org)), just to name but a

few. The first results of this renewed motivation begin to appear. Since 2005 a scale-up malaria-control programme introduced a number of antimalarial interventions that produced a noticeable reduction in both the disease morbidity and mortality. The newly introduced or improved control measurements produced a visible down trend in the malaria incidence. In January 2007 the ascertainment of the United Arab Emirates in the registry of malaria-free country represented the first formerly-endemic country since the 1980s to be certified by the WHO. Since then, two additional countries have been added to the list: Morocco (May 2010) and Turkmenistan (October 2010), bringing the total number of malaria endemic countries to 106. In 2010, the cumulative financial aid on a yearly base reached US\$1.8 millions (WHO, 2010a).

Additionally, the new millennium has brought remarkable milestones on the path towards new effective antimalarial tools. The elucidation of the complete genome sequences of *Homo sapiens* (Venter et al., 2001) and some of the species of the parasite and vector that cause malaria in humans (Gardner et al., 2002; Holt et al., 2002) have pushed the scientific knowledge towards the desirable long-term aim of a global eradication of this disease, which now seems closer than ever.

## 1.2 Biology of malaria parasite

Malaria is a tropical parasitic disease spread worldwide by a protozoan parasite called *Plasmodium*. The parasite is transmitted across its human reservoir by the bite of an *Anopheles* female mosquito following a host-vector-host (human-mosquito-human) cycle. Notably, humans are not the only species affected by *Plasmodia*. Indeed the parasite mammalian hosts range from mice to apes. Even birds and snakes have their own form of malaria. Among the 175 species present in the *Plasmodium* genus, described in 1885 by Marchiafava and Celli, only four were historically considered able to infect human beings: *Plasmodium falciparum* (*Pf*), *Plasmodium vivax*, *Plasmodium ovale* and *Plasmodium malariae*. Although *Plasmodium vivax* is the most common species to afflict mankind, *Pf* (Figure 1.2a) is by far the most virulent, being responsible for the majority of malaria deaths. Prevalently present in sub-Saharan Africa, it accounts for 80% of all human malarial infections and for 90% of the malarial deaths (Miller and Greenwood, 2002). Recently, in certain forested areas of South-East Asia, an additional fifth species of *Plasmodium*, the *Plasmodium knowlesi*, has been shown to be able to infect humans in addition to their pre-

ferred host; the long-tailed macaque monkeys (Cox-Singh et al., 2008; Singh et al., 2004).



**Figure 1.2: Electron microscopy images of *Plasmodium falciparum* and *Anopheles gambiae*.**

(a) Image of two sickle-shaped *Pf* malaria parasites (■) floating within human red blood cells (■). Type of Image: SEM. Magnification: x2450 at 6x7cm size. Source: Science Photo Library Ltd, ©2011 (Image colors were modified). (b) Female Anopheline mosquito vector carrier of malaria. Type of Image: SEM. Magnification: 4x, based on a 35mm slide image of 24mm in the narrow dimension. Source: Dennis Kunkel Microscopy, Inc. ©2009.

The five species of *Plasmodium* capable of inducing malaria in humans are all transmitted by female mosquitoes of the genus *Anopheles* (Subbarao and Sharma (1997); Grassi (1898)). However, among the >400 species of *Anopheles* mosquitoes known, just 40-50 varieties are able to carry and transmit the parasite. Of these, the best known and most well-studied is *Anopheles gambiae* (Figure 1.2b) due to its predominant role in the transmission of *Pf*. Adult female anopheline mosquitoes transfer parasites to the human body during their blood meal. The majority of *Anopheles* mosquitoes are crepuscular (active at dusk or dawn) or nocturnal (active at night), thus acting when their targets are usually more sedentary and easier to reach. A bite from an infected mosquito can inoculate the target with up to 20-30 parasites, however, infection with even a single parasite can, in theory, be enough to cause malaria and consequently be fatal. The effect of malaria infection upon the human population depends on numerous epidemiological and ecological factors. Indeed they play a substantial role in determining the incidence of malaria and the intensity of disease transmission.

Clinical symptoms of malaria may vary. These include recurrent fever (every 3-4 days), bouts of chills, nausea, shivering, headache, joint pain and vomiting. Those can be followed by spleen and liver enlargement, anaemia and icterus. Death can potentially occur as a result of clogging of the cerebral vessels, anaemia or by general debility. Hence, a correct and early diagnosis

followed by rapid treatment is essential (malERA Consultative Group, 2011a; Girard et al., 2007).

### 1.2.1 The *Plasmodium falciparum*

Among the varieties of *Plasmodium* able to use human beings as a reservoir, *Pf* is certainly the one that causes the most severe symptoms and the highest number of fatalities. Indeed, *Pf* is responsible for nearly all (90%) malaria mortality in humans due to its high level of virulence. The high virulence can be attributed to both the capacity of the parasite to reach a higher parasitaemia and the ability to be peripherally sequestered in the microvascular system, limiting the parasite clearance. The pathogenesis, morbidity and mortality of malaria are determined by the species-specific parasite-host interaction and the parasite life cycle, during which it establishes an intimate relationship with its host and vector cells at both cellular and molecular levels.

#### The *Pf* genome

From the genomic point of view, the most relevant findings have been reached since 2002, when the complete genome sequence of *Pf* was first published (Gardner et al., 2002). This research showed that *Pf* has a nuclear genome consisting of 14 linear chromosomes, with sizes ranging between 0.64 to 3.29 Mb and an overall size of 22.8 Mb (see Figure 1.2). Notably, 80.6% of the genome is composed of adenines (A) and thymines (T) and these can approach 90% of the base pairs in introns and intergenic regions. This AT rich composition is unusually high and contrasts with those of *Homo sapiens* (~60%, Venter et al. (2001)), *Saccharomyces cerevisiae* (~60%, Gardner et al. (2002)) or *Escherichia coli* (~50%, Blattner et al. (1997)). Sequencing of the genome revealed 5,268 predicted genes with an average gene density of 1 gene every 4,338 base pairs (bp). More than 1 out of 2 (53%) genes contain at least one intron with an average number of exons of 2.4 per gene. The open reading frames showed a mean length of 2.3 kb, which is quite large compared to other organisms (e.g. *E. coli*, 0.9 kb, or *S. cerevisiae*, 1.4 kb). Significant limited redundancy of tRNA genes was also highlighted with 43 tRNAs identified, which is only half of the tRNA genes available in *E. coli* (n. 85) and from 6 to 10 times less than the number of tRNAs available in *S. cerevisiae* (n. 275) and *H. sapiens* (n. 446). Additionally, among these 43 tRNAs, only 2 anti-codons occur twice (CAT, methionine and CCT, glycine).

Since the publication of the *Pf* genome, a significant amount of downstream

Overview	<i>P. falciparum</i>	<i>E. coli</i>	<i>S. cerevisiae</i>	<i>H. sapiens</i>
Size (bp)	22,853,764	4,630,000	12,495,682	2,907,000,000
GC content (%)	19	51	38	38
N. of genes	5,268	4,383	5,770	39,114
Mean gene length (bp)	2,283	938	1,424	27,000
N. of tRNA genes †	43	85	275	446
<b>Exons</b>				
Number	12,674	NA	ND	233,785
N. per gene	2	NA	NA	8.8
GC content (%)	24	NA	28	ND
Mean length (bp)	949	NA	ND	170
Total length (bp)	12,028,350	NA	ND	39,841,315
<b>Introns</b>				
Number	7,406	NA	272	207,344
GC content (%)	14	NA	NA	ND
Mean length (bp)	179	NA	NA	5,419
Total length (bp)	1,323,509	NA	ND	1,123,657,235
<b>Intergenic regions</b>				
GC content (%)	14	NA	ND	ND
Mean length (bp)	1,694	118	515	ND

**Table 1.2: *Pf* genome summary and comparison to other organisms.**

Table modified from: Gardner et al. (2002). *E. coli* genome information was retrieved from Archer et al. (2011) and Blattner et al. (1997). *H. Sapiens* genome data were obtained from Sakharkar et al. (2004) and Venter et al. (2001). (†) tRNA figures for *S. cerevisiae* and *H. sapiens* were sourced from Goodenbour and Pan (2006). N.: number. NA: Not available. ND: Not determined.

data has been produced. New and interesting findings at both the transcriptome and proteome level have provided significant new knowledge about this organism (Florens et al., 2002; Le Roch et al., 2003). Data on protein-protein interactions have also recently been provided (LaCount et al., 2005). It has been shown that 61% of the predicted proteome is composed of hypothetical proteins that do not possess any orthologues in other organisms (see Table 1.3).

The lack of orthologues genes can potentially be explained by both the great evolutionary distance between *Plasmodium* and other eukaryotes and by its highly A+T rich genome. Relevant from the point of view of potential antigenicity (i.e. parasite proteins directly exposed to the immune system Rodriguez-Ortega et al. (2006)), 31% of the predicted proteins were shown to possess either one or two transmembrane domains. Additionally 17% had putative signal peptides or signal anchors. All this information has served to

<i>Pf</i> Proteome	Number	(%)
<b>Overview</b>		
Total predicted proteins	5,268	
Hypothetical proteins	3,208	60.9
InterPro matches	2,650	52.8
Pfam matches	1,746	33.1
<b>Gene Ontology</b>		
Process	1,301	24.7
Function	1,244	23.6
Component	2,412	45.8
<b>Targeted to</b>		
Apicoplast	551	10.4
Mitochondrion	246	4.7
<b>Structural features</b>		
Transmembrane domain(s)	1,631	31
Signal peptide	544	10.3
Signal anchor	367	7
Non-secretory protein	4,357	82.7

**Table 1.3: *Pf* proteome summary.**

The table was retrieved from: Gardner et al. (2002).

confirm the doubtless complexity of the parasite's life cycle.

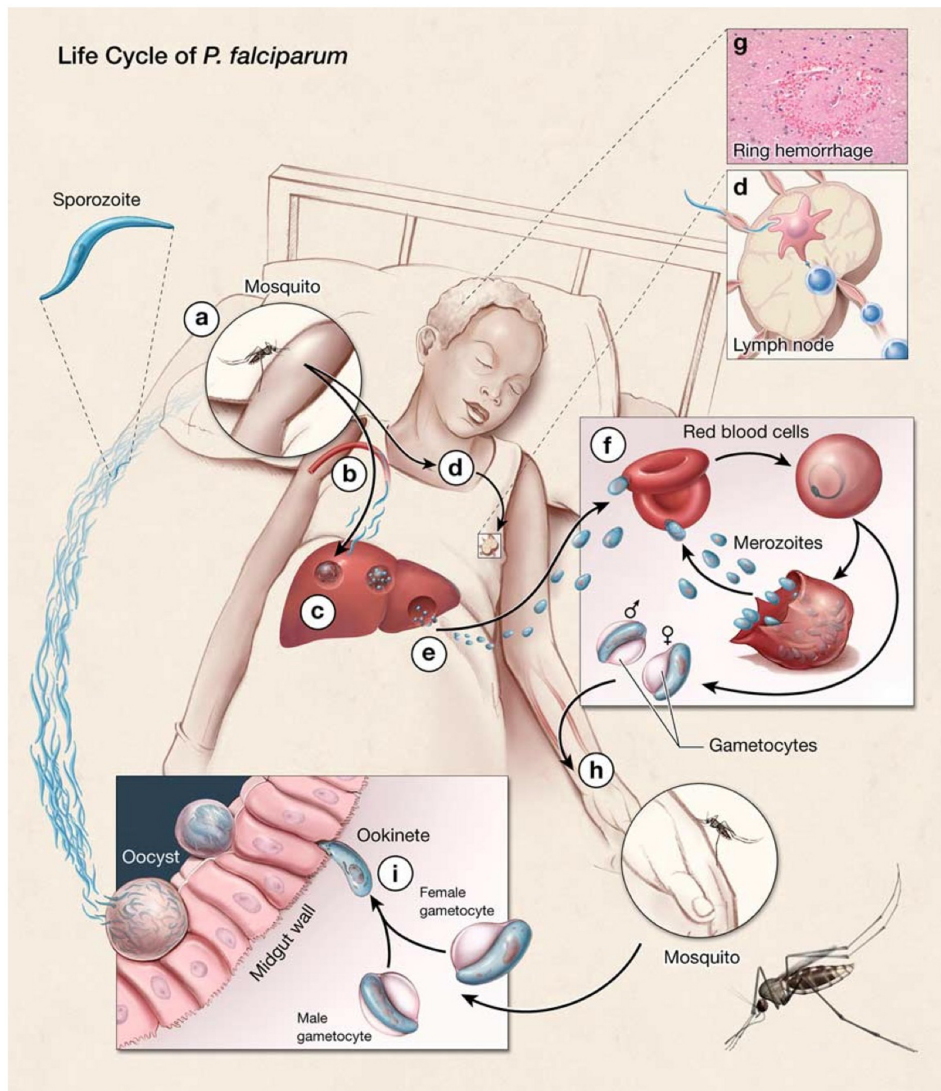
### The life cycle of *Pf*

The parasite life cycle involves constant and fine controlled differentiation of the *Plasmodium* both in the asexual and sexual stages. The former being carried out in the human host and the later in the female anopheline mosquito. Each step of the parasite development is characterised by different morphology, activity and protein expression patterns. Throughout its life cycle the parasite is able to rapidly adapt itself to the different host and vector environments, establishing specific and transient interactions both with the cell membrane and with the intracellular machinery (Florens et al., 2002).

*Pf* is transmitted to its human reservoir by the bite of an infected female mosquito. During the mosquito bite (Figure 1.3a) haploid sporozoites are released from the salivary glands of the mosquito into the human lymph system or blood stream (on average 20-30 sporozoites). Sporozoites that are transmitted to the draining lymph node trigger both the innate and adaptive immune sys-



tem (Figure 1.3d). Whereas those able to successfully reach the blood stream rapidly target the liver and invade the hepatocytes within a few minutes (Figure 1.3b). Once inside, partially hidden from the host's immune system, they establish strong and specific interactions with the hepatocytes and undergo asexual replication. During this liver stage, which can last between 6 to 16 days, each schizont (i.e. sporozoite-hepatocyte cell system) is able to generate several thousand merozoites (40,000 per hepatocyte cell, Figure 1.3c). Once nutrients present in the host cell have been completely exhausted, the schizont ruptures releasing merozoites into the blood stream (Figure 1.3e). In the circulation, the merozoites are able to quickly invade the red blood cells (RBCs), where the individual merozoite undergo a progressive maturation and asexual replication from the characteristic ring-stage parasite, through the trophozoite stage to an erythrocytic schizont (Figure 1.3f). At this point, infected erythrocytes synchronously (e.g. every 3-4 days) rupture releasing between 8-24 new merozoites, which can go on to subsequently invade other neighbouring RBCs. It is during this phase that the infected subjects show the typical clinical malaria symptoms (i.e. recurrent fever, bouts of chills, shivering, headache etc.). In the case of severe malaria, the infected RBCs (iRBCs) tend to adhere to the endothelial cells of the peripheral tissues. In particular, when there is a sequestration of iRBCs in the brain, the subject may suffer alterations of the mental status and coma, the typical symptoms of cerebral malaria (Figure 1.3g). Some of the merozoites instead of invading RBCs start the parasite's sexual sub-cycle. These differentiate into micro- and macro-gametocytes, through a process called gametocytogenesis (Figure 1.3h). The gametocytes circulate in the blood stream until another mosquito bites the host taking up not only blood, but also some of the mature sexual forms of the parasite. The gametocytes that successfully reach the mosquito's midgut form a motile zygote (i.e. ookinete, Figure 1.3i). The newly formed ookinetes migrate through the midgut barrier, replicate in the form of oocysts and develop into sporozoites. These will then start to migrate towards the salivary glands and eventually enter the salivary duct lumen. When the infected mosquito next takes a blood meal, infective sporozoites are transmitted from the mosquito back into the host (Figure 1.3a), thus completing the transmission cycle.

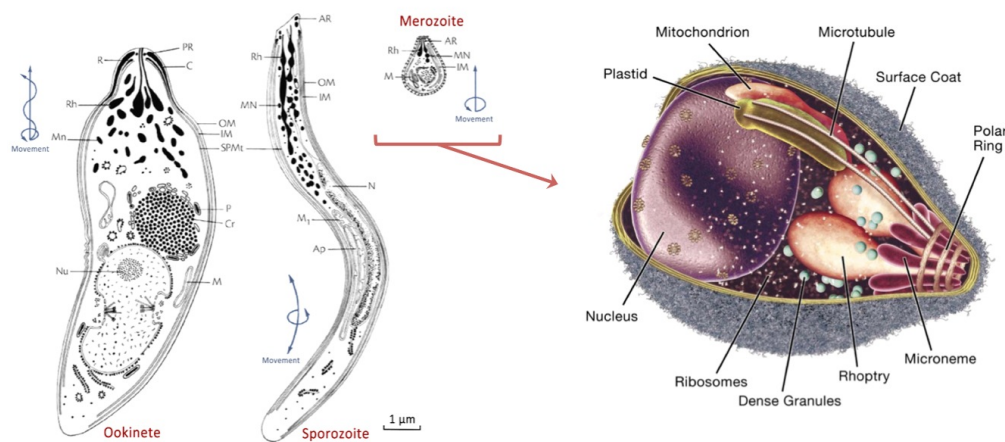


**Figure 1.3: *Pf* life cycle.**

The *Pf* life cycle is described in detail in Section 1.2.1. The image was source from Pierce and Miller (2009). A comprehensive video reconstruction of the malaria parasite website is available at the following link: <http://youtu.be/szlfndj0TFE> .

## Immune response to *Pf*

Malaria falls into the class of antigenically hypervariable microorganisms together with HIV and serogroup B meningococcus. Indeed, the *Plasmodium* parasite escapes the host's immunity thanks to its genetic variability. In particular, *Pf* not only expresses hyper-variable antigens that are able to adapt to the immunological pressure but it also shows a high degree of genetic diversity within species isolates (i.e. strains). Additionally, it is characterised by a complex life cycle, which stages are distinguished by a different morphology and behaviour (Figure 1.4). Due to such high degree of variability, a natural



**Figure 1.4: Morphology of *Pf* stages.**

The diagrams describe the three major morphologies of *Pf* life cycle: Ookinete (sexual stage), sporozoite (asexual/pre-erythrocyte stage) and merozoite (asexual/erythrocyte stage). The B/W images were retrieved from the Oxford Textbook of Medicine 5<sup>th</sup> (©Oxford University Press 2010). The merozoite image obtained from Cowman and Crabb (2006). **AR**: apical ring (electron dense ring, which is the backbone of the apical complex). **OM**: outer membrane. **IM**: Inner membrane. **N**: nucleus. **Ap**: apicoplast/plastid (non-photosynthetic plastid thought to be involved in the lipid metabolism. It is essential for the formation of the parasitophorous vacuole membrane needed for the host cell invasion). **M**: mitochondrion. **MN**: microneme (protein rich organelles essential for the parasite mobility and host invasion). **Rh**: Rophtry (secretory organelles that release enzymes during the host cell invasion). **SPMt**: sub-pellicular microtubules (longitudinal fibrils that support the external cell structure). **PR**: polar ring (see AR). **R**: lower ring (see AR). **C**: electron dense collar (secretory organelle essential for the invasion of the mosquito mid-gut). **Cr**: crystalloid (Structures involved in the protein trafficking and sporozoite transmission). **P**: pigment vacuole (vacuole containing haemozoin crystals derived from the haemoglobin catabolism).

acquired immunity (NAI) towards the parasite is only established slowly and it is never sterile, allowing the parasite to survive at low concentrations in the bloodstream of malaria asymptomatic individuals. Generally, humans with no exposure to the *Plasmodium* become infected during their first encounter with the parasite. This is the case for young children or tourists in malaria endemic

regions. However, thanks to NAI newborn (<3-6 months) and older children (>5 years), together with adults tend to be less susceptible to the severity of the illness (Roca-Feltrer et al., 2010; Gupta et al., 1999).

The initial encounter between the parasite and the human immune system happens soon after the inoculation of the sporozoites. Indeed, around a third of the injected parasites will move towards the lymphatic system within 6 hours from the mosquito bite. In the regional lymph nodes the dendritic cells will internalise the parasite and expose the sporozoite antigens to the T cells (Amino et al., 2006). Similarly, an early humoral response could inhibit the sporozoite invasion of the hepatocytes or target the infected hepatocytes, which provide a source of *Pf* antigens (Leiriao et al., 2005). However, due to the limited exposure of the parasite to the host immune system during these early (pre-erythrocytic) stages, the initial innate immune response towards the parasite has an important but limited role in inducing and modulating the adaptive malaria immunity (Stevenson and Riley, 2004). Longitudinal studies in holoendemic malaria regions demonstrated that adults heavily exposed to the parasite, although they are able to limit the malaria symptoms and severity, their immune system was not able to establish significant defences against *Pf* re-infections (Owusu-Agyei et al., 2001). Hence, although in theory the host's immune system is able to target the parasite at any given stage of its life cycle, the humoral immune response towards the plasmodial erythrocytic stage is believed to play a central role in NAI.

The hypothesis that the antibody response towards the blood stage of the parasite (i.e. merozoite and iRBC) is a key element of the NAI has been corroborated by several studies (Table 1.4). Arguably the most pivotal study was carried out by Cohen et al. (1961). In particular, the research team was able to transfer the NAI from asymptomatic adults to symptomatic children via a simple transfer of purified IgG. Similar findings were obtained by Sabchareon et al. (1991). During this latter study a decrease of 728 fold in the parasitemia was observed in patients injected with IgG purified from a pool of malaria protected individuals. Passively acquired immunity from mother to newborn *in utero* was also observed in the field by McGregor et al. (1970) and Logie et al. (1973), however the protective role of these antibodies is still debated (Riley et al., 2001). Finally in the late 90's early 00's a link between the *Pf* resistance and women pregnancies was established (Fried and Duffy, 1996; Duffy et al., 2001). In particular, it was noted that women in their third or fourth pregnancy were less susceptible to malaria compared to their first and second gestations. The increase in parasite clearance is due to the acquisition

Year	Observation	Method	PI
1900	Protection against malaria was acquired only after heavy and uninterrupted exposure to the parasite.	Cross-sectional studies on parasitemia in High Vs. Low endemic populations	Kock R.
1910's	i) Effective in adults after lifelong heavy exposure.	Cross-sectional microscopic studies in Asia and Africa	Varios
1920's	ii) Lost upon cession of exposure.		
	iii) Species/stage specific.		
	iv) Acquisition proportional to degree of exposure.		
1960's	Marked reduction in parasitemia was observed following passive administration of IgG from immune adult.	Antibody transfer experiments	Cohen S. McGregor I.
1970's	Passive transfer of antimalarial antibodies (IgG) from mother to child across the placenta.	Observations on the protective role of maternally derived antibodies in newborns.	Billewicz W.Z.
1991	Stage-specific, non-sterilizing effect of protective IgGs in decreasing the asexual parasitemia. Improvement in the patient clinical conditions and parasitic clearance were equal of better than the ones observed in patients treated with antimalarial drugs.	Antibody transfer experiments	Druilhe P.
2000's	Susceptibility and protection from malaria in pregnant women is antigen specific.	Studies an plasma antibodies from pregnant women exposed to the <i>Pf</i>	Duffy P.E. Rogerson S. J.

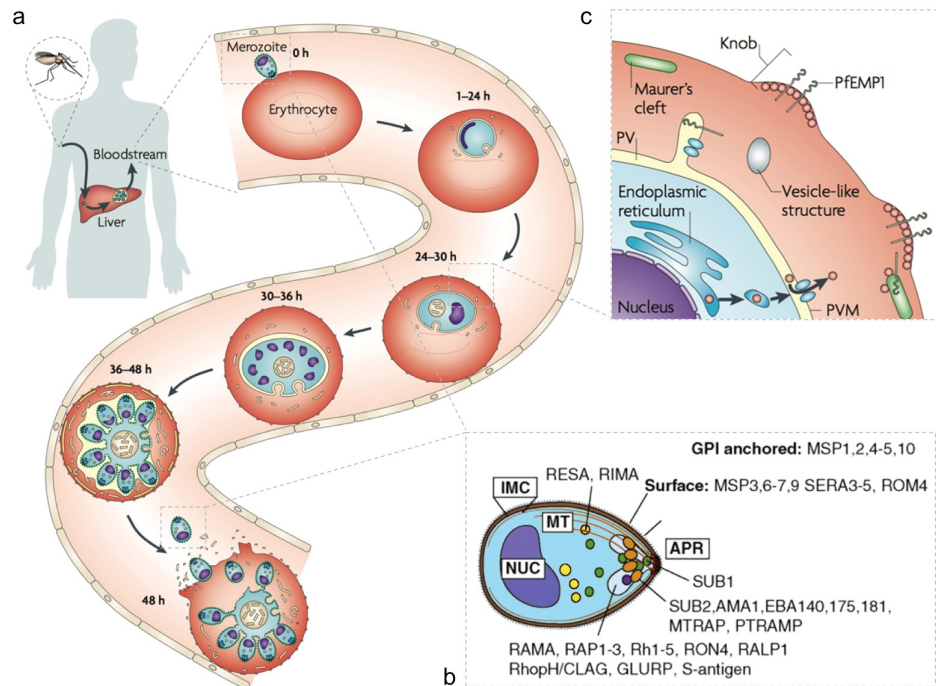
**Table 1.4: Scientific basis for NAI.**

of protective antibodies towards a conserved ligand expressed on the surface of the placenta-binding parasite.

If the importance of antimalarial antibodies in the NAI is now clear, the effector mechanisms by which these elements protect the host from parasite infection is still a matter of debate. The merozoite and iRBC parasite antigens exposed to the extracellular space represent the immediate targets of these humoral response (Figure 1.5).

Indeed, between the possible protective actions of NAI antibodies the most probable are at least three (Langhorne et al., 2008):

1. Opsonisation of the free merozoites floating in the bloodstream, which could inhibit the RBC invasion and increase the parasite clearance (e.g. antibodies against the apical membrane antigen 1 (AMA-1), Figure 1.5b).
2. Opsonisation of the iRBC, which should induce the macrophage phagocytosis of the infected erythrocyte and increase the spleen iRBC clearance.
3. Block the iRBC cytoadherence to the endothelial walls or towards other uninfected RBC (i.e. a phenomenon called rosetting that facilitate the parasite invasion of the erythrocytes) by targeting the knob-associated



**Figure 1.5: Merozoite and iRBC, targets of NAI.**

a) Overview of the *Pf* erythrocytic life cycle, starting from the merozoite invasion of a RBC and ending 48 hours later with iRBC rupture and the release of a new generation of merozoites. b) Diagram of the merozoite structure and major antigens. Image was modified from Baum et al. (2008). *NUC*: nucleus. *IMC*: inner membrane complex. *MT*: sub-pellicular microtubules. *APR*: apical polar rings. c) Diagram of the iRBC membrane, which is remodelled by the parasite during this intracellular erythrocytic stage. The parasites buried within the RBC tend to modify the cell structure by transporting antigens involved in the binding of the other cells to the exterior of surface membrane. The process produces an electron-dense excrescences called knobs (Atkinson and Aikawa, 1990). The RBC membrane alteration starts with the secretion of knob-associated proteins via a specific export pathway, which initiates at the endoplasmic reticulum level and progress into the parasitophorous vacuole (*PV*) first and then into the parasitophorous vacuolar membrane (*PVM*). From here the secreted proteins are shuttled to the erythrocyte membrane via vesicular structures (e.g. Maurer's clefts). Images a and c were modified from Goldberg and Cowman (2010).

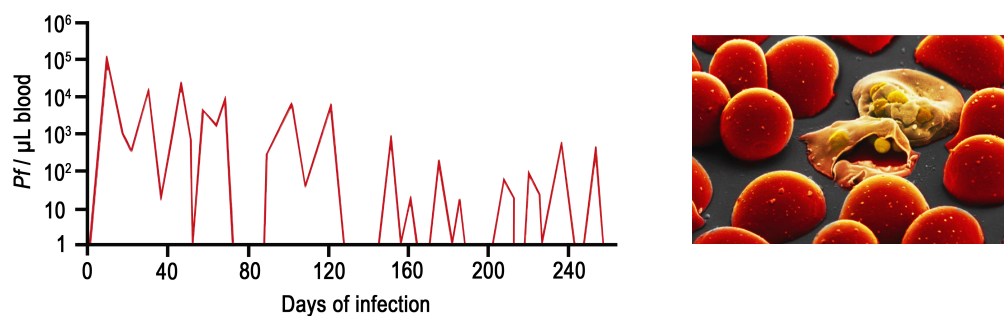
antigens (i.e. *Pf* erythrocytic membrane protein 1 (PfEMP1), Figure 1.5c).

Additionally, antibodies against immature gametocytes can inhibit the segregation from the circulation of the sexual form of the parasite, allowing a more efficient clearance of the micro/macro-gametocytes. Moreover, the opsonisation of mature forms of the gametocytes together with the components of the complement block the fertilisation and development of the parasite ookinete in the mosquito mid-gut. These two actions, although have a limited effect on the clinical status of the patient may play a significant role in reducing parasite transmission rate, if enough antibodies are produced.



Despite all this, the *Pf* parasite, is still able to escape most of these immunological mechanisms using its remarkable immune evasion strategies, which are mostly driven by the parasites genetic diversity and antigenic variability. Recent studies (Volkman et al., 2007; Mu et al., 2007) on comparative genomics, aimed at understanding the *Plasmodium* population structure expressed in gene copy number variation, insertion, deletions, single nucleotide polymorphism (SNP) and frequency/linkage between alleles within specific populations, have highlighted how the parasite strain genomes are highly recombinogenic when aligned to the 3D7 *Pf* reference genome (e.g. SNPs average density is equal to 1 every 214bp, twice the one observed in the *H. sapiens* genome; Carlton et al. (2008)).

Such evolutionary agility is further exacerbated by the antigenically hypervariable nature of the parasite, which makes any attempt to interfere with the parasite life cycle nearly impossible. This immunological challenge is well described by the observations made by Miller et al. (1994) on the parasite density of a malaria infected patient. They reviewed data from both laboratory and epidemiological studies and described how a single parasite infection could last for up to 480 days. They also highlighted how such persistent infection is characterised by a synchronised fluctuation of the parasite density in peripheral blood, which results in periodic peaks (Figure 1.6).



**Figure 1.6: Fluctuation of *Pf* parasitemia in an infected individual.**

The left panel describes the parasite density (i.e. number of *Pf*/ $\mu\text{L}$  of blood) detected in the bloodstream of an infected patient during the  $\sim 250$  days of *Pf* infection. Each peak in the parasitemia corresponds to the expression of a different PfEMP1 variant on the surface of iRBCs. In response of each antigen expression switch, the immune response actively produces new antibodies specific for the new PfEMP1 variant. Eventually, successive expression of antigens continues until the parasite exhaust the available variants and the host's immune system is capable of effectively clear the *Pf* parasite. The parasitemia graph was modified from Miller et al. (1994). The image on the right shows a new generation of merozoites leaving two iRBCs. The image was modified from National Geographic, photograph by Albert Bonniers Forlag.

It is now known that each wave of high parasitemia is mainly caused by the fact that the parasite is expressing a different variant of the PfEMP1 knob-associated antigen (see Figure 1.5c). PfEMP1 genes encode for heterogeneous surface receptors that are responsible for the cytoadherent, virulent and antigenic properties of the specific iRBC in which they are expressed (Smith et al., 1995). The activation and silencing of these virulence proteins is tightly regulated at the DNA level in a mutually exclusive expression pattern (Dzikowski et al., 2006). Around 60-70 PfEMP1 genes can be found in each parasite genome and they encode for a family of highly polymorphic genes named *VAR*. The clonally variant PfEMP1 family, however, it is not the only one in the *Pf* genome. The gene members of the *RIFIN* family (>130 genes) are also present, together with the *STEVOR* (~30 genes), *SURF* (~10 genes) and *Pfmc-2TM* (~10 genes), (Scherf et al., 2008).

The successive expression of clonally variant antigens on the external membrane of the iRBC, however is only one of the parasite mechanisms used to escape the human immune response. Indeed, the *Plasmodium* parasite developed a number of evasive strategies that include: i) intracellular parasitism, which allows the parasite to hide inside the host cells (e.g. hepatocytes and erythrocytes); ii) the cytoadherence properties of the iRBCs, which not only minimise the probability of the parasite to be cleared in the spleen, but also via the rosetting formation increases the chances to infect contiguous RBCs without being detected by the host's immune system; iii) antigen polymorphism, which helps decreasing the immune pressure derived from both antibodies and T cells targeting specific epitopes of the antigens, which may be changed by expressing mutated alleles of the same gene.

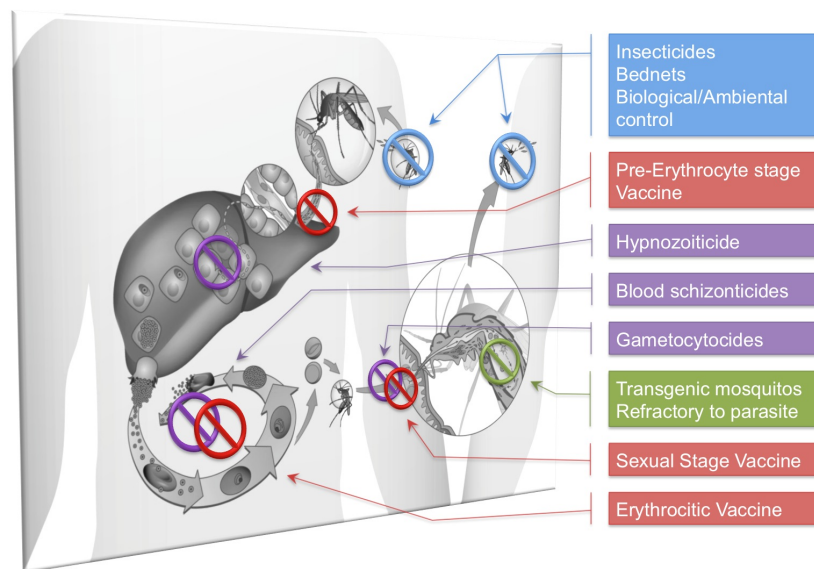
Not unsurprisingly, the parasite so far has been able to escape most of immunological and chemotherapy attempts to target it. Such ability has caught the attention of the scientific community. Indeed, since the first description of the *Plasmodium* parasite in the peripheral blood and it was shown to correlate with recurrent fever in malaria (Laveran A. 1880, Marchiafava and Celli, 1880-1885), *Pf* has been deeply studied and characterised. Such diversity clearly adds a number of challenges to any attempt aimed at interfering with the parasite development. In fact, a strategy targeting one specific stage of the *Plasmodium* life cycle will not necessarily be able to hinder parasite development or differentiation (Richie and Saul, 2002). Yet, on the other hand, such great complexity provides us with a few opportunities to interfere with its life cycle. Indeed, the anti-malarial drugs available are exploiting this complexity and most of the methods applied for the reduction of malaria transmission are



targeting the parasite all along its life cycle.

### 1.3 Malaria control and treatments

The unexpected regional failure of the WHO global eradication program carried out in the 50's-60's showed clearly that malaria could not be addressed as a single and uniform problem. Since that moment the scientific community has understood that malaria can not be addressed by one global control strategy and its control was indeed a demanding task (Collins and Paskewitz, 1995). The failure to eradicate malaria can be explained by the fact that malaria control has to address three organisms: the parasite *Plasmodium*, the host (humans), and the vector (*Anopheles*), see Figure 1.7. As stated by the



**Figure 1.7: Malaria, the controls and treatments.**

Existing and potential tools that can interfere with the life cycle of the parasite and decrease the malaria burden, by stopping the interaction between parasite and vector. Blue box: Vector controls available for mosquito populations; Purple box: Antimalarials drugs available for *Pf*. The blood schizonticides are the most important ones in the chemotherapy of malaria: chloroquine, quinine, mefloquine, halofantrine, pyrimethamine, artemisinin etc. A common gametocytocide is primaquine. Artemisinin and primaquine also have a hypnozoiticide effect against the latent or dormant form of the parasite; Green box: Vector controls not yet available, which aim to stop the interaction between parasite and vector; Red box: Three main vaccine strategies, pre-erythrocyte stage vaccine, gametocytes stage or sexual stage vaccine, red blood stage or erythrocytic vaccine. Malaria cycle image modified from [www.imm.ul.pt/html/Ciclo\\_Vida.bmp](http://www.imm.ul.pt/html/Ciclo_Vida.bmp).

United Nations Secretary-General in April 2008, the international community is aiming to stop the malaria deaths by increasing the malaria interventions until “universal coverage of malaria interventions” is achieved (World Malaria

Day, New York 2008). To reach such global coverage, two are the major strategies currently applied: a) vector control; b) early diagnosis and effective treatment; hence targeting directly both mosquito populations and parasite infections.

### 1.3.1 Prevention of malaria transmission: vector control

The vector control strategy has a two fold aim. One is reducing the longevity, distribution and the contact of the vector with the host; the other one is reducing the incidence and intensity of the infection at the community level. Both objectives can be achieved by using a combination of insecticides-treated mosquito nets (ITNs also known as long-lasting insecticide-treated mosquito nets, LLINs) and indoor residual spraying (IRS) programs. In particular, it has been shown that the use of ITNs reduces the malaria clinical episodes up to 62% (Lengeler, 2004); whereas, the IRS may reduce significantly the malaria transmission (protective efficacy ranging from 34% to 92%, Pluess et al. (2010)) by shortening the lifespan of the mosquitoes. This is achieved by spraying insecticides at the inner surfaces of a house or an animal shelter, which are common places where the *Anopheles* rests after a blood meal (WHO, 2006).

An efficient control of the vector, however, can only be achieved with a sustained widespread coverage of the affected areas with both ITNs and IRS. Additionally, to not compromise the effectiveness of these interventions, they have to be carried out continuously and according to specific procedures. Indeed, as observed in the past, these two strategies have suffered several setbacks due to the low population coverage, the mosquitoes high mobility and adaptability and the unpredictable ability of the parasite to develop resistance towards insecticides (WHO, 2006). Moreover, the presence of bed nets increases the parasite transmission rate driven by outdoors feeding mosquitoes (e.g. *Anopheles arabiensis*), as a shift in the parasite vector preferences has been reported in certain areas covered by ITNs (Russell et al., 2011). However, thanks to "the largest scale-up of a malaria intervention in Africa's history" (Dr. Margaret Chan, Director-general WHO, Geneva 2010) driven by the rise in funding, the ITNs coverage in the sub-Saharan Africa is increasing. Since 2008, 66% of the population at risk of infection has received ITNs, with a predicted coverage of nearly 80% by the end of 2010 (WHO (2010a), data not yet available). As today, over 40% of the households in the African region possess at least one ITN, with more than 1 in three children (35%) sleeping under a

bed net. A similar tendency has been observed for the IRS. From 2005 to 2009 the number of African people covered by IRS increased 5.8 fold, going from 13 to 75 millions ( $\sim 10\%$  of the population at risk in the country). Currently, four classes of chemicals (i.e. carbamates, organochlorines, organophosphates and pyrethroids) are recommended as insecticides by the WHO, with pyrethroids being the most broadly used. Pyrethroids are potent neurotoxic chemicals that induce paralysis in intoxicated insects. All ITNs and part of the IRS are based on such insecticide class. These are potent neurotoxic chemicals that induced paralysis in intoxicated insects. An other very common chemical used in IRS is DichloroDiphenylTrichloroethane (DDT) as it is the cheapest and the most resilient option (6-12 months, compared to the pyrethroids, which last for 4-6 months, WHO (2006)).

Overall, the application of both ITNs and IRS proved to be of primarily importance for reducing or interrupting the transmission of malaria. However, the two strategies suffer of similar long-term problems. They both require a high level of constant organisation, planning, implementation and financial commitment. ITNs distribution and IRS programmes have to be carried out following a rational approach based on the epidemiological data and on a profound understanding of both vector behaviour and disease transmission. The lifespan of LLINs is around three years. If not replaced in time, both malaria transmission and vector distribution can rise quickly (WHO, 2010a). The same is true for the IRS programmes. The effectiveness of insecticides tend to be from 2 to 12 months depending on the surface in which it was spread, the compound and the dosage utilised. Hence, effective spraying programmes can require from 1 to 6 IRS cycles per year, which has both operational and financial repercussions. Additionally, insecticides should be used on a periodic rotation system (WHO, 2006). Therefore, it is not always possible to use the most cost-effective compound. Indeed, the use of the same insecticide during multiple IRS cycles may lead to the the spread of insecticide resistance, which ultimately may compromise the efficacy of both ITNs and IRS strategies (WHO, 2010a).

If on the one hand the dependence of the malaria parasite on the mosquito as a vector has been exhaustively characterised, on the other hand, targeting mosquito populations with both insecticides and ITNs is not the only way to control the vector distribution. One of the strategies that is now gaining credibility is the SIT (Sterile Insect Technique). The concept behind this approach lies with the fact that sterile male mosquitoes are still able to mate with virgin females that will consequently lay sterile eggs. Importantly, *Anopheles* males

tend to mate several times in their short life (1.5-3 weeks), but females are usually refractory to re-insemination (Thailayil et al., 2011). Thus if enough sterile male mosquitoes are released over a constant period of time, local eradication of the vector population can occur, hopefully preventing the transmission of the parasite concurrently. The feasibility and applicability of the SIT was successfully proven a few years ago in Central America during the local eradication of New World screwworm, *Cochliomyia hominivorax*, (Wyss, 2000). This data, in addition to the recent findings about the feasibility of sex determination in *Anopheles* mosquitoes (Nolan et al., 2011; Catteruccia et al., 2005) shows that alternatives to insecticides exist. However, the introduction of transgenic mosquitoes in malaria endemic areas with the consequent suppression of the wild type mosquito population rises many efficacy and safety issues, for instance the parasite could switch to a different *Anopheles* vector. Concerns of unexpected alterations of both transgenic and wild type populations together with unpredictable epidemiological effects have also been discussed (Benedict and Robinson, 2003).

### 1.3.2 Diagnosis of malaria

The diagnosis of a malaria infection is a pivotal component of the disease control strategies. Patients with malaria can not only be severely affected by the infection but they also represent an active reservoir of *Plasmodia* parasite. Hence an early diagnosis, immediately followed by an effective treatment with antimalarial medicines is essential. Additionally, the unjustified administration of an antimalarial drug to a patient with misclassified malaria tends to exacerbate the problematic level of multi-drug resistance already observed in the malaria endemic regions (malERA Consultative Group, 2011a).

It is for these reasons that at the beginning of 2010, the WHO further modified the initial guidelines for the diagnosis of malaria to allow the inclusion of children less than five years old. As today, every single person suspected of having malaria, no matter of what age or epidemiological background, should be diagnosed through parasitological confirmation via either visual microscopy or a rapid diagnostic test (RDTs, WHO (2010a)). Currently the golden standard for the clinical evaluation of a malaria infection is a polymerase chain reaction (PCR) assay that detects the presence the parasite DNA. This procedure, however, is limited to laboratories with trained personnel and fitted with expensive hardware. Consequently, the test is usually run only in a restricted number of malaria surveillance centres. In resource-limited settings

the RDT systems most broadly used are the dipsticks. These devices are progressively replacing the light microscopy, which until recently was the most common method of diagnosis in these situations. The dipsticks are small, single-use devices based on a lateral-flow immunochromatographic antibody-sandwich assays. By detecting the presence of plasmodial antigens these assay can reach analytical sensitivity levels similar to the ones routinely observed in the light microscopy (i.e. analytical sensitivity of 100-200 parasites/ $\mu\text{L}$ , Bell et al. (2006)). The fact that they are easy to use (no training required compared to both PCR-based assays or microscopy) and that their performances are not operator-dependent is now helping their introduction at the community level for both case management and parasite prevalence.

Such shift towards a better diagnosis and subsequent treatment, produced a 100 fold increase in RDTs demand (from  $<200,000$  in 2005 to  $\sim 30$  millions in 2009). Furthermore, these figures are only an underestimation of the real number of RDTs according to the WHO. Indeed, only half of the African endemic countries logged these data in 2009. In regards of the number of microscopic examinations, over 150 millions stained blood smears were carried out worldwide in 2009. Globally, the percentage of malaria patients that were tested for parasitemia increased from 67% (2000) to 73% (2009). However for half of the African countries the testing rate is  $<20\%$ .

The current prediction of the WHO is that if on one hand the number of microscopic examinations are meant to decrease as the practise is substituted by RDTs, on the other hand the number of rapid test will increase in parallel with the embrace of the new WHO guidelines (WHO, 2010a). Indeed, as the burden of malaria declines more emphasis will be placed on both passive (symptomatic) and active (asymptomatic) case detection (malERA Consultative Group, 2011a). Hence a race towards new screening tools capable of detecting either the whole parasite (e.g. digital microscopy), or a parasite by-product (i.e hemozoin, a derivate of the *Pf* metabolism that can be detected via refraction and/or absorbance of laser light), or parasite antigens (e.g. dipstick, ELISA or serum profiling) or even the parasite DNA (e.g. PCR-based assays) has started. Hopefully, devices similar to the one recently developed by Columbia University (New York, USA, Chin et al. (2011)) for the rapid-low-cost detection of HIV and syphilis, will soon be available for diagnosis of malaria. The device, called 'mChip' (a.k.a. mobile microfluidic chip for immunoassay on protein markers), is able to combine the use of microfluidics ( $<1\mu\text{L}$  of unprocessed blood) into a multiplexed lost-cost ( $\$0.55$  material and reagents, compared to the  $\$1.8-6.2$  ELISA cost) assay, which reproduce the

ELISA system but in a rapid (20 min versus  $\sim 210$  min of the ELISA) miniaturised single-use device. As for the mChip, the implementation of any newly developed RDTs, however, will depend on the cost-effectiveness and performances of the assay, which eventually will determine if the newly developed RDTs will be adequate for a point-of-care (POC) diagnostics aimed at the early detection of the disease in resource-limited settings.

It is envisaged that the availability of an efficient and cost-effective POC tool, will eventually save some of the money spent in antimalarial treatment and more importantly will decrease the risk of developing new drug resistances, which is arguably the biggest threat in a effective malaria control.

### 1.3.3 Current malaria treatments: antimalarial drugs

For nearly four centuries the only available antimalarial drug was quinine. Then chloroquine (a quinine synthetic analogue), chloroguanide (proguanil), amodiaquine and pyrimethamine were introduced during the second World War. Yet, soon after their first global application during the 60's, several strains of *Plasmodium* developed resistance to these treatments. Consequently, a few additional compounds, such as mefloquine and halofantrine, were developed to protect the US military troops during the war in Vietnam (White, 2008, 1996). Unfortunately, since then not many new antimalarial drugs have been discovered. Actually of the 1223 drugs registered between 1975 and 1996, just three were antimalarials (Trouiller and Olliaro, 1998) highlighting evidence of a lack of a rational and coherent treatment policy. On top of this, new single and multi-drug resistances towards chloroquine, which is the antimalarial drug most widely used, have being observed worldwide. Consequently monotherapy and some of the antimalarial combinations started to be ineffective (see Table 1.5, Wongsrichanalai et al. (2002)).

In the late 80s early 90s, the artemisinin-based combination therapy (ACT) was introduced (White, 2008). The main active ingredient in the formulation is based on a natural compound, the *artemesia annual*, and is combined with a second antimalarial drug. The antimalarial properties of artemisinin (known as qinghaosu, a traditional Chinese medicine) were discovered in the 70's by a team of scientist in China (CRGQ, 1979). The compound reduces the chance of parasite recrudescence by killing *Pf* both in its asexual and sexual stage (ter Kuile et al., 1993; Chen et al., 1994). This results in a decreased immunological selection pressure within the patient and in prevention of the parasite transmission. Until its international ban in 2007 (resolu-

Antimalarial drug	Effect	Introduced	Reported Resistance
Quinine	Blood schizonticide and gametocide	1632	1910
Chloroquine	Blood schizonticide and gametocide	1945	1957
Proguanil	Blood schizonticide	1948	1949
Sulfadoxine-pyrimethamine	Blood schizonticide	1967	1967
Mefloquine	Blood schizonticide	1977	1982
Atovaquone	Blood schizonticide	1996	1996
ACTs (Cambodia region)	Hypnozoiticide, blood schizonticide, trophozoiticide, gametocide	2001	2004

**Table 1.5: Antimalarial drug resistances.**

Summary of the dates of introduction and first reports of resistances for the major antimalarial drugs. Table modified from Wongsrichanalai et al. (2002), figures for the ACTs were added from Dondorp et al. (2010). ACT: Artemisinin-based combination therapy. The ACT was introduced in Cambodia in 2001 as the first-line treatment. However, unregulated artemisinin based mono-therapies were available in the region since the middle 70's.

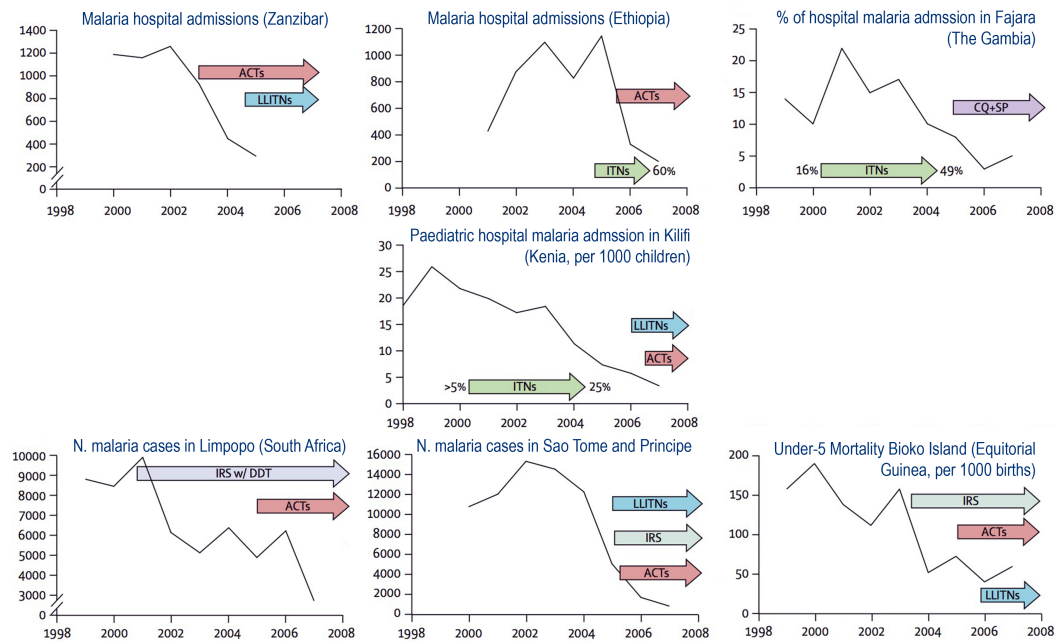
tion n. WHA60.18, WHA 2007), artemisinin was also used as an oral-based monotherapy. The drastic ban was due to the threat that this single-drug formulation was posing in the insurgence of artemisinin resistances. Currently five are the combinations most widely used in conjunction with artemisinin: artemether–lumefantrine, artesunate–amodiaquine, artesunate–mefloquine, artesunate–sulfadoxine–pyrimethamine and dihydroarte-misinin–piperaquine (WHO, 2010a). Worryingly, recent evidence regarding resistance to ACTs has emerged in a few isolated areas in southeast Asia (Cambodia-Thailand border, Afonso et al. (2006); Jambou et al. (2005)). Nevertheless, the WHO is promoting the use of ACT as a leading antimalarial treatment since 2001, as it is still considered "the most effective antimalarial drug available today" (WHO, Geneva, 2005, WHO (2010b)). Of the 39 clinical trials reviewed by the international organisation, in which over 6000 patients were treated with ACTs, only 7 recorded some failures of the treatment (in 1-7% of patients) within the first two weeks (WHO, 2010c). The pharmaceutical properties of the ACTs formulation are based on the fact that initially the artemisinin or any of its derivatives destroys the majority of the parasites, then the small remaining number of *Plasmodia* are killed by the drug-partner of choice. Such high efficiency in cure rate, however, comes at a price. Indeed the ACTs treatment can cost up to 10 times more than the chloroquine (\$ 0.10 per course of treatment for an adult) and it requires regimens covering many days. If not subsidised, the treatment could be sold at 5 times that price in African pharmacies, a cost that is far too expensive for the poorest living in endemic areas (Arrow et al., 2004).

In conclusion, even if different combinations of antimalarial strategies can help to overcome *Pf* ability to evade single drug therapies we are still missing a prompt and effective treatment. ACTs have to be administered within 24-48 hours from the onset of the parasitic infection symptoms to be highly effective. Hence, a wide coverage of the affected area and affordability of the formulation are crucial. The ideal treatment would be based on the use of efficient, safe and affordable antimalarial drugs. The effectiveness of such a therapy could decrease not only the substantial figures of mortality and morbidity, but could also potentially resolve issues relating to parasite resistance.

Unfortunately, none of the current treatments have been able to fully overcome the antigenic hypervariability and virulence of *Pf*. ITNs are not always available and if they are, they are not always refreshed with new insecticide. Furthermore, in many occasions life saving treatments are not easily accessible within the first 24 hours after malaria symptoms arise (UNICEF, 2007). Although mortality and morbidity figures shown in the latest world malaria report (WHO, 2010a) are the best of the last decades, a recent review shows that it has not always been possible to link the decline of malaria cases with the introduction of control measurements (O'Meara et al., 2010). In this review the authors analysed a number of studies that reported the incidence and/or prevalence of the malaria in the Africa sub-Saharan region. In particular they describe how, in certain occasions, the decline of malaria cases and related hospitalisations started to decline even before the implementation of specific malaria control measurements. They noted how in Zanzibar, The Gambia and Kenya the reduction in malaria hospital admissions started to decline before an effective ITNs program was introduced (Figure 1.8). In Ethiopia the decline observed in malaria cases was only accentuated by an outbreak of infections occurred between 2002-2005, and now the malaria incidence is back to historical levels (Figure 1.8). In contrast the remaining three studies (South Africa, Sao Tome and Principe, and Bioko Island) show a clear link between the introduction of malaria control measures and the decrease in the malaria indicators (bottom Figure 1.8). Noteworthy, however, is the fact that the Sao Tome and Principe graph ends in 2008. Indeed, since 2009 a substantial increase in malaria incidence ( $\sim 140\%$ ) was reported (WHO, 2010a).

What is clear from these studies is that for an effective and sustainable malaria control, antimalarial combination therapies, insecticide spraying programs and the use of treated bed nets are not enough. The current malaria control and eradication program is a complex combination of initiatives designed to complement one another and insecticides, ITNs and antimalarial





**Figure 1.8: Variations in the malaria indicators from 1998 to 2008 in relation to the introduction or increase of control measures.**

Malaria hospital admissions, cases and mortality in different countries of the African sub Saharan region are superimposed with the introduction of prevention measures or treatments for malaria. ACT: artemisinin combination therapies. LLITNs: long lasting insecticide-treated bednets. ITNs: insecticide-treated bednets. CQ+SP: chloroquine with sulphadoxine plus pyrimethamine. IRS: indoor residual spring. DDT: DichloroDiphenylTrichloroethane. Figure modified from O'Meara et al. (2010).

drugs appear to be nothing more than a piece of the final solution (Greenwood and Targett, 2009). Of this opinion is the Malaria Eradication Research Agenda (malERA) Consultative Group, a group of independent scientist with expertises in different technological and scientific areas focus on malaria control. In one of their recent White Papers (malERA Consultative Group, 2011b) they highlight how the decline observed in recent years in malaria incidence is only the initial step towards the eradication of this disease. Indeed, the closer we will get to the disappearance of *Plasmodia* parasite the more difficult it will get to eliminate any latent or resilient form of the parasite. Some of the plasmodia can remain dormant for years in either the liver (*Plasmodium vivax* and *Plasmodium ovale*) or in the blood (*Plasmodium malariae*). In addition, low concentrations of the active parasite can survive, and therefore transmit the disease, without being detected by the current RDT systems. Hence, the eradication process will be a race against the clock to avoid further increase in drug and insecticide resistances, which may occur even in areas with low malaria transmission (Cortese et al., 2002).

If thanks to the international effort and the few but effective malaria control tools, the malaria burden has declined in the last few years, such reduction in malaria cases is, however, not homogeneous. On the contrary, for at least three countries (Sao Tome and Principe (see Figure 1.8), Rwanda and Zambia), a sharp resurgence in malaria incidence was reported by the WHO in 2009, (WHO, 2010a). In the case of Sao Tome and Principe for instance, one of the African countries with the highest coverage of ITNs (78%, 2007), although the number of malaria cases were down by 84% in the period between 2005 and 2008 compared to 2000-04, in 2009 the annual number of confirmed malaria cases was up  $\sim 140\%$  (from 1647 to 3893), which resulted in an increase of 44% in both malaria related hospitalisations and deaths. The source of these opposite trend is still not clear. This confirms ones again that the battle against these deadly parasite is not easy. The progress achieved so far is very fragile and it has to be constantly strengthen with new strategies, which can effectively tackle the evasive behaviour of the *Plasmodium*. Host, vector and agent will all have to be tackled if a complete interruption of the malaria transmission is to be achieved.

### 1.3.4 Rational for a malaria vaccine

From the observations reported in the latest world malaria report, although malaria morbidity and mortality are declining globally, there are still areas in which the illness seems to be out of control. Hence, it is clear that we are still missing one of the most important links in this intricate chain of malaria control - a malaria vaccine. Indeed, it has been previously demonstrated that the use of an effective vaccine can be pivotal for the eradication of diseases like smallpox (Henderson, 1987), polio (John, 2009) and measles (Moss and Griffin, 2006). Considered one of the most formidable scientific challenges and medical quests of the last 40 years, the search for a malaria vaccine has seen hundreds of different research teams fighting to solve its puzzle. Being relatively cheap (once developed), easy to use and applicable through existing clinical protocols, this milestone could help us to improve remarkably our armoury against malaria. This is why the development of a vaccine able to reduce both morbidity and mortality is one of the strategic goals of Roll Back Malaria Partnership (the global framework for coordinated action against malaria). The current target set by the organisation is the development of a vaccine with 50% efficacy by 2015 and one with 80% efficacy by 2025 (RBM, 2011).

Afflicted by persistent failures, the factors that lie behind the inability to

develop a malaria vaccine include many elements of the parasite and the host. On one hand the parasite immunological complexity and genetic variability represent a complex matrix of variable to solve. On the other hand the limited knowledge about the parasite-host interactions and the NAI together with the lack of appropriate animal models for vaccine trials create a gap of knowledge that has to be filled (Aide et al., 2007; Parekh and Richie, 2007). Ultimately, both the logistics and costs of vaccine development - around 10-12 years of development and \$ 500 millions of investment before the vaccine can be marketed (Bonn, 2005), represent the last but not the least of the issues. Despite all these downsides, the scientific and the socio-political community strongly believe that a malaria vaccine is feasible and can still be achieved (malERA Consultative Group, 2011c). There are four key reasons fuelling such optimism.

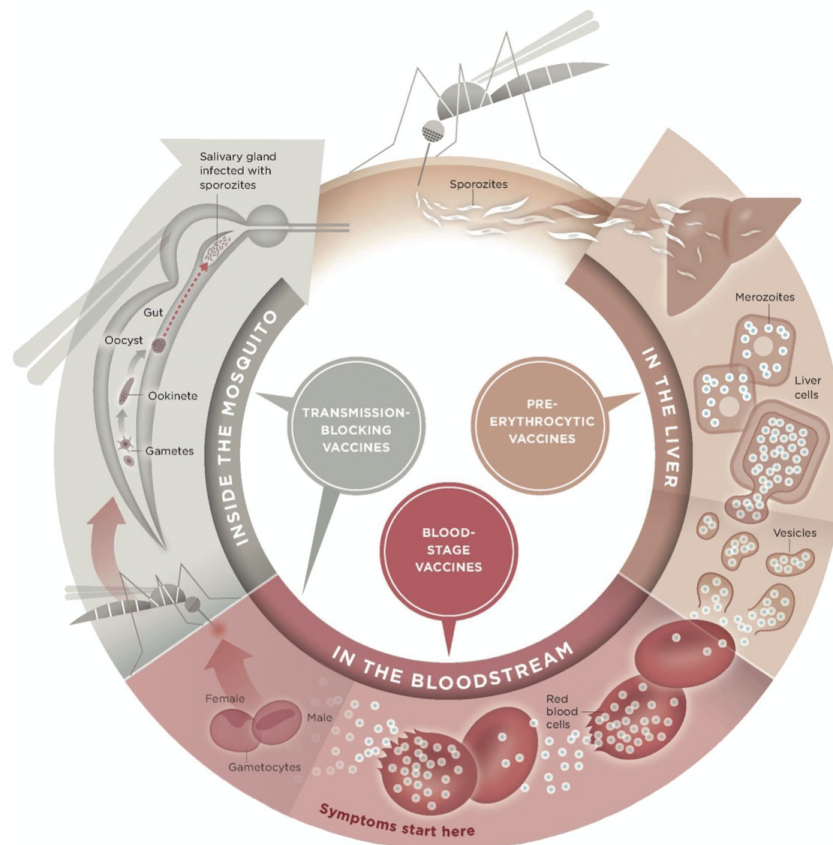
The first one is based on evidence that NAI towards the parasite has been observed in the field. Indeed, people living in heavily-exposed areas first develop partial immunity and then if challenged again, they progressively acquire the immune specificity to decrease the parasitemia down to undetectable levels ( $10^4$ - $10^6$  fold lower than infected naive subjects, Druilhe and Perignon (1997)). However, it can take more than a decade to develop an optimal NAI and it is not a complete sterilizing immunity. Hence, these individuals still become infected but do not show or develop any typical malarial symptoms. Moreover, they represent reservoirs that can transmit malaria to other individuals.

The second argument is based on the fact that the passive transmission of immunity between a mother and her newborn infant has been observed (Ballou et al., 2004). This idea has been further proven by the evidence that immunity can be transferred from protected individuals to unprotected individuals through administration of purified immunoglobulin IgG from the protected patient (Cohen et al., 1961; Sabchareon et al., 1991).

Thirdly, in the 70's, experiments carried out using healthy volunteers exposed to UV-irradiated (i.e. weakened) sporozoites demonstrated that these patients were able to develop an immunity of nearly 90% when re-challenged with parasites (Rieckmann et al., 1979; Clyde et al., 1973). Moreover these results have been successfully replicated recently using genetically modified *Plasmodia* in rodents (Hoffman et al., 2010; Pinder et al., 2010; Mueller et al., 2005).

The fourth and last argument is the evidence obtained in experimental trials in both adults and children that have shown efficacy, although limited (40-50%), of various candidate vaccines phase III clinical trials (Olotu et al., 2011; Aponte et al., 2007; Alonso et al., 2005, 2004).

These data clearly support the concept that a malaria vaccine can be achieved. Furthermore the multi-stage life cycle of *Plasmodium* and its constant interaction with either host and vector allow us to select different antigenic targets. Indeed, *Plasmodium* can be targeted in at least three moments during its development, dividing the efforts of the scientific community towards the vaccine into three different approaches, see Figure 1.9.



**Figure 1.9: The different stages of the life cycle of the parasite that can be targeted by a malaria vaccine.**

Image modified from Vogel (2010). Image credits to: Samuel Velasco, 5W INFOGRAPHICS for Malaria Vaccine Initiative/PATH.

The first potential vaccine is called Pre-Erythrocytic Vaccine (PEV) and it aims to block the parasite before it is able to reach its erythrocytic stage. This vaccine is designed to target both the sporozoites that flow through the bloodstream after the mosquito's bite, and those already within the hepatocytes. This approach is mainly aimed at non-immune individuals that need 100% 'sterilizing' protection, i.e. tourists. In principle, an effective PEV should be able to reduce the parasitemia and possibly block the parasite reproduction cycle in the host and its transmission. In fact, it should be possible to neutralize the parasite before it induces any sort of clinical symptoms (malERA

Consultative Group, 2011c). The feasibility of a PEV vaccine is supported by studies that used irradiated sporozoites, which successfully induced a nearly complete protection ( $\sim 90\%$ ) in both animal models and in humans (Pinder et al., 2010; Rieckmann et al., 1979; Clyde et al., 1973). An similar result was also recently achieved by Sanaria, a biotech company that used metabolically active, non-replicating *Pf* sporozoites (Hoffman et al., 2010). However, these two studies using whole parasite *Pf* sporozoite (PfSPZ) vaccines both face very difficult short- and long-term challenges ahead. The implementation of attenuated organisms as malaria vaccines requires a large-scale manufacturing and quality control processes that have to be combined with an effective method for stabilizing and delivering the vaccine (Hoffman et al., 2010; Pinder et al., 2010). Arguably the most renowned antigen-based PEV is the RTS,S. The formulation of RTS,S is based on *Pf* circumsporozoite protein (CSP), that in combination with the hepatitis B surface antigen molecules and a proprietary GSK adjuvant system form the GSK RTS,S AS01/AS02. This compound is currently in phase III clinical trial, after showing a 40-50% reduction in the clinical episodes of malaria over a 15-18 months follow-up period (Olotu et al., 2011; Aponte et al., 2007; Alonso et al., 2005, 2004). Overall, however, the PEV approach has shown different downsides. So far 36 different formulations have been tested and 20 were rejected (see Table 1.6). It is hard to believe these formulations or any other based on the current pre-erythrocytic antigen portfolio can have a significant impact on the parasitic gametocyte prevalence. The results collected from both clinical trials and *in vivo* experiments show that the necessary efficacy of 100% is nearly impossible to reach, at least with the preparations available today. Such a high efficiency is needed due to the fact that just one *Pf* parasite able to escape the vaccine pressure is sufficient to start the erythrocytic stage, eventually killing the host and perpetuating the transmission cycle (Doolan and Hoffman, 2000).

The second vaccine strategy aims at targeting the parasites once they move into the erythrocytic stage (Erythrocytic Vaccine, EV) and along all the stages involved in this sub-cycle. Numerous clinical studies and experimental data have described NAI as being predominantly directed against asexual blood stage parasites (e.g. merozoites) and mainly mediated by antibodies. Further evidence to support this has been revealed by analysis of the parasite genome sequence. Here, it is clear that both the pre-erythrocytic and sexual-stage parasite antigens are much less polymorphic than erythrocytic equivalents, promoting the idea that there is far greater immune pressure on the blood stage than on any of the other stages (Struik and Riley, 2004). The EV vaccine

mainly targets merozoite parasites either in the bloodstream or hidden inside the RBC, with the additional ability to block the invasion of erythrocytes by the merozoites. In particular, the vaccine can stimulate the immune system to target the merozoite antigens present on the cell surface and the ultrastructures derived from the transformation of the iRBC membrane (i.e. knobs, see Figure 1.5). This should result in an increase of parasite clearance and thus a consequent decrease in both parasitaemia and the severity of symptoms.

The parasites buried within the erythrocytes tend to modify the cell structure by transporting antigens involved in the binding of the cells to the exterior of surface membrane. The process produces an electron-dense excrescences called knobs (Atkinson and Aikawa, 1990). Although necessary for parasite cell invasion, these ultrastructures also represent an appealing target for the immune system. Although the EV does not prevent infection (e.g. it is not sterilising) it can certainly add efficacy to the partially acquired immunity commonly observed in children and adults, with an overall reduction of the parasite burden, morbidity and mortality (Chilengi and Gitaka, 2010; Good et al., 1998). The history of the EV vaccine begins when a synthetic multi-epitope, multi-stage peptide called SPf66 was developed in Colombia in 1992 (Valero et al., 1993). Unfortunately, however, this compound showed efficiency that was close to zero in several Phase III trials (D'Alessandro et al., 1995; Nosten et al., 1996). Nowadays, one of the most investigated EV vaccines is the so-called 'Combination B' that contains a mixture of merozoite surface protein (MSP) -1, MSP-2 and a ring stage infected erythrocyte surface antigen (RESA) combined with a Montanide adjuvant. During a Phase I/IIb trial in Papua New Guinea this combination showed an average of 62% reduction of parasitaemia in 120 children (Genton et al., 2002). However, while demonstrating a substantial increase in immune protection towards malaria, this response was unfortunately strain specific (Fluck et al., 2007). Indeed, a significant increase in parasite strains different to those used for vaccination was observed. It is likely, however, that the use of new adjuvants and antigens with limited allelic variants could overcome these issues and increase the efficacy of the Combination B vaccine. Recently a second attempt focusing on the entire MSP-1 molecule has been reported. During a Phase I randomised trial in the USA, Kenya and Mali, a formulation of this antigen combined with AS02 adjuvant was found to be significantly immunogenic and safe (Stoute et al., 2005). Consequently the vaccine is now undergoing further investigations in Kenya using children at great risk of malaria. In addition, other merozoite surface proteins are undergoing continuing assessment, e.g. AMA-1, MSP-3 and RESA

(Girard et al., 2007). AMA-1 in particular is the blood stage antigen most studied so far. It has been tested in at least 16 formulations (see Table 1.7). A formulation based on this protein stimulates the production of inhibitory antibodies that can reduce the ability of the parasites to infect the erythrocytes (Remarque et al., 2008; Kocken et al., 2002). However, only marginal levels of immunogenicity responses were observed during clinical testing (Saul et al., 2005; Malkin et al., 2005; Ockenhouse et al., 1998). Interestingly, immunogenicity was increased when AMA-1 was used in combination with MSP-1, as the fusion of multiple malaria antigens seems to enhance the specificity of the immune response (Faber et al., 2007; Pan et al., 2004). Other formulations based on both MSP-3 and GLURP proved to be effective in promoting the synthesis of cytophilic antibodies (IgG1 and IgG3), which in combination with white blood cells were able to kill the parasite (Roussillon et al., 2007; Audran et al., 2005; Theisen et al., 1998; Oeuvray et al., 1994). As today, vaccines based on the EV approach represent nearly half (48%, 25/52, see Table 1.6) of all the pre-clinical and clinical malaria vaccine trials. During these studies, 38 different formulations have been tested and almost all are based on a few major antigens (AMA-1, MSP-1, MSP-3 and GLURP, see Table 1.7).

A third vaccine approach aims to block parasite transmission by targeting antigens exposed by the *Plasmodium* during its sexual stage. This approach in theory will have little effect on the severity of the disease but it should offer the best chance to progressively reduce and stop the infections. Indeed, the implementation of a sexual stage vaccine (SSV), which targets the gametocytes and/or the minority of infected erythrocytes containing the schizonts, will have a decisive impact on the *Plasmodium* life cycle, as it can disrupt the natural parasite passage from the vertebrate to the invertebrate host (malERA Consultative Group, 2011c). The reduction of the number of infectious mosquitoes will be proportional to the efficacy of the formulation, which depends on the effective disruption of the fertilisation of a female gamete by the male gametocyte inside the midgut invertebrate host.

Thus, the main vaccine targets are represented by those antigens exposed by the parasite before the fertilisation in the gametocytes (Pfs230 and Pfs48/45) and after the zygote formation (Pfs25 and Pfs28). The idea is to trigger the production of antibodies against these antigens in the host. Then, these antibodies will be carried across to the mosquito blood meal. In the vector these immunoglobulins will be able to inhibit parasite development (Chowdhury et al., 2009; Carter et al., 2000). So far only 5 SSV formulations have been tested, 2 of which have been rejected and 3 are in pre-clinical trials. Due

to the complexity of the parasite developmental process and the little impact of the SSV in both malaria morbidity and mortality, the SSV still represent the minority (5/91) of the formulations developed until now (see Table 1.6). Recently, however, the current malaria control strategies started to have an impact on both morbidity and mortality, hence, the role of a transmission blocking vaccine becomes even more important. The use of this approach in conjunction with either PEV or EV techniques and an effective multi-stage antimalarial drug could not only stop the transmission of the disease but also eliminate the appearance of new mutant strains (malERA Consultative Group, 2011b,c).

Vaccine type	Development Status			
	<i>Pre-clinical</i>	<i>Clinical</i>	<i>Rejected</i>	<i>Total</i>
<b>Pre-erythrocytic</b>	7	9	20	36
<i>Whole parasite (PfSPZ)</i>	1	1	0	2
<b>Blood-stage</b>	8	17	13	38
<b>Sexual stage/mosquito antigen</b>	3	0	2	5
<b>Multiple Stage</b>	3	3	4	10
<b><i>Total</i></b>	22	30	39	91

**Table 1.6: Vaccines against malaria: types and development status.**

The pre-clinical projects are the ones that reached a pre-clinical process development stage with a reasonably high chance of reaching clinical evaluation. Vaccines listed as "clinical" have been utilised in the initial immunisations of the first study participants. Vaccines counted as discontinued had previously been under clinical evaluation. Data and development status definitions retrieved from the *Tables of Malaria Vaccine Projects Globally*, WHO, last updated December 2010.

Since the discovery of NAI towards the *Plasmodium* and the possibility to transfer it to unprotected individuals in 1961, the malaria vaccine development went through a long and bumpy route. The supporting arguments towards each specific approach (e.g PEV, EV and SSV) are many. However each vaccine strategy has to be successfully implemented with the malaria control strategies already in place. Therefore, the scientific community has to focus on understanding the parasite interaction with both the human and the mosquito hosts and developing newer and more innovative strategies. One of such involves the use of antigens from different stages of the *Pf* life cycle (multi-stage vaccine, PEV-EV-SSV combined). This approach aims at triggering a sequential immune response towards antigens exposed during the different stage of the parasite life cycle. In theory this will produce a boost in the immunity and



hopefully eliminate all possible routes of escape for the parasite. A multi-stage vaccine will not only limit the severity of the disease but will also be able to block the transmission (malERA Consultative Group, 2011c). However, it is important to consider that this increase in sensitivity could have implications in terms of raising complexity as well as cost and undesired side effects. Recently new findings related to the immunity towards malaria and its antibody-mediated response have suggested that naturally-induced immune protection may not be mediated by a single critical antibody targeted against one dominant antigen. But instead, that antimalarial immunity is a multi-factorial response triggered by the production of specific antibody repertoires against not only a single major antigen but against a pattern of them (Trieu et al., 2011; Gray et al., 2007; Mahanty et al., 2003; Perlmann and Troye-Blomberg, 2002).

Thus the ideal vaccine should probably combine multiple key targets in order to induce the desired breadth of immune response. On the contrary, the current generation vaccines are generally based on just a few well-known and well-characterised antigens (see Table 1.7). At the moment only 7 of the most studied *Pf* antigens (CSP, AMA-1, MSP-1, MSP-2, TRAP, GLURP and EBA175) are used in clinical vaccines trials (n.22), and represent 80% (28 out of 35) of the specific antigen panel. The limited amount of antigens available is highlighted in the total number of trials done so far (n.91), in which a totality of 23 antigens were utilised. Of these panel of antigens, the top 10 candidates have been tested 81% (93/115) of the times. These figure calculated using the Tables of Malaria Vaccine Projects Globally, WHO, (last updated December 2010) showed how we spent the last 30 years of research focusing on less than 0.2% (10/5268 putative genes) of the *Pf* genome/proteome. Hence, it is not unexpected that the vaccine trials carried out until now showed only a partial increase in immune protection (20-50%) (Girard et al., 2007). Most of the new vaccine formulations are still based on the same antigens that failed before. Nearly 48% (11/23) of the antigens currently tested in clinical and pre-clinical evaluations have already failed in previous attempts. An example is the CSP antigen, that has been utilised 23 times so far and has been rejected 12 times. Nevertheless, it still forms the core of the most advance malaria vaccine currently tested, the GSK RTS,S AS01/AS02.

The limited knowledge of the exact NAI targets, combined with the high polymorphism and variability of *Pf* antigens still represents a major challenge to the development of an efficacious and durable vaccine. A deeper understanding of the various immune mechanisms that underlie protection would

N.	Antigen	Development Status				Stage
		<i>Clinical</i>	<i>Pre-Clinical</i>	<i>Rejected</i>	<i>Total</i>	
1	CSP	6	5	12	23	PRE
2	AMA-1	8	3	5	16	BLOOD
3	MSP-1	4	2	5	11	BLOOD
4	LSA1	1	1	8	10	PRE
5	SSP2/TRAP	2		6	8	PRE
6	EXP1	1		5	6	PRE
7	LSA3	1		5	6	PRE
8	STRAP	1		4	5	PRE
9	MSP-3	4			4	BLOOD
10	Pfs25		2	2	4	SEXUAL
11	GLURP	2		1	3	BLOOD
12	pb9- <i>Pb</i> CSP	1		2	3	PRE
13	CelTOS		2		2	PRE
14	EBA175	2			2	BLOOD
15	MSP-2			2	2	BLOOD
16	Pfs16			2	2	SEXUAL
17	Var2-CSA		2		2	BLOOD
18	PfSPZ	1	1		2	WHOLE
19	P27A		1		1	BLOOD
20	Pfs 48/45		1		1	SEXUAL
21	RESA			1	1	BLOOD
22	SERA5	1			1	BLOOD
23	NA		1		1	NA
<b>Total</b>		35	20	60	115	<b>Overall</b>
		13	8	42	63	<b>PRE</b>
		21	8	14	43	<b>BLOOD</b>
		0	3	4	7	<b>SEXUAL</b>
		1	1	0	2	<b>WHOLE</b>

**Table 1.7: *Pf* antigens utilised in the malaria vaccine development.**

PRE: Pre-erythrocytic antigen expression. BLOOD: Erythrocytic antigen expression. SEXUAL: Sexual antigen expression. Development status definitions see Table 1.6. The information was retrieved from the *Tables of Malaria Vaccine Projects Globally*, WHO, last updated December 2010.

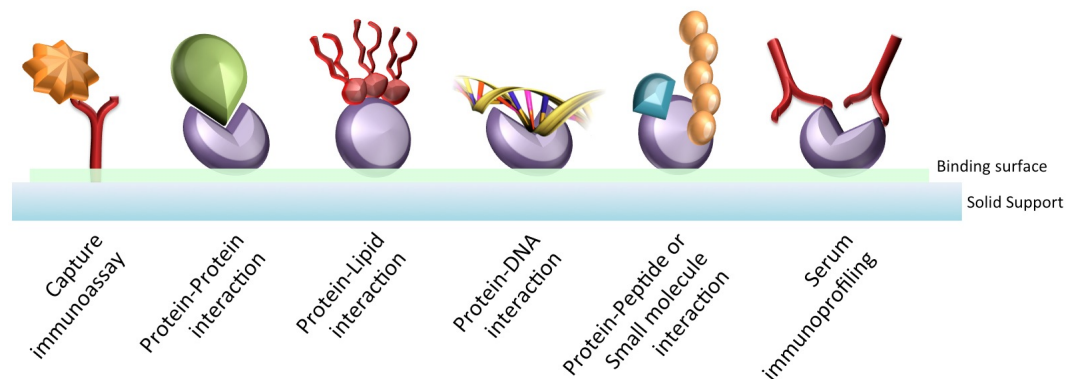
certainly boost the development of a broadly effective malaria vaccine. Of the three major vaccine approaches (PEV, EV and SS), the EV is perhaps the strategy that most likely replicates successfully the NAI (Bull et al., 1998) and reduces mortality and morbidity in malaria-endemic regions. However, regardless of the vaccine strategy adopted, the main and critical step still remains the search for the appropriate key antigens involved in generating protection. Now that the *Pf* genome has been revealed and its information is available world-

wide, a multiple antigen vaccine with higher efficacy is probably closer. The combination of such precious information, together with the high-throughput experimental tools that are now available, allows the assessment of protein expression, function and antigenicity at the proteome scale. Such wider approach is fully embraced by the recent advances in the reverse vaccinology field (Sette and Rappuoli, 2010; Rappuoli, 2000). This is an emerging technology that combines the genomic and immunological knowledge to reveal antigens that can be utilised in diagnostic and vaccination. The milestones achieved in the last decade, both in genomic (e.g. genome sequence capabilities) and in proteomic (e.g. high-throughput protein expression, protein microarray etc.) technologies, enabled us to exploit the pathogens proteomes and to identify novel antigens, without the need to grow the pathogen: anthrax (Ingram et al., 2010), *Pf* (Doolan et al., 2003), *Mycobacterium tuberculosis* (McMurry et al., 2005); or a homology-free method for the prediction of novel protective antigens (Magnan et al., 2010). This paradigm shift, which moves away from the old Pasteur's rules of obtaining a vaccine by isolating, inactivating, and injecting the causative microorganism, is progressively enabling us to overcome many of the difficulties and limitations that until now have slowed down the quest for this essential malaria control tool.

## 1.4 Protein microarrays

Traditionally, complex cellular systems and the biochemical activity of proteins have been studied by individually investigating the respective components or proteins. In a post-genomic era, however, where scientists have access to the properties of a full organism's proteome, this approach is clearly sub-optimal. The understanding of a proteome and its complexity requires high-throughput methods that can allow the simultaneous, parallel analysis of thousands of parameters/interactions within a discrete micro-environment. Since its introduction the microarray technology (DNA microarrays: Schena et al. (1995); protein microarrays MacBeath and Schreiber (2000); carbohydrate microarrays Wang et al. (2002)) has emerged as one of the most promising and challenging innovations for a wide variety of new applications including genomic/proteomic research, clinical diagnostics, disease biology, drug target identification and biomarker profiling (see Figure 1.10).

The use of either contact robotic arraying (Dottorini et al., 2011; Di Cristina et al., 2010; Gray et al., 2007; Zhu et al., 2001; MacBeath and Schreiber, 2000)



**Figure 1.10: Protein microarray applications.**

The protein microarray platform can be used in a number of different applications. Antibody arrays are mostly used as capture immunoassays in clinical diagnosis and for the analysis of food and environmental safety. Functional protein arrays instead, are commonly used to generate complex serum profiles, to characterise enzyme substrates or evaluate protein activities (e.g protein/protein, protein/lipid, protein/DNA, protein/drug, protein/peptide interactions).

or non-contact robotic arraying (Jones et al., 2006; Delehanty, 2004) allows the spotting of minute amounts (pL-nL scale) of aqueous biomolecules onto a solid surface (typically a glass microscope slide or a nitrocellulose membrane) in an ordered and high density manner ( $10^2$ - $10^3$  samples in  $<1$ - $2$  cm<sup>2</sup>). It is clear that via its intrinsic capacity for miniaturisation, the microarray format can finally overcome the spatial, logistical and economic limitations that have made truly parallel genome- and proteome-wide investigations impossible in the past. Indeed, the microarray system allows the screening of thousands of natively spotted proteins for drug discovery, protein-protein interaction analysis, antibody profiling and disease biomarker assessment. By now hundreds of organisms have been successfully sequenced, allowing fields such as genetics and proteomics to achieve new and unprecedented findings. Especially in the last few years, large-scale protein analysis has begun to play a main role in this new and exciting era, focusing the attention towards a deeper and wider understanding of the incredible complexity of biological networks and systems biology. Since 2001 with the first yeast proteome-wide microarray (Zhu et al., 2001), over 15 different platform have been developed and utilised for proteome-wide studies (Yang et al., 2011). Even a human "ProtoArray" containing over 9000 purified human proteins, has been recently (i.e. 2010) developed and commercialised by Invitrogen (ProtoArray<sup>®</sup> Human Protein Microarray, Life Technologies Corporation).

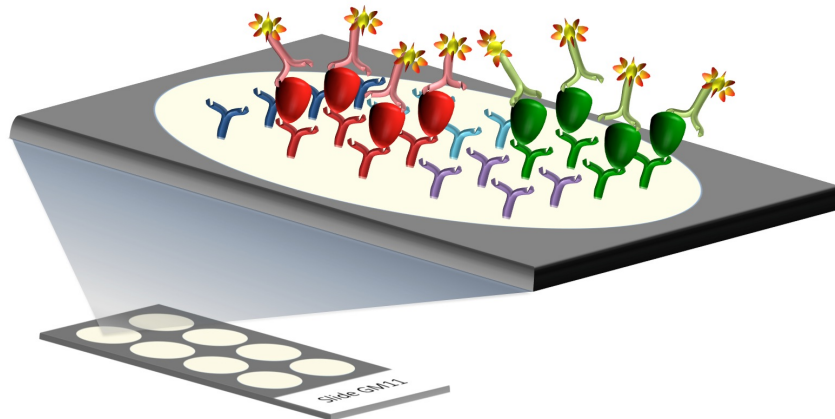
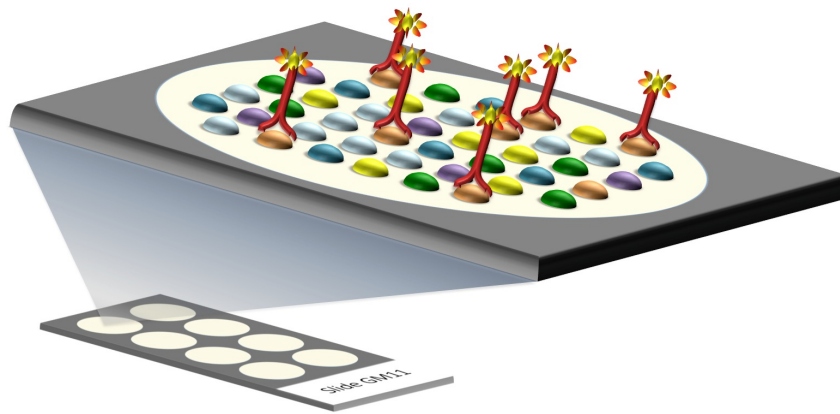
### 1.4.1 Protein microarrays: sub-classes

The protein microarray technology embraces features like miniaturisation and parallelism to allow the simultaneous analysis of large protein repertoires within a high-density and ultra-sensitive (zeptomol range) common micro-environment format. Depending on the assay structure, two main sub-classes of protein microarrays have emerged for use in characterizing the activity, expression or antigenicity of proteins: forward- or reverse-phase microarray (see Figure 1.11 Hartmann et al. (2009)).

**Forward-phase** This format of microarray is usually based on a panel of capturing agents that are printed on a planar surface and incubated with a single sample. Depending on the characteristic of the capturing molecules the assay can be classified as follows (Hall et al., 2007):

1. Analytical protein microarrays, which are commonly used to characterize complex protein mixtures in terms of biomolecular affinities, specificities and expression levels. Usually a library of antibodies, affibodies (i.e. non-immunoglobulin binding protein based on combinatorial protein engineering of the small and robust  $\alpha$ -helical structure of *Staphylococcus aureus* protein A), aptamers or antigens are printed onto a solid surface and probed with the sample of interest (Dottorini et al., 2011; Renberg et al., 2007; Bacarese-Hamilton et al., 2004; Mezzasoma et al., 2002).
2. Functional protein microarrays, which are typically used to assess a broad spectrum of biochemical activities of specific protein repertoires or proteomes. The technique is based on a chip containing protein domains or full-length functional proteins that can provide insight into their individual bimolecular function (e.g. kinase activity) (Hall et al., 2004; Zhu et al., 2001).

**Reverse-phase** This format of arrays contains antigens or whole cells, which are isolated from the tissue of interest, lysed and spotted onto a solid support. The chip is subsequently probed with labelled-antibodies specific against the protein target. This array format provides us with an innovative tool for discovering new disease biomarkers or antigens for diagnostic assays (Gilabert et al., 2010; Yu et al., 2010).

(a) *Forward-phase microarray*(b) *Reverse-phase microarray***Figure 1.11: Protein microarray classes.**

### 1.4.2 Protein microarrays: use in serum profiling

The recent findings on immune reactivity profiles for allergy and asthma predisposition (Dottorini et al., 2011), SARS infection (Zhu et al., 2006) and for malaria vaccine candidates (Trieu et al., 2011; Gray et al., 2007) using protein chips have shown how this promising technique is becoming one of the most powerful tools in the fields of diagnostics and serum profiling. The protein microarray technology enables the simultaneous assessment of a patient's serum for hundreds-thousands of distinct antigen-antibody interactions within a single test that lasts less than two hours and requires only a few microlitres of serum sample. The key qualities of true-parallelism, low antigen requirement and simultaneous antigen assessment make protein microarrays an ideal platform on which to build a system for rapidly identifying antigenic

proteins (Bacarese-Hamilton et al., 2002). ‘Virtual’ microbial proteomes, constructed from genome sequence information, have recently revealed a range of proteins encoded by a number of microbes of high medical importance. From the perspective of controlling disease, there now exists an urgent need to identify novel techniques that can exploit this added information by permitting high-throughput assessment of microbial proteins for their antigenicity. An example of this strategy is the use of a high-throughput *in vitro* transcription and translation system (e.g. *E. coli*) coupled with robotic cloning, expression and spotting procedures to create a proteome-wide microarray platform that contain a panel of antigens previously selected via an *in silico* antigenicity analysis (see Table 1.8). This approach has been developed and optimised by P.L. Felgner and its team at the University of California (Irvine, CA, USA). Since its first proteome-chip, which contained the entire proteome of the *Vaccinia* virus (i.e. 185 viral proteins, Davies et al. (2005)), the group has developed a high throughput antigen screening system. This procedure allows the reproduction of entire proteomes given the genomic sequence of the organism of interest. Supported by the use of robotic workstations, the team can rapidly clone hundreds of PCR-products into standard expression vectors that are eventually used to synthesise hundreds of proteins *in vitro* per week. The newly expressed candidates can then be spotted as crude extracts on a microarray format and utilised for serum profiling. Using this platform Felgner’s group tested large portions of proteomes belonging to a wide range of microorganisms (*Vaccinia*, *F. tularensis*, *C. burnetii*, *B. pseudomallei*, *B. melitensis* and *Pf*, Table 1.8).

Target organism	N. proteins	Proteome coverage	Group	Reference
<i>Vaccinia</i>	185	~99%	P.L. Felgner	Davies et al. (2005)
<i>Francisella tularensis</i>	1741	~99%	P.L. Felgner	Eyles et al. (2007)
<i>Coxiella burnetii</i>	1988	~97%	P.L. Felgner	Beare et al. (2008)
<i>Burkholderia pseudomallei</i>	1205	~21%	P.L. Felgner	Felgner et al. (2009)
<i>Chlamydia trachomatis</i>	225	~27%	L.M. de la Maza	Molina et al. (2010)
<i>Candida albicans</i>	363	~6%	H.P. Liu	Mochon et al. (2010)
<i>Brucella melitensis</i>	1406	~44%	P.L. Felgner	Liang et al. (2010)
<i>Mycobacterium tuberculosis</i>	4099	~99%	M.L. Gennaro	Kunnath-Velayudhan et al. (2010)
<i>Plasmodium falciparum</i>	1204	~23%	P.L. Felgner	Crompton et al. (2010)

**Table 1.8: List of *in vitro* expression-based proteome microarrays**

Their microarray screening of *in vitro* expressed proteomes is now at the core of the company recently founded by Felgner, Antigen Discovery Inc. (Irvine, CA, USA), which aims to play a leading role in the future of vaccine

discovery. Indeed, the combination of empirical information, bioinformatics and high-throughput antigen screening may, in the long-term, lead to vaccine preparations with enhanced efficacy as a result of placing greater emphasis on unbiased, merit-based selection of potential vaccine candidates. Attention also has to be placed, however, on the quality of the information obtained from such cutting-edge technology. The constant race towards automation and high-throughput production may, in some occasions, impose limitations on the quality of the unpurified proteins synthesised *in vitro*. It is for these reason that a number of scientists are focusing their attention and efforts on producing a proteome-microarray based on purified natively-folded proteins. Indeed, the use of a purified protein candidate, although much more complicated to produce, allows not only a qualitative analysis of the protein interactions (e.g. positive/negative binary response) but also the quantification of the protein-related reactions. Due to the complexity in the fabrication of such high-quality platform, the number of studies based on this kind of chips is limited. The organisms covered so far are only a few (e.g. *N. meningitidis* Steller et al. (2005), SARS-CoV, (Zhu et al., 2006) and the Human ProtoArray<sup>®</sup> from Invitrogen). A proof-of-principle study on the *Pf* parasite was carried out in our laboratory by J.C. Gray (Gray et al., 2007). Using an in house protein microarray platform, Gray et al. were able to simultaneously screen and assess the humoral response of 189 African children towards 18 recombinant proteins obtained from some of the major erythrocytic stage *Pf* antigen candidates (i.e. AMA-1, MSP-1, MSP-2, MSP-3). In this study, the authors demonstrated the complexity of the naturally-induced humoral immune response. Indeed, they provided evidence that protection against malaria may rely on the ability of an individual to raise antibodies towards a combination of parasite antigens, thus questioning the idea that NAI can be achieved with an antibody response towards a specific single antigen (Gray et al., 2007; Parekh and Richie, 2007). Clustering analysis performed on the 189 serum profiles demonstrated that basing studies on the assessment of antigens in isolation can overlook antigen involvement in possible “higher order” protective patterns. The idea that not one but multiple antigens correlate with *Pf* immune protection is also supported by the work of Trieu et al., which was only recently published (Trieu et al., 2011). In this study the group compared the serum profiling of patients without and with sterile protection, induced by irradiated *Pf* sporozoites, using a microarray platform (developed by the Felgner’s group) containing 1200 *Pf* proteins expressed via an *E. coli in vitro* expression system. The analysis highlighted the fact that the reactivity towards the single antigens did not



correlate with protection. However when the reactivity of 19 different antigens were clustered, a significant association between the antigen signature and the protective immunity was observed.

Overall, the data gathered so far underline the validity of the protein microarray platform as a powerful tool that can play a pivotal role in the investigation of the antibody-driven immune response towards large sets of protein targets. Proteome wide repertoires of putative antigens can easily be screened in either small, medium or large epidemiological studies (Trieu et al., 2011; Gray et al., 2007). It is envisaged that through the optimisation of this technology, innovative and ground-breaking diagnostic devices will be soon made available for both basic research, small POC and large clinical laboratories (Yu et al., 2010; Kricka et al., 2010; Hartmann et al., 2009).

## 1.5 Aim of my PhD project

The objective of my PhD was to develop a high-throughput platform for the selection, production and immune screening of new *Pf* vaccine candidates. The new platform should facilitate the discovery of *Pf* blood-stage antigens that, either alone or in combination, are able to induce the acquired immunity to malaria, which is usually developed naturally by the host as a result of repeated exposure to the pathogen. Recent studies not only have demonstrated the importance of simultaneously measure the *Pf* antigenic signatures to analyse the malaria immune protection (Trieu et al., 2011; Crompton et al., 2010; Gray et al., 2007; Davies et al., 2005), but also have highlighted the challenges and limitations that rise due to the complexity of both parasite and human. On these bases, we aimed to synergistically combine innovative approaches, such as reverse vaccinology, *in silico* protein modelling and protein microarrays, with well establish technologies such as heterologous *E. coli* expression and immunoaffinity chromatography to provide a valid platform for the multiplexed antigenic analysis of hundreds of high quality *Pf* proteins by comparing malaria protected and unprotected immune profiles. The milestones to be achieved along the project were four: the initial bioinformatic screening of the *Pf* proteome for blood-stage expressed proteins with antigenic properties; the selection and implementation of a high-throughput cloning system; the development of a suitable expression platform for the production of challenging parasite proteins; and the evaluation of a multiplex assay capable of performing a high-throughput serum profiling analysis on sample collected in

malaria endemic regions.

My PhD project was part of a European research project (FightMal) that has been sponsored by the European Community (FP7 Marie Curie). Skills, knowledge and resources were mobilised between a consortium of 5 members of the industry (Microtest Matrices Ltd. (MtM, UK), PX' Therapeutics (PX', France), Med Biotech Laboratories (MBL, Uganda)) and academia (University of Perugia (UNIPG, Italy) and London School of Hygiene and Tropical Medicine (LSHTM, UK)). Different research groups led or participated in different steps of the project; however a high degree of synergy was maintained throughout the study with the establishment of mixed working groups. The scope of this PhD, therefore, span all the disciplines while maintaining a particular focus on the initial selection/production of candidate antigens, the construction, validation and processing of protein microarray immunoassays and the analysis of microarray data and associated serum reactivity profiles.

## Chapter 2

# MATERIALS AND METHODS

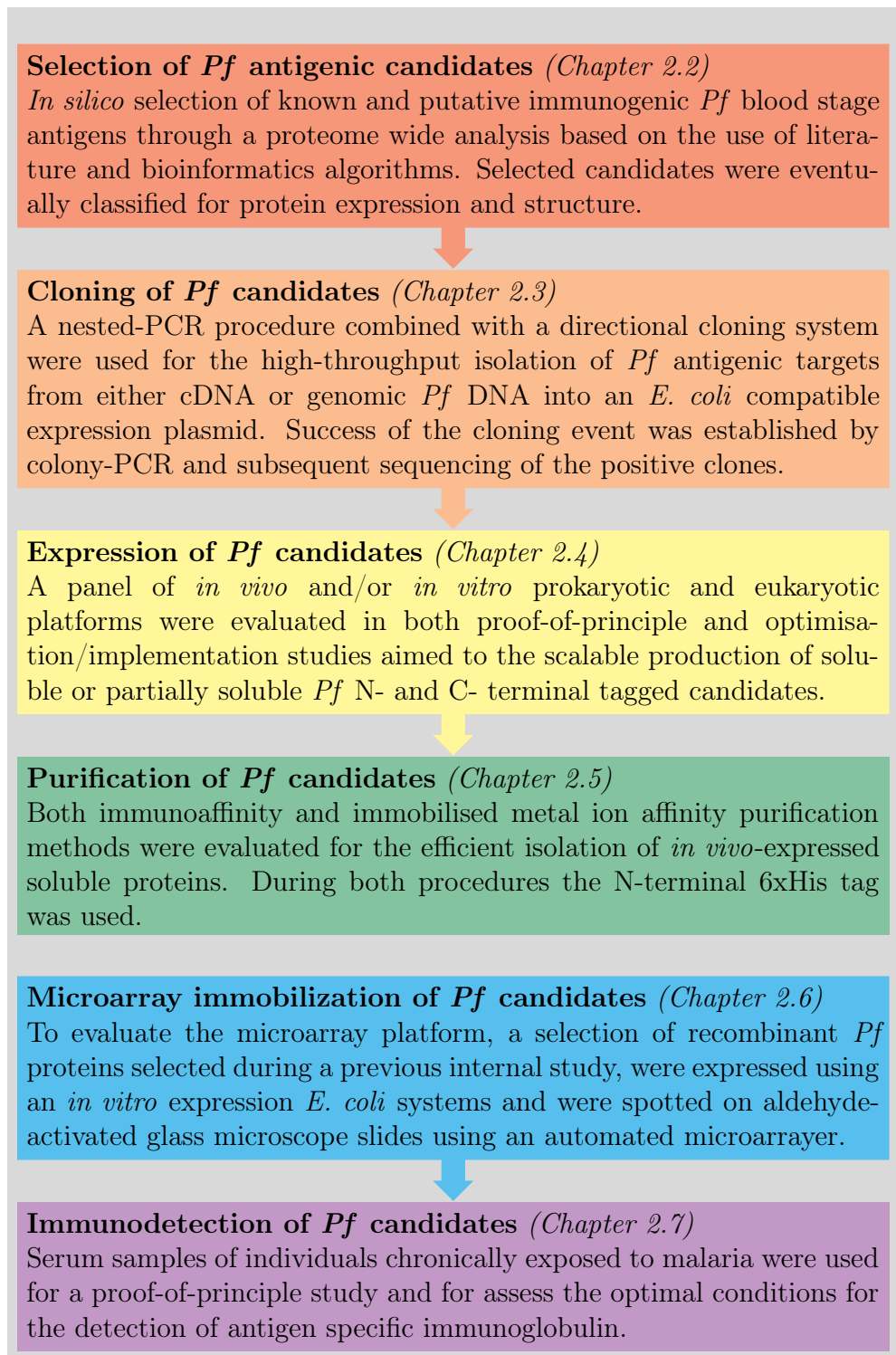
### 2.1 Overview

During my PhD, a series of innovative and well established techniques, developed in the fields of bioinformatics, molecular biology, proteomics and immunology, were identified and evaluated in either proof-of-principle or design/implementation studies. An overview of methodology used is briefly summarised in Figure 2.1.

### 2.2 Bioinformatic selection of *Pf* antigens

One of the major deliverables of the FightMal project was the production of three protein microarray chips, each of which containing a specific set of *Pf* antigens. CHIP 1 includes a panel of already well known antigens, CHIP 2 covers a set of putative antigenic candidates selected using a reverse vaccinology approach and CHIP 3 contains conserved domains of highly variable surface antigens. The first milestone of my project was the *in silico* selection of both known and potentially immunogenic *Pf* antigens to be included within the three arrays. Following an exhaustive search of the literature and available databases (see below), a panel candidates, was drawn up with the help of Dr. Tania Dottorini from the UNIPG consortium partner. Scientific hypotheses, biological evidences and bioinformatics tools (see below) were used to identify surface exposed *Pf* antigens expressed during the blood stage, and hence potential targets of protective anti-malarial immunity. The criteria used to select candidate antigens for the three individual microarray chips are described below.

**CHIP 1 The vaccine candidate antigens chip** This microarray chip will



**Figure 2.1: Methodology Overview.**

Description of the techniques applied during this PhD. Each section contains a brief summary of the activities and methodologies utilised in the specific task. The chapter number of each section is also available to facilitate the reader.

contain the *Pf* antigens (expressed both during the liver and blood stage) currently being investigated in the literature for their ability to elicit protective immunity in animals and humans. These antigens were identified following a comprehensive search of the web and the literature. The main web sources used were: the WHO, the Malaria Vaccine Initiative (MVI) and the National Center for Biotechnology Information (NCBI, [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Further information was retrieved from our collaborator Dr. Patrick Corran at the LSHTM.

**CHIP 2 Putative antigenic targets chip** This microarray chip will contain recombinant *Pf* proteins that, on the basis of their structure or experimental evidence, are predicted to be either secreted or exposed on the surface of this malaria parasite. DNA sequences encoding putative antigens were retrieved from the *PlasmoDB* ([www.plasmodb.org](http://www.plasmodb.org), Kissinger et al. (2002)) and GeneDB ([www.genedb.org](http://www.genedb.org), Hertz-Fowler et al. (2004)) databases, which retain genome sequence data for the 3D7 *Pf* isolate (Gardner et al., 2002). A combination of biological results and *in silico* predictions was used to narrow down the number of antigen candidates to 5-6% of the *Pf* proteome. Experimental evidence was obtained for: merozoite surface proteins (Cowman and Crabb, 2006); GPI-anchored proteins (Gilson et al., 2006); parasite-infected erythrocyte surface proteins, PIESPs (Florens et al., 2004); and *PlasmoDB* (annotated antigens). Bioinformatics analysis for the prediction of secretion and/or extracellular exposure was performed using the following algorithms: PEXEL (Marti et al., 2004), HT (Hiller et al., 2004), Export-Pred (Sargeant et al., 2006), PlasmoAP (Foth et al., 2003), SignalIP 3.0 (Bendtsen et al., 2004) and *PlasmoDB* ([www.plasmodb.org/plasmo/](http://www.plasmodb.org/plasmo/)), see Table 2.1.

Furthermore, results from two previous internal studies carried out under an FP6 EU grant by Dr. Dottorini (i.e. Dottorini FP6) at UNIPG and a second study performed by Dr. Patrick Corran at the LSHTM (i.e. Corran 2008) were added. The resulting list of annotated antigens were first organised according to the source of origin and subsequently depleted for multiple entries. RIFIN and STEVOR multi-gene family members were also excluded. Remaining candidates were investigated for structural and biological characteristics using the following algorithms: TMHMM2 (Krogh et al., 2001), Phobius (Kall et al., 2007), Interpro Domain-Pfam ([www.ebi.ac.uk/interpro/index.html](http://www.ebi.ac.uk/interpro/index.html)), SSpro and SSpro8 (Pollastri et al.,

EXPORTOME and/or PARASITE SURFACE ATTRIBUTES		
PREDICTOR	INFO	AUTHORS
PEXEL	A conserved pentameric sequence that plays a central role in protein export into the host cell and predicts the exported proteome in <i>Pf</i> .	Marti et al. (2004)
HT	An 11-amino acid signal peptide required for the secretion of proteins from the <i>Pf</i> vacuole into the human erythrocyte.	Hiller et al. (2004)
ExportPred	Uses a generalised hidden Markov model (GHMM) to model simultaneously the signal sequence and PEXEL motif features, which are required for protein export.	Sargeant et al. (2006)
PlasmoAP	An amino acid sequence constituting transit peptides which target proteins to the relict plastid (apicoplast) of malaria parasites.	Foth et al. (2003)
SignalIP 3.0	Two different predictors, based on neural network and hidden Markov model algorithms, able to predict the presence of a signal peptide sequence and cleavage site.	Bendtsen et al. (2004)

**Table 2.1: Exportome and/or parasite surface attributes, list of algorithms.**

List of algorithms used for the prediction of relevant attributes present in exported/extracellular *Pf* candidates.

2002), DIpro 2.0 (Cheng et al., 2005), DGPI ([www.gpi.pathbot.com](http://www.gpi.pathbot.com)), GPI-SOM ([www.gpi.unibe.ch](http://www.gpi.unibe.ch)) and big-PI Predictor (Eisenhaber et al., 1999), see Table 2.2.

For membrane proteins, to increase the probability of producing soluble products, known or possible hydrophobic regions, such as transmembrane domains, signal peptides or GPI anchors were excluded from the region believed to be exposed extracellularly. Finally, a priority score was applied to further narrow down the selection (see Table 2.3).

Experimental evidence of blood stage expression from Mass Spectrometry (MS) and mRNA microarray (MA) analysis was utilised to score each gene. Priority was given to proteins expressed during the erythrocytic stage of *Pf* as determined by MS. Antigens having (or predicted to have) multiple transmembrane domains with short extracellular loops (<100 Aa, SEL) were negatively scored (-6). The final scoring ranged from a maximum of 6 to a minimum of -6. An arbitrary cut off of  $\geq 5$  was chosen. Candidates with score of 4 were evaluated on a one by one base.

CANDIDATE STRUCTURE ANALYSIS		
PREDICTOR	INFO	AUTHORS/ SOURCE
TMHMM2	A membrane protein topology prediction method based on a hidden Markov model.	Krogh et al. (2001)
Phobius	A prediction method for transmembrane topology and signal peptides from the amino acid sequence of a protein.	Kall et al. (2007)
Interpro Domain-Pfam	A database of protein families, domains, regions, repeats and sites in which identifiable features found in known proteins can be applied to new protein sequences.	www.ebi.ac.uk/interpro
SSpro- SSpro8	Protein secondary structure prediction based on an ensemble of 100 1D-RNNs (one dimensional recurrent neural networks).	Pollastri et al. (2002)
DIpro 2.0	Cysteine disulfide bond predictor based on 2D recurrent neural network, support vector machine, graph matching and regression algorithms.	Cheng et al. (2005)
DGPI	Detection/prediction of GPI cleavage site (GPI-anchor) in a protein (DGPI).	www.dgpi.pathbot.com http://129.194.185.165/dgpi (V.2.03)
GPI-SOM	Identification of GPI-anchor signals by a Kohonen Self Organizing Map.	www.gpi.unibe.ch Fankhauser and Maser (2005)
big-PI Predictor	GPI modification site prediction.	www.mendel.imp.ac.at/sat/gpi Eisenhaber et al. (1999)

**Table 2.2: Candidate structure analysis, list of algorithms.**

List of algorithms used for the analysis of the predicted structure of the selected *Pf* candidates.

**CHIP 3 Variant surface antigens chip** This microarray chip will contain individual domains (i.e. NTS, C2, DBL, CIDR) of the 74 PfEMP1 VSA proteins encoded by the multi-gene VAR family. A panel of antigens and their specific domains was obtained from the *PlasmoDB* database (www.plasmodb.org, release 5.4 and 5.5). The key words used for the queries were: text *PfEMP1*, no pseudogenes, genes specific for *Pf*. RIFIN and STEVOR family members were also taken into consideration as they were the most populated VSA groups during the computational analysis performed for CHIP 2.

## 2.3 Isolation and cloning of *Pf* candidates

A high-throughput PCR cloning and expression platform was developed to clone and express the selected *Pf* antigens. In particular, a nested-PCR pro-

**SCORE A EXPRESSION EVIDENCE**

<b>6</b>	MS and MA Blood stage evidence
<b>5</b>	MS Blood and MA not available (NA)
<b>4</b>	MS Blood and MA Liver/Sexual
<b>3</b>	MS NA and MA Blood
<b>2</b>	MS NA and MA NA
<b>1</b>	MS Liver/Sexual and MA Blood
<b>0</b>	MS Liver/Sexual or NA and MA Liver/Sexual or NA

**SCORE B ADDITIONAL NEGATIVE ATTRIBUTES**

<b>-6</b>	Short Extracellular Loops (SEL)
-----------	---------------------------------

**Table 2.3: Priority selection score for CHIP 2.**

The table contains the score utilised to prioritize the selection of putative candidates according to blood stage expression evidence (A) obtained with Mass Spectrometry (MS) and Microarray (MA). Additionally, candidates were scored for “negative attributes” (B) like extracellular loops shorter than 100 Aa. The final score was calculated as the sum of A+B. The cut-off for the selection was arbitrarily set at 5 (green). Candidates scoring 4 (yellow) were individually evaluated and where possible included in the final list.

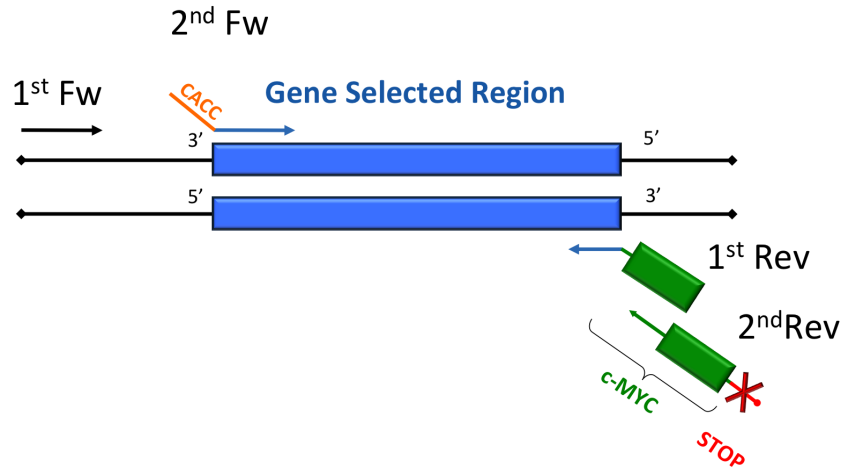
cedure was used to amplify the selected gene or portion of gene (section 2.3.1). PCR products were ligated into a TOPO<sup>®</sup>-directional expression vector (section 2.3.2), pET160/GW/D-TOPO<sup>®</sup> (Invitrogen), which was then inserted and propagated in a TOP10 *E. coli* strain (section 2.3.3 and 2.3.5). Colony PCR (section 2.3.4) was used to identify colonies carrying full-length, correctly orientated DNA targets, which were eventually fully sequenced and characterised (section 2.3.6).

**2.3.1 Polymerase Chain Reaction (PCR)**

Sequences of interest were amplified using a nested-PCR technique. As a source of DNA, four cDNA blood-stage libraries or genomic DNA (MRA-296, MRA-297, MRA-298, MRA-299 and MRA-102G, source MR4: ATCC, Manassas, VA, USA) derived from *Pf* 3D7 parasites were used. cDNA libraries were pooled together equally [v/v], diluted 1:10 using ddH<sub>2</sub>O, or used individually according to the MS expression data. Genomic DNA was used at a concentration range of 5 to 90 ng. DNA was boiled for 5 minutes (min) and used as a template for amplification of the selected ORFs. The procedure consisted of a double amplification step allowing incorporation of the TOPO<sup>®</sup>-directional entry sequence CACC at the 5'-end and a sequence encoding the c-MYC tag at the 3'-end. This 3'-end sequence consists of 39 nt and encodes a 2 Aa spacer



(-AS-), the 10 Aa c-MYC epitope (-EQKLISEEDL-) and the final stop codon. During the first amplification step a forward primer, annealing upstream to the selected sequence, and a reverse primer containing the first portion of the 3'-end sequence were used (Figure 2.2).



**Figure 2.2: Nested PCR strategy and primers.**

The figure describes the strategy used to amplify the gene selected region using a nested PCR procedure. DNA of interest (■); Fw: forward primer; Rev: reverse primer; 1st: first PCR step; 2nd: second PCR step; CACC: TOPO-directional entry sequence (■); c-MYC: c-Terminus tag (■) containing a 2Aa spacer and followed by a stop codon (■).

The resulting PCR products were further amplified during the second amplification step using a forward gene-specific primer (containing the 5'-CACC entry sequence) and a universal reverse primer, capable of annealing with the first half of the 3'-end sequence and containing the remaining encoding information for the epitope and stop codon. The design of the PCR primers to amplify the DNA of interest is shown in figure and Table 2.4.

The first step of the PCR was carried out using 2  $\mu\text{L}$  of DNA (either 1:10 of cDNA library pool or genomic DNA), 0.5  $\mu\text{M}$  of the 1st forward and 1<sup>st</sup> reverse primers, 200  $\mu\text{M}$  of dNTP, 1X Phusion HF buffer and 0.02 U/ $\mu\text{L}$  Phusion high fidelity DNA Polymerase (Finnzymes) in a final volume of 25  $\mu\text{L}$ . PCR products were column-purified using the QIAquick PCR Purification kit according to the manufacturer's instructions (Qiagen) and eluted in 25  $\mu\text{L}$  of sterile water. The second PCR step was carried out using 2  $\mu\text{L}$  of the purified product from the first amplification step, 0.5  $\mu\text{M}$  of the 2<sup>nd</sup> forward and 2<sup>nd</sup> reverse primers, 200  $\mu\text{M}$  of dNTP, 1X Phusion HF buffer and 0,02U/ $\mu\text{L}$  Phusion high fidelity DNA Polymerase (Finnzymes) in a final volume of 25  $\mu\text{L}$ . The final products were purified as described above and eluted in 50-100  $\mu\text{L}$  of sterile water. The nested-PCR amplification steps were performed using the following criteria: 5

---

**FIRST STEP OF AMPLIFICATION**


---

1<sup>st</sup>For 5'-nnnnnnnnnnnnnnnnnn-3'

1<sup>st</sup>Rev 5'-CAGTTTCTGTTTCGGACGCnnnnnnnnnnnnnnnnnn-3'

---

**SECOND STEP OF AMPLIFICATION**


---

2<sup>nd</sup>For 5'-**CACC**nnnnnnnnnnnnnnnnnn-3'

2<sup>nd</sup>Rev 5'-**TTA**TAAATCTTCTTCAGAAATCAGTTTCTGTTTCGGACGC-3'

**Table 2.4: Sequences of the Nested PCR primers.**

Template sequences used to create the sequence-specific primers for the double-step nested-PCR. Fw: forward primer; Rev: reverse primer; 1st: first PCR step; 2nd: second PCR step; nnn: gene specific nucleotides; CACC (■): TOPO-directional entry sequence; GGACGC (■): 2 Aa Spacer Ala-Ser; TAAAT...TGTTTC (■): 10 Aa c-MYC, Glu-Gln-Lys-Leu-Ile-Ser-Glu-Glu-Asp-Leu; TTA (■): Stop codon. The sequence underlined in the 1<sup>st</sup> and 2<sup>nd</sup> reverse primers is the overlapping/annealing sequence of the two primers.

min at 95°C, 25 cycles (10 seconds (sec) at 95°C, 30 sec Temperature Gradient, 30/kb sec extension at 72°C), 10 min at 72°C. The annealing temperature was calculated by subtracting 3°C from the melting temperature of the annealing portion of the primers. The latter was calculated with a customised Finnzymes Tm calculator available at [www.finnzymes.fi/tm\\_determination.html](http://www.finnzymes.fi/tm_determination.html).

Final PCR products were analysed on an agarose gel (1% [w/v] DNA-grade agarose in 0.5 x TAE buffer (20 mM Tris-Acetate, 0.5 mM EDTA [pH 8.0], [pH 8.3]) at 90-120 V. Ethidium bromide (0.5 µg/mL) was used to visualise the DNA. Quality analysis of the bands was performed under ultraviolet (UV) light of 132 nm using a transilluminator. Estimation of the size of the amplified PCR products was carried out using 5 µL of Hyperladder I<sup>TM</sup> (Bioline), which was run alongside the samples.

### 2.3.2 Ligation into the pET160/GW/D-TOPO<sup>®</sup> vector

Plasmid vectors were supplied linearised with 3'-GTGG overhangs with a topoisomerase I from Vaccinia virus covalently bound to specific sites (CC-CTT) (Shuman, 1991), see Figure 3.10. PCR products were ligated into the pET160/GW/D-TOPO<sup>®</sup> expression vector (Invitrogen) according to the TOPO<sup>®</sup> Cloning reaction protocol supplied by the manufacturer. Briefly, TOPO<sup>®</sup> Cloning reactions were set up adding from 0.2 to 2 µL of the PCR product to a reaction solution containing: 10ng pET160/GW/D-TOPO<sup>®</sup>,

8.33% glycerol, 8.33 mM Tris-HCl [pH 7.4], 0.2 M NaCl, 10 mM MgCl<sub>2</sub>, 0.166 mM EDTA, 0.166 mM DDT, 0.0166% Triton X-100 and 16.66  $\mu\text{g}/\text{mL}$  BSA, giving a final reaction volume of 3  $\mu\text{L}$ . The ligation procedure was generally performed keeping a vector to insert ratio of about 1:1, but ratios ranging from 2:1 to 1:4 were also tested. The ligation solution was gently mixed and left at room temperature (RT) for 30 min. Once the incubation was completed, the reaction was left on ice and used immediately to transform *E. coli* TOP10 cells.

### 2.3.3 Transformation of bacteria

The transformation reactions were performed on One Shot<sup>®</sup> TOP10 Chemically Competent *E. coli* cells using a heat-shock protocol, as suggested by the manufacturer's instructions. The full ligation volume, 3  $\mu\text{L}$ , was gently mixed with 50  $\mu\text{L}$  of TOP10 *E. coli* cells and incubated for 30 min on ice. Once ready, the cells were heat-shocked for 30 sec at 42°C and immediately placed on ice for at least 2 min. After the addition of 250  $\mu\text{L}$  of SOC medium (2% Tryptone [w/v], 0.5% yeast extract [w/v], 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM MgSO<sub>4</sub>, 20 mM glucose) a recovery incubation was performed at 37°C for 1 hour at 200 rpm shaking. Finally, the total volume of transformation reaction (300  $\mu\text{L}$ ) were plated on pre-warmed ampicillin (100  $\mu\text{g}/\text{mL}$ , amp100) plates and incubated at 37°C overnight (O/N).

### 2.3.4 Colony PCR

Colony PCR was performed to identify colonies containing plasmids with the *Pf* DNA sequence in the correct orientation. This procedure is based on the fact that using a common T7 forward primer, able to anneal 220 base pairs (bp) ahead of TOPO<sup>®</sup> recognition site 1, and a universal reverse primer annealing to the final portion of the c-MYC and the first 6 bp of the vector, it is possible to identify inserts that are both correctly orientated in the plasmid and carrying a full length c-MYC tag. This step also allows the evaluation of the insert's size by electrophoresis. To carry out the colony PCR, colonies (3 to 10 per transformation) were picked and resuspended in 10  $\mu\text{L}$  of ddH<sub>2</sub>O. A PCR reaction was then set up mixing 0.4  $\mu\text{M}$  of each primer with 0.5  $\mu\text{L}$  of the resuspended colony in a final volume of 15  $\mu\text{L}$  containing 1xPlatinum<sup>®</sup> PCR SuperMix High Fidelity (Invitrogen). The PCR reaction was performed following a protocol of: 5 min at 95°C, 25 cycles (30 sec at 95°C, 30 sec at

51°C, 1min/kb extension at 68°C). The resulting amplification products were analysed on an agarose gel as described in section 2.3.1.

### 2.3.5 Miniprep of plasmid DNA

Colonies containing a product of the expected size, when evaluated by colony PCR, were grown in 20 mL LB for 12-16 hours, at 37°C with shaking. Miniprep procedure was used to isolate the plasmids. For this purpose the commercially available miniprep kits from either Promega or Qiagen were used according to the manufacturer's instructions. Purified plasmids were eluted in 30 to 50  $\mu\text{L}$  nuclease-free water and stored at  $-20^\circ\text{C}$ . Due to the low-copy-number origin of replication (pBR3222) of the pET160 plasmid, relatively low concentration were usually obtained after miniprep purification (e.g. 50-100ng/ $\mu\text{L}$ , 3-5 $\mu\text{g}$  total). Plasmid concentration and quality were assessed by measuring the absorbance of plasmid DNA at 260 nm and 280 nm on a NanoDrop ND-1000 Spectrophotometer. Nuclease-free water was used as a blank.

### 2.3.6 Sequencing analysis

Purified plasmids were sequenced to ensure the integrity of the DNA insert and the c-MYC sequence. The sequencing reactions were performed using the T7 universal forward (5'-TAATACGACTCACTATAGGG-3') and T7 universal reverse (5'-TAGTTATTGCTCAGCGGTGG-3') sequencing primers. For sequences longer than 1200-1300 bp an internal gene specific primer was designed. Each sequencing mixture contained 1  $\mu\text{g}$  of plasmid DNA and 30  $\rho\text{M}$  of each primer in a final volume of 15  $\mu\text{L}$  ddH<sub>2</sub>O. The preparations were initially sent to MWG for sequencing. However, with the progress of the cloning activity and the increasing number of positive clone, the Applied Biosystems 3730 XL Genetic Analyser provided by Cogenics (Beckman Coulter Genomics) was used. Usually three 96 wells plates were analysed at a time as described in ABI's protocols. Each well contained 5  $\mu\text{L}$  plasmid (80-100ng/ $\mu\text{L}$ ), 10  $\mu\text{L}$  T7 universal forward primer (5  $\mu\text{M}$ ) and 10  $\mu\text{L}$  T7 universal reverse primer (5  $\mu\text{M}$ ). Once the sequence reactions are complete we purify them using Agen-court's CleanSeq protocols. Finally data were analysed using ABI's Sequence Analysis Software (V. 5.3.1).

Identity, integrity, orientation and reading frame of the cloned sequences, at both nucleotide and polypeptide consensus levels were analysed and compared during a sequence evaluation step. The analysis was performed by comparing

the sequencing result with the annotated *Pf* (3D7) genome sequence using either Lasergene Seq Man (DnaStar), Vector NTI 10 (Invitrogen) or DNA dynamo (BlueTractorSoftware Ltd.). Where necessary, resulting sequences were further investigated by performing BLAST similarity algorithms with *PlasmoDB* BLAST ([www.plasmodb.org/plasmodb/servlet/sv?page=blast](http://www.plasmodb.org/plasmodb/servlet/sv?page=blast)) and aligning them either with ClustalW ([www.ebi.ac.uk/clustalw](http://www.ebi.ac.uk/clustalw)) or T-coffee ([www.ebi.ac.uk/Tools/t-coffee](http://www.ebi.ac.uk/Tools/t-coffee)).

### 2.3.7 Plasmid storage

To preserve and guarantee the long-term availability of the plasmid library created, both purified plasmids and glycerol stocks of transformed *E. coli* clones containing vectors with *Pf* verified ORFs were prepared. To prepare the glycerol stocks, 0.5 mL of sterile 100% glycerol was added to 0.5-1 mL of the bacterial culture in a cryo-tube, gently mixed and left on ice for 5 min. Stocks were stored at  $-80^{\circ}\text{C}$ .

## 2.4 *In vivo* expression of the *Pf* candidates in *E. coli*

### 2.4.1 Transformation of Rosetta-Gami<sup>TM</sup>B(DE3)

Transformations were performed through heat shock of chemical competent Rosetta-Gami<sup>TM</sup>B(DE3) (RGB) *E. coli* strain (Novagen<sup>®</sup>, Merk). One  $\mu\text{L}$  of the purified ligated product (1-10 ng, see Section 2.3.5) was added to 20  $\mu\text{L}$  of freshly thawed *E. coli* cells. The solution was gently mixed and incubated for 5 min on ice. Bacteria were subsequently heat-shocked for 30 sec at  $42^{\circ}\text{C}$  in a water bath and immediately placed on ice for 2 min. To speed up cell recovery, 80  $\mu\text{L}$  of SOC medium was added to the bacteria, which were then left shaking (215 rpm) for 1 hour at  $37^{\circ}\text{C}$ . The total volume of each transformation was spread onto pre-warmed ( $37^{\circ}\text{C}$ , for 30 min) LB-agar plates supplemented with antibiotics (ampicillin (Amp) 50  $\mu\text{g}/\text{mL}$ , chloramphenicol (Cam) 34  $\mu\text{g}/\text{mL}$ , tetracycline (Tet) 12.5  $\mu\text{g}/\text{mL}$  and kanamycin (Kan) 15  $\mu\text{g}/\text{mL}$ ) to guarantee a constant selective pressure on the RGB strain and its key features (see Table 2.5). Colonies were subsequently picked, resuspended and grown O/N in 5 mL for glycerol stock preparation or 10-20 mL of LB (Kan+Tet+Cam+Amp) for protein expression.

Marker	Rosetta-Gami B <sup>TM</sup> (DE3) <i>E. coli</i> strain: key features	R <sup>r</sup>	Derivation
DE3	Chromosomal copy of the T7 RNA polymerase gene under control of the lacUV5 promoter. This is needed for production of the target genes when T7-driven expression vectors are used.	None	Host is a lysogen of $\lambda$ DE3
lacZY	LacZY deletion mutant of BL21. This allows induction with IPTG to occur in a true concentration-dependent manner, which should be uniform throughout the culture.	None	Turner <sup>TM</sup> , BL21 (DE3)
pRARE	Presence of 6 rare tRNA codons (AGG, AGA, AUA, CUA, CCC and GGA) for an improved translation that would otherwise be limited by the <i>E. coli</i> usage of codons.	Cam	Rosetta <sup>TM</sup> , BL21 (DE3)
trxB	Abolition of the thioredoxin reductase. As a consequence disulfide bond formation is strongly enhanced in the cytoplasm of <i>E. coli</i> .	Kan	Origami <sup>TM</sup> , BL21 (DE3)
gor	Abolition of the glutathione reductase. As trxB, this favors disulfide bond formation in the cytoplasm of <i>E. coli</i> .	Tet	Origami <sup>TM</sup> , BL21 (DE3)

**Table 2.5: Rosetta-Gami<sup>TM</sup> B (DE3) *E. coli* Key Feature.**

List of the main RGB strain attributes. R<sup>r</sup>: selectable marker for bacterial resistance gene. Cam: chloramphenicol resistance. Kan: kanamycin resistance. Tet: tetracycline resistance.  $\lambda$ DE3: lambda DE3 phage construct that expresses T7 RNA polymerase. Source: Novagen<sup>®</sup>.

During the evaluation studies of the bacterial expression performances, four additional strains were tested: BL21 Star<sup>TM</sup> (Star, Invitrogen), BL21 BLR (BLR, Novagen<sup>®</sup>, Merk), BL21 Codon Plus RIL<sup>®</sup> (RIL, Stratagene), Rosetta-Gami<sup>TM</sup>2 (RG2, Novagen<sup>®</sup>, Merk). On these regards, transformation of cells was obtained following either the protocol used for RGB or the one utilised for TOP10 cells (depending on the strain manufacturer's suggestions).

## 2.4.2 Medium preparations for *Pf* protein expression

To express heterologous *Pf* proteins in the prokaryotic *E. coli* host four medium preparations were evaluated during the project: LB (Luria-Bertani Medium), AIM (Auto-Induced Medium), 2xYT and TM (Tania Medium). All media were always supplemented with antibiotics for the selective growth of RGB *E. coli* strain (Novagen<sup>®</sup>, Merk), e.g. Amp 50  $\mu$ g/mL, Cam 34  $\mu$ g/mL, Tet 12.5  $\mu$ g/mL and Kan 15  $\mu$ g/mL as recommended by the manufacturer. Each preparation was sterilised either for 30 min at 121°C in autoclave or through 0.22  $\mu$ M filtration, according to necessity.

LB medium was prepared as following: 25 g (or otherwise specified by the

supplier) of LB Broth (LB broth, MILLER from Merk) were solubilised in 1 L of ddH<sub>2</sub>O. The medium was sterilised in autoclave and used within 3 weeks. AIM medium was produced adding 1:50 [v/v] of 5052 50X (glycerol 250 g/L, glucose 25 g/L,  $\alpha$ -lactose 100 g/L, autoclave sterilization), 1:20 ratio of NPS 20X ((NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 60 g/L, KH<sub>2</sub>PO<sub>4</sub> 136 g/L, Na<sub>2</sub>HPO<sub>4</sub> 142 g/L in ddH<sub>2</sub>O, autoclave sterilization) and MgSO<sub>4</sub> (MgSO<sub>4</sub>, 7H<sub>2</sub>O 1 mol/L in ddH<sub>2</sub>O, autoclave sterilization) to a solution of 5 g/L yeast extract, 10 g/L casamino acid in ddH<sub>2</sub>O. 2xYT medium was obtained mixing 1.6% Bacto-Tryptone [w/v], 1% Bacto-yeast extract [w/v], 0.5% NaCl [w/v] and 1% sterile glucose in ddH<sub>2</sub>O (autoclave sterilization). TM medium, named after one of Fight-Mal project collaborators, who suggested the initial formulation, is a derivate of the nutritionally rich 2xYT medium. At the core of the medium, components of the 2xYT medium (1.6% Bacto-Tryptone [w/v], 1% Bacto-yeast extract [w/v], 0.5% NaCl [w/v]) were supplemented with 100 mM potassium-phosphate buffer (stock 1 M KH<sub>2</sub>PO<sub>4</sub>/1 M K<sub>2</sub>HPO<sub>4</sub> pH 7.2-7.4) and 2 g/L (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (autoclave sterilization). The final expression media was obtained by adding further essential components that had either been autoclaved or filter sterilized. Autoclaved components were: 1000X CaCl<sub>2</sub> (2% [w/v]), 1000X MgSO<sub>4</sub> (10% [w/v]), 100X ZnCl<sub>2</sub> (0.7% [w/v]). Filtered components were: 1000X FeSO<sub>4</sub> (0.2% [w/v]), 50% glucose [w/v], 2000X thiamin (10 g/L), 1000X biotin (3 g/L in H<sub>2</sub>O 11 mM NaOH). Two amino acids mixtures were prepared as described in Table 2.6 and referred to 50X Aa mixture A and 50X Aa mixture B. Both mixtures were added (1:50, [v/v]) to the TM to complete the preparation. Complete TM medium was always freshly prepared under sterile conditions just before expression. Complete TM medium was stored for not longer than a week at +4°C whereas the TM core solution was stored at RT for up to 3 weeks.

<b>Aa Mixture A (50X)</b> <i>Solubilised in ddH<sub>2</sub>O</i>		<b>Aa Mixture B (50X)</b> <i>Solubilised in HCl(3.8%)</i>	
<b>Aa</b>	<b>g/L</b>	<b>Aa</b>	<b>g/L</b>
Arg	25.0	His	30.0
Cys	5.0	Ile	37.5
Lys	37.5	Leu	37.5
Met	5.0	Gln	20.0
Trp	5.0	Asp	6.25
Asn	5.0	Ser	37.5
Ala	12.5	Tyr	10.0
Gly	12.5	Glu	5.0
Val	5.0		
Thr	25.0		
Phe	25.0		
Pro	25.0		

Table 2.6: Amino acid mixtures composition

### 2.4.3 *In vivo* expression of *Pf* candidates

Initial *in vivo* expression feasibility studies were carried out using five different *E. coli* strains: BL21 Star<sup>TM</sup> (Star, Invitrogen), BL21 BLR (BLR, Novagen<sup>®</sup>, Merk), BL21 Codon Plus RIL<sup>®</sup> (RIL, Stratagene), Rosetta-Gami<sup>TM</sup>2 (RG2, Novagen<sup>®</sup>, Merk) and RGB (Novagen<sup>®</sup>, Merk) in LB, 2xYT and AIM. However, after the preliminary strain evaluation all subsequent experiments were carried out using RGB *E. coli* strain growth in either 2xYT medium and/or TM medium at 37°C or RT. Bacterial growth was monitored through measurements of culture optical density (OD<sub>600</sub>) using a spectrophotometer (NanoDrop<sup>®</sup> ND-1000). Generally, the O/N culture, derived from freshly transformed bacteria or a glycerol stock (see Section 2.4.1), was pelleted and resuspended to normalise the bacterial density to 5 OD<sub>600</sub> in expression medium (see Section 2.4.2). Normalised cultures were diluted 1:100 to 0.05 OD<sub>600</sub> in 5 mL (mini-scale) or 2 L (large-scale) of expression medium. To reach induction point (0.8-1.0 OD<sub>600</sub>), bacteria were grown for 3-4 hours at 37°C shaking at 150-225 rpm, depending on the volume of the culture. Protein expression was induced using 0.8 mM IPTG. Induced bacteria were grown for additional 3 hours at 37°C or at RT (~ 22°C) for 16-18 hours shaking at 150-225 rpm. By the end of the incubation the OD of the bacterial culture was recorded and aliquots of 1-2 mL were collected for glycerol stock preparation and one-dimensional (1-D) sodium dodecyl sulfate-polyacrylamide discontinuous gel electrophoresis analysis (SDS-PAGE). Glycerol stocks were prepared by mixing 250 μL



of bacterial culture with 250  $\mu\text{L}$  of sterile 100% glycerol [v/v] and immediately transferred at  $-80^\circ\text{C}$ . Remaining bacteria from the expression culture were eventually harvested by centrifugation for 15 min at  $+4^\circ\text{C}$  and 4000 rpm. Resulting pellets were stored at  $-80^\circ\text{C}$ .

#### 2.4.4 Bacterial lysis

Pellets derived from induced bacteria, stored at  $-80^\circ\text{C}$ , were defrosted on ice. Cell lysis was carried out by resuspending the bacterial pellet to 16  $\text{OD}_{600}$  (mini-scale) or in 40 mL per liter of starting culture (large-scale) of freshly prepared lysis buffer A (50 mM HEPES pH 7.2 or 1X PBS pH 7.5-7.8, 0.3-0.5 M NaCl, 2 mM  $\text{MgCl}_2$ , 10% glycerol [v/v], 0.5% Tween-20 [v/v], 5 mM  $\beta$ -mercaptoethanol, 1.0 mg/mL lysozyme, 0.001 U/ $\mu\text{L}$  benzonase and urea (400 mM, optional), final solution was filtered (0.22  $\mu\text{M}$ ). For large-scale expressions the lysis buffer was supplemented with a complete protease inhibitor cocktail (Complete Protease Inhibitor Cocktail Tablets, 1 tablet per 50 mL of buffer, Roche) and with 40 mM imidazole (optional) to increase stringency in the first step of the 6xHis-tag binding to the resin (see Section 2.6.4). Lysis reactions were initiated for 1-1.5 hours at  $+4^\circ\text{C}$  on a horizontal shaker (150 rpm) in pre-chilled 1.5 mL tubes (mini-scale) or 50 mL tubes (large-scale). Lysis was continued with three freeze-thaw cycles, each consisting of 30-60 sec in liquid nitrogen and 3-6 min in a  $37^\circ\text{C}$  water bath. Lysis of the large-scale expressions was further increment with three cycles of sonication on ice (7 min/30 mL) using a 9 mm probe equipped sonicator at 6 microns output (Soniprep 150, MSE). Samples were eventually homogenised using a 1 mL or 20 mL syringe equipped with a 25G needle. During the entire procedure samples were either kept on ice or processed inside a cold room. For the mini-scale analysis, 300-500  $\mu\text{L}$  of the resulting lysate solution (total fraction), were transferred in a pre-chilled 2.0 mL tube and centrifuged at maximum speed (5000-14000 rpm) at  $+4^\circ\text{C}$  until optimal separation between soluble and insoluble fractions was achieved (15-30 min). The soluble fraction was collected in a pre-chilled 1.5 mL tube and stored in ice. The remaining pellet (insoluble fraction) was washed and resolubilised in lysis buffer with a volume equal to the one used in the total fraction. For the large-scale analysis, the full volume of the lysate was centrifuged at maximum speed (5000-14000 rpm) in an ultra-centrifuge (SORVALL RC 5B Plus Centrifuge) until optimal separation between soluble and insoluble fractions was achieved (30 min). The soluble fraction was then collected and either kept on ice until the purification procedure was started or

flash frozen and stored at  $-80^{\circ}\text{C}$ .

For each fraction obtained during the above procedure a  $60\ \mu\text{L}$  aliquot was prepared for SDS-PAGE by mixing it with  $20\ \mu\text{L}$  of 4X loading dye (0.25 M Tris pH 6.8, 2% SDS [w/v], 0.1M DTT, 4 M Urea, 40% glycerol [v/v], 0.1% Bromophenol blue [w/v]) and boiled for 3 min at  $96^{\circ}\text{C}$ . The remaining volume was stored at  $-80^{\circ}\text{C}$  for further assessments.

## 2.5 *In vitro* expression of the *Pf* candidates

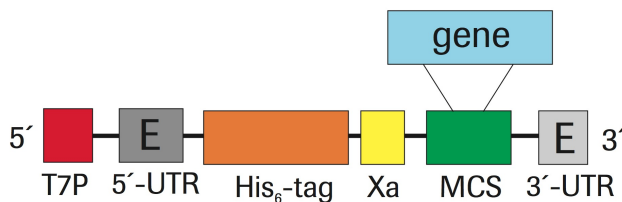
### 2.5.1 Expressway<sup>TM</sup> Plus *E. coli* system

A panel of 40 ORFs, that I successfully cloned during a previous study in our laboratory, was *in vitro* transcribed and translated using a commercially available kit, the Expressway<sup>TM</sup> Plus Expression System (Invitrogen UK). The kit contains an optimised *E. coli* slyD– extract (Zubay, 1973) capable of promoting the full expression of recombinant active proteins from specific DNA constructs, e.g. pEXP5-NT/TOPO<sup>®</sup>. Indeed, once properly ligated, the plasmid contains all the regulatory elements necessary to exploit the cellular machinery present in the *E. coli* extract. Following the manufacturer’s instructions,  $1\ \mu\text{g}$  of plasmid DNA was mixed with  $42\ \mu\text{L}$  of the expression cocktail (slyD-*E. coli* extract, 2.5X IVPS Plus *E. coli* reaction buffer, 50 mM Aa (-Met), T7 enzyme mix and 75 mM methionine mixed together at a ratio of 20:20:1:1:1, respectively) and placed into a 2 mL screw-cap tube. Using DNase/RNase-free water, the mixture was brought to a final volume of  $50\ \mu\text{L}$  and shaken either with a thermomixer at 14,000 rpm (Eppendorf) or with an orbital incubator at 250 rpm (SI 50, Stuart Scientific) for 30 min at  $30^{\circ}\text{C}$ . After this first incubation  $50\ \mu\text{L}$  of a “feed buffer” (2.0X IVPS Plus *E. coli* Reaction Buffer, 50 mM Amino Acid, 75 mM Methionine and DNase/RNase free water mixed together at a ratio of 25:1.25:1:22.75 respectively) was added to each sample and vials were left to shake at  $37^{\circ}\text{C}$  for another 5.5 hours. Finally, following completion of the expression step the reaction mixtures were briefly centrifuged and stored at  $-20^{\circ}\text{C}$ , ready for use in the production/spotting of the microarray chips (see Section 2.8.1).

### 2.5.2 RTS wheat germ & rabbit reticulocyte system

Alternative eukaryotic systems for production of *Pf* proteins that were insoluble or not express in *E. coli*, two Promega T7 TNT<sup>®</sup> Coupled Systems were

tested: the wheat germ extract and rabbit reticulocyte lysate. Due to the differences in the transcription and translation machineries present in these two platforms, however, an alternative expression plasmid had to be selected, pIVEX 1.4 WG (Roche). It is specifically designed for high expression level of histidine tagged proteins in eukaryotic cell-free systems and particularly in wheat germ extract (see Figure 2.3).



**Figure 2.3: pIVEX 1.4 WG vector, functional elements.**

Description of the functional elements present in the pIVEX 1.4 WG vector. T7 P: T7 Promoter. UTR: untranslated regulatory regions containing translational enhancer (E) elements upstream (5') and downstream (3') the gene of interest. His<sub>6</sub>-tag: N-terminal tag formed by six histidine residues in straight succession. Xa: Factor Xa restriction protease cleavage site. MCS: Multiple cloning site in three different reading frames for the insertion of the target gene

### pIVEX plasmid optimization and subcloning

As the pIVEX vector did not have the same functional elements as the pET160 (i.e 5'-6xHis-Lumio<sup>TM</sup>-TEV-attB1-CACC-attB2-3'), and to facilitate the DNA subcloning between the two vectors, the pIVEX 1.4 WG plasmid was partially modified. A newly synthesised DNA cassette of 844 bp (GeneART, see Figure 2.4), which contained some of the main features of pET160 plasmid, was inserted into the pIVEX vector by ligation following double digestion with HindIII and XhoI restriction enzymes (New England Biolabs).

Reaction conditions were assessed by analytical digestion following the manufacture's instructions (New England Biolabs). In the preparative digestion 30  $\mu\text{L}$  of the two plasmids, pIVEX 1.4 WG (250 ng/ $\mu\text{L}$ , Roche) and pET160\_PFcassette (100 ng/ $\mu\text{L}$ , GeneART) were digested for 2h at 37°C with 2  $\mu\text{L}$  HindIII, 2  $\mu\text{L}$  XhoI (New England Biolabs) in 12.5  $\mu\text{L}$  of autoclaved H<sub>2</sub>O, 5.0  $\mu\text{L}$  Buffer 2 (New England Biolabs) and 0.5  $\mu\text{L}$  BSA (100X). Digested products were run on an agarose gel as described in Section 2.3.1. DNA bands of interest were extracted and solubilised using a DNA gel extraction kit (Qiagen). Ligation was performed mixing the open pIVEX vector and the PFcassette at different ratios (from 5:1 to 1:5 in 5  $\mu\text{L}$  of solution) with 5  $\mu\text{L}$  of Quick Ligase Buffer and 0.5  $\mu\text{L}$  DNA Ligase (New England Biolabs). Mixtures

```

T7 promoter          HindIII          5' Enhancer
TAATACGACTCACTATAggcctAAGCTTaaaaATACTCCCCACAACAGCTTACAATACTCCCCACAC
                                START          6xHIS+Gly
AGCTTACAAATACTCCCCaacaacagcttgtcgaaccATGtctggttctCATCATCATCATCATGg
Lumio™ tag          TEV recognition site
cagcggcTGTGTCTCGCTGTTGCGggtggcggcGAAAACCTGTATTTTCAGGGAattatcACAAGTTT
attB1          NotI   Gene+cMYC+STOP   AscI          attB2
GTACAAAAAGCAGGCTccGCGGCCGcccccttCACCaagggtgGCGGCCGgACCCAGCTTTCCTTGTGA
                                XhoI          STOP          3' Enhancer
CAAAGTGGTgataatcggCTCGAGatccggTAAactaactaaggtacccAGCTCTTCGGTTTTGGTTTTGG
ACCTCTGGTCCTGCAACTTGAGGTAGTCAAGATGCATAATAAATAACGGATTGTGTCCGTAATCACACG
TGGTGCGTACGATAACGCATAGTGTTTTTCCCTCCACTTAAATCGAAGGGTTGTGTCTTGGATCGCGCG
GGTCAAATGTATATGGTTTCATATACATCCGCAGGCACGTAATAAAGCGAGGGGTTTCGAATCCCCCGTT
ACCCCCGGTAGGGGC

```

Figure 2.4: pIVEXpET160 PFcassette MAL7p1.177.

were left at RT for 30 min prior transformation in DH5 $\alpha$  *E. coli* strain. For each condition, 2  $\mu$ L of the ligation reaction were added to 50  $\mu$ L of competent cells. Bacteria were left on ice for 30 min and heat shocked in a water bath at 42°C for 45 sec. Vials were cooled down on ice for 2 min and 125  $\mu$ L of SOC medium was added to each reaction. Cells were incubated for 1h at 37°C shaking at 180 rpm and subsequently plated on pre-warmed LB-agar plates containing 100 $\mu$ g/mL ampicillin. After an O/N growth at 37°C colonies were picked and inoculated in 5 mL of LB with 100  $\mu$ g/mL ampicillin. Inocules were left growing at 37°C and used for maxi-prep plasmid preparation (2x100 mL of LB+Amp). The new pIVEX-PF hybrid plasmid was isolated using an EndoFree Plasmid Maxi Kit (Qiagen) following the manufacture's instructions.

To subclone the *Pf* candidates into the new pIVEX-PF a double restriction enzyme digestion was assessed as an alternative to the Gateway<sup>®</sup> platform. The digestion reaction used two octanucleotide-recognizing sites up- and downstream of the gene of interest specific for two restriction enzymes, NotI and AscI, respectively (NewEngland BioLabs). The recombinant form of NotI, NotI-HFTM, was selected for compatibility reasons over the digestion buffer. Preparative digestion and ligation were carried out following the same procedure as above, with the exception that the NEBuffer 4 was used during the digestion reaction to achieve optimal endonuclease activity for both enzymes.

### Expression using the two eukaryotic systems

Performances of two eukaryotic *in vitro* expression systems (TNT<sup>®</sup> T7 Coupled wheat germ extract system and TNT<sup>®</sup> Coupled reticulocyte lysate system, Promega) were assessed during a panel of small-scale expression tests

(50  $\mu\text{L}$ ). Hybrid pIVEX-PF plasmids containing five *Pf* candidates (MAL7P1 177, PFB0100c\_P1, PFI1785w, PF14\_0013 and PFE1605w) were expressed following the manufacturer's instructions.

Briefly, the cell extract was quickly thawed by hand and placed on ice. Remaining reagents were thawed at RT and placed on ice together with the TNT<sup>®</sup> RNA Polymerase. Any precipitation present in the TNT<sup>®</sup> Reaction Buffer was rapidly resuspended by vortexing. The expression reaction was assembled on ice by adding 25  $\mu\text{L}$  of cell extract to 2  $\mu\text{L}$  of reaction buffer, 1  $\mu\text{L}$  of TNT<sup>®</sup> T7 RNA Polymerase, 0.5  $\mu\text{L}$  of amino acid mixture minus leucine and 0.5  $\mu\text{L}$  of amino acid mixture minus methionine, 1  $\mu\text{L}$  of RNasin<sup>®</sup> Ribonuclease Inhibitor (40u/ $\mu\text{L}$  Promega), 2  $\mu\text{L}$  of linearised or circular pIVEX-PF plasmid and nuclease-free water to reach a final reaction volume of 50  $\mu\text{L}$  in Rnase free tubes. The reaction was carried out at 37°C for 1h and 30 min shaking at maximum speed (e.g. 160-210 rpm). For each reaction positive and negative controls were run. As a positive control, a luciferase vector provided in the expression kit (Luciferase T7 Control DNA, Ampr, with luciferase gene 1649 bp) was used according to the manufacturer's instructions. A control luciferase assay was produced and a qualitative measurement of the luciferase activity was assessed by visualization in a dark room. In the negative control neither circular nor linear plasmid was added. *Pf* recombinant protein expression was evaluated by SDS-PAGE and Western blot analysis as described in paragraph 2.7.1.

## 2.6 Protein purification procedures

Due to the miniaturisation level achieved in the protein microarray technology, minimal amounts (less than 100-150  $\mu\text{g}$ ) of protein/antigen are required for the successful production of hundreds of immunoassays. However, high purity (90-95%) and integrity of the spotted antigens are essential to guarantee optimal levels of assay sensitivity and specificity. Therefore, two different affinity purification approaches were assessed during my PhD: 1) immunoaffinity chromatography; 2) immobilised metal ion affinity chromatography. Both procedures aimed to efficiently purify our 6xHis-tagged recombinant proteins without altering the integrity and folding of the target protein. For this soluble fractions produced using the large-scale expression procedure (0.5-2 L bacteria culture) were utilised. In particular, purification priority was given to the clones showing the highest expression and solubility level during the initial

mini-scale analysis.

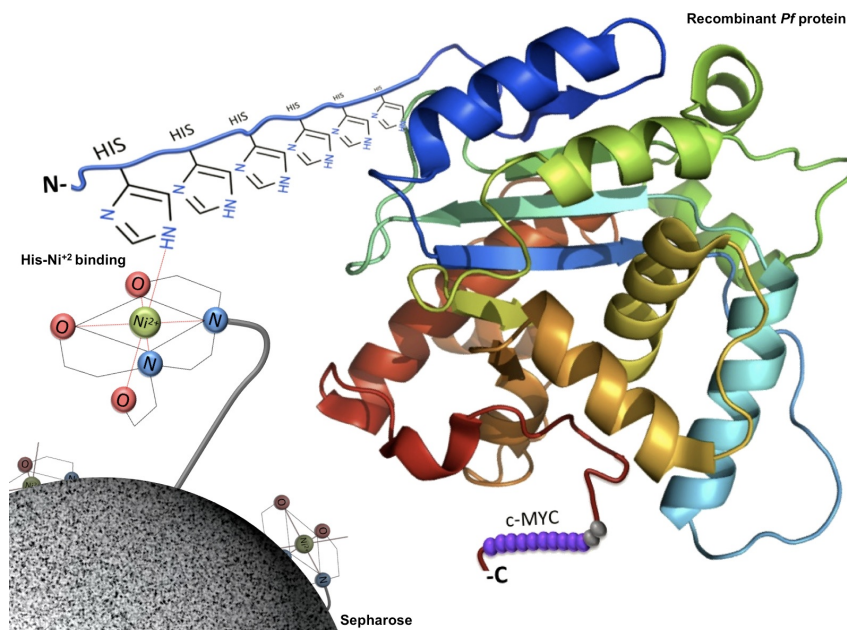
### 2.6.1 Immunoaffinity chromatography: Magnetic beads

Due to the simplicity, specificity and the fact that it allows to purify multiple antigens in parallel, the immunoaffinity chromatography system, based on magnetic beads coated with an anti-His antibody, was the first platform to be evaluated. The beads, Dynabeads<sup>®</sup> MyOne<sup>™</sup> Tosylactivated (Dyna<sup>®</sup> Biotech) were first covalently conjugated with a mouse anti-6xHis mAb (4D11 clone, kindly provided by Prof. Andrea Crisanti) and then used for batch method purification. The optimal concentration of dynabeads was empirically calculated at 40  $\mu\text{g}/\mu\text{L}$  whereas the optimal antibody concentration was set at 40  $\mu\text{g}/\mu\text{L}$  (Total volume of reaction: 1250  $\mu\text{L}$ ; in 3 M ammonium sulphate: 415  $\mu\text{L}$ ; 100 mg/mL beads pellet: 100  $\mu\text{L}$ ; 1.5 mg/mL 4D11 mAb: 250  $\mu\text{L}$ ; coating buffer 0.1 M sodium borate, pH 9.5: 485  $\mu\text{L}$ ). After the preliminary coating step at 37°C using slow-tilting rotation for 16-24 hours, the beads were magnetically attracted to the side and the supernatant was replaced with blocking buffer (PBS pH 7.4 with 0.5% BSA and 0.05% [v/v] Tween-20), which was incubated with the beads using slow-tilting rotation O/N at 37°C. Once the blocking step was completed the beads were washed 3 times with 1 mL of washing buffer (PBS pH 7.4 with 0.1% BSA and 0.05% [v/v] Tween-20). As a next step, an optimal amount of recombinant protein (50-250  $\mu\text{g}$  per 10 mg of conjugated beads - calculated following the manufacturer's instructions) was mixed with the beads. After leaving the mixture for an hour at RT under slow-tilting rotation the beads were magnetically moved aside and the supernatant ( $\sim 300$   $\mu\text{L}$ ) was collected. Seventy-five  $\mu\text{L}$  of this were used for analysis SDS-PAGE loading and the remaining was flash frozen and stored at  $-80^\circ\text{C}$ . Beads were subsequently washed and incubated with 30  $\mu\text{L}$  of 1 M glycine-HCl pH 2.5 to elute the recombinant protein. To ensure that neither the beads nor the recombinant proteins were affected by the low pH, both were brought to physiological pH (7.4) immediately after elution.

### 2.6.2 Immobilised Metal Affinity Chromatography

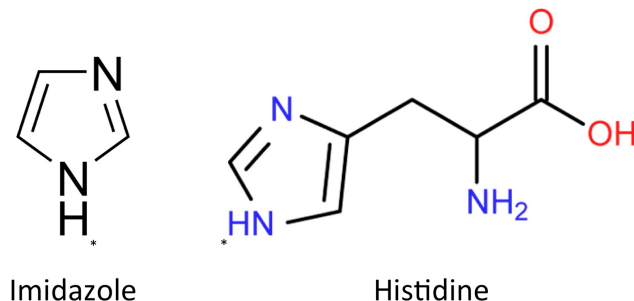
A number of Immobilised Metal Affinity Chromatography (IMAC) purification trials were initially performed to assess the efficiency and suitability of this approach for the isolation of *Pf* recombinant proteins. The technique is based on the ability of metal ions to covalently bind certain amino acids, particularly

histidines. Therefore, in the IMAC platform, the solid phase is composed by agarose or sepharose resins conjugated to chelating molecules, which are strongly bound to metal ions ( $\text{Ni}^{2+}$ ,  $\text{Zn}^{2+}$ ,  $\text{Co}^{2+}$ ). In particular,  $\text{Ni}^{2+}$  is the metal ion considered to have the strongest affinity to hexa-histidine-tagged proteins (see Figure 2.5).



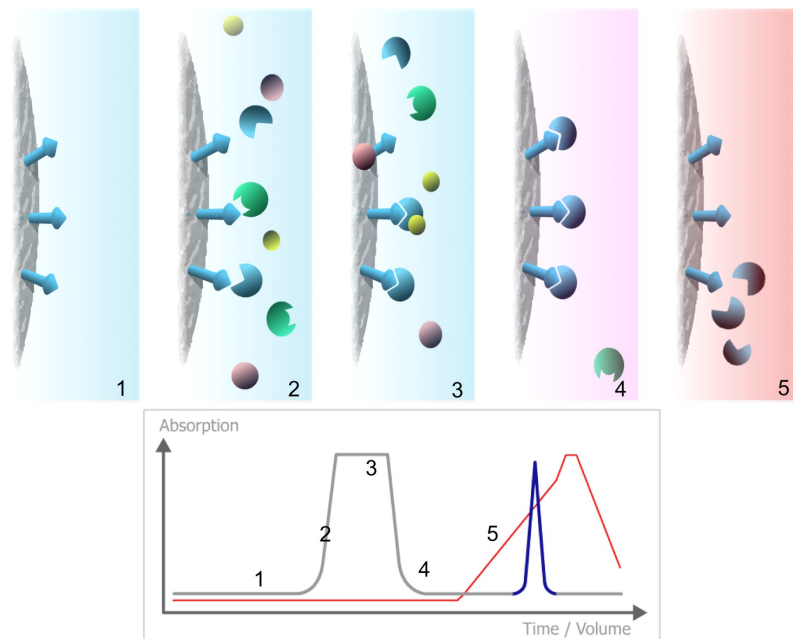
**Figure 2.5: Immobilised metal ion affinity chromatography (IMAC).** The figure describes the binding of N-6xHis-tagged *Pf* protein to a  $\text{Ni}^{2+}$  charged resin (sepharose).

$\text{Ni}^{2+}$ -bound proteins can eventually be released by competitive elution using imidazole, a molecule that contains the same aromatic heterocyclic ring structure found in the histidine amino acid (see Figure 2.6). During the initial feasibility studies both gravity column chromatography and fast protein liquid chromatography (FPLC) were evaluated.



**Figure 2.6: Structure comparison, imidazole versus histidine.**

An graphic overview of the principles and the key steps of an IMAC purification procedure is available in Figure 2.7.



**Figure 2.7: IMAC purification procedure.**

**1)** Equilibration, during which the solid/stationary phase ( $\text{Ni}^{2+}$  charged resins) is equilibrated to desired start conditions (e.g. 50 mM imidazole). **2)** Sample application/loading, the main objective is the to bind the target molecules (His-tagged proteins) to the solid phase. The material not bound will flow through the column eliminating some of the unspecific molecules (contaminants). **3)** Wash, this step aims to wash away all the remaining material that is weakly/unspecifically bound to the solid phase . **4)** Additional washes/gradient elution. The stringency (imidazole concentration, 50-150mM) of the washing buffer is increased linearly (gradient) or at intervals (step) until all contaminants are fully removed. **5)** Sample elution. By reaching an optimal imidazole concentration (200-500mM), the target molecules are released from the ligand/solid phase into the elution buffer. Image modified from GE Healthcare

### 2.6.3 Gravity column chromatography

Gravity column chromatography was carried out by packing Poly-Prep Chromatography columns (Bio-Rad) with 1 mL  $\text{Ni}^{2+}$  Sepharose<sup>TM</sup> 6 Fast Flow medium (GE Healthcare). The procedure was accomplished following the GE Healthcare packing recommendations (11-0008-87 Edition AA). Packed columns were either stored in 20% ethanol or immediately equilibrated with 10 column volumes (CV) of lysis buffer A (50 mM HEPES pH 7.2, 0.5 M NaCl, 2 mM  $\text{MgCl}_2$ , 10% glycerol [v/v], 0.5% Tween-20 [v/v], 5 mM  $\beta$ -mercaptoethanol, and 40 mM imidazole). During equilibration columns were mounted on a column holder and checked for gravity flow performance (at least 0.5 ml/min flow rate) and re-packed if necessary. Subsequently, the soluble expression fraction (20-40 mL), pretreated in lysis buffer A, was loaded. The flow through was collected in a 50 mL tube for analysis. The column was washed with 10-20 CV of lysis buffer A, and the elution steps were carried out using



5-10 CV elution buffer (same composition of lysis buffer A) with increasing imidazole concentrations. The first washing step (50 mM imidazole) aimed to remove unspecific bound molecules; in the second elution step (300 mM imidazole) the histidine-tagged protein was eluted. Finally in last stripping step (1 M imidazole), all remaining bound molecules were removed from the resin.

#### 2.6.4 Protein Liquid Chromatography

FPLC purification trials and feasibility studies using ÄKTA Explorer or ÄKTA Purifier (GE Healthcare) were performed at PX' Therapeutics facilities (Grenoble, France). Validation and final purification procedures were assessed and established at Microtest Matrices Ltd. (London, UK) using an ÄKTA Prime-Plus instrument (GE Healthcare). Pre-packed 1-5 mL HisTrap<sup>TM</sup> FF Columns (GE Healthcare) and Glass XK 16/20 column (GE Healthcare) packed using 0.5-2 mL Ni<sup>2+</sup> Sepharose<sup>TM</sup> 6 Fast Flow resin (GE Healthcare) were used. Packaging procedure was carried out following the manufacturer's instructions (GE Healthcare, 11-0008-87 Edition AA) and a bench top peristaltic pump was used to facilitate the procedure. After packing and/or equilibrating the solid phase, the columns were either mounted in a FPLC ÄKTA instrument or stored in 20% ethanol at +4°C. All FPLC machines were operated according to the manufacturer's instruction following either pre-installed purification methods or customised protocols. Upon installation of the Ni<sup>2+</sup> Sepharose<sup>TM</sup> column in the FPLC, the ethanol was washed away from the column with 10 CV of ddH<sub>2</sub>O at 4 mL/min flow rate. Equilibration was carried out with 10 CV of lysis buffer A at 4 mL/min. At completion, the pretreated soluble expression fraction (40-80 mL, see Section 2.4.4) was loaded using the sample-loading valve at a flow rate of 1-2 mL/min. During loading, the sample was always kept on ice. The column was washed using lysis buffer A until the A<sub>280</sub> absorbance reached baseline (5-10 CV). The 6xHis-tagged protein was recovered through a step elution procedure using 15 CV of elution buffer (same as equilibration buffer) containing increasing imidazole concentrations (100, 150, 300, 500 mM). For the gel filtration procedure, a Superdex 75 column (GE Healthcare) was used following the manufacturer's instructions. The size-exclusion chromatography procedure was also used as buffer exchange step to eliminate the presence of imidazole from the eluted fraction. To do so, the column was equilibrated (1.5-2 CV) using buffer A depleted of the imidazole. During all the FPLC chromatography procedures, fraction collection was completely automated and performed either at +4°C or on ice.

## 2.7 Protein characterization procedures

### 2.7.1 SDS-PAGE protein analysis

To validate the overall expression and purification procedures, newly synthesised/isolated recombinant proteins were separated using SDS-PAGE technique (4-12% gel gradient). Gel evaluation and analysis of expression and solubility levels were assessed in parallel using Coomassie Brilliant Blue staining (Coomassie), Lumio™ staining when possible and Western blotting (WB). Each gel was loaded with 5  $\mu$ L of pre-treated samples (total, soluble, insoluble and, when available, flow-through and elute fractions) for Coomassie staining, or 13  $\mu$ L for WB. A ladder to help estimate protein size was also included in each gel. A pre-stained molecular weight standard visible under UV light for Lumio™ staining (BenchMark™ Fluorescent Protein Standard, 11-155 kDa, Invitrogen) and a corresponding standard for Coomassie staining and WB analysis (SeeBlue® Plus2 Pre-stained Standard, 4-250 kDa, Invitrogen or Precision Plus Protein Kaleidoscope Standard, 10-250 kDa, Bio-Rad) were used in this study. Gels were run at 200V for 40 min using a premixed running buffer solution provided by BioRad (25 mM Tris, 192 mM glycine, 0.1% SDS [w/v], pH 8.3) and were subsequently stained using Bio-Safe™ (BioRad) a Coomassie G-250 stain without methanol and acetic acid. If protein were analysed using Lumio™ staining, the gel was assessed under UV light (254nm) to evaluate the presence of the Lumio™ signal and only subsequently stained by Coomassie. Gels were dried at RT and digitally acquired using a digital camera or a scanner. In parallel, a second gel containing the same samples and ladders was run and used for immunoblot analysis.

Generally it was observed that several proteins run either a few kDa lower or higher than expected. This is an already known behaviour of plasmodial proteins during the SDS-PAGE analysis. It can be explained by the amount of positive charge on the protein surface. In particular, proteins with a low isoelectric point,  $< 7.0$ , which represents the majority in *Pf*, tend to run slower with an apparently higher molecular weight. The operator has to take this protein behaviour into account before analysing the samples to avoid the determination of false negative or false positive values.

### 2.7.2 Immunoblot analysis

After a preliminary SDS-PAGE separation, proteins were electroblotted onto a nitrocellulose membrane (blotting buffer: 25 mM Tris-HCl, 192 mM glycine,

10% methanol [v/v]) using a wet transfer unit provided by BioRad (Critereon<sup>TM</sup>) run at 100V for 30-40 min. Once the samples were successfully transferred onto the membranes, an O/N blocking step at +4°C was performed (blocking buffer: TBS-T, (20 mM Tris-HCl pH 7.2-7.6, 300 mM NaCl, 0.3% Tween-20) containing 2% BSA and 3% non-fat dry milk). At completion, membranes were probed at RT for 1h using an anti-6xHis mouse IgG mAbs (4D11 clone, kindly provided by Prof. Andrea Crisanti) as primary antibody, which was diluted in blocking buffer to a concentration of 1  $\mu\text{g}/\text{mL}$ . Bound primary antibodies were subsequently detected using a secondary anti-mouse IgG antibody conjugated to horseradish peroxidase (HRP) (Goat Anti-Mouse IgG, Fc $\gamma$  fragment specific, Jackson Laboratories). The anti-mouse IgG-HRP was diluted 1:15000 in blocking buffer and incubated at RT for 1h. Three washing cycles, each of 3 min, consisting of TBS-T were performed after each antibody incubation to guarantee that all unbound reagents were eliminated. Finally, the presence of HRP activity was visualised via incubation with a developing substrate (ECL, GE Healthcare), used according to the manufacturer's instructions. Chemiluminescence detection was performed exposing an X-ray film (Amersham Hyperfilm ECL, GE Healthcare) to the membrane for 30 sec to 15 min, depending on the strength of the signal.

## 2.8 Protein microarray development

### 2.8.1 Construction of protein microarray chips

For array construction, either *in vitro* crude expression reactions or *in vivo*-expressed purified proteins were resuspended in PBS containing Tween 20 (0.2 mL/L), to a concentration from 5 to 50  $\mu\text{g}/\text{mL}$ . Microarrays were then printed onto either 8-well-masked or standard aldehyde-activated glass microscope slides (CEL Associates). Printing of the slides was accomplished using two high-throughput automated microarrayers: MicroGrid II, BioRobotics Genomic Solutions, for the 8-well-masked slides containing *in vitro* crude expression reactions; or the Nano Plotter<sup>TM</sup> NP 2.1/E, GeSim, for the standard slides containing the *in vivo*-expressed purified proteins. The samples were transferred from 384-well microtiter plates to glass slides by use of either a stainless steel solid pins of 200  $\mu\text{m}$  diameter (MicroGrid II) or a piezoelectric pipetting tip (Nano Plotter<sup>TM</sup> NP 2.1/E). Each pin/tip is estimated to transfer  $\sim 1.5$  nL of sample to the slide and produces spots with a pitch of  $\sim 0.6$  mm. After deposition of each solution, the pin/tip were washed either

sequentially with ddH<sub>2</sub>O and 70% EtOH (MicroGrid II) or with ddH<sub>2</sub>O only (Nano Plotter<sup>TM</sup> NP 2.1/E), and finally dried. Printing was performed at 18-23°C and at 55-60% humidity. The microarrays employed in this study were designed as n x m matrices (8x8, 8x12, 10x12) discrete spots with an average diameter ranging from 200 to 350  $\mu\text{m}$ ) allowing the incorporation of the candidate antigens in duplicates. Between 4 to 10 replicates of either slyD-*E. coli* extract (50  $\mu\text{g}/\text{mL}$ ) or Bovine serum albumin (BSA, 20 g/L) were used as negative controls. Additional replicates (n.6) of the printing buffer were utilised as background spots. For those arrays in which IgG subclass was determined, 2 replicates of human IgG1 and human IgG3 were used as reference spots (200  $\mu\text{g}/\text{mL}$ ). Various assay formats were used for different experiments depending on the particular aim. On completion of the printing process, slides were kept inside the robot printer cabinet for at least 12 hours. This stable and controlled environment (55 to 60% humidity and 18 to 23°C) is conducive to the binding of samples to the slide surface (Mezzasoma et al., 2002), and so an O/N maturation step was routinely employed. On removal from the robot, printed slides were stored in the dark in boxes containing desiccant silica gel bags (Sigma). It is critical that the slides are kept dry as slides that have been exposed to moisture during storage give imprecise and varying signal.

## 2.8.2 Processing of protein microarray chips

### Preparation of human serum samples

Over 30 (n.33) serum samples, collected from individuals chronically exposed to malaria, were used to assess and characterize the possible antigenicity of printed candidates. The serum samples were received as part of a longitudinal survey of anti-malarial immunity in Gambia. Due to the high levels of parasite exposure experienced by these adults, the serum samples were used to identify antigenic parasite proteins. Additionally, as negative controls, 13-16 serum samples obtained from European donors who had never been exposed to malaria were used. Prior the microarray processing, samples were mixed with a serum dilution buffer (2XPBS containing 2% BSA and 0.01% Tween 20) to achieve a final working serum dilution of 1:100.

These samples were also run to assess the specificity and sensitivity of the system. In particular for the *in vitro* chip, which contained unpurified putative antigens, sera (2  $\mu\text{L}$ ) were first pre-absorbed with an 80% *E. coli* extract solution (8  $\mu\text{L}$  slyD-*E. coli* Extract), for 90 min at 37°C to achieve the best signal:noise ratios and to reveal specific antigen-antibody interactions.

Subsequently, samples were centrifuged at 2500 *g* for 10 min. The supernatant was recovered and added to 190  $\mu\text{L}$  of serum dilution buffer (working dilution of 1:100).

### Processing of the printed slides

To block non-specific binding to the activated slide surface, the printed slides were incubated for 60 min at RT with a blocking solution containing Tween 20, 0.1 mL/L, and BSA, 20 g/L, before any downstream incubation. The slides were then rinsed with a washing solution containing PBS and Tween 20, 0.1 mL/L (washing solution) for 3 min on a rotator set for gentle agitation.

To assess the expression levels of the candidate proteins expressed *in vitro*, the arrays were then incubated with an anti-6xHis mouse IgG antibody, 4D11, at a concentration of 100  $\mu\text{g}/\text{mL}$  for 30 min at RT. The detection step was carried out using an anti-mouse IgG Alexa 555 or Alexa 647 conjugate, 20  $\mu\text{g}/\text{mL}$  (15 min at RT). After each incubation slides were washed as describe above. At completion of the processing slides were dried for 15-30 min at 37°C before fluorescence was measured at 532 and/or 635 nm by confocal scanning (PerkinElmer S5000, GenePix 4100a). Images were generated with the ScanArray<sup>TM</sup> software, according to the manufacturer's Instructions. All the slides were scanned under fixed instrument settings (e.g. 90% of laser power and 60% of photomultiplier tube (PMT) gain for PerkinElmer S5000, and 400-450 gain for GenePix 4100a), unless excessive saturation was observed - at which point laser and PMT gain settings were reduced). The images were subsequently stored as both TIFF and JPG files.

Performances of the chip containing the *in vivo* expressed purified antigens, the arrays were processed using either an anti-6xHis mouse IgG antibody, 4D11 (100  $\mu\text{g}/\text{mL}$ ) or anti-cMYC mouse antibody, 9E11 (20  $\mu\text{g}/\text{mL}$ ), for 30 min at RT. Detection of the primaries antibodies was achieved by incubating the chips for 15 min at RT with anti-mouse IgG Alexa 555 conjugated 20  $\mu\text{g}/\text{mL}$ . Additionally, the "*in vivo*" microarrays were also tested using an amplified detection system, which was based on a secondary anti-mouse antibody conjugated with HRP (goat anti-mouse IgG, Fc $\gamma$  fragment specific, Jackson Laboratories, 15 min at RT) and tertiary detection reagent containing tyramide-Alexa555 diluted 1:200 in a proprietary amplification buffer (Invitrogen, 15 min at RT).

As for the detection of antigenic reactivity, after the initial blocking of the

slides, the microarray were incubated for 30 min with a distinct human serum (see Section 2.8.2). Slides were rinsed and total human IgG was detected using a goat anti-human IgG Alexa 555 conjugated secondary antibody (Invitrogen) at a dilution of 1:100 (20  $\mu\text{g}/\text{mL}$  in 2XPBS containing 0.01% Tween 20 and 2% BSA). For the detection of IgG1 and IgG3 subtypes, a mouse anti-human IgG1 Alexa 555 and a mouse anti-human IgG3 Alexa 647 were used at a concentration of 10  $\mu\text{g}/\text{mL}$  in 2XPBS containing 0.01% Tween 20 and 2% BSA. In this case, the Alexa fluorophores were conjugated to secondary antibodies using an Alexa Fluor<sup>®</sup> Monoclonal Antibody Labeling Kit (Invitrogen) following the manufacturer's instructions. After incubation for 15 min at RT, unbound secondary antibody was removed with PBS/Tween and slides were dried at 37°C.

### Quantitative analysis of image data

Quantification of the spot reactivities was achieved using Pro Scan Array Express<sup>™</sup> (V.3.0) software according to the manufacturer's instructions. Data were stored as Microsoft Excel files. The elaboration of the raw data consisted firstly of normalizing the sample signals against background spots (average of the signal recorded from spotting buffer dots). Secondly, the replicate candidate antigen signals were averaged and corrected against the mean of the reading coming from the *E. coli* extract spots or the BSA spots (noise signal). The resulting data were eventually plotted as Signal/Noise ratio. A signal to noise ratio of 5 indicated that specific recombinant protein was detected by human IgG in the serum with an intensity 5 times greater than the intensity recorded from the *E. coli* expression extract alone.

The specific reactivity was then evaluated by taking into consideration the reactivity recorded from European serum samples, which have never been exposed to the parasite. Here, the mean plus 3 Standard Deviations (SD) of the Signal/Noise ratios of 13-16 negative control sera was applied as cut-off for assigning a positive or negative reactivity against a selected ORF candidate. Digital profiles (positive/negative) were then generated for each serum by charting positive or negative reactions against the 40 printed ORFs.

## Chapter 3

# RESULTS and DISCUSSION

### 3.1 Overview

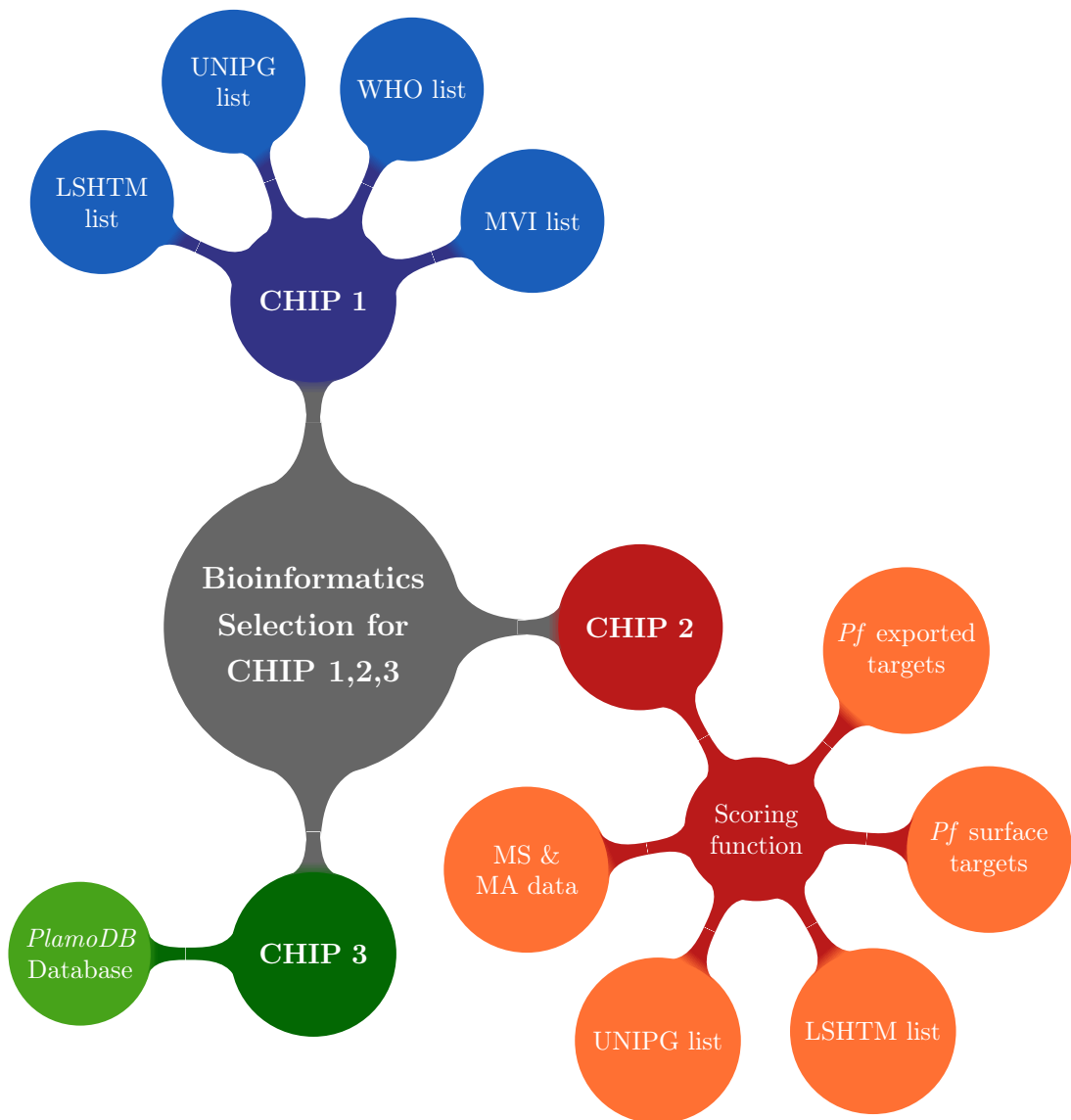
The activities described in this thesis helped to lay the foundations for a consortium-based project named FightMal (see Section 1.5), which aims to identify new putative vaccine candidates for the *Pf* malaria parasite using a microarray-based antigenicity screening platform. Hence, both scientific and technical achievements obtained throughout this PhD coincided with the initial milestones of the FightMal projects: a) Bioinformatic selection of *Pf* candidate antigens; b) Cloning, expression and purification of *Pf* targets; c) Protein microarray production and immunoassay development.

Due to the complexity of the tasks, some of the activities were first addressed via a small scale proof-of-principle study and, if successful, implemented using an optimised large-scale procedure. Otherwise, if the performances of the initial evaluation study were sub-optimal, additional effort was placed to evaluate alternative methods, until satisfactory results were obtained. Such work-flow allowed us to assess time lines, procedural complexities and optimal experimental conditions; information that was then used to draw a more generic and comprehensive plan for the application of the newly obtained methodology into the final scale-up activity.

Results obtained at each evaluation and implementation stage were analysed and presented in front of the consortium members that regularly met every 6-12 months. The totality of the results described in this chapter have been achieved by myself, if not otherwise specified, in the last 40-42 months with the crucial help and support of all the FightMal consortium members, in particular University of Perugia (UNIPG) which as member leader supervised all the activities, PX' Therapeutics (PX') and Microtest Matrices Ltd. (MtM).







**Figure 3.2: Structure of the bioinformatic analysis for CHIP 1, 2, 3.**

A combination of experimental evidence and bioinformatic algorithms were implemented together to create a complex filtering network that was eventually employed to screen the *Pf* proteome. As a result, three lists (CHIP 1, 2, 3) of well known and/or putative extracellular candidates, which are either anchored on the parasite/iRBC surface or secreted by the parasite, were produced. **CHIP 1:** contains vaccine candidates biologically assessed and currently employed in both pre-erythrocytic and erythrocytic stages vaccine trials. The selection was produced by combining together four list of vaccine candidates obtained from (LSHTM, UNIPG, WHO 2007 report and MVI 2001 report). **CHIP 2:** contains putative extracellular candidates expressed by the parasite during the blood stage. The selection derived from the application of an internal scoring function to a large list of genes possessing a specific set of antigenic attributes. (MS= mass spectrometry expression analysis and MA= microarray gene expression analysis). **CHIP 3:** contains members of the multiple-gene VAR family, a sub population of the VSA. The list was obtained by querying the *PlamoDB* database for VAR-specific attributes.

### 3.2.2 CHIP 1 candidate list

A total of 25 antigen candidates were selected. The list was produced after an extensive search throughout the existing literature for vaccine candidates that have already been biologically assessed. Both WHO (Malaria Vaccine Rainbow Tables, 2007; [www.who.int/vaccine\\_research/links/Rainbow](http://www.who.int/vaccine_research/links/Rainbow)) and MVI (Filip Dubovsky, 2001) databases, together with the National Center for Biotechnology Information (NCBI, [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) website were used to retrieve information. Additional genes were obtained internally from two of the members of the FightMal consortium, who previously performed similar analysis (e.g. Dr. Patrick Corran, LSHTM and Dr. Tania Dottorini, UNIPG). To facilitate the analysis and highlight possible redundancies, the genes were grouped into four lists, one for each source of information (WHO, MVI, UNIPG and LSHTM). Once all the material was gathered together (Table 3.1), it was filtered for redundant entries. Finally, a "CHIP 1" list was produced containing 25 immuno-targets.

Gene ID	Name-Info <i>Plasmo DB</i> access 190908	WHO 07'	MVI 01'	UNIPG	LSHTM
<i>MAL13P1.60</i>	Erythrocyte binding antigen 140,EBA-140/BAEBL	✓			✓
<i>MAL7P1.176</i>	Erythrocyte binding antigen,EBA-175 alias <i>PF07_0128</i>	✓	✓	✓	✓
<i>MAL7P1.231</i>	Histidine rich protein II, hrpii,His rich prot II, HRP-2				✓
<i>PF10_0344</i>	Glutamate-rich protein, GLURP	✓	✓	✓	✓
<i>PF10_0345</i>	Merozoite surface protein 3, MSP-3	✓	✓	✓	✓
<i>PF10_0346</i>	Merozoite surface protein 6, MSP-6			✓	
<i>PF11_0224</i>	Circumsporozoite-related antigen, Exp-1		✓	✓	
<i>PF11_0344</i>	Apical membrane antigen 1, AMA1	✓	✓	✓	✓
<i>PF11_0486</i>	MAEBL, putative	✓			✓
<i>PF11_0507</i>	Antigen 332, putative			✓	
<i>PF13_0197</i>	Merozoite Surface Protein 7 precursor, MSP7			✓	✓
<i>PF14_0102</i>	Rhoptry-associated protein 1, RAP1	✓	✓	✓	✓
<i>PFA0110w</i>	Ring-infected erythrocyte surface antigen, RESA	✓	✓	✓	
<i>PFB0300c</i>	Merozoite surface protein 2 precursor, MSP-2	✓	✓	✓	✓
<i>PFB0305c</i>	Merozoite surface protein 5, MSP-5	✓	✓	✓	✓
<i>PFB0310c</i>	Merozoite surface protein 4, MSP-4	✓	✓	✓	✓
<i>PFB0360c</i>	Cysteine protease, putative, SERA	✓	✓	✓	
<i>PFE0080c</i>	Rhoptry-associated protein 2, RAP2	✓	✓	✓	✓
<i>PFE0120c</i>	Merozoite Surface Protein 8, MSP8, (RMP-1) alias MSP-8	✓		✓	✓
<i>PF11475w</i>	Merozoite surface protein 1, precursor, MSP-1	✓	✓	✓	✓
<i>PFL1385c</i>	Merozoite Surface Protein 9, MSP-9, ABRA alias MSP-9	✓			✓
<i>PF10_0356</i>	Liver stage antigen-1, LSA-1	✓	✓		✓
<i>PF13_0201</i>	Thrombospondin-related anonymous protein, TRAP, alias SSP2	✓	✓		✓
<i>PFB0915w</i>	Liver stage antigen 3, LSA-3	✓	✓		✓
<i>PFC0210c</i>	Circumsporozoite (CS) protein, CSP	✓	✓	✓	✓

**Table 3.1: List of candidates for CHIP 1.**

CHIP 1 candidate selection was achieved throughout an investigation of the existing literature based on the putative vaccine candidates that have already been biologically assessed. WHO: Malaria Vaccine Rainbow Tables, 2007; accessible via [www.who.int/vaccine\\_research/links/Rainbow](http://www.who.int/vaccine_research/links/Rainbow)). MVI: Filip Dubovsky (2001). Two additional lists were obtained internally from Dr. Patrick Corran (LSHTM) and Dr. Tania Dottorini (UNIPG). Identification codes of the genes retrieved from the database are listed in the first column, followed by the official gene name. In the last four columns, the ✓ indicates whenever that specific gene was present in the selection list.

### 3.2.3 CHIP 2 candidate list

The entire *Pf* genome (source: *PlasmoDB* database) was analysed for putative antigens synthesised and exported to the cell membrane of the infected erythrocyte during the blood stage of the parasite life cycle. This was achieved through a multi-step analysis (see Section 2.2), which consisted of three linearly progressive computational genome-screenings using each time different attributes (see Table 2.1 and Table 2.2 for more details). The initial filter aimed to flag all the possible proteins predicted to form the *Pf* parasite "secretome". The search aimed to identify features known to be typically present on proteins exposed to the extracellular space: signal peptide, GPI-anchor, transmembrane domain and export motifs, with the latter representing a distinctive feature in *Pf*. In the attempt to determine the entire *Pf* secretome and to avoid an underestimation of it via miss-predictions produced by different algorithms, we decided to implement a computational analysis based on both experimental and *in silico* approaches. The three different secretome algorithms PEXEL, ExportPred and Host Targeting-motif (HT) identified 372, 413 and 229 genes, respectively (Table 3.2). These were complemented with a panel of additional predictors (e.g. SignalIP 3.0, PlasmoAP) and experimental evidences. Hence, the "secretome" selection was eventually implemented with the addition of the proteins predicted to have a transmembrane domain (n.1692, source: *PlasmoDB*), proteins containing a signal peptide (n.866) and proteins identified as GPI-anchored (n.30, Gilson et al. (2006)). Further 47 entries were retrieved from a comprehensive scientific review by Alan F. Cowman and Brendan S. Crabb (Cowman and Crabb, 2006), which listed a number of important *Pf* merozoite proteins. The panel of candidates was completed with the addition of two internal antigen lists (LSHTM n.62 and UNIPG n.318) previously produced by the respective consortium members during a similar study. The overall analysis yielded a large set of genes (n.4470, see Table 3.2). However, this preliminary selection was largely redundant and it included proteins that were off-target. Indeed, within the 4470 proteins there were some (n.517) that, although possessing a bipartite N-terminal leader sequence, use the secretory pathway to go into the a membrane-bounded plastid called the apicoplast, which is an organelle essential to the parasite during both the pre-erythrocytic and erythrocytic stages. In the initial list there were also proteins annotated as members of the STEVOR (n.219) and RIFIN (n.222) VSA families, which were eventually considered for the CHIP 3 analysis. When we compared the PEXEL and the ExportPred groups we found a high degree of

Source (Reference)	N. genes
PEXEL (Marti et al., 2004)	372
HT (Hiller et al., 2004)	229
ExportPred (Sargeant et al., 2006)	413
Transmembrane Proteins (Database)	1692
Signal Peptide (NN-HMM)	866
GPI-anchored (Gilson et al., 2006)	30
Merozoite, MSPs (Cowman and Crabb, 2006)	47
LSHTM (Dr. Patrick Corran)	62
UNIPG (Dr. Tania Dottorini)	318
Apicoplast (Database)	517
STEVOR (Database)	219
RIFIN (Database)	222
Initial Input, Total	4470

**Table 3.2: CHIP 2 antigen screening, initial pool of candidates.**

Composition of the initial pool of candidates selected for CHIP 2 (the chip containing putative immune targets). During the process, our attention was focused on the identification of a number of features commonly found on secreted or extracellularly exposed proteins. The parasite genome was therefore filtered for proteins predicted to containing export motifs, transmembrane domain, signal peptide and GPI-anchor. This *in silico* analysis was complemented with a number of publications (e.g. Cowman and Crabb (2006)) and internal studies (i.e. LSHTM and UNIPG) which listed several genes proved or predicted to be secreted and exposed at the merozoite stage and/or at the erythrocytic stage of the parasite life cycle. Combined together these selection criteria produced an initial redundant pool of candidates of 4470 genes.

redundancy, consisting of over 300 common proteins (n.330), which were eventually eliminated. Of the 229 HT predicted entries, only 15 were unique to the Host Targeting-motif when overlapped with the PEXEL group. Additionally, we depleted both the PEXEL and ExportPred list of both RIFIN and STEVOR family members of the proteins imported into the apicoplast. These allowed the creation of two smaller groups containing n.159 PEXEL and n.67 ExportPred unique entries. Interestingly, of the 67 ExportPred candidates, 54 were VAR genes and therefore they were moved into the CHIP 3 list of VSA genes. To the remaining 13 ExportPred members we added other 36 entries that were found in the literature but were not annotated into the database.

Noteworthy, is the fact that during the computational analysis the *PlasmoDB* database and consequently the genes annotated in it were constantly updated. This created a discrepancy between the initial list of proteins predicted to be imported into the apicoplast. For some of these proteins the gene

annotation was modified and their apicolplast import signal became either a PEXEL or an ExportPred signal, which forced us to create a new group, named PexApic containing 73 newly annotated entries. At completion of this second step of the analysis a list of 666 unique positive entries was produced (see Table 3.3).

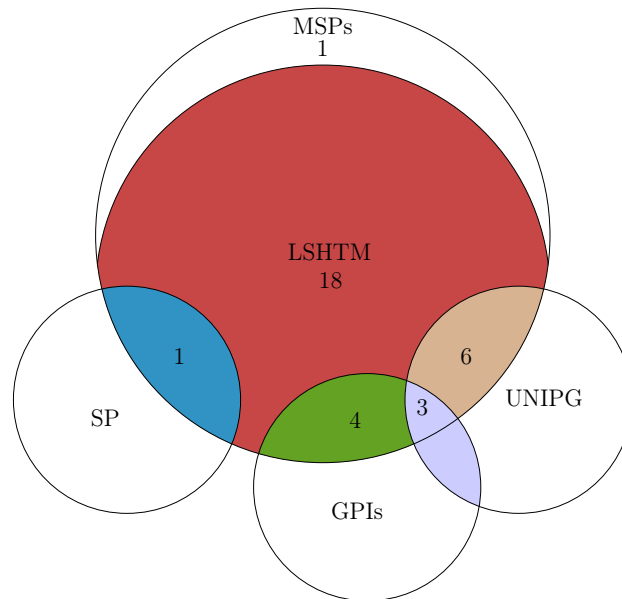
Source, Bioinformatic analysis	N. genes
PEXEL, PEXEL-(STEVOR $\cup$ RIFIN $\cup$ Apicoplast)	159
HT, HT genes not present in PEXEL	15
ExpPred, ExportPred-(stevor $\cup$ rifin $\cup$ Apicoplast $\cup$ Sel PEXEL)	13
ExportPred, Exportpred not included in database	36
PexApic, (PEXEL-(STEVOR $\cup$ RIFIN)) $\cap$ Apicoplast where Sel Exp Apicoplast - Sel Pex Apicoplast $\Rightarrow \emptyset$	73
SP, Signal peptide, blood stage expressed - redundant entries	201
Merozoite (MSPs), (Cowman and Crabb, 2006)	26
GPI-anchored, (Gilson et al., 2006)	20
UNIPG, Bioinformatics study carried out by Dr. Dottorini 2006	123
Preliminary analysis, Total	666

**Table 3.3: CHIP 2 antigen screening, preliminary selection.**

The list of candidates was produced after depleting the initial pool of selected genes (n.4470) from redundant entries. The procedure reduced the number of targets to 666, which corresponds to a 7<sup>th</sup> of the initial number.

An example of the high level of redundancy between the different antigen groups is described in Figure 3.3. Of the 33 refined entries of the merozoite list (MSPs, Cowman and Crabb (2006)), 32 were already present in the LSHTM list. Of these, 18 were only common in these two initial lists (MSP and LSHTM), whereas the remaining 14 were shared within SP (n.1), GPI (n.4), LSHTM (n.18) and UNIPG (n.6) and GPI  $\cap$  UNIPG (n.3) lists.

A third step of analysis was eventually implemented to further reduce the size of the antigen panel (n.666). The new filtering criteria were based on a blood stage expression analysis, which used experimental evidences from Mass Spectrometry (MS) and Microarray data (MA), both retrieved through the literature and/or accessible via the *PlasmoDB* database. This third filter (see Table 2.3 for more details) aimed to selected and prioritise those genes which possessed experimental evidence of expression during the merozoite, throphozoite and ring stages of the parasite life cycle, penalizing gene expressed during the erythrocytic and/or sexual stage. Additionally, the list was cleaned from genes with no evidence of expression (neither in MA or MS), as they could



**Figure 3.3: Example of redundancy in the preliminary antigen selection.**

The figure highlights the level of redundancy present in the initial pool of genes selected for their antigenic properties. In particular, this Venn diagram describes the relations between the MSPs (n.19), LSHTM (n.62), SP (n.866), GPI (n.30) and UNIPG (n.318) collection of genes. Each list was produced via a unique analysis based on specific criteria. However, when the selections were compared one with the other a high level of common entries was found (e.g. all the genes present in the LSHTM list were already present in the other groups). The dimension of each circle does not reflect the size of the related group of antigens and it was used for graphical purposes only.

represent a subgroup of genes that lost their ability of coding for proteins (i.e. pseudogenes). Finally, each entry was analysed for primary, secondary and tertiary protein structure. The analysis took in consideration the expression feasibility in terms of solubility, prediction of functional domains, presence of short external loops (typical of transmembrane proteins and known to be difficult to express) and possible disagreement of the MS and MA data. Gene and/or protein attributes were individually evaluated and scored against a priority scoring function. An example of such analysis is visible in Table 3.4.

Through the application of this new filter 260 unique entries were selected to form the final list of candidates selected for CHIP 2 (see Table 3.5).

GENE	<i>MAL7P1.170</i>	<i>PF11_0023</i>	<i>PFE0710w</i>
Information from PlasmoDB	Ring Stage Expressed Protein	Hypothetical Protein	Hypothetical Protein, conserved
Length (bp)	882	927	2604
Exportome or Secretome	PEXEL motif present	ExportPred motif present	GPI-anchor motif present
Other Criteria	present in UNIPG and LSHTM	None	present in GPI-anchored and LSHTM
Mass Spec (MS) Expression	FW Trophozoite-Schizonts/ iRBC membrane, Trophozoite (PlasmoDB), Trophozoite/Ring (Florens et al., 2002)	None	None ( <i>PlasmoDB</i> ), Sporozoite (Florens et al., 2002)
Microarray (MA) Expression	Early Rings / Late Rings / Early, Trophozoite / Late Trophozoite / Early Schizonts / Late Schizonts / Merozoites	All stages	Sporozoite / Gametocyte / Early Rings / Late Rings / Early Trophozoite / Late Trophozoite / Early Schizonts / Late Schizonts
Additional Negative Attributes	None	Short External Loops (SEL)	None
Final Decision	<b>SELECTED</b>	<b>EXCLUDED</b>	<b>EXCLUDED</b>

**Table 3.4: CHIP 2 antigen screening, snapshot of the internal scoring function.** View of the internal scoring system used to reduce the number of candidates selected for CHIP 2. In this example three genes are scored against a series of attributes described in the first column of the table. The red color indicates a negative attribute. A summary of the internal scoring function is described in Table 2.3.

Source	N. genes
PEXEL	55
HT	6
ExportPred	6
PexApic	15
Signal Peptide	138
Merozoite (MSPs)	26
GPI-anchored	8
UNIPG	6
Total unique entries	260

**Table 3.5: CHIP 2 antigen screening, the final results.**

A summary representing the bioinformatic analysis performed for CHIP 2 candidates is available in Figure 3.4. To maximise cloning efficiency and downstream protein expression, each of the 260 genes was re-analysed for primary, secondary and tertiary protein structure using the information obtained during the previous step. In accordance with the literature and following the suggestions received from the PX' team, the major attribute that was evaluated was

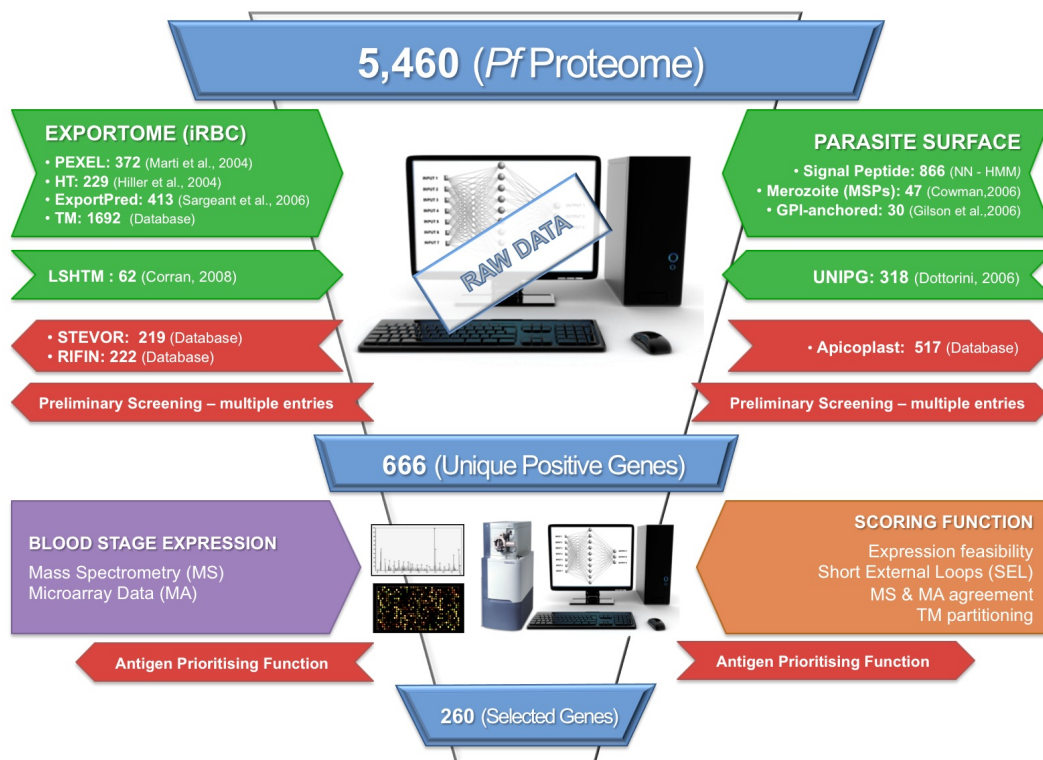
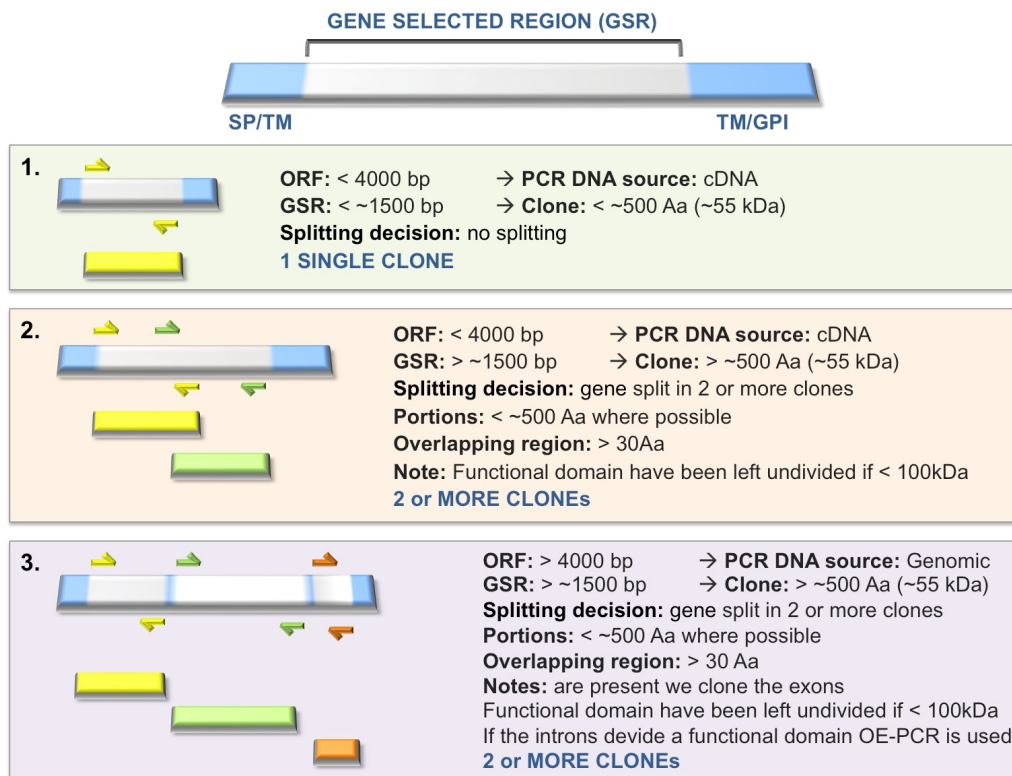


Figure 3.4: Summary of the bioinformatic analysis for CHIP 2.

the protein size, as proteins bigger than 50-60 kDa may result challenging for heterologous expression in *E. coli*. Where necessary, large genes (over  $\sim 55$  kDa size) were split in smaller segments ( $< 40$ -50 kDa) as described in Figure 3.5. Where possible, predicted functional domains and disulfide bonds were kept intact to guarantee optimal protein folding capability. In the case of genes bigger than 4000 bp (over 120 kDa) primers were designed to target specific exons, due to the restriction of our cDNA libraries, which according to the supplier contained genes with average size smaller than 4000 bp. This post-selection analysis of the 260 entries produced 448 portions of genes.

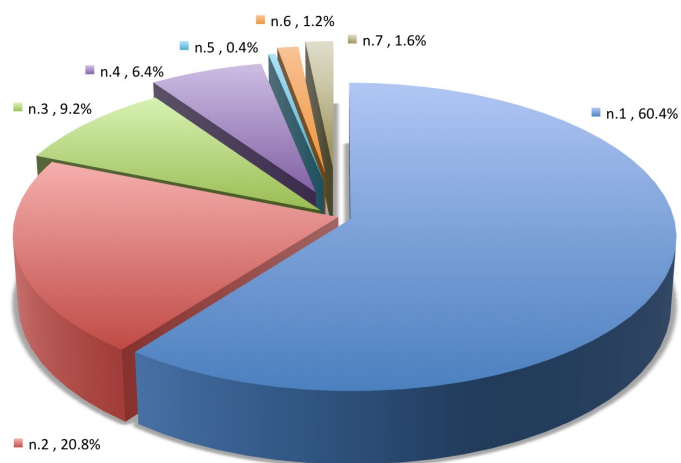




**Figure 3.5: Summary of the gene splitting criteria.**

Brief overview of the strategy used to decrease the length of "oversized" genes. The splitting of these genes should improve the expression efficiency when *E. coli* is used as a host. GSR: Gene Selected Region, which is predicted to encode for the portion of the protein exposed to the extracellular space. SP: Signal Peptide sequence. TM: Transmembrane domain. GPI: Glycosylphosphatidylinositol anchor region. ORF: Open Reading Frame, which consists in the gene-encoding sequence from the start to the stop codon. OE-PCR: overlap extension polymerase chain reaction, which represented one of the alternative strategies designed to overcome the presence of introns when the genomic DNA was used as source of DNA (not yet applied to any of the genes).

The average value of the splitting was  $\sim 1.8$  portions per gene, meaning that out of 10 genes, 18 gene portions were formed. The majority of the genes (60%) were maintained in one single sequence. Out of the 40% "oversized" genes, the majority was split into two portions (20.8% of the 39.6%). However, in some rare occasions genes were divided into more than three (n.26), four (n.16), five (n.1) or six (n.3) segments (see Figure 3.6). A maximum of seven portions were created for 4 genes, one of which (*PFB0405w*) encodes for a protein of 3,135 amino acids (AAs) and over 340 kDa of MW.



**Figure 3.6: Results of the gene splitting activity.**

The figure summarises the results of the gene splitting activity for the 260 *Pf* candidates. The main group (n.1) is formed by those genes left intact. These are the genes encoding for an extracellular region smaller than  $\sim 50\text{-}60$  kDa or  $\leq 100$  kDa if a functional domain was present. The second main group is represented by those genes, which were split into two portions (n.2), as their sequence of interest was longer than the suggested cut-off. Only a limited number of genes had to be divided in more than three (n.3) or four (n.4) portions. Overall, the splitting activity produced 448 portions, out of the 260 candidates selected for CHIP2.

### 3.2.4 CHIP 3 candidate list

According to the consortium guidelines the third chip to be produced during the project should contain the members of the multiple-gene VAR family, a sub population of the VSA. These genes encode for a very interesting set of variable surface antigens that are expressed by the parasite in a mutually exclusive fashion. Indeed, such antigenic variation enables the parasite to evade the host immunity during the invasion of the RBCs, as each infected erythrocyte exposes a different variant type of VAR protein (see Section 1.2.1 and Figure 1.6). The list of VAR genes was produced by querying the *PlasmoDB* database for VAR-specific attributes (see Section 2.2). Using the information available in both the literature (Dzikowski et al., 2006; Gardner et al., 2002) and centralised databases (*PlasmoDB* and *GeneDB*), we carried out an *in silico* analysis for *Pf* VSA proteins. The aim was to identify each of the erythrocyte membrane protein 1 (PfEMP1) VAR gene family members using information present in the *PlasmoDB* database. The search led to a list of 74 VAR genes (see Table 3.6). Members of RIFIN and STEVOR families were also taken into consideration, as they were the most populated VSA groups during the computational analysis performed for CHIP 2 (n.222 RIFIN and n.219 STEVOR). However, due to the complexity and variability of VSA protein sequences the consortium decided to focus only on the 74 PfEMP1 family members, leaving the remaining candidates aside for a possible future project.

N.	Gene ID	Gene Description
1	<i>MAL13P1.1</i>	erythrocyte membrane protein 1
2	<i>MAL13P1.356</i>	erythrocyte membrane protein 1
3	<i>MAL13P1.405</i>	erythrocyte membrane protein pfemp3, putative
4	<i>MAL13P1.6</i>	erythrocyte membrane protein 1-like
5	<i>MAL7P1.186</i>	VAR-like erythrocyte membrane protein 1
6	<i>MAL7P1.187</i>	erythrocyte membrane protein 1
7	<i>MAL7P1.212</i>	erythrocyte membrane protein 1
8	<i>MAL7P1.50</i>	erythrocyte membrane protein 1
9	<i>MAL7P1.55</i>	erythrocyte membrane protein 1
10	<i>MAL7P1.56</i>	erythrocyte membrane protein 1
11	<i>MAL8P1.207</i>	erythrocyte membrane protein 1, PfEMP1
12	<i>MAL8P1.209</i>	VAR-like protein
13	<i>MAL8P1.220</i>	erythrocyte membrane protein 1
14	<i>PF07_0048</i>	erythrocyte membrane protein 1
15	<i>PF07_0049</i>	erythrocyte membrane protein 1
16	<i>PF07_0050</i>	erythrocyte membrane protein 1
17	<i>PF07_0051</i>	erythrocyte membrane protein 1
18	<i>PF08_0103</i>	erythrocyte membrane protein 1
19	<i>PF08_0106</i>	erythrocyte membrane protein 1
20	<i>PF08_0107</i>	erythrocyte membrane protein 1
21	<i>PF08_0140</i>	erythrocyte membrane protein 1
22	<i>PF08_0141</i>	erythrocyte membrane protein 1
23	<i>PF08_0142</i>	erythrocyte membrane protein 1
24	<i>PF10_0001</i>	erythrocyte membrane protein 1
25	<i>PF10_0011</i>	erythrocyte membrane protein 1 , truncated, degenerate
26	<i>PF10_0012</i>	erythrocyte membrane protein 1 , truncated
27	<i>PF10_0361</i>	hypothetical protein
28	<i>PF10_0406</i>	erythrocyte membrane protein 1
29	<i>PF11_0007</i>	erythrocyte membrane protein 1
30	<i>PF11_0008</i>	erythrocyte membrane protein 1
31	<i>PF11_0521</i>	erythrocyte membrane protein 1
32	<i>PF13_0003</i>	erythrocyte membrane protein 1
33	<i>PFA0005w</i>	erythrocyte membrane protein 1
34	<i>PFA0765c</i>	erythrocyte membrane protein 1
35	<i>PFB0010w</i>	erythrocyte membrane protein 1
36	<i>PFB0045c</i>	erythrocyte membrane protein 1, truncated
37	<i>PFB0974c</i>	erythrocyte membrane protein 1, truncated, degenerate
38	<i>PFB0975c</i>	erythrocyte membrane protein 1, truncated
39	<i>PFB1025w</i>	erythrocyte membrane protein 1, truncated, degenerate
40	<i>PFB1045w</i>	erythrocyte membrane protein 1, truncated
41	<i>PFB1055c</i>	erythrocyte membrane protein 1
42	<i>PFC0005w</i>	PfEMP1
43	<i>PFC1120c</i>	VAR (3D7-varT3-2)
44	<i>PF0005w</i>	erythrocyte membrane protein 1
45	<i>PF0020c</i>	erythrocyte membrane protein 1
46	<i>PF0615c</i>	erythrocyte membrane protein 1

N.	Gene ID	Gene Description
47	<i>PFD0625c</i>	erythrocyte membrane protein 1
48	<i>PFD0630c</i>	erythrocyte membrane protein 1
49	<i>PFD0635c</i>	erythrocyte membrane protein 1
50	<i>PFD0995c</i>	erythrocyte membrane protein 1
51	<i>PFD1000c</i>	erythrocyte membrane protein 1
52	<i>PFD1005c</i>	erythrocyte membrane protein 1
53	<i>PFD1015c</i>	erythrocyte membrane protein 1
54	<i>PFD1235w</i>	erythrocyte membrane protein 1
55	<i>PFD1245c</i>	erythrocyte membrane protein 1
56	<i>PFE0005w</i>	erythrocyte membrane protein 1
57	<i>PFE1640w</i>	erythrocyte membrane protein 1, truncated
58	<i>FFF0010w</i>	erythrocyte membrane protein 1
59	<i>FFF0020c</i>	erythrocyte membrane protein 1-like protein
60	<i>FFF0845c</i>	erythrocyte membrane protein 1
61	<i>FFF1580c</i>	erythrocyte membrane protein 1
62	<i>FFF1595c</i>	erythrocyte membrane protein 1
63	<i>PFI0005w</i>	erythrocyte membrane protein 1
64	<i>PFI1820w</i>	erythrocyte membrane protein 1
65	<i>PFI1830c</i>	erythrocyte membrane protein 1
66	<i>PFL0005w</i>	erythrocyte membrane protein 1
67	<i>PFL0020w</i>	erythrocyte membrane protein 1
68	<i>PFL0030c</i>	erythrocyte membrane protein 1
69	<i>PFL0935c</i>	erythrocyte membrane protein 1
70	<i>PFL1950w</i>	erythrocyte membrane protein 1
71	<i>PFL1955w</i>	erythrocyte membrane protein 1
72	<i>PFL1960w</i>	erythrocyte membrane protein 1
73	<i>PFL1970w</i>	erythrocyte membrane protein 1
74	<i>PFL2665c</i>	erythrocyte membrane protein 1

**Table 3.6: Final list of the CHIP 3 antigens.**

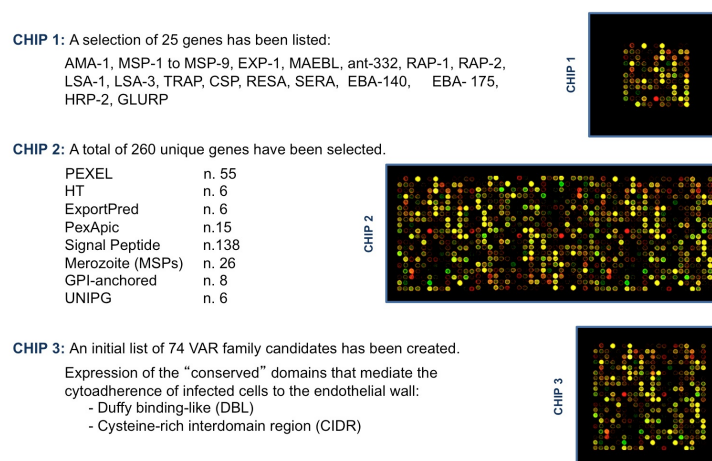
CHIP 3 candidate selection was produced combining the information available in the literature with the dataset available on both *PlasmoDB* and *GeneDB* databases. Overall 74 gene entries were selected, all of which are members of the multiple-gene VAR family. These sub population of the VSA are essential element for the evasion of the parasite from the host immune system during the invasion of erythrocytes.

### 3.2.5 Discussion

Before the publication of the *Pf* genomic sequence (Gardner et al., 2002) and the elucidation of its proteome (Florens et al., 2002), no more than 20-25 vaccine candidates were identified (Filip Dubovsky, 2001) and tested with little success (Moorthy et al., 2009; Greenwood et al., 2008; Wipasa and Riley, 2007; Doolan and Stewart, 2007). However, the post-genomic era brought the opportunity to utilise genome-wide high-throughput screenings aimed at identifying novel *Pf* antigens for both diagnostic and vaccine development. Hence, increasing emphasis has been put into enriching the antigen portfolios by most of the funding organisations (e.g. PATH MVI). This process will not only allow a better understanding of the complex relation between the human host and the parasite but also ensure a smooth flow of targets through the vaccine development pipeline. However, none of the bioinformatic tools currently available can be utilised in combination with the genomic information alone to effectively identify correlates of protection capable of promoting a protective immunity. The *in silico* prediction needs to be proven by experimental evidences. It is under these bases that the FightMal project sponsored the genome-wide bioinformatic evaluation of the *Pf* parasite. The analysis successfully produced three lists of characterised and putative parasite antigens (see Figure 3.7). One for each of the protein microarray chips, which will form a robust platform for the high-throughput serum immune profiling of Malaria-protected individuals.

The rationale behind the antigen selection was based on the use of reverse vaccinology (Sette and Rappuoli, 2010; Rappuoli, 2000) approach and it aimed to highlight the portion of the *Pf* proteome potentially exposed to the action of antibodies during the *Pf* infection of the human host. Candidates present in each list shared similar antigenic properties (e.g. blood stage expression, exposure to or secretion into the extracellular space, etc). However, they were divided into three different groups due to either their availability in the scientific community (e.g. CHIP 1 list, current malaria vaccine candidate) or for their simplicity of production (e.g. CHIP 2, putative immune targets versus highly variable surface antigens, CHIP 3).

Antigen sourcing was just one of the many challenges faced during this initial step of the project. The lack of a uniformed nomenclature for gene annotation and the high number of hypothetical proteins (58%, 3168/5460 source: *PlasmoDB*, 2007), for which little, if any, is known in terms of biological and structural data were two other crucial issues that had to be addressed



**Figure 3.7: Summary of the bioinformatic selection for CHIP 1,2,3.**

The images of the CHIP 1, 2 and 3 are only a graphic representation of the expected arrays.

during this complex analysis.

For CHIP 1, the WHO and the MVI websites as well as a number of scientific reviews have been investigated in order to create a comprehensive list of 25 target antigens which represent an exhaustive number of biologically characterised malaria vaccine candidates. Some of these have been extensively utilised in clinical trials (see Table 1.7). Currently the most clinically advanced malaria vaccine candidate (GSK RTS,S AS01/AS02, see Section 1.3.4) is based on one of these antigens. AMA-1 is also a leading blood-stage vaccine candidate, see Table 1.7 and Section 1.3.4. Additional blood stage candidates like MSP2, MSP4 and MSP5 are currently under investigation to evaluate their ability to induce a protective immune response and their safety (MVI-PATH, 2007).

Critically, since the majority of these antigens are currently being explored as vaccine candidates in clinical trials, the consortium intends to source high quality, purified preparations of each antigen in CHIP 1 from the broader malaria research community. The LSHTM is already in possession of some of these antigens and they will act as an intermediary to access the remaining antigens from other laboratories in the field. To date, the consortium has gathered more than 10 of the CHIP 1 candidates. In the event that we are unable to collect particular candidates, they will be inserted into our *in vivo* expression pipeline established for the production of CHIP 2 targets.

For the chip containing putative immune targets (CHIP 2), a list of 260 genes ( $\sim 5\%$  of the *Pf* proteome) has been produced using an *in silico* genome/

proteome screening. These genes represent mainly exported/surface-expressed parasite proteins expressed during the merozoite and/or iRBC stage. However, in order to not exclude potentially relevant candidates, genes possessing experimental evidence of blood stage expression, antigenic properties and/or putative transmembrane domains were included. Generally, we aimed to not underestimate the importance of experimental evidence coming from the literature (Cowman and Crabb, 2006; Gilson et al., 2006), but also tried to complement this relatively limited pool of candidates with bioinformatic tools capable of highlighting the presence of transmembrane domains or predicted exporting/surface-targeting signals. To reach the cytosol or the iRBC membrane these proteins need to cross both the parasite plasma membrane and the parasitophorous vacuole membrane (PVM). Although, the first crossing event requires only a standard ER-type signal sequence, the second passage, through the PVM is dependent on a peptide sequence known as either *Plasmodium* EXport ELement (PEXEL, Marti et al. (2004) or as Host Targeting-motif (HT, Hiller et al. (2004)). These two sequence motifs can be identified using two separate algorithms. However, both computational tools recognise the same N-terminal core sequence consisting in RxLxE/Q/D. In addition, the PEXEL-based prediction program was recently expanded to include the ability to identify not only the export element sequence but even recessive signal peptide sequences (ExportPred, Sargeant et al. (2006)). According to their creators, each version of the 3 secretomes account for over 250 proteins. However, as shown by Christiaan van Ooij (van Ooij et al., 2008) the overlap between the 3 groups is limited. Only 59 proteins are common between the HT and PEXEL secretomes, highlighting the differences in specificities in the prediction programs.

As a result of this initial analysis of the *Pf* proteome, over 4900 entries were annotated. Due to the high degree of overlap between the different groups a filter was applied in order to deplete the raw data from multiple entries. Additionally, the numerous members of the RIFIN and STEVOR families were excluded (for use in a designated multi-copy gene families array at some point in the future) together with the genes that were representative of CHIP 1 and 3 resulting in a list of  $\sim 1/8$  of the original. To further narrow down the size of this selection a priority scoring function was applied. The ability to score the genes for positive attributes (MS and MA blood stage expression) and negative attributes (liver and or sexual stage expression; Short External Loops, SEL) finally enabled us to redefine the putative candidate list for CHIP 2 to a  $\sim 1/3$  of the initial selection. It has to be noted, however, that this figure does



not correspond to the final number of discrete putative candidates that will be cloned, expressed, purified and spotted onto the array. Each gene had to be individually analysed prior its cloning and expression. Where necessary, it was split into smaller overlapping fragments to improve the chances of obtaining a soluble protein, as observed by (Vedadi et al., 2007). Critically, however, an overlapping sequence of at least 30-40 AAs was maintained between two adjacent portions to avoid the destruction of linear epitopes in the region of division. Candidates predicted to carry large functional domains, were left untouched up to 100 kDa, as suggested by our French partner PX'. This approach was employed to guarantee a higher degree of solubility during the *in vivo* expression. Under these circumstances, from the list of 260 unique genes we doubled the number of clones to be expressed ( $\sim 1.8$  clones per gene).

Finally for CHIP 3, the VSA antigen array, 74 genes were listed. This selection contains genes annotated as members of a clonally variant surface antigen family collectively termed PfEMP1 or VAR family. These large genes (average ORF size  $> 6000$ bp), although highly variable and polymorphic, do possess conserved motifs in their extracellular region, which are classified according to their sequence similarity. These are defined as Duffy Binding Like (DBL) and Cysteine-rich InterDomain Regions (CIDR) and they are the motifs that are more likely to be expressed for this chip. Currently no action has been taken to start the production of CHIP 3 candidates. However it is envisaged that the chip will be produced and assessed using a set of 20 domains of a VAR gene provided by Dr. Odile Puijalon from the Unité d' Immunologie Moléculaire des Parasites, Institut Pasteur, Paris-France.

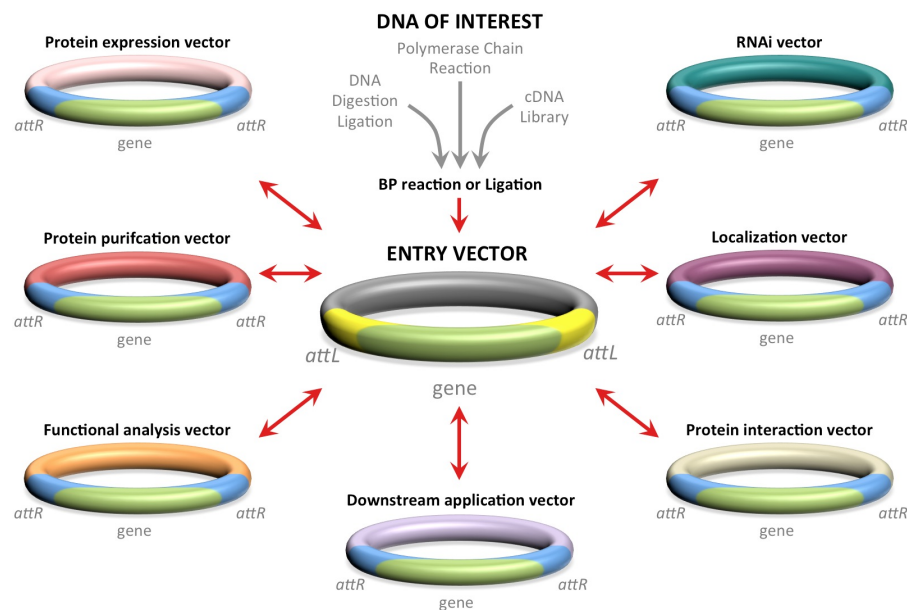
Overall, this initial *in silico* analysis produced a feasible list of interesting candidates for all the 3 chips. From the high level of hypothetical proteins (58%, 3168/5460), for which little, if any, is known in terms of biological and structural features, to the absence of a unified gene annotation ontology, which led many scientific papers and databases to annotate with different names the same genes, many were the challenges that we faced. Biological evidences and scientific hypotheses together with complex bioinformatic algorithms were successfully implemented to create a complex network of screening tools that can be fine tuned according to the specific requirements. The application of this reverse vaccinology approach allowed us to specifically select hundreds of genes encoding for proteins that are either secreted or anchored on the parasite/infected host cell surface during the blood stage of *Pf* life cycle. The speci-

ficity of the bioinformatic analysis was highlighted by the positive selection of many genes already known to be secreted or anchored in the extracellular space. However, it will be possible to assess the power of this computational analysis only after the completion of the serum immune profiling. In the meantime, a database containing structural and biological attributes of selected parasite proteins is currently being created.

### 3.3 Cloning of *Pf* CHIP 2 candidates

#### 3.3.1 Selection of the cloning platform

The design of a suitable cloning strategy was one of the initial tasks to be addressed in the project. Now a days, the insertion of a PCR-amplified gene into a plasmid of choice is considered a relatively simple and routine task, however, when the procedure has to be applied to hundreds genes of an AT-rich organism in different laboratories the unforeseen issues can be many. This is why, since the beginning, I started investigating, which standardised high-throughput cloning system was available on the market. To look for the most appropriate system, the criteria used were: flexibility, adaptability, consistency, efficiency, cost of implementation and widespread use. Several technologies were taken into consideration, and Invitrogen's Gateway<sup>®</sup> system resulted as the most suitable, see Figure 3.8.



**Figure 3.8: Gateway<sup>®</sup> Technology.**

The Gateway<sup>®</sup> platform allows the cloning of DNA inserts into and back out of multiple plasmids through a site-specific recombination, overcoming some of the traditional limitation of typical cloning procedures (i.e. enzymatic digestion, clean-up steps, low-efficiency recovery etc.). Once the gene of interest is inserted into a Gateway<sup>®</sup> vector, it can theoretically be accessed by any expression system or used in different downstream applications. As both clone orientation and reading frame are maintained during the subcloning step there is no need of re-sequencing the DNA insert.

The Invitrogen platform allows the cloning and exchange of DNA inserts from one Gateway<sup>®</sup> vector to another, in a high-throughput manner and main-

taining orientation and reading frame without the need of restriction enzyme digestion, ligation and re-sequencing. DNA inserts can be moved within any expression, cloning, RNAi or two hybrid system vectors. Moreover, the technology is widespread with over 1,500 scientific references, which will allow common and standardised access to the FightMal constructs once the cloning activity is completed. Indeed, it is envisaged that once the production of the 448 DNA constructs (representing the initial 260 genes selected for CHIP 2) is finalised, all the plasmids harbouring the inserts will be submitted to the Malaria Research and Reference Reagent Resource Center (MR4) database and made available to the public.

The Gateway<sup>®</sup> technology is based on a well-characterised bacteriophage  $\lambda$  site-specific recombination sequence. In the phage, there is a short sequence of DNA called attP (the P comes from Phage), and in *E. coli*, there is a stretch of DNA called attB (the B stands for Bacteria). Following a  $\lambda$  phage infection the phage DNA recombines with the equivalent bacterial DNA via the att by a  $\lambda$ -encoded integrase enzyme and the *E. coli*-synthesised integration host factor. The recombination product is the integration into the bacterial genome of the  $\lambda$  DNA. As a result of such event, two att sequences (attL and attR, Left and Right respectively) are flanking the phage DNA. The integration is reversible and the excision of the  $\lambda$  DNA will regenerate the original attP and attB sequences. The specific recognition of a 7-bp core att sequence avoids cross-reactivity and it ensures the directionality of the reaction. The Gateway<sup>®</sup> technology is able to replicate this natural integration pathway via either a BP or LR *in vitro* recombination reactions (i.e. BP= PCR fragment + donor vector = entry vector; LR = entry vector + destination vector = expression clone), which in 5 min recombines an insert containing attLs with one containing attRs sites. This reversible event recreates the original attBs and attPs integration and it guarantees the correct orientation of the DNA into the new plasmid.

Once the cloning platform was selected, I focused my attention to the numerous plasmids available for this technology. My focus was on the destination vector family which, overall, contained 72 expression plasmids, see Figure 3.9.

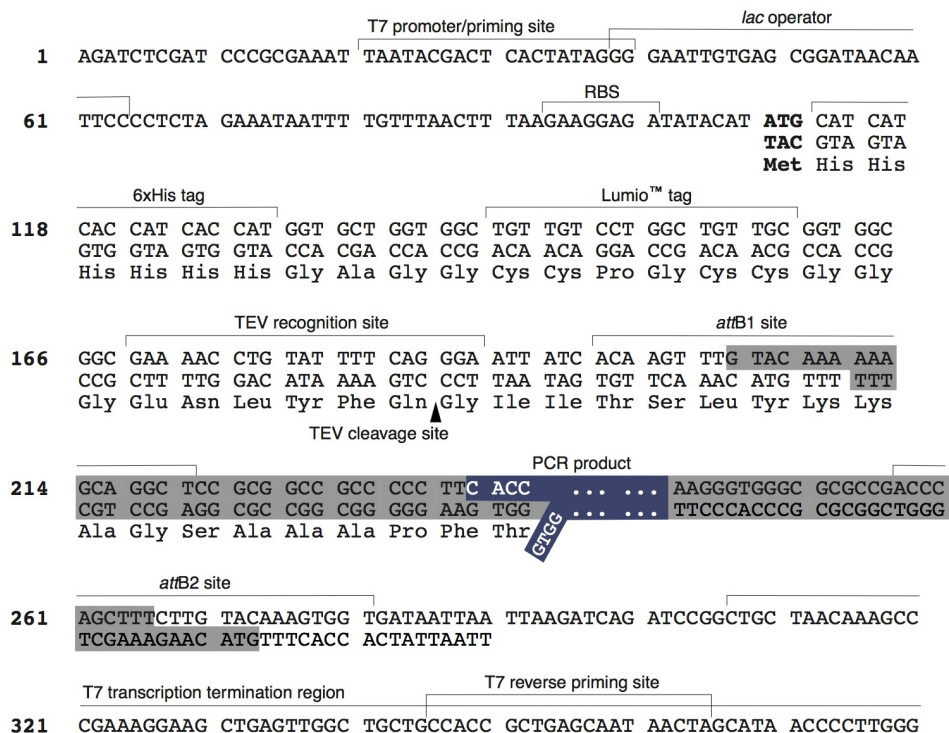
The Gateway<sup>®</sup> plasmids were analysed using the following criteria: suitability for *in vivo* expression in *E. coli* host, promoter sequence, presence of an inducer, N- and C-terminal tags, selection marker, presence of multiple cloning sites and a directional cloning site (see Figure 3.10). Of the existing

Vector Type	Host System	Cloning System	Inducer
Expression (72)	Mammalian (51)	Gateway (93)	IPTG (9)
Cloning (18)	Bacteria (10)	TOPO (3)	Tetracycline (4)
RNAi (3)	Insect (9)	TOPO adapted	galactose (1)
Two Hybrid (1)	In vitro (4)	Gateway (1)	copper sulfate (1)
	Yeast (2)		Arabinose (1)
Promoter	Selection marker	N-term tag	C-term tag
CMV (35)	Ampicillin (72)	6xHis (9)	V5 (36)
T7 (13)	Kanamycin (11)	V5 (8)	6xHis (14)
Polyhedrin (7)	Blasticidin (7)	lacZ (5)	Lumio (4)
EF1alpha (4)	Zeocin (5)	Lumio (4)	None (4)
Lac (4)	Gentamicin (4)	None (4)	BioEase (2)
Tet operator (4)	Spectinomycin (4)	GST (3)	capTEV (His (2)
promoterless (4)	Tetracycline (2)	BioEase (3)	GFP (2)
lac operator (3)	Gentamycin (1)	capTEV (His (2)	EmGFP (1)
ADH1 (1)	Tetracyclin (1)	GAL4DBD (1)	GST (1)
GAL1 (1)		Xpress (1)	YFP (1)
MT (1)		GAL4AD (1)	bla(M) (1)
OplE2 (1)		EmGFP (1)	
UbC (1)		YFP (1)	<b>N-term Protease cleavage</b>
araBAD (1)		bla(M) (1)	TEV (8)
none (1)		gene10 leader (1)	None (4)
		GFP (1)	EK (3)
		HP-Thioredoxin (1)	

**Figure 3.9: List of vectors and features available with Gateway<sup>®</sup> Technology.** Plasmids information was sourced from the Initrogen website: [www.invitrogen.com](http://www.invitrogen.com)

10 *E. coli* expression vectors, pET160/GW/D-TOPO<sup>®</sup> was chosen. Its main features are:

- *T7lac* promoter for IPTG-inducible expression in *E. coli*.
- *lacI* gene for the expression of the lac repressor, which is essential for the reduction of basal transcription from the *T7lac* promoter in the vector and from the *lacUV5* promoter in the *E. coli* host chromosome.
- Ampicillin resistance marker, which is needed for plasmid selection in the *E. coli* host.
- Two N-terminus tags, 6xHIS tag and Lumio<sup>™</sup> tag, followed by a TEV recognition/cleavage site (a sequence specific for a cysteine protease that is found in the Tobacco Etch Virus.)
- Two recombination sites, *attR1* and *attR2*, for recombinational cloning from a Gateway<sup>®</sup> entry plasmid clone.
- Directional TOPO cloning site, with CACC-nt as an entry sequence.
- pBR322 origin for low-copy replication and maintenance in *E. coli*.



**Figure 3.10: pET160/GW/D-TOPO® Cloning site.**

T7 promoter site: consensus promoter sequence required for an effective binding of the T7 DNA polymerase. *lac* operator sequence: region needed for the binding of lactose repressor, which is important for the repression of T7 RNA polymerase-induced basal transcription of the gene of interest. RBS or ribosome binding site: it promotes the efficient and accurate translation of mRNA. ATG: start codon. 6xHis tag sequence: it encodes for a N-terminal tag formed by six histidine residues in straight succession. Lumio™ tag sequence: it allows the expressed fusion protein to be specifically recognised by a biarsenical labeling reagent. TEV recognition site: a cleavage sequence site that enable the specific binding of a cysteine protease derived from Tobacco Etch Virus. attB1 and attB2 sites: they are required for moving DNA inserts into and back out of multiple Gateway® plasmids through a site-specific recombination (see also Figure 3.8). The region in gray corresponds to sequences that could be transferred from the this vector into a Gateway® entry plasmid following a recombination reaction. GTGG/CACC-PCR product: it is a four nucleotide sequence that allows the directional cloning of the of any DNA insert containing a 5'-CACC bp segment. T7 transcription termination region: consensus sequence required to facilitate the stop of RNA transcription. T7 priming sites: represent two consensus sequences, upstream and downstream of the cloning site used to assess the sequence of the cloning product with a forward and reverse T7 universal primers. The pET160/GW/D-TOPO® cloning site image was obtained from Invitrogen UK.

### 3.3.2 Selection of N- and C- terminal tags

The search for the optimal cloning platform continued with the selection of the most suitable C-terminus tag to be used for mediating recombinant protein purification and post-processing array normalisation. More than 12 different epitopes were considered (see Table 3.7). The initial list of candidate-epitopes was

Tag	N. residues	AA Sequence	Size (kDa)	pI	GRAVY
Poly-Arg	5-6	RRRRR(R)	0.80	12.70	-4.500
6xHis	6	HHHHHH	0.84	7.21	-3.200
Flag	8	DYKDDDDK	1.01	3.97	-3.325
Strep-tag II	8	WSHPQFEK	1.06	6.75	-1.825
c-MYC	11	EQKLISEEDL	1.20	4.00	-1.010
S-tag	15	KETAAKFERQHMS	1.75	6.76	-1.393
HAT-	19	KDHLIHNVHKEFHAAHANK	2.31	8.54	-1.368
3xFLAG	22	DYKDHDGDYKDHDIDYKDDDDK	2.73	4.16	-2.741
Calmodulin-binding peptide	26	KRRWKKNFIAVSAANRFKISSGAL	2.96	12.32	-0.565
Cellulose-binding domains	27-189	Domains	3.00-20.00	NA	NA
Thioredoxin-tag (TrxA)	108	Protein	11.70	4.67	0.007
Maltose-binding protein (MBP)	396	Protein	40.00	5.53	-0.251

**Table 3.7: List of possible C-terminal tags.**

The C-terminal tag list was based on the information available in the literature. In particular, two publications were mainly used Terpe (2003); Keesey (1996). pI and GRAVY values were calculated using "Protein Identification and Analysis Tools on the ExPASy Server", (Gasteiger et al., 2005). NA: not available.

analysed to identify an epitope long enough to guarantee a specific and unique recognition site for the antibody detection, but short enough to minimise any interference with the folding and reactivity of the recombinant protein. Additionally, we analysed both the epitope's surface charge (as the majority of the proteins of our most probable expression host (*E. coli*) range from a pI of 4.00 to 7.00, Kiraga et al. (2007)) and hydrophobicity (i.e. GRAVY: Grand average of hydropathicity index, which predicts the solubility of a peptide/protein. GRAVY > 0 the sequence is hydrophobic, GRAVY < 0 the sequence is hydrophilic; Kyte and Doolittle (1982)). Finally, I also evaluated the availability of an antibody directed against the tag, the affinity purification matrix and elution condition, the widespread use in multiple host systems and the ownership. Of the initial 12 epitopes, only 3 tags were selected for further investigation (Flag, c-MYC and Strep-tag II), see Table 3.8.

The c-MYC tag was eventually chosen for its short length, polarity charge (12 AAs, 1.2 kDa, pI 4.0 and GRAVY -1.010), absence of ownership and widespread use.

Tag	Advantages	Disadvantages
Flag	<ul style="list-style-type: none"> <li>- Small hydrophilic epitope, 8 AA long</li> <li>- It can be detected using monoclonal, polyclonal and conjugated antibodies</li> <li>- System used in a variety of cell types: bacteria, yeast and mammalian</li> <li>- Non-denaturing conditions for protein purification</li> <li>- Can be located either at the N- or C- terminus</li> </ul>	<ul style="list-style-type: none"> <li>- Registered by Sigma</li> <li>- The mAb purification matrix is not as stable as others (Ni-NTA or Strep-Tactin)</li> <li>- Detection is improved when using a 3x FLAG system, hydrophilic 22 AAs long epitope</li> <li>- The epitope can be removed by treatment with enterokinase (specific for the 5 C-terminal AAs of the peptide seq)</li> </ul>
c-MYC	<ul style="list-style-type: none"> <li>- Not registered, the sequence derives from a human gene (<i>Myc</i>)</li> <li>- Widespread use</li> <li>- It can be detected using monoclonal, polyclonal and conjugated antibodies</li> <li>- System used in a variety of cell types: bacteria, yeast, insect and mammalian cells</li> <li>- Successfully used for Western-blot, immuno-precipitation and flow cytometry</li> <li>- Tagged proteins can be affinity purified by coupling Mab 9E10 to divinyl sulphone-activated agarose</li> <li>- Washing condition are physiological. Elution can be achieved by dropping the pH</li> <li>- Can be located either at the N- or C- terminus</li> </ul>	<ul style="list-style-type: none"> <li>- Longer than the Flag tag, 11 AAs</li> <li>- Rarely applied for purification</li> </ul>
Strep-tag II	<ul style="list-style-type: none"> <li>- Short balanced AA composition, generally no effect on protein structure or activity</li> <li>- It can be detected using Strep-Tactin conjugated to HRP or AP or monoclonal antibodies against Strep-tag II</li> <li>- Highly selective and easily controllable binding properties</li> <li>- Rapid one-step purification under physiological conditions, parallel purification possible (96 wells plate)</li> <li>- System used in a variety of cell types: bacteria, yeast and mammalian cells</li> <li>- Can be located either at the N- or C- terminus</li> </ul>	<ul style="list-style-type: none"> <li>- Registered by IBA</li> <li>- Close system that requires expensive reagents not widely available</li> </ul>

**Table 3.8: Attributes of Flag, cMYC and Strep-tag II.**



### 3.3.3 Selection of the optimal amplification procedure

The correct amplification of the parasite blood stage putative proteins was crucial for the success of the project. I first assessed which source of DNA was suitable to use. Of the two most used sources of DNA, genomic DNA and cDNA libraries, the latter was chosen as a primary source as we wanted to avoid possible issues due to the presence of introns and pseudogenes. The MR4 was identified as a suitable supplier of the 3D7 DNA. Within its catalog four *Pf* cDNA libraries were selected for their specific content. Indeed, each of the four libraries were produced at different time points of the *Pf* blood stage (16h, 26h, 36h, asynchronous). However, empirical experimentation has shown that ORFs larger than 4000 bp are more likely to be found in the genomic DNA material than in the cDNA library (information supplied by MR4). Therefore, to guarantee the availability of all selected genes, two genomic DNA sources (3D7 and 3D7A) were identified as secondary DNA source representing a suitable back-up for the amplification of large and/or low expressed genes.

After an initial amplification assessment, which successfully demonstrated the suitability of both cDNA and genomic sources (data not shown), a panel of polymerase enzymes were assessed. The *Pf* parasite has a highly polymorphic genome, rich in repetitive regions and extremely poor in GC content. Consequently, to avoid the introduction of alterations and to guarantee a high degree of specificity for amplified sequences, a panel of 3 proofreading polymerase enzymes was tested (Pfx50<sup>TM</sup> DNA Polymerase (Invitrogen), Phusion<sup>TM</sup> High-Fidelity DNA Polymerase (Finnzymes) and PfuUltra<sup>TM</sup> II Fusion HS DNA Polymerase (Stratagene)), see Table 3.9.

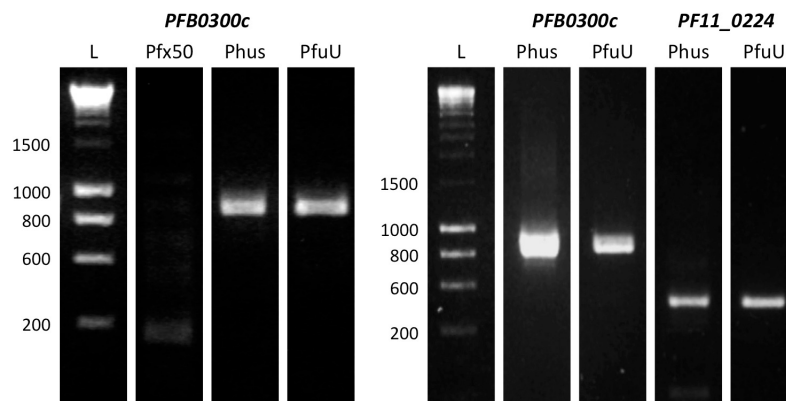
DNA Polymerase	Error Rate	Processivity	Cost per run	Supplier
Pfx50 <sup>TM</sup>	1/260,000	15sec/kb	£0.80	Invitrogen
Phusion <sup>TM</sup> High-Fidelity	1/770,00	15sec/kb	£0.75	Finnzymes
PfuUltra <sup>TM</sup> II Fusion HS	1/2,500,000	5sec/kb	£2.05	Stratagene

**Table 3.9: Panel of polymerase enzymes: performance and costs.**

Performance data were obtained from each enzyme manufacturer and from a technical note provided by Stratagene: "High-Fidelity PCR enzymes: properties and error rate determination" (TN5-03/06). Cost per run was obtained from the 2008 manufacturer's catalogues.

Each of the three polymerases is widely available and produced in large quantities from all the suppliers. In terms of cost, PfuUltra<sup>TM</sup> II Fusion HS DNA Polymerase is the most expensive (£82 for 40 runs), whereas Phusion<sup>TM</sup> High-Fidelity DNA Polymerase is the cheapest with a price tag of £75 for 100

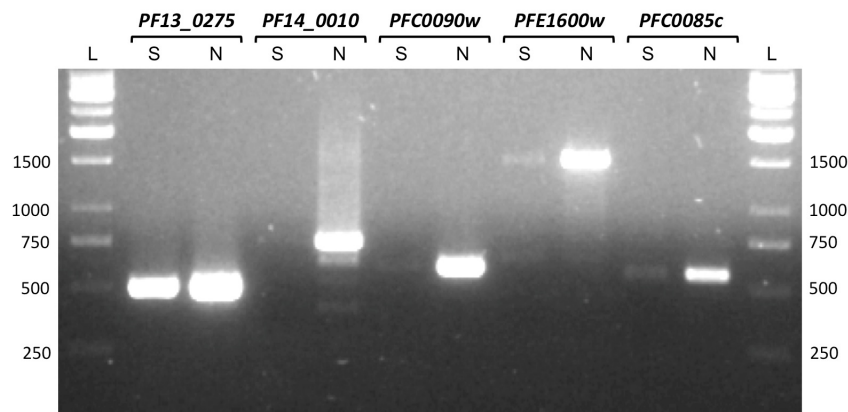
runs. Due to the AT-rich content of the *Pf* genome our focus was on performance, which were assessed in house according to the individual manufacturer's instructions using a panel of randomly selected ORFs. Each polymerase produced amplicons after the standard PCR procedure (Figure 3.11) but only two (Phusion<sup>TM</sup> High-Fidelity and PfuUltra<sup>TM</sup> Fusion HS) generated a satisfactory result with clear sharp bands of the correct size. Pfx50<sup>TM</sup> generated weak specific bands and/or multiple unspecific fragments. On the contrary both Phusion<sup>TM</sup> High-Fidelity and PfuUltra<sup>TM</sup> Fusion HS, produced single sharp bands. Noteworthy is the fact that the Phusion<sup>TM</sup> performed similarly to the more expensive PfuUltra<sup>TM</sup> in terms of both sensitivity and specificity.



**Figure 3.11: Polymerase enzyme evaluation tests.**

Two *Pf* genes were randomly selected for the assessment of enzymes performance: PFB0300c (862 bp) and PF11\_0224 (532 bp). The *left panel* shows the PFB0300c amplicons obtained using the three enzymes. As visible the second lane the Pfx50<sup>TM</sup> had difficulties amplifying the parasite DNA. The *right panel* compares the performances of the two best polymerases, Phusion<sup>TM</sup> (Phus) and PfuUltra<sup>TM</sup> (PfuU). Both enzymes successfully amplified the target DNA, however, the Phus seemed to outperform the PfuU in several occasions. (e.g. PFB0300c).

Performances of the Phusion<sup>TM</sup> and PfuUltra<sup>TM</sup> polymerases were further evaluated by comparing the efficiency of ligation. Positive amplicons, derived from the two different amplification reactions, were ligated into the pET160/GW vector. Subsequently, colonies were screened by colony PCR. Both ligation-products produced visible colonies, however, the Phusion amplified product generally produced more positive colonies than the PfuUltra counterpart. For example, when the PF11\_0224 amplicons were ligated and transformed, 8 out of 10 tested colonies containing the Phusion amplified product resulted positive. Whereas, for the PfuUltra product only 4 out of the 10 picked colonies were positive. Similar results were obtained when additional genes were assessed. Overall, the Phusion<sup>TM</sup> High-Fidelity DNA Polymerase



**Figure 3.12: Standard PCR Vs nested PCR.**

Five randomly selected targets were amplified using both standard (S) and nested (N) PCR amplification protocols. *PF13\_0275*, 478bp; *PF14\_0010*, 730bp; *PFC0090w*, 577bp; *PFE1600w*, 1552bp; *PFC0085c*, 550bp. DNA target length is expressed in base pair and includes both 5' and 3' additional primer sequences (CACC and c-MYC).

was shown to be the most convenient due to its high specificity and low cost.

As a next step, optimal conditions for PCR amplification were assessed using 5 randomly selected ORFs. Each gene was amplified with both a standard PCR procedure and with a nested PCR procedure (see Section 2.3.1). In both protocols the c-MYC C-terminal tag was added using two reverse primers (see Figure 2.2), whereas the two methods differentiated in the use of a unique forward (standard) or two forward primers (nested). Using the standard procedure, 1 (*PF13\_0275*) out of 5 genes gave a clearly visible band when run on an agarose gel. Three out of 5 resulted in a very faint band. One out of 5 was negative (*PF14\_0010*). Using the nested PCR protocol, 4 out of 5 genes produced a clear sharp band, and one resulted in a strong band with additional fragments/smear (*PF14\_0010*), see Figure 3.12.

The highest cloning efficiency was given by the nested PCR procedure, which was ultimately chosen. Additionally, it was observed that to guarantee the highest ligation efficiency and specificity, two extra steps of PCR purification had to be added after each round of amplification. These helped to avoid contamination of the second phase of the amplification from the first step PCR primers.

The overall applicability of the cloning strategy was assessed by performing a small pilot study on 36 genes. Using the procedure described above 97% (35/36) of the genes were successfully amplified. Amplicons were subsequently ligated into the expression vector and screened using a colony PCR assay. All of them produced positive clones, indicating that ligations were successful. As

for the DNA sequence quality, all the amplified targets showed a nucleotide sequence similar, if not identical, to the *PlasmoDB* reference database.

### 3.3.4 High-throughput cloning of 260 gene targets

Once the initial evaluation of the cloning procedure was completed, myself and two other PhD students started to implement the procedure in two different laboratories (UNIPG and MtM). Each member of the team was responsible for a specific set of candidates, consisting of  $\sim 150$  DNA portions each. Primers were designed as described in Section 2.3.1 and listed on an excel file shared over the network to allow multiple-site access and continuous up-to-date information. Overall, more than 1400 primers were designed, with an average of  $\sim 3$  primers per DNA target (i.e. 260 genes split into 448 DNA targets). The nested-PCR protocol was firstly optimised in terms of number of cycles, temperature, extensions time, DNA purification and reagent concentrations. To maximised the number of successful clones, DNA sources were utilised in a gene dependent manner. In particular, the first 3 amplification attempts were carried out using a mixture of the four cDNA libraries. If unsuccessful, a single cDNA library, selected according to the MS data of the gene expression level, was used to enrich starting DNA material. If this second strategy failed, increasing quantities of genomic DNA were used. Then, if this last amplification was unsuccessful, the candidate target was eventually discarded.

PCR protocols were shared through the network and implemented according to the instrumentation available in the each site. Where possible, attention was given to eliminate any uncontrollable variable. Track of reagents and batch to batch comparison were recorded. Notably, in two separate and independent occasions drop of cloning efficiency was observed between pET160/GW cloning kits. The causes of the efficiency drop were unknown, however, when the two kits were replaced by the supplier the efficiency rate went back to normality. So far, the high-throughput cloning activity successfully produced 394 amplicons, with a PCR success rate of 88% (394/448). Only 8.5% of candidates (38/448) failed to be correctly amplified. Of these, 32 resulted negative at least on 3 separate PCR attempts, even when candidate-specific conditions were used. Whereas, for 6 of the 38 negative targets, it has been impossible to design a suitable pair of primers due to the high content of AT nucleotides and long repeated regions. Consequently all these targets (i.e. 32+6) were discarded and depleted from the cloning list. Currently, 16 clones (16/448, 3.5%) resulted negative during the initial amplification round. Further attempts will

be carried out using target-optimised conditions.

As for the ligation efficiency, 325 of the 394 amplified targets were successfully inserted into the pET160/GW vector, providing an efficiency rate of 82.5%. Twenty-four DNA targets (24/394, 6%) failed for three times to ligate and were discarded. For 45 amplicons (45/394, 11.5%) only one unsuccessful attempt was carried out. These targets will eventually be reassessed together with the 16 PCR-negative candidates. Notably, no noticeable impact on the ligation efficiency was observed when the quantity of reagents in the pET160/GW kit were halved.

Eventually, the efficiency of the entire cloning procedure was assessed by sequencing the clones successfully ligated (n.325). This final step enabled us to verify not only the correct identity and quality of the DNA target, but also the specificity of the designed primers and the efficiency of the Phusion<sup>TM</sup> High-Fidelity DNA Polymerase, which, together with the PCR procedure and the ligation reaction, were at the core of this cloning activity. Purified plasmids were sequenced using the 3730 XL DNA Analyzer ABI's sequencing platform, which allowed us to fully sequence (i.e forward, reverse and if necessary internally) up to 32 plasmids per plate, for a total of 96 parallel reactions. Sequence analysis was carried out manually. Each plasmid sequence was analysed for DNA identity and presence of mutations, insertions and/or deletions. Considering that even the most up-to-date database contains errors, wrongly annotated genes and unique sequences for highly-polymorphic genes, the acceptance criteria used were: either 100% sequence homology or limited point mutations and/or in frame insertions/deletions (see Section 2.3.6). As today, 269 out of the 325 (83%) DNA targets showed a sequence identical or similar to the database. Four plasmids (4/325, 1%) were discarded, as their content failed to be correctly sequenced in at least 3 separate occasions using 3 different plasmid clones. A common reason for rejection was the presence of either an insertion and/or a deletion in poly-adenosine regions, which eventually resulted in the presence of a stop-codon in the coding sequence. For 52 plasmids (52/325, 16%) the DNA sequence could not be immediately accepted due to a) low quality or absence of internal coverage; b) poor sequencing reaction with a general low quality coverage; c) ambiguous sequence coverage with possible frame-shift in highly-repeated regions or long poly-adenosine sequences; d) problems with primers used during the amplification or mutations on the c-MYC C-terminal tag. Of these 52 problematic plasmids, 24 will be directly re-sequenced using either the same clone or a new one from a different colony. Whereas the remaining 28 vectors will be added to the 16 negative

PCR-targets and 45 failed-ligation products. They will form a group of problematic DNA targets that will eventually be reassessed before the end of the FightMal project. A summary of the cloning activity is described in Figure 3.13.

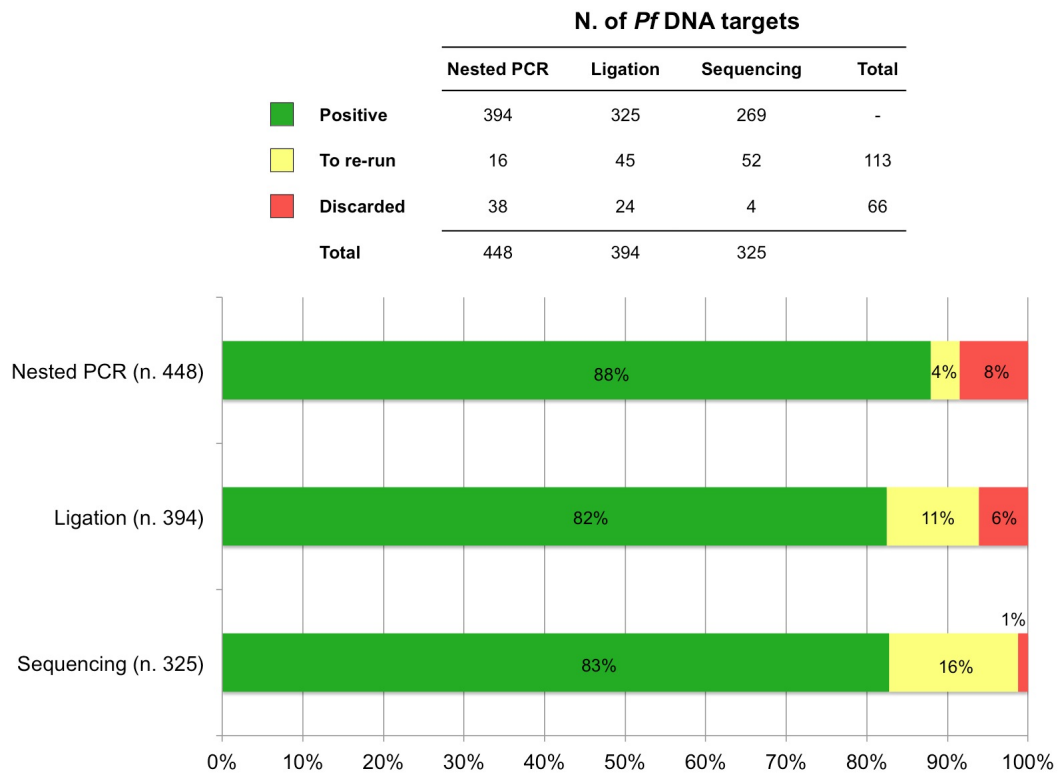


Figure 3.13: Overall results of the cloning activity.

### 3.3.5 Discussion

In parallel with the bioinformatic analysis, a number of feasibility studies were undertaken to establish an efficient cloning system. At first, an appropriate cloning platform was investigated. From those commercially available, Invitrogen's Gateway<sup>®</sup> was identified as the most suitable. It is based on bacteriophage lambda recombination into bacterial genome and it overcomes the difficulties of large-scale traditional cloning. Indeed, the system allows the flexible interchange of the insert between different vectors of different downstream applications or expression systems without the need of re-sequencing the construct. It is envisaged that such flexibility will come to hand if we will need to further characterise any of newly discovered *Pf* antigens, or if there will be the need to shift some of the challenging targets into a different expression platform, an operation that could be quickly achieved by the use of automation.

As for the selected Gateway<sup>®</sup> plasmid, pET160/GW/D-TOPO<sup>®</sup>, it also allows a quick (5 min) directional ligation, eliminating the need for ligation enzymes, and pushing the productivity to a medium (e.g. up to 10-20 ligations per day per operator) or high-throughput levels (e.g. 48-96 ligations per day using a robotic workstation). Additionally, the pET160 plasmid possesses a few features, which were essential for our project (see Section 3.3.1): a *T7lac* promoter for high-level IPTG-inducible expression in *E. coli* and two N-terminal tags, 6xHIS and Lumio<sup>™</sup> for specific detection/purification of the recombinant protein. It also contained a low copy number origin of replication, the pBR322 that allows the synthesis of  $\sim 40$  copies of plasmid per cell (1/10 of a standard high copy number origin of replication, Lee et al. (2006)). This, although it increases the plasmid stability during the propagation, created a few problems in the cloning step. Indeed, the yields obtained from the plasmid expansion and purification were very low ( $< 10\text{-}20 \mu\text{g}/\text{mL}$ ), and many times it caused the failure of the sequencing reaction. Although very rich in features the pET160 plasmid backbone was modified to contain an additional C-terminus tag. This was considered an essential element to have because it will enable the detection and normalisation on the microarray of full-length expression products. However, having already two small N-terminal tags we did not want to further affect any relevant characteristic or function of the protein target. Therefore, of the investigated tags only the c-MYC tag fulfilled our selection criteria. It was chosen for its relatively short length (11 AA), its hydrophilic properties (negative GRAVY value), its simple detection with a wide range of

commercial monoclonal and polyclonal antibodies, and its widespread use in a variety of different expression systems (bacteria, yeast, insect and mammalian cells, (Terpe, 2003)).

After selecting the cloning platform and plasmid, the attention was shifted towards the selection of a polymerase enzyme suitable for the amplification of the *Pf* coding regions. Indeed, the recent elucidation of *Pf* entire genome showed how this organism tends to be extremely rich in AT content. The AT average composition is over 80% and it can rise up to 90% in the intron and intergenic regions, while dropping to 60-70% in coding regions (Gardner et al., 2002). Consequently, PCR amplification of *Pf* genes is challenging (Moll et al., 2008; Doolan et al., 2008). In addition, the *Pf* parasite possesses a high number of polymorphisms and genetic variability. Hence, to avoid any further artefacts or sequence alteration during the PCR amplification, the use of a proofreading polymerase is required. On this bases, using two selected ORFs I assessed the performance of a panel of three proofreading polymerase enzymes: (Pfx50<sup>TM</sup>, Invitrogen; Phusion<sup>TM</sup> High-Fidelity, Finnzymes; PfuUltra<sup>TM</sup> II Fusion HS, Stratagene). While Invitrogen's polymerase had difficulty amplifying the correct target, (generating multiple bands and smears), the Phusion<sup>TM</sup> and the PfuUltra<sup>TM</sup> polymerases performed well, amplifying bands of the correct size. The PfuUltra<sup>TM</sup> polymerase, however, generated similar results to the Phusion<sup>TM</sup> polymerase, which is less than half of its price tag. So the Phusion<sup>TM</sup> High-Fidelity DNA Polymerase was ultimately chosen for its overall performance and low cost. It is envisage, however, that PfuUltra<sup>TM</sup> II Fusion could eventually be utilised to rescue those problematic genes that presented highly repetitive regions or in rich in AT content, as it has a very low error rate in the DNA processivity.

Using the Phusion<sup>TM</sup> High-Fidelity polymerase a suitable PCR set-up was sought. In particular we looked for a protocol with high efficiency and specificity but that was also capable of adding a C-terminal tag to the amplicons. The insertion of the c-MYC tag was successfully achieved using a double amplification step. This procedure was chosen over the single amplification to successfully add the relatively large C-terminal c-MYC together with its spacer and stop codon. As a consequence, products were generated using gene-specific primers that incorporated DNA encoding for the 4 base pairs necessary for directional cloning and the C-terminal tag+spacer (40 nt). When both standard and nested PCR protocols were tested in parallel with a panel of five antigens, the latter proved to be optimal with a success rate of 100%. The essential difference in the two protocols was that in one the ORFs were amplified using



an identical forward primer in both amplification rounds (standard), while two different forward primers were used in the nested protocol. In particular, the second forward primer was annealing to a sequence down-stream the annealing site of the first forward oligonucleotide (see Section 2.3.1). This approach produced not only a higher quality of the amplicons with clear, sharp bands but it was also able to improve the amplification efficiency, rescuing several ORFs that were difficult to amplify with the standard PCR. These improvements can be explained by the fact that in the nested PCR the probability that an incorrect region is mistakenly amplified is very low thanks to the two forward primers, which increase the coverage of DNA target sequence. This results in a higher specificity, avoiding the waste of nucleotides and enzymatic activity on unspecific products. Such capacity is even more important when a genome rich in AT content and repetitive regions like the *Pf* one is used. Presence of the target bands however is not enough to fully prove the overall quality of the procedure. Indeed, when the nested PCR was initially tested on a small scale, although the majority of the targets resulted positive, only 42% of these possessed a full length c-MYC sequence (data not shown). We hypothesised that the problem was caused by a contamination during the second round of amplification from the primary reverse primer, which contains only a partial c-MYC sequence (see Figure 2.2). This created a competition with the secondary reverse primer (containing the second half of the c-MYC), which favoured the production of the shorter version of the amplicon with only half of the c-MYC encoding sequence. The issue was solved by adding a PCR purification step between the two runs of amplification. The suitability of this final procedure was ultimately confirmed when 98% (35/36) of a panel of candidates were successfully amplified and cloned into the expression vector.

Optimal experimental conditions and cloning procedures were then implemented into the production line, which consisted in the employment for 18 months of three PhD students (including myself) mainly based at MtM. The cloning activity, which aimed to screen and, when possible, produce plasmids containing the 448 entries for CHIP 2, is still undergoing. When this thesis went in press, 60% of the list candidates (269/448) had successfully been cloned. During the first round of cloning 88% (394/448) of the entries were correctly amplified. Due to the high level of hypothetical/putative proteins present in the list (178/250, 71.2%), it is expected that out of the 54 negative PCR products some of them will be caused by the fact that the targeted sequence was representing a pseudogene and is therefore not present in our cDNA library. In fact, I believe that 32 of these, which failed to be ampli-

fied after 3 or 4 PCR attempts (using the gene-specific optimised condition), are likely to be either dysfunctional genes or sequences wrongly annotated in the database (used as reference during the primers design). Six out of the 54 entries were rejected during the DNA sequence analysis due to their highly repetitive regions, which prevented us from designing primers unique enough to guarantee a specific amplification. Of the remaining 16 negative genes, some may have complex secondary DNA structures and are therefore not easily accessible for the polymerase enzyme. Others genes may possess relatively high repetitive regions, which leads to the amplification of unspecific amplicons. On this bases, the real amplification efficiency could be raised from 88% to a maximum estimation of 96% (394/410). A percentage very similar to the one observed during the initial pilot study (98%) and in concordance with the 96% efficiency observed by Doolan et al. during the amplification of 960 *Pf* targets.

During the initial amplification run, 360 (80%, 360/448) clones were successfully obtained using either cDNA or genomic DNA. Negative entries were then re-assessed using both a gene dependent DNA source and a higher concentration of  $MgCl_2$ , which could increase the polymerase processivity. Customisation of the DNA source was accomplished by enriching the cDNA mix with the specific library in which the gene is supposedly expressed (according to MS and MA data). At times the bias towards a certain library was two to four times the standard DNA concentration. Possible inhibitory effects due to a high DNA concentration were minimised by maintaining the overall DNA amount at the same level. Such modifications in the amplification procedure enabled us to rescue 34 candidates, bring the overall number of available amplicons to a total of 394.

The implementation of the ligation protocol was relatively straight forward. The insertion of the amplicons into the linearised pET160 vector was achieved via a 5 min incubation on ice. This helped to reduce the length of the procedure and increase its high-throughput. In less than 90 min, 10 to 20 PCR products can be efficiently ligated and transformed into *E. coli* strain (including the plating), saving time and effort. Compared to a standard ligation protocol, where both DNA and vector have to be individually cut to produce complementary ends and ligated by DNA ligase, this method could at least halve the time and the money spent in the procedure. Indeed, although more expensive than a standard circular Gateway<sup>®</sup> plasmid, pET160/GW/D-TOPO<sup>®</sup> can efficiently be ligated with our DNA target following a specific orientation even when the ligation reagents volumes were halved (from 6  $\mu L$  to 3  $\mu L$ ). This allowed us to save a relevant amount of money considering the

fact that each kit is sold for 20 ligations with a price tag of £520, and that throughout the cloning activity we have probably performed over 500 ligation attempts. Additionally, we successfully showed that the colony PCR technique is a reliable tool to quickly assess the success of ligations, the identity of the amplicon and the integrity of the c-MYC tag. Indeed, using a universal reverse colony primer, which is able to anneal both the second half of the c-MYC tag and the vector sequence, we were able to amplify only those clones correctly orientated in the plasmid but also containing the right amplicon (right size) and the full length c-MYC.

Overall, 325 out of the 396 (82.5%) PCR products were successfully ligated into the pET160 vector. Although, the value falls short when compared to the initial feasibility study (100%) or the work of Doolan et. al work (97%, Doolan et al. (2008)), it is not far behind the other ones found in the literature (79%, Mehlin et al. (2006) or 84%, Aguiar et al. (2004)) or in the supplier's manual (~90%). In particular, it is interesting to note how our ligation efficiency is far more similar to the one obtained by Aguilar using the Gateway<sup>®</sup> platform to clone plasmidial genes than the one supplied by Invitrogen. This can probably be explained by the fact that plasmidial genes have peculiar DNA sequence that are more prone to form secondary DNA structures, which can impede the correct ligation and orientation of the insert into the linearised plasmid. Following this assumption, attempts to disrupt these structures and linearise the cDNA were made by heating up the PCR products to 90°C for 5 min. Additionally, different ratios insert:vector (from 1:1 to 4:1) were tested. However, neither of the two approaches helped to improve the ligation efficiency. Similarly to the PCR approach, a cut-off of 3 ligation attempts was applied to determine which of the challenging amplicons had to be discarded. So far, 24 inserts failed to be inserted into the pET160 vector in 3 or more ligation reactions. For the remaining 45, there is still a possibility of running one or more additional test. Considering that not all of these will be able to produce a usable clone, however, it is estimated that the final ligation efficiency can reach between 82.5% to 94%.

Up to 3 colonies for each of the positive clones were then grown, the DNA was purified and sent to sequence. The analysis aimed to uncover possible alteration of the target DNA (e.g. insertions, deletions, alternative splicing), which in some of the cases led to the rejection of the clone. Out of the many sequencing service providers evaluated for the sequencing service, Beckman Coulter Genomics resulted the most cost-effective and reliable. They employ a DNA Analyzer platform, which in our case (AT-rich sequences) was capable of

sequencing up to 1000bp per reaction, with an average reading capability approaching the 600-700bp. Overall, a relatively high percentage (67%, 218/325) of entries showed 100% homology with their *PlasmoDB* database-counterpart obtained at the time of the bioinformatic analysis. This was considered a satisfactory result, considering the high level of polymorphisms and genetic variability present in the *Pf* genome. A high level of DNA variability was visible for 51 candidates, which all possessed either small point mutations or limited in-frame insertion/deletion (up to 30 AAs). Only a minority of the clones (17.2%, 56/325) contained considerable DNA alterations that forced us to temporarily discard them. Overall, although the the sequencing verification felt short in comparison to the one observed by Doolan et al. (>99%, Doolan et al. (2008)), it correlates very well with the study carried out by Mehlin et al. (67% Mehlin et al. (2006)), which also highlights the high rate of mutations and insertions present despite the use of a proofreading polymerase. Further assessments will be required to establish if these mutations are artefacts of the cloning system or correct sequences that differ from the database due to the high degree of genetic diversity of the parasite (see Section 1.2.1). At least 4 DNA targets were rejected due to the insertion of 1 or 2 adenines in long repeated AT-rich sequences (e.g. 8-10 or even more consecutive adenines), which are mainly caused by the failure in the processivity of the polymerase. Additionally, entire introns were found in a number of rejected sequences, even when cDNA was used as DNA source. A plausible cause of this discrepancy with the original database sequence could rely on a possible mistake in the annotation of the gene or the presence of alternative splicing forms. So far, 269 (82.7%, 269/325) candidates resulted ready for the expression.

In conclusion, although the percentage of correct clones obtained during the scale-up production of the 448 candidates was lower than the one observed during the initial feasibility study, overall the cloning efficiency was still satisfactory and similar to the one described in the literature. The discrepancy observed between the initial feasibility study and the subsequent scale-up production can be at least in part attributed to the higher number of hypothetical proteins within the 260 gene list (68%, 178/260) in comparison to those in the feasibility study (50%, 14/28). It is predicted that we will be able to rescue some of the 113 outstanding clones. In particular, we are expecting to recover most of the 52 clones, which were initially put aside due to some incongruence in DNA sequences, and some of the ones that were not successfully ligated into the vector (n.45). If this estimation will be confirmed during the second round

of cloning (currently in progress at UNIPG), the final cloning efficiency could possibly rise to  $\sim 90\%$ , which is very similar to the ones found in the literature (Doolan et al., 2008; Aguiar et al., 2004). Following the same estimation, we are setting the current target for the completion of the cloning activity to 310-330 clones.

## 3.4 *In vivo* expression of *Pf* candidates in *E. coli*

### 3.4.1 Selection of the optimal *in vivo* expression platform

#### Evaluation of the most suitable *E. coli* strain

The heterologous production of *Pf* proteins in *E. coli* is known to be challenging due to the evolutionary divergences of the two organisms. In particular, there are two main reasons correlated with the challenging expression of soluble *Pf* proteins. The first is the bias towards the use of certain tRNAs, that are abundant in *Pf* and rare in *E. coli*. The second challenge is to express exogenous proteins to *E. coli* that are correctly folded. Generally, the engineering of new recombinant bacterial strains has proven to be a useful tool to address similar issues. Indeed, some of the new genetically modified strains available in the market allow the use of rare tRNA codons commonly used by *Pf* or the enhancement of disulphide bond formation, which is needed for the correct folding of the protein.

Hence, the first objective was the evaluation of which of the available *E. coli* strains was the most suitable for the production of *Pf* proteins. To do so, a list of five different *E. coli* strains was initially created: BL21 Star<sup>TM</sup> (Star), BL21 BLR (BLR), BL21 Codon Plus RIL<sup>®</sup> (RIL), Rosetta-Gami<sup>TM</sup>2 (RG2) and Rosetta-Gami<sup>TM</sup>B (RGB). A summary of the strains main features is available in Table 3.10.

The expression of *Pf* antigens in the five different bacterial strains was evaluated by WB using a total of 77 *Pf* candidates. Initially, the Star and RGB strains were evaluated on a small scale at UNIPG. Of the five clones assessed none was expressed in Star while four out of five were expressed in RGB. In parallel, my colleague Florence Wolff at PX' performed a small pilot study in bacterial strains that were not available at UNIPG: BLR, RIL and RG2. Her results showed an expression efficiency of 1/3 in BLR, 2/4 in RIL and 11/22 in RG2. An additional larger study on the Star strain was

Strain	Derivation	Features
Star <sup>TM</sup>	BL21	It is a widely used strain for protein expression. It increases mRNA stability and it is deficient in the <i>lon</i> and <i>ompT</i> proteases.
BLR	BL21	It improves plasmid monomer yield and target plasmid stability. It is deficient in the <i>lon</i> and <i>ompT</i> proteases.
Codon Plus RIL <sup>®</sup>	BL21	It enhances the yield of proteins that contain codons rarely used in <i>E. coli</i> .
Rosetta-Gami <sup>TM</sup> 2	K-12, Origami <sup>TM</sup> 2 and Rosetta <sup>TM</sup> 2	It enhances both disulfide bond formation and the yield of proteins that contain codons rarely used in <i>E. coli</i> .
Rosetta-Gami <sup>TM</sup> B	BL21, Turner <sup>TM</sup> , Origami <sup>TM</sup> , Rosetta <sup>TM</sup>	It allows to adjust and fine tune the protein expression level. It enhances both disulfide bond formation and the yield of proteins that contain codons rarely used in <i>E. coli</i> .

**Table 3.10: Panel of *E. coli* strains evaluated for the heterologous production of *Pf* proteins.**

A selection of five common *E. coli* strains was obtained via a literature search. The strain column indicates the commercial name of the host. The derivation column shows the original host background from which the strain was obtained. The features column highlights the key elements/capabilities of each bacterial strain.

also performed at PX'. This evaluation produced an overall efficiency rate of 26 expressed targets out of the 37 tested. To consolidate these preliminary findings we evaluated, at PX', the expression of *Pf* clones on a larger scale using the bacterial host showing the highest success rate, RGB. The overall results from the initial small scale expression and the subsequent larger scale expression showed an expression efficiency rate of 33% in BLR, 50% in RLI, 50% in RG2, 62% in Star and 72% in RGB (Table 3.11).

Similar trends in the strain performance were observed when the expression results of specific plasmids were directly compared between the different hosts. In particular, of the 77 plasmids used in these studies, 18 were expressed in Star, RGB and RG2, namely the three strains most used. Star produced an expression rate of 56% (10/18), not far from the original 62%. RGB expressed 78% (14/18) of the clones, a value slightly higher than the initial 70%. A similar improvement was visible on the RG2, which showed an efficiency of 61% (11/18). Remarkably, only in two occasions the clones positive in RG2 resulted negative in RGB; whereas in one occasion a clone positive in Star was negative in RGB, indicating that the RGB was outperforming both RG2 and Star.

Culture conditions of the different strains were those previously optimised in the specific laboratories where the expression was carried out (see Section

Strain	N. of expression	Expression Pos/Neg	Location
Star <sup>TM</sup>	5	0/5	UNIPG
	37	26/37	PX'
	<b>42</b>	<b>62%, 26/42</b>	<b>Overall</b>
Rosetta-Gami <sup>TM</sup> B	22	17/22	UNIPG
	66	46/66	PX'
	<b>88</b>	<b>72%, 63/88</b>	<b>Overall</b>
BLR	3	33%, 1/3	PX'
Codon Plus RIL <sup>®</sup>	4	50%, 2/4	PX'
Rosetta-Gami <sup>TM</sup> 2	22	50%, 11/22	PX'

**Table 3.11: *E. coli* strain evaluation studies.**

To define a clear expression efficiency rate of the selected bacterial strains two independent studies were carried out at UNIPG (2xYT medium) and PX' (AIM medium).

2.4.3). Consequently, comparisons between the expression efficiency of the strains might, in part, also be influenced by the culture conditions. However, results clearly show that Star, RG2 and RGB outperformed the other strains, with RGB showing the highest expression efficiency. As a consequence RGB was selected as the strain of choice and culture conditions of this strain were optimised (see following section).

### Selection of the optimal expression medium

In addition to the host capabilities, the expression conditions are also important for an optimal production of recombinant proteins. Therefore, two studies were carried out at PX' and UNIPG to assess which were the best culture conditions for both protein synthesis and folding. In each study, the initial selection of the expression protocol depended on previous internal studies carried out either at PX' or UNIPG. On these bases, at PX' 66 plasmids were tested mainly with the AIM medium, however, two additional media were also tested (i.e. LB and TB). Whereas, 22 plasmid were tested at UNIPG using the 2xYT medium.

At PX', after the initial evaluation of 16 constructs expressed in LB, TB and AIM by F. Wolff, in which the AIM outperformed the other two (see Table

3.12), I decided to assess the validity of the AIM using a protocol previously optimised in the company, in which the bacteria were inoculated and incubated at 37°C for 24h. Using these procedures only one noticeable difference was

Target (n.16)	pI	Size	Coomassie			WB			Best Medium
			LB	TB	AIM	LB	TB	AIM	
<i>PF14_0010</i>	5.3	32.4	+	+	+	+	+	+	LB/TB/AIM
<i>PFE0395c</i>	6.8	39.7	-	-	-	+/-	+	+	TB/AIM
<i>PF13_0196</i>	5.3	46.3	-	-	-	-	+	+	TB/AIM
<i>PF14_0075</i>	5.2	49.6	-	-	-	+	+/-	+	LB/AIM
<i>PFA0300c</i>	7.1	51.1	-	NA	-	+	NA	+	LB/AIM
<i>PF11_0509_p1</i>	5.4	48.7	-	-	-	-	-	+	AIM
<i>PFB0100c_p2</i>	9.4	44.9	-	-	+	-	+	+	AIM
<i>PF08_0137_p2</i>	4.6	53.3	-	-	+	+	+	+	AIM
<i>PFC0120w_p4</i>	9.0	37.1	-	-	-	+/-	-	+	AIM
<i>PFI0265c_p1</i>	6.2	44.0	-	-	-	-	-	+	AIM
<i>MAL7P1.170</i>	7.1	39.7	-	-	-	-	-	+/-	AIM
<i>PFE0040c_p3</i>	4.5	47.0	-	-	-	+	-	+/-	LB
<i>PF14_0201_p2</i>	4.5	60.8	-	-	-	-	-	-	NE
<i>PFI1785w</i>	5.7	45.3	-	-	-	-	-	-	NE
<i>PFI0265c_p2</i>	8.9	26.5	-	-	-	-	-	-	NE
<i>PFB0345c_p3</i>	5.6	32.8	-	-	-	-	-	-	NE

**Table 3.12: Expression medium PX' study: LB vs TB vs AIM.**

The table summarise the initial study carried out in the PX' facilities, during which AIM resulted the most performing formulation out of the three tested. pI: Isoelectric point. Size: Protein size expressed in kDa. Coomassie: Coomassie Brilliant Blue staining positive (+) or negative (-). WB: Western blot analysis positive (+) or negative (-). NE: Not expressed.

recorded between the AIM medium, LB and TB in terms of bacterial growth. It was observed that the final OD<sub>600</sub> of the cultures in AIM was significantly higher (5-6 OD<sub>600</sub>) than the same bacteria grown in LB or TB (1-2 OD<sub>600</sub>). As part of the evaluation of AIM, I also optimised a 48h expression-analysis protocol which enabled me to screen 12-24 constructs in parallel. Overall, I expressed 66 plasmids using this procedure (i.e. AIM, 24h at 37°C). Forty-six out of 66 plasmids were expressed resulting in an expression efficiency of 70%. Of these, 6 expressed in very low quantities and it was not possible to assess their solubility by WB analysis. Of the remaining 40 candidates, 6 were over-expressed and were soluble, 7 were mainly soluble while 27 were insoluble, resulting in 32% (13/40) of the expressed proteins being soluble, Table 3.13.

In terms of productivity, it was possible to express and analyse up to 24-36 proteins per week, with a possible projection of 10-15 weeks for a complete



mini-scale expression (4 mL) analysis of 300-330 plasmids using only one operator. However, as expected this mini-scale protocol produced only limited quantities of recombinant proteins (<10-20  $\mu\text{g}$ ) which were not sufficient for the subsequent purification step (see Section 3.6). Consequently, I implemented a large-scale version of the procedure consisting of 1-2 L of culture per clone. Overall 8 clones were successfully expressed, producing between 100 to 500  $\mu\text{g}$  of recombinant protein.

At UNIPG, we successfully expressed 77% (17/22) of plasmids tested. The *E. coli* culture showed a stable and fast growth at 37°C. Bacteria lag phase lasted around 2-3h, while the log phase was eventually stopped after 3h following induction, which was carried out when bacteria reached 0.8/1.0 OD<sub>600</sub> using IPTG. Final culture OD<sub>600</sub> readings were in the range of 1 to 2 OD<sub>600</sub>, which was similar to the one observed with LB and TB at PX', but still lower than the one obtained with the AIM. Of the 17 plasmids successfully expressed, 10 showed a high degree of solubility (soluble  $\geq$  insoluble), 4 resulted mainly insoluble (insoluble > soluble) and 1 was completely insoluble. For the remaining 2 constructs solubility values were not available due to inconsistencies in the WB data. Overall, in terms of 2xYT solubility performance, 67% (10/15) of the successfully expressed plasmids produced soluble or partially proteins, see Table 3.13.

Medium	Sol > Ins	Ins > Sol	Low Exp./Neg.	Total	Solubility (%)	Location
AIM	13	27	6/20	66	32% (13/40)	PX'
2xYT	10	5	2/5	22	67% (10/15)	UNIPG

**Table 3.13: Expression medium evaluation studies.**

The table summarise the two medium evaluation studies carried out in the UNIPG and PX' facilities. The protein solubility was established via Coomassie and WB analysis. Sol > Ins: the intensity of the protein band is higher in the soluble fraction compared to the insoluble one. Ins > sol: protein band mainly present in the insoluble fraction. Low Exp.: indicates the presence of a very faint/dubious band in WB. Neg.: negative, no band visible in WB. Solubility (%): percentage of Sol>Ins / (Sol>Ins + Ins>Sol).

At completion of these preliminary studies I rejoined the MtM team. Here, together with Dr. Trivelli and under the supervision of Dr. T. Dottorini, I re-analysed the expression efficiency data obtained so far (AIM/PX': expression rate = 70%, solubility level = 32%; 2xYT/UNIPG: expression rate = 77%, solubility = 67%). As the 2xYT medium resulted the most performing, we decided to look into the possibility of creating a new expression media based on the 2xYT medium, with the hope to improve both the protein expression

and solubility. The new formulation, named TM, is a minimal medium that combines the basic components of 2xYT with additional nutrients (e.g. metals, vitamins and bivalent cations). This high-energy formulation should enhance both protein synthesis and biomass values. The medium uses the same induction system of 2xYT (IPTG), and allows a sustainable bacterial growth for 18h at RT (i.e. 20-22°C). Together with metals, vitamins and bivalent cations, we included the addition of 20 AAs crucial for both the bacterial growth (AAs with high metabolic cost) and the synthesis of *Pf* proteins (AAs common in *Pf* and rare in *E. coli*). Using this new formulation we assessed the expression yield and solubility level of 8 proteins previously expressed in both AIM and 2xYT. The constructs were re-expressed using both 2xYT and TM. Expression values were also retrospectively compared to the AIM expression (see Table 3.14).

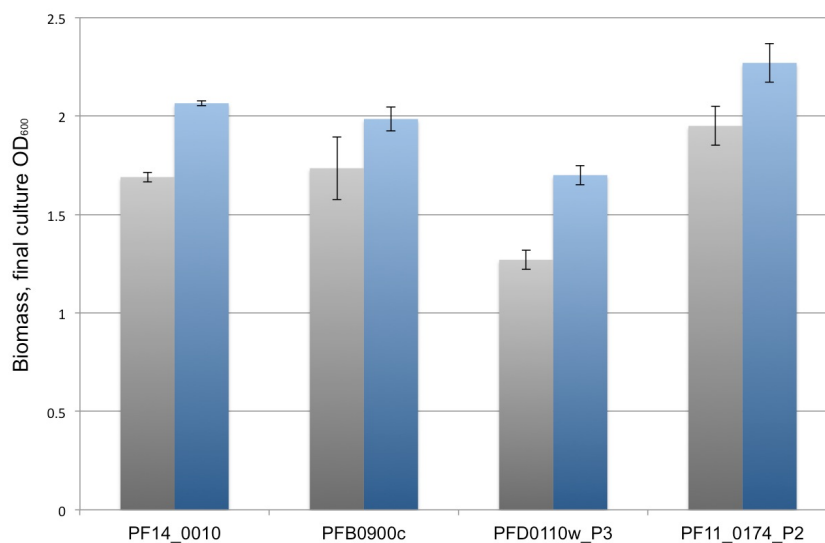
Target (n.8)	pI	Size	Coomassie			WB			Best Medium
			2xYT	TM	AIM	2xYT	TM	AIM	
<i>PF14_0010</i>	5.3	32.4	+	+	+	Sol>Ins	Sol>Ins	Sol>Ins	2xYT/TM/AIM
<i>PF13_0275</i>	7.8	23.1	-	-	-	Sol>Ins	Sol>Ins	Sol>Ins	2xYT/TM/AIM
<i>PFB0900c</i>	6.0	35.8	-	-	-	Ins>Sol	Ins>Sol	Ins>Sol	2xYT/TM/AIM
<i>MAL7P1.174</i>	8.6	39.0	+	+	-	Ins>Sol	Ins>Sol	Ins>Sol	2xYT/TM
<i>PFD1170c</i>	7.1	34.3	-	-	-	Ins>Sol	Sol>Ins	Ins>Sol	TM
<i>PFE0395c</i>	6.8	39.7	-	-	-	Ins>Sol	Sol>Ins	Ins>Sol	TM
<i>PFD0110w_p3</i>	5.6	65.7	-	-	-	Insoluble	Ins>Sol	Insoluble	TM
<i>PF11_0174_p2</i>	5.4	48.1	-	-	-	Insoluble	Ins>Sol	Insoluble	TM

**Table 3.14: Expression medium MtM study: final feasibility study 2xYT Vs TM Vs AIM.**

A summary of the medium evaluation study carried out at MtM. Performances of both 2xYT and TM media were evaluated. Results were retrospectively compared with the ones obtained at PX' using the AIM. pI: Isoelectric point. Size: Protein size expressed in kDa. Sol: Soluble fraction. Ins: Insoluble fraction. Coomassie: Coomassie Brilliant Blue staining. WB: Western blot analysis.

All the 8 plasmids were successfully expressed using the TM. More interestingly however, the TM outperformed both 2xYT and AIM media in at least four occasions. For the remaining four the new formulation matched the performances observed for the other two media. Bacteria biomass was generally higher when TM medium was used (average final OD<sub>600</sub> was 2.0 for TM against 1.6 for 2xYT, see Figure 3.14), but they were still lower than the ones previously observed with AIM.

As for solubility level, two of the soluble proteins produced in both 2xYT and AIM, remained soluble when expressed in TM. A noticeable improvement in terms of solubility was observed for 4 of the 6 remaining constructs. In-



**Figure 3.14: Biomass Comparison with 2xYT Vs TM.**

The histograms describe the final biomass values (OD<sub>600</sub>) of the RGB cultured in triplicates with 4 out of the 8 plasmids tested using either 2xYT (■) or TM (■). The selection of plasmids is representative of the general behaviour observed during the *in vivo* expression in the RGB strain. The error bars indicate  $\pm$  three standard deviations.

deed, although they were insoluble or mainly insoluble in 2xYT and AIM, these 4 proteins resulted soluble or partially soluble when expressed using the new medium. For the remaining 2 constructs no improvement was observed, however both maintained a high expression yield, resulting over-expressed (e.g visible in Coomassie) when expressed in 2xYT and TM. Finally, no negative effects on the bacterial culture were observed when the new medium was used. The improved solubility observed under the TM medium growth conditions was eventually confirmed by expanding the retrospective analysis on 35 plasmids, for which I observed a limited solubility level when expressed in AIM (see Table 3.15). In 69% of the cases (24/35) the TM outperformed AIM. For 20% (7/35) of the plasmids, the use of a different media did not produce any significant difference and in 11% of the cases (4/35) the AIM produced better results than the TM. The use of the new medium not only produced better solubility levels but in some cases it increased the expression yield. In particular, in 18 of the 24 cases in which TM outperformed AIM, protein solubility was increased. For the remaining 6, protein yield was improved.

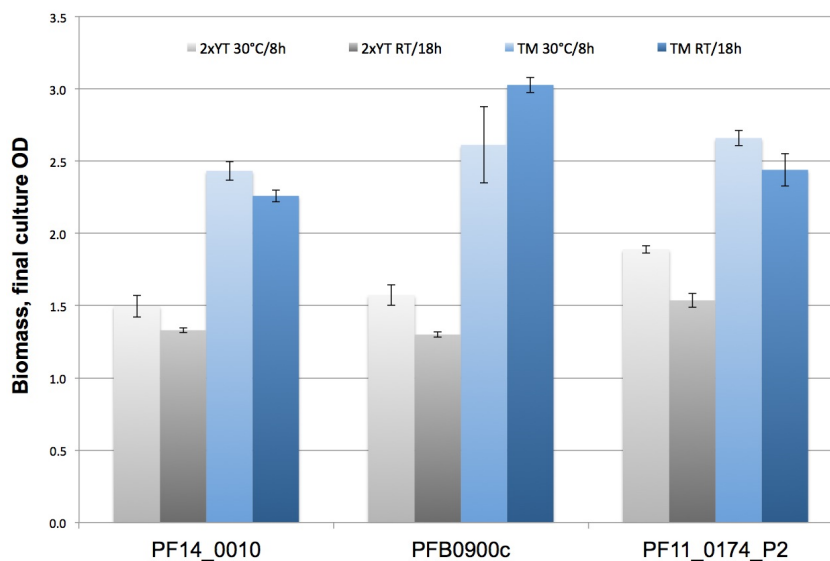
To investigate whether the enhanced solubility observed for TM in comparison with the 2xYT or the AIM was due to medium composition more than expression conditions (i.e. temperature 37°C for 2xYT versus 20-22°C for TM, and induction window, 3h for 2xYT versus 18h for TM), an additional trial comparing the two most performing media, TM and 2xYT, was carried

Target (n.35)	pI	Size	AIM			TM			Best Medium
			Coomassie	WB	Solubility	Coomassie	WB	Solubility	
<i>PF14_0010</i>	5.2	32.4	+	+	Soluble	+	+	Soluble	AIM/TM
<i>PF14_0018</i>	5.5	53.5	-	+	Insoluble	-	+	Insoluble	AIM/TM
<i>MAL7P1.177</i>	8.5	26.0	-	+	Insoluble	-	+	Insoluble	AIM/TM
<i>PF08_0137_p2</i>	7.8	60.4	+	+	Insoluble	+	+	Insoluble	AIM/TM
<i>MAL8P1.4</i>	4.4	54.2	-	+	Insoluble	-	+	Insoluble	AIM/TM
<i>PFI1370c</i>	8.7	43.2	-	+	Insoluble	-	+	Insoluble	AIM/TM
<i>PFL0070c_p1</i>	5.9	59.8	+	+	Insoluble	+	+	Insoluble	AIM/TM
<i>MAL7P1.172_p1</i>	6.3	60.7	-	+	Sol=Ins	+	+	Sol=Ins	TM
<i>MAL7P1.172_p2</i>	5.8	56.9	-	+	Insoluble	-	+	Ins>Sol	TM
<i>MAL8P1.6</i>	7.7	17.1	-	+	Insoluble	-	+	Sol=Ins	TM
<i>PF10_0021</i>	8.3	35.3	-	+	Insoluble	-	+	Sol=Ins	TM
<i>PF11_0037_p2</i>	5.7	42.0	-	-	NA	-	+	Insoluble	TM
<i>PF11_0503</i>	8.8	46.7	-	+	Sol=Ins	+	+	Sol>Ins	TM
<i>PF11_0509_p1</i>	5.4	48.8	-	+	Sol=Ins	-	+	Sol>Ins	TM
<i>PF13_0275</i>	7.8	23.1	-	+	Sol=Ins	-	+	Sol>Ins	TM
<i>PF14_0075</i>	5.2	49.6	-	+	Ins>Sol	-	+	Sol>Ins	TM
<i>PFA0125c_p1</i>	8.6	53.5	-	+	Insoluble	+	+	Insoluble	TM
<i>PFA0125c_p3</i>	4.4	59.3	-	+	Insoluble	-	+	Sol>Ins	TM
<i>PFA0135w</i>	5.5	37.7	-	+	Insoluble	-	+	Sol>Ins	TM
<i>PFA0300c</i>	7.1	51.1	-	+	Insoluble	+	+	Sol>Ins	TM
<i>PFB0090c</i>	6.6	50.2	-	+	Sol>Ins	+	+	Sol>Ins	TM
<i>PFB0900c</i>	6.0	35.8	+	+	Insoluble	+	+	Sol>Ins	TM
<i>PFC0075c</i>	4.5	34.8	-	+	Ins>Sol	-	+	Sol>Ins	TM
<i>PDF0240c</i>	8.0	46.6	-	+	Insoluble	-	+	Sol>Ins	TM
<i>PDF0295c_p1</i>	6.2	46.4	-	+	Insoluble	+	+	Insoluble	TM
<i>PDF1170c</i>	7.1	34.3	-	+	Ins>Sol	-	+	Sol=Ins	TM
<i>PFE0395c</i>	6.8	39.7	-	+	Ins>Sol	-	+	Sol>Ins	TM
<i>PFE1605w</i>	9.5	59.4	+	+	Insoluble	+	+	Ins>Sol	TM
<i>PFI0265c_p1</i>	6.2	44.0	-	+	Ins>Sol	-	+	Sol>Ins	TM
<i>PFI1445w_p1</i>	7.6	48.9	-	+	Insoluble	+	+	Insoluble	TM
<i>PFL0040c</i>	8.3	59.6	-	+	Insoluble	-	+	Sol>Ins	TM
<i>MAL7P1.174</i>	8.7	39.0	-	+	Ins>Sol	+	+	Insoluble	AIM
<i>PFB0100c_p1</i>	6.3	38.8	-	+	Sol=Ins	+	+	Insoluble	AIM
<i>PFF0615c</i>	8.2	39.6	-	+	Ins>Sol	-	+	Insoluble	AIM
<i>PF13_0196</i>	5.1	46.5	-	+	Insoluble	-	-	NA	AIM

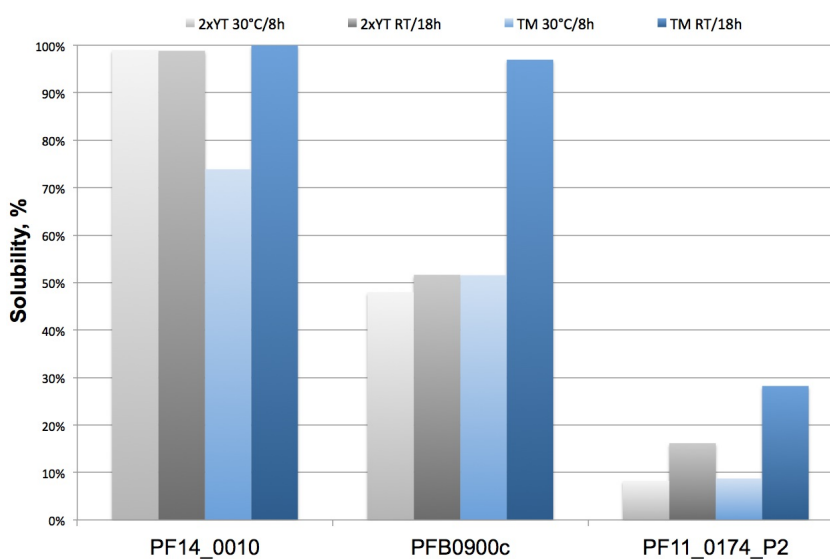
**Table 3.15: Expression medium MtM study, a retrospective evaluation of TM Vs AIM.**

The results of 35 proteins expressed in TM were retrospectively compared with the performances obtained at PX' using the AIM. Overall, the new medium formulation outperformed the AIM 69% (24/35) of the times. With only 4 plasmids performing better in AIM. Overall the TM matched or improved the solubility levels in 89% (31/35) of the cases. Additionally in 6 occasions the TM seemed to improved the protein yield, allowing the visualisation the newly synthesised protein on Coomassie. pI: Isoelectric point. Size: Protein size expressed in kDa. Sol: Soluble fraction. Ins: Insoluble fraction. Coomassie: Coomassie Brilliant Blue staining. WB: Western blot analysis.

out at different temperatures and induction conditions. The analysis was accomplished testing the performances of 3 plasmids in both 2xYT and TM at different temperatures (i.e. 30°C or RT) and for different induction phases (i.e. 8h or 18h). Plasmids selection was based on the performance of the constructs observed so far, as we wanted to test the widest possible array of behaviours (i.e. one over-expressed protein, one soluble protein and one insoluble protein). *PF14\_0010* was selected because it was synthesised as a soluble over-expressed protein independently from the expression conditions (2xYT, AIM and TM). *PFB0900c* was selected as it had previously produced partially soluble proteins in both 2xYT and AIM. Finally, *PF11\_0174\_p2* was selected as it had been insoluble in all prior expression conditions. Using the new conditions, it was observed that the use of TM improved both biomass productions (Figure 3.15a) and solubility levels (Figure 3.15b), independently from the temperature and induction conditions. This is clearly visible for *PFB0900c*, where both the biomass and the protein solubility double when bacteria were grown in TM compared to the 2xYT. As for *PF11\_0174\_p2*, although the improvement is smaller compared to *PFB0900c*, it was still noticeable. Indeed, it could lead to the production of enough soluble protein for the purification during the large-scale expression. As expected, no significant difference was observed for the solubility of *PF14\_0010*, a part from the lower percentage of solubility observed for the 8h at 30°C in TM. The biomass production of the *PF14\_0010* in TM produced a nearly 2-fold increase, confirming the general trend observed in the trial.



(a) Effect on biomass



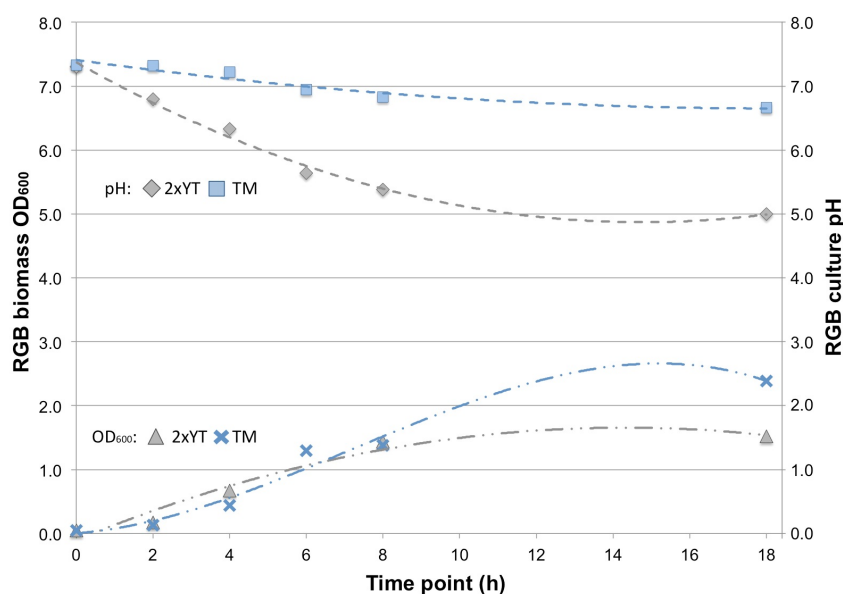
(b) Effect on solubility

**Figure 3.15: Influence of temperature and length of incubation on biomass and solubility, 2xYT vs TM.**

**a)** The histograms describe the final biomass values ( $OD_{600}$ ) of the RGB cultured in triplicates with 3 plasmids tested in triplicates using either 2xYT (■) or TM (■) and incubated for 8h (lighter color) or 18h (darker color). *PF14\_0010* is always expressed as a soluble over-expressed protein independently from the expression conditions (2xYT and TM). *PFB0900c* produced partially soluble proteins with both 2xYT and AIM. *PF11\_0174\_p2* produced insoluble protein in 2xYT and AIM. The error bars indicate  $\pm$  three standard deviations. **b)** The histograms describe the percentage of solubility for the three proteins, which represent the intensity of the soluble fraction / band intensity of the total extract observed during the WB analysis.

### Characterisation of the TM expression medium

The use of TM medium together with a low expression temperature (i.e. RT) were eventually chosen as the most suitable conditions for the heterologous production of *Pf* proteins in *E. coli*. However, we decided to further characterise the new medium before implementing it into the mini-scale screening of the over 300 available constructs. Between the many features of the new formulation its buffer capacity and the presence of the AA mixture were considered essentials. Indeed, acidification of the expression medium can cause growth inhibition and high metabolic stress. To avoid such event during the formulation of the TM medium we decided to include a potassium buffer system for the maintenance of a neutral pH. Correlation of culture growth and pH variation was assessed through the expression of *PF14\_0010* using both 2xYT and TM. The variation of pH was monitored every 2 h for the first 5 time points, the 6<sup>th</sup> time point was taken just before harvesting the cells after 18 hours. Together with the pH values, the biomass variation was monitored via the measurement of the optical density during each time point. The analysis of both pH and OD<sub>600</sub> value is described in Figure 3.16.



**Figure 3.16: Analysis of pH and biomass, 2xYT Vs TM.**

The *PF14\_0010* clone was expressed for 18h in both 2xYT (OD<sub>600</sub>:♦; pH:▲) and TM (OD<sub>600</sub>:■; pH:×) media. Every 2h (time points, X axis) an aliquot of culture was aliquoted and analysed for biomass content (OD<sub>600</sub>) and pH, Y axis.

A noticeable variation of pH between the two media was observed (TM  $\Delta_{pH} = -0.6$ ; 2xYT  $\Delta_{pH} = -2.3$ ). If on one hand the bacteria cultured in TM produced an 8.2% decrease of pH from 7.3 at 0h to 6.7 at 18h; on the other

hand the use of 2xYT resulted in a significant drop of pH (31.5%), with a reduction from 7.3 to 5.0 after 18h. Additionally, bacteria in TM showed a sustainable growth up to  $OD_{600}$  2.4, 1.6 times higher than the bacteria cultured in 2xYT ( $OD$  1.5), which reached the stationary phase soon after the pH was lower than 6.0.

The availability of AAs is also important for a successful heterologous expression, particularly when host (*E. coli*) and target organism (*Pf*) have different AA usage. Notably, highly expressed genes in *E. coli* manifest greater frequencies for less costly AAs, where the AA cost is expressed as "the number of high-energy phosphate bonds, -P, carried in ATP and GTP, plus the number of available hydrogen atoms, H, carried in NADH, NADPH, and FADH<sub>2</sub>, that would have been gained if the metabolite had remained in energy-producing pathways minus the numbers of these molecules gained before diversion", (Akashi and Gojobori, 2002). Interestingly, when we analysed the data from two separate studies in which the authors determined the AA usage frequencies for *E. coli* (Lobry and Gautier, 1994) and *Pf* (Chanda et al., 2005), we noticed an organism dependent AA usage. In particular, we found that *Pf* possesses a higher usage (1.4-fold average compared to the *E. coli*) of AA with a high metabolic cost, see Table 3.16 and Figure 3.17.

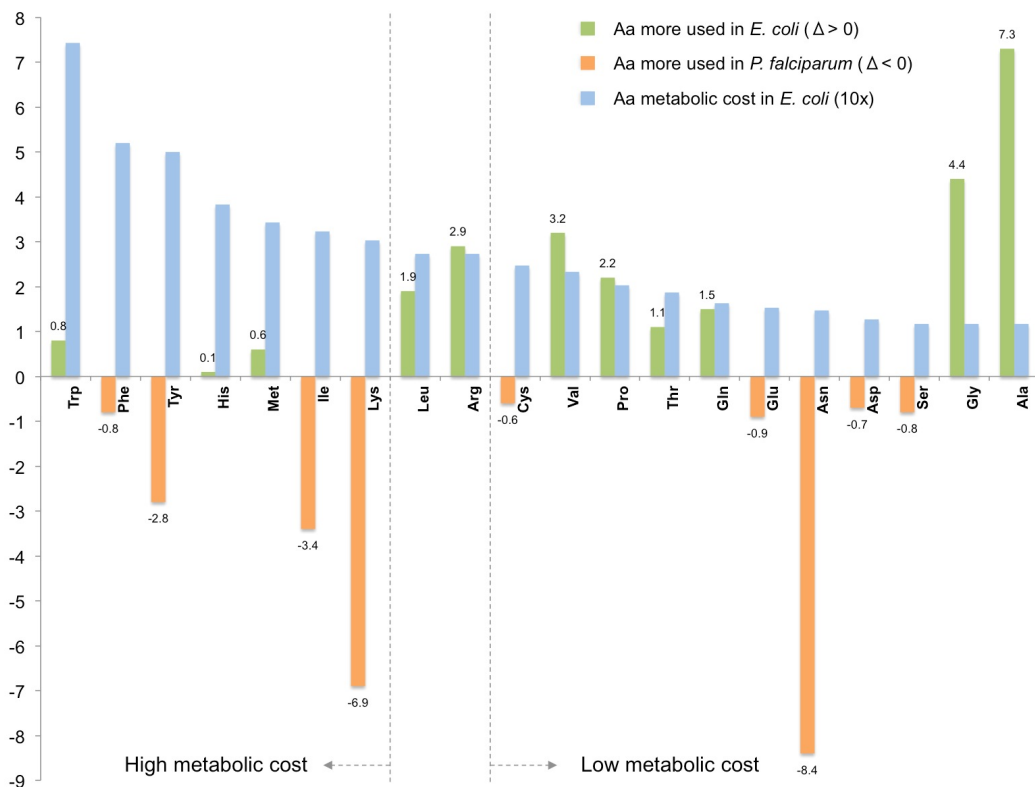


Figure 3.17: Amino acids usage *E. coli* Vs *Pf*.



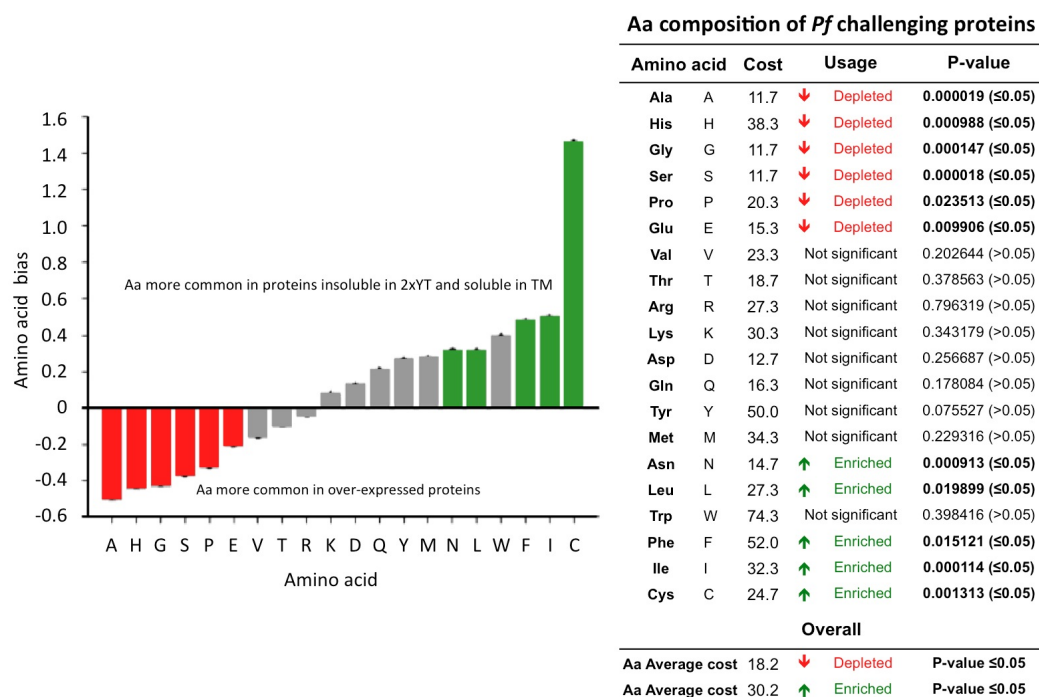
	AA	AA Cost	AA frequency (%±SD)		Δ%	Ratio
			<i>Pf</i>	<i>E. coli</i>		
	Theoretical	27.6	5.0	5.0	0.0	1.0
High metabolic cost	Trp	74.3	0.5±0.6	1.3±1.0	0.8	0.4
	Phe	52.0	4.6±2.0	3.8±1.6	-0.8	1.2
	Tyr	50.0	5.5±1.9	2.7±1.3	-2.8	2.0
	His	38.3	2.2±1.0	2.3±1.2	0.1	1.0
	Met	34.3	2.2±0.9	2.8±1.2	0.6	0.8
	Ile	32.3	9.3±2.4	5.9±1.9	-3.4	1.6
	Lys	30.3	11.6±3.2	4.7±2.3	-6.9	2.5
	Leu	27.3	8.1±2.2	10.0±2.7	1.9	0.8
	Arg	27.3	2.9±1.6	5.8±2.2	2.9	0.5
Low metabolic cost	Cys	24.7	1.8±1.0	1.2±1.0	-0.6	1.5
	Val	23.3	4.1±1.6	7.3±1.9	3.2	0.6
	Pro	20.3	2.2±1.3	4.4±1.6	2.2	0.5
	Thr	18.7	4.2±1.4	5.3±1.5	1.1	0.8
	Gln	16.3	2.8±1.2	4.3±1.8	1.5	0.7
	Glu	15.3	7.0±2.7	6.1±2.3	-0.9	1.1
	Asn	14.7	12.2±5.0	3.8±1.4	-8.4	3.2
	Asp	12.7	6.0±2.1	5.3±1.8	-0.7	1.1
	Ser	11.7	6.3±1.9	5.5±1.7	-0.8	1.1
	Gly	11.7	3.1±1.8	7.5±2.1	4.4	0.4
	Ala	11.7	2.4±2.0	9.7±2.6	7.3	0.2

**Table 3.16: Amino acids usage, *E. coli* Vs *Pf*.**

AA cost: Metabolic cost in *E. coli* for the AA synthesis. AA frequency (%): Average AA composition of an *Pf* protein (Chanda et al., 2005) or *E. coli* protein (Lobry and Gautier, 1994). Δ%: AA frequency of *E. coli* - AA frequency of *Pf*. Ratio: AA frequency of *Pf* / AA frequency of *E. coli*, >1 higher usage in *Pf*, <1 higher usage in *E. coli*. Average metabolic cost of an AA = 27.6. Theoretical frequency of use is 5.0%.

The AAs with the highest frequency of use (approximately 2-fold higher than the 5.0% theoretical use) in *Pf* were those encoded by codons with at least two A/T bases: Aparagine (Asn, GAT), Lysine (Lys, AAA) and Isoleucine (Ile, ATT). Notably, the same AAs do not have similar frequencies in *E. coli* (with an average use value close to 5.0%). Additionally, when we looked into the metabolic cost for each of the 20 AAs in *E. coli* (Akashi and Gojobori, 2002) we noticed that out of the 9 most costly AAs in the bacteria (average metabolic cost  $\geq 27.6$ ), 4 (Ile, Lys, Leucine (Leu) and Tyrosine (Tyr)) were used more than averaged in *Pf* (usage frequency  $\geq 5.0\%$ ), with Phenylalanine (Phe), which is the second most costly AA in *E. coli*, having a usage just below the threshold (4.6%). On these bases it is expected that proteins containing more costly AA will be expressed less efficiently than those proteins containing

less costly AAs, as the latter will put less pressure on the *E. coli* during their synthesis. This hypothesis was eventually confirmed when I looked at the AA composition of the over-expressed soluble proteins in 2xYT versus those with low-medium expression. From the *in silico* analysis I noticed that in the problematic proteins there was a significant bias towards the use of cysteines and AAs with a high cost (see Figure 3.18).



**Figure 3.18: Amino acid composition of *Pf* challenging proteins (insoluble in 2xYT and soluble in TM).**

The figure above describe the AAs enrichment and depletion patterns between a datasets of protein insoluble in 2xYT and soluble in TM and a dataset of proteins over-expressed in both media. The comparison highlights how the proteins insoluble in 2xYT possess a higher usage of costly AAs (e.g. enrichment of AAs with average metabolic cost of 30.2 and depletion of low-cost AAs) and are particularly rich of cysteines ( $\sim 1.5$  times more) than the proteins over-expressed in both media). The dataset of the insoluble (2xYT)  $\rightarrow$  soluble (TM) protein contains the unique sequences of 6 candidates for a total of 2,247 AAs. Whereas, the dataset of the over-expressed proteins contains the sequences of 11 candidates, for a total of 2,475 AAs. The two datasets were compared via Composition Profiler software ([www.cprofiller.org](http://www.cprofiller.org) Vacic et al. (2007)). The statistical significance p-value  $\leq 0.05$  is estimated using the two-sample t-test between the two dataset.

Consequently, to avoid any stress related to the expression of those plasmidial proteins we decided to directly supply the *E. coli* host with a customised mixture of AAs. The decision was based on the fact that most of the media commercially available use tryptone-peptone as an AA supply, which is a mixture of peptides derived from milk digested via trypsin. However, it is not possible to control or manipulate the AA composition of the medium other

than supplement it with an external AA mixture. During the of the in-house AAs solution, priority was given to both AAs with high metabolic cost in *E. coli* and to the AA highly used in *Pf*. All the remaining AAs were added at a lower concentration to increase the energy supply and decrease the overgrowth burden. Care was taken to avoid any inhibitory effect due to highly concentrated compounds. AA solubility issues were overcome via the preparation of two separated mixtures with the addition of HCl or NaOH (Gatewood Brown and Rousseau, 1994; Zumstein and Rousseau, 1989).

To assess the influence of AA availability and consumption in TM, we compared the performances of medium with and without the addition of an AA supplement. Both our internally formulated mixture and two commercially available solutions of essential and non-essential AAs (MEM Amino Acids Solution, Invitrogen UK) were tested. Notably, our mixture uses specific AAs at a concentration higher than the one present in the commercially available mixture (from twice to 70 times higher concentrations, see Table 3.17)

Three were the conditions tested: 1, TM without AAs; 2, TM with the two MEM AAs mixtures; 3, TM with our own AA solutions. A total of 7 plasmids were tested. Two already known to produce soluble proteins (*PF14\_0010* and *PFA0430c\_p3*), 4 from which we previously obtained partially soluble or insoluble proteins (*PFB0900c*, *PF08\_0006*, *PFD0110w\_p3* and *PF11\_0174\_p2*) and 1 that resulted negative in the previous tests (*PFL2505c\_p3*). Proteins were expressed for 18h at RT in two different expression runs. As shown in Figure 3.19 both bacteria biomass and protein solubility were influenced in most cases by the presence or absence of the AA supplements. In terms of biomass accumulation, the addition of the supplements increased the final OD<sub>600</sub> in all bacterial cultures, but OD<sub>600</sub> values were up to 2.5 times higher with internal AA mix (*PFA0430c\_p3*), Figure 3.19a. In all but one case our own formulation outperformed the Invitrogen MEM. For *PFB0900c* final OD<sub>600</sub> values were similar if not identical, 3.1 OD<sub>600</sub> for the MEM and 3.0 OD<sub>600</sub> for the in-house mix. Most importantly, the addition of AAs did not only improve the final quantity of bacteria, but it did also improve the solubility of the recombinant proteins, Figure 3.19b. For the 4 constructs considered problematic the influence of AAs availability not only was evident but it was clone dependent. The enhancement in solubility ranged from 1.5-fold for *PFD0110w\_p3* to a more impressive 14.7-fold for *PF11\_0174\_p2*, with an average increase in solubility of 1.6-fold if we exclude the latter, or 4.9-fold if we include it, Figure 3.20.

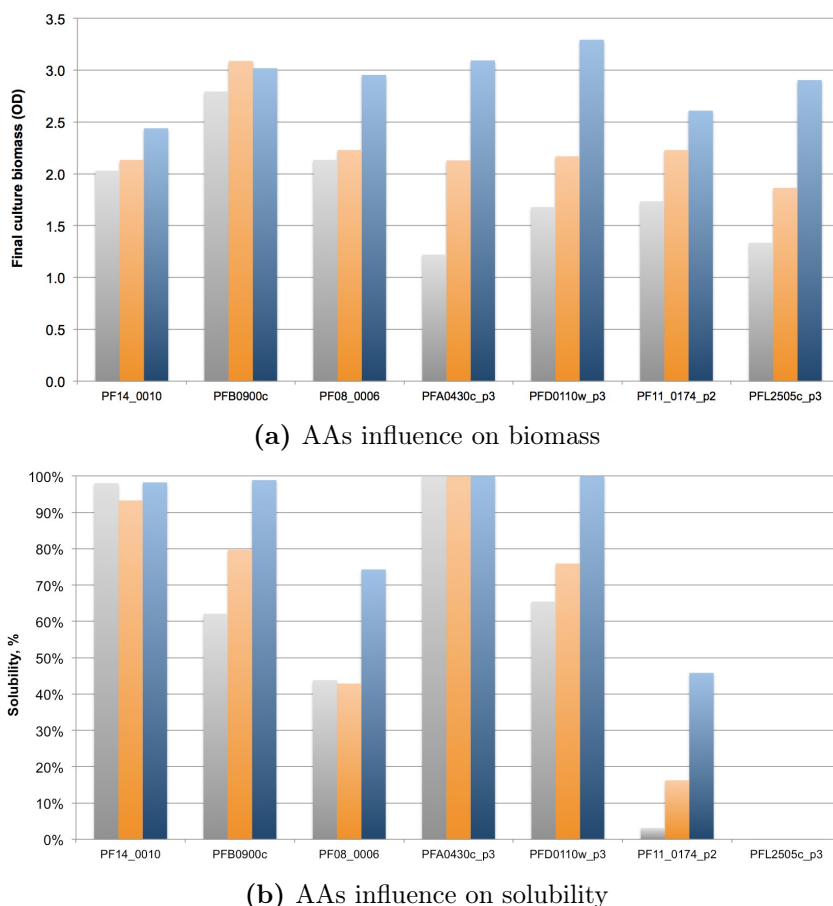
On the contrary, the solubility of *PF14\_0010* and *PFA0430c\_p3* did not

AA Solution (1X)	MEM	Internal	Internal/MEM
Amino acid	mg/L	mg/L	Ratio
Alanine	8.9	250	28.1
Arginine	126.4	500	4.0
Asparagine	13.2	100	7.6
Aspartic acid	13.3	125	9.4
Cystine	24.0	100	4.2
Glutamine	NA	<1	NA
Glutamic Acid	14.7	100	6.8
Glycine	7.5	250	33.3
Histidine	42.0	600	14.3
Isoleucine	52.4	750	14.3
Leucine	52.4	750	14.3
Lysine	72.5	750	10.3
Methionine	15.1	100	6.6
Phenylalanine	33.0	500	15.2
Proline	11.5	500	43.5
Serine	10.5	750	71.4
Threonine	47.6	500	10.5
Tryptophan	10.2	100	9.8
Tyrosine	36.0	200	5.6
Valine	46.8	100	2.1

**Table 3.17: TM amino acid supplements, Invitrogen vs internal.**

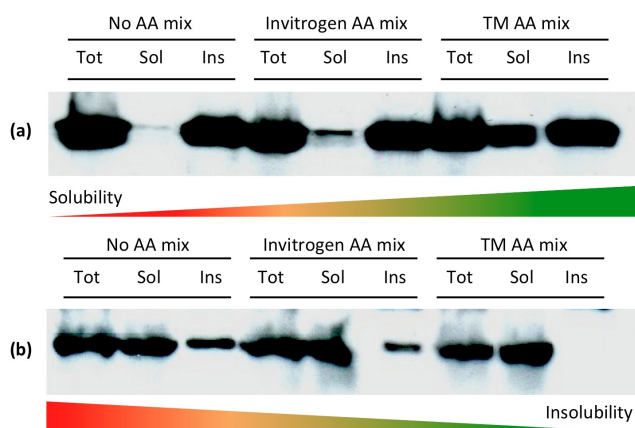
Two AAs sources were tested using TM. One was based on the Invitrogen MEM products that contained both essential and non-essential AAs with the exclusion of glutamine; and one based on an in-house formulation, which contained all 20 AAs. The concentration of reagents present in the internal mixtures were generally higher than in the Invitrogen counterpart (from 2.1 to 71.4 times higher).

show any AA dependence, which was supported by the fact that they already produced soluble products when expressed in 2xYT and/or AIM. No improvement was observed for *PFL2505c\_p3* construct, which remained negative in all the three expression conditions. Overall, the internal AA mix produced better results than the Invitrogen MEM for all expressed clones.



**Figure 3.19: Influence of amino acid compositions on bacterial biomass and protein solubility.**

Overall seven plasmids were expressed under three different conditions: 1, TM without additional AAs (■); 2, TM with the MEM AAs mixtures (■); 3, TM with our own AA solutions (■). Resulting cultures were assessed for bacteria biomass in terms of final OD<sub>600</sub> (a), and for percentage of solubility calculated as soluble fraction/insoluble fraction, with fraction band intensity obtained from WB analysis (b).



**Figure 3.20: Amino acid influence on solubility, WB.**

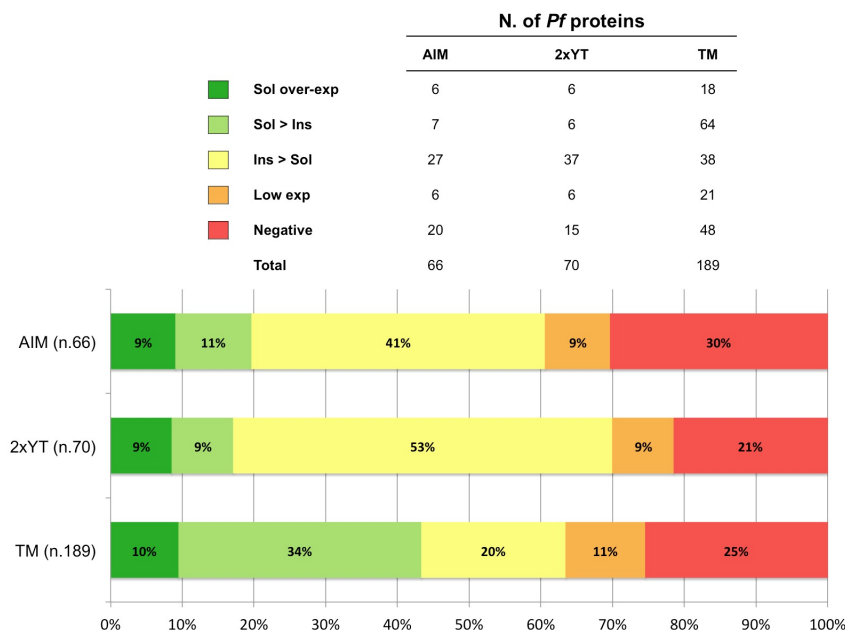
a) *PF11\_0174\_p2* and b) *PFD0110w\_p3*.

### 3.4.2 Mini-scale solubility analysis of 189 *Pf* candidates expressed in TM medium

Following the promising results obtained with the TM medium, we reassessed the expression and solubility levels of those challenging proteins that so far resulted either partially soluble or insoluble when expressed using either the AIM medium or 2xYT. Low priority was given to those candidates that already showed over-expression and high solubility in 2xYT and AIM, as we assumed that they will behave similarly when expressed in TM, as seen for the *PF14\_0010* and *PFA0430c\_p3* candidates. Similarly, those highly challenging proteins (negative or lowly expressed in 2xYT and AIM) were left aside after observing that the TM generally is not capable of improving protein yield to a level sufficient for their purification. Hence, a mini-scale expression and solubility analysis was run on 189 constructs, containing mainly the partially soluble proteins observed with 2xYT and/or AIM and additional plasmids not tested before. Of the 189 plasmids, 10% (18/189) produced over-expressed soluble proteins. More interestingly, 34% of the expressed proteins (64/189) resulted soluble or mainly soluble, which is a 3-fold improvement from the 11% or 9% figures observed for the AIM and 2xYT respectively. As for the insoluble proteins, their fraction was halved, from 53% (2xYT) or 41% (AIM) to 20%. No substantial difference was observed for the those portion of the proteins resulting low or not expressed, Figure 3.21.

### 3.4.3 Large-scale solubility analysis of 189 *Pf* candidates expressed in TM medium

The protein yield observed so far in the mini-scale studies were far from being sufficient for an effective purification step. Indeed, an estimation of the yields obtained in the last TM expression and solubility analysis indicated that in a 4-5 mL volume of bacterial culture we were able to produce between 0.5 to 5  $\mu\text{g}$  of plasmodial protein, which is below the 100-150  $\mu\text{g}$  needed for our microarray study. However, the low volumes used in the mini-scale study allowed us to quickly screen hundreds of available targets and to prioritise them according to the their solubility behaviour. It is thanks to this knowledge that with the help of F. Trivelli, I started developing a new large-scale (1-2 L) expression procedure based on the optimised TM medium using only the most promising candidates (i.g. soluble or partially soluble). The focusing on a subset of candidates was mainly due to the fact that handling litres of bacte-



**Figure 3.21: Mini-scale studies, overall results: AIM vs 2xYT vs TM.**

Distribution of over-expressed, mainly soluble, mainly insoluble, low expressed and not expressed proteins as function of the medium used for expression study. The total of the soluble proteins is represented by the two green bars.

rial cultures reduced drastically our throughput power. Compared to the TM mini-scale protocol, where between 12 and 24 candidates were run in parallel, the large volumes required in the new procedure limited us to express up to 3 proteins per week. However, if on one hand the expression throughput was decreased of a factor of 4 or 8, on the other hand the protein yield, estimated using the Coomassie staining, increased 100-200-fold to 0.5-1 mg of recombinant proteins. To this date, 25 *Pf* candidates have been expressed using the new large-scale procedure (see Table 3.18). The majority (76%, 19/25) of the constructs produced soluble or partially soluble proteins, 15 of which were expressed to a level sufficient for a downstream purification step. Interestingly, all the 6 remaining plasmids seemed to be somehow affected by the increase in volume. Indeed, all but one of them produced soluble or partially soluble proteins in the previous studies. Additional work will be carried out on these plasmids to assess whether the altered expression behaviour was due to the conditions in which the bacteria were cultured or a technical problem during the preparation of the medium.

Target	pI	Size	Mini-scale (4-5 mL)		Large-scale TM (1-2 L)
			2xYT/AIM	TM	
<i>PF14_0010</i>	5.3	32.4	Soluble (AIM,2xYT)	Soluble	Soluble
<i>PFB0100c_p2</i>	9.4	44.9	Soluble (AIM)	NA	Soluble
<i>PF11_0164</i>	6.3	25.0	Soluble (AIM)	NA	Soluble
<i>PF10_0208</i>	7.2	32.5	Soluble (AIM)	NA	Soluble
<i>PFI1755c</i>	4.5	38.6	Soluble (AIM)	NA	Soluble
<i>PFB0090c</i>	6.6	50.2	Sol > Ins (AIM)	Soluble	Soluble
<i>PF11_0503</i>	8.8	46.7	Sol > Ins (AIM)	Soluble	Soluble
<i>PFB0900c</i>	6.0	35.8	Ins > Sol (AIM)	Soluble	Soluble
<i>PFL0035c_p2</i>	9.2	31.0	Sol > Ins (2xYT)	Sol > Ins	Sol > Ins
<i>PF14_0723</i>	5.1	54.6	Ins > Sol (2xYT)	Sol > Ins	Sol > Ins
<i>PFL2505c_p1</i>	5.7	34.3	Ins > Sol (2xYT)	Sol > Ins	Sol > Ins
<i>PFI1730w_p1</i>	6.5	33.2	NA	Sol > Ins	Sol > Ins
<i>PF14_0435_p1</i>	6.8	40.0	NA	Sol > Ins	Sol > Ins
<i>PF1180w</i>	6.1	42.5	NA	Sol > Ins	Sol > Ins
<i>PF08_0091_p1</i>	6.5	61.3	Ins > Sol (2xYT)	Ins > Sol	Sol > Ins
<i>PF14_0417_p2</i>	5.3	76.6	Ins > Sol (2xYT)	Soluble	Soluble, low-expressed
<i>PFA0430c_p3</i>	6.3	40.1	NA	Soluble	Soluble, low-expressed
<i>PF1150c_p1</i>	8.3	57.2	Ins > Sol (2xYT)	Sol > Ins	Soluble, low-expressed
<i>PFE0370c_p2</i>	7.0	50.5	Ins > Sol (2xYT)	Sol > Ins	Soluble, low-expressed
<i>PF110w_P3</i>	5.4	72.2	Ins > Sol (2xYT)	Soluble	Ins > Sol, low expression
<i>PFB0405w_p6</i>	7.8	60.6	Ins > Sol (2xYT)	Sol > Ins	Ins > Sol, low-expressed
<i>PFL1300c</i>	7.6	45.6	Ins > Sol (2xYT)	Ins > Sol	Ins > Sol, low-expressed
<i>PFB0405w_p3</i>	5.9	64.3	Ins > Sol (2xYT)	Sol > Ins	Not expressed
<i>PFB0926c</i>	6.4	36.5	Soluble (AIM)	NA	Not expressed
<i>PF11_0037_p1</i>	5.7	35.8	Soluble (AIM)	NA	Not expressed

**Table 3.18: Large-scale expression of 25 candidates in TM.**

Candidates were expressed in *E. coli* culture of 1-2 L of TM to produce enough recombinant proteins to allow a downstream purification procedure. The overall expression procedure last between 3-4 days. Bacteria were routinely induced with IPTG at OD<sub>600</sub> between 0.48 and 1.0. Final OD<sub>600</sub> readings ranged between 1.0 and 3.5 after 18h of expression at RT. Solubility analysis was performed evaluating the expression pattern of the Coomassie and/or WB staining. NA: Not available.

### 3.4.4 Discussion

One of the most important milestones achieved during my PhD was the creation of a suitable expression tool for the production of what so far has been defined very challenging, the heterologous expression of *Pf* genes in a *E. coli* host (Crompton et al., 2010; Vedadi et al., 2007; Mehlin et al., 2006; Aguiar et al., 2004). Both *in vivo* and *in vitro* *E. coli* systems were considered. Sample quality and purity were set as key parameters for the successful use of our proteins in a microarray immunoassay. We therefore selected the *in vivo* approach as our primary platform, leaving the *in vitro* expression only as a



back-up option. There are many *E. coli* strains, which are currently available in the market, engineered for the expression of heterologous proteins, but only a few are specifically optimised for the synthesis of challenging proteins like the *Pf* ones. We chose to evaluate 5 of the most common *E. coli* strains: BL21 Star<sup>TM</sup>, BL21 BLR, BL21 Codon Plus RIL<sup>®</sup>, Rosetta-Gami<sup>TM</sup> B and Rosetta-Gami<sup>TM</sup> 2. To capitalise on the capabilities and expertise shared within the consortium, we decided to carry out two independent studies, one at PX' and one at UNIPG, in which I was directly involved. In both studies two were the main strains tested, Star and RGB, as they represented the two extreme sides of the 5 selected strains. Star<sup>TM</sup> is the most widely used and less engineered bacteria strain, whereas the Rosetta-Gami<sup>TM</sup> series are one of the most complex and re-elaborated strains available. The challenging aspect of the *Pf* proteins was immediately clear when we tested the Star<sup>TM</sup> strain at UNIPG. Four out of 5 of the selected plasmid failed to show any expression. Such high level of failure has probably a two-fold explanation. One is the small number of plasmids tested (only 5) and the other one is the limited expertise we shared at that time. Indeed, when F. Wolff repeated the expression in the Star<sup>TM</sup> strain at PX' using 37 plasmids, 26 of the targets resulted positive, with a few showing limited or minimal expression level.

The main factor affecting the expression performances, which ranged from 33% in BLR to 72% in RGB, is the presence of a codon bias factor. Eukaryotic organisms in general and *Pf* in particular often use codons that are rarely used in *E. coli*, (Nakamura et al., 2000). Indeed, eukaryotic organisms, especially those with AT-rich genomes (see Section 1.2.1), utilise codons that are rarely used by bacteria such as *E. coli* to encode specific AAs such as Arg, Ile and Gly (AGA/AGG, AUA and GGA) (Zhou et al., 2004; Saul and Battistutta, 1988). Consequently, due to the lack of these specific tRNAs in the expression machinery, the stall of translation may occur during the production of *Pf* proteins in *E. coli* (Zhou et al., 2004; Zahn, 1996). This effect should be even more prominent with large ORFs, where the cumulative presence of these rare codons is more likely to exhaust the available source of these limited bacterial tRNAs. An example of such effect is the drop in expression and solubility observed by Vedadi et al. (2007). In particular they observed a drop of 1/3 in solubility when the average size of the protein to be expressed was increased from <20 kDa to 50 kDa. Different strategies have already been applied in the literature to overcome these problems. Applications such as re-engineering genes to contain only codons commonly used in *E. coli* (Zhou et al., 2004) and co-transforming plasmids encoding for rare tRNAs (Baca

and Hol, 2000) are two such approaches that may partially help solve this bias, although some issues may still remain as observed in the use of the RIL strain by Vedadi et al. (2007). For this project, however, while re-engineering individual genes offers the most effective method for overcoming this problem, the high costs involved and the time required make this impossible for large-scale expression. Consequently, to overcome the codon bias issue without moving to a more complicated and expensive expression system (e.g. yeast or baculovirus) we opted for the methodology suggested by Baca and Hol (2000). This utilises an engineered *E. coli* strain able to provide rare tRNAs codons as well as enhance disulfide bond formation. These two features taken together should guarantee a high yield of expression as well as enhanced folding of the expressed protein, crucial for its solubility and use in measuring immune reactivity. These concepts were the reasons why out of the 5 bacterial strains selected, 4 possessed the ability of synthesise rare tRNA (BLR, RIL, RG2 and RGB), with the Rosetta gami series capable of enhancing also the formation of disulfide bonds. Experimental prove of these hypothesis were observed during both expression studies. At UNIPG, out of the 22 plasmids tested in RGB (including the 5 tested with the BL21 strain) 17 were successfully expressed, increasing the expression rate to 77%. When WB assays were performed they not only confirmed expression and solubility levels but they also demonstrated that proteolytic activity was relatively low in our assays (6xHIS detection vs c-MYC, data not shown). These findings were further corroborated by the results obtained at PX'. There, a total of 66 plasmids were tested in RGB with an overall expression rate of 70%, which matches with a considerable concordance the initial trial at UNIPG. As for the three additional strains (BLR, RIL and RG2) tested by F. Wolff as backup host, the size of the evaluation test was limited and not sufficient for a comprehensive strain evaluation, but it was enough to highlight some of the possible limitations of the three hosts. Indeed, neither of the systems was able to exceed 50% of expression rate. Interestingly, when we evaluated the RG2, a strain very similar to the RGB, only 11 of the 22 plasmids tested produced enough protein to be visible on WB. This could be explained by one essential difference between the RGB and RG2, the strain of origin. As explained in Table 2.5 the RGB have a BL21 background, whereas the RG2 have a K-12 background. Overall, both studies clearly confirmed the idea that engineered bacterial strains capable of using rare tRNA and enhancing disulfide bond formation can improve the expression efficiency. Although, not all recombinant proteins were successfully produced and the expression level was relatively low ( $\sim 0.5-1 \mu\text{g/mL}$ ), a 70-

80% expression efficiency was comparable to the study of Vedadi et al. (2007) and much higher than what observed by Mehlin et al. (2006) (34%) or Aguiar et al. (2004) (7-16%). Interestingly, the use of an engineered strain seemed to have only a minor effect on the solubility of the proteins. Indeed, the initial solubility percentage obtained without a systematic optimisation of the expression conditions (e.g 2xYT, 18% or AIM, 20%) was similar to the 19% reported in the literature by (Mehlin et al., 2006) and  $\sim 25\%$  obtained by (Vedadi et al., 2007).

Consequently, once the *E. coli* host systems was chosen (RGB) the attention was shifted towards the optimisation of expression conditions (i.e. medium, temperature, medium additives etc.). Indeed, the performances obtained so far were satisfactory in terms of expression but the percentage of soluble proteins obtained was suboptimal. Hence, now that we were confident with the expression host, we tried to improve the expression conditions. In particular, an enhanced expression medium together with alternative incubation protocols were investigated to obtain higher yields of correctly folded recombinant proteins.

The first variable under scrutiny was the composition of the medium used for the expression. Interestingly, of the five different formulations tested (LB, TB, 2xYT, AIM and TM), those highly rich in energy and carbon source constantly outperformed the others in terms of protein yields ( $TM \gg AIM \geq 2xYT \gg TB \geq LB$ ). Such tendency was observed in different occasions (i.e. at UNIPG, PX' and MtM) and with a multiple and unbiased selection of recombinant proteins. Moreover, when the three best performing media were compared (TM Vs 2xYT Vs AIM) it was clear that the optimised formulation of the TM (a formulation based on the 2xYT with the addition of metals, vitamins and AAs) was better suited for the expression of difficult *Pf* proteins (e.g. those insoluble in 2xYT and AIM). This can be explained by the fact that none of the other media was specifically designed for the expression of *Pf* proteins which expression was difficult. Indeed, the idea behind the TM was to chose the best performing medium available and to customise it to favour the challenging expression of *Pf* targets. The need for an optimised medium is clear when the current studies on the large-scale expression of *Pf* proteins are taken into account (Vedadi et al., 2007; Mehlin et al., 2006; Aguiar et al., 2004). Protein yields and bacterial growth are limited when cultured in basal broths like LB. As shown by Gunsalus and Stanier (1960) and reviewed by Shiloach and Fass (2005), the availability of essential elements like carbon, vitamins and nitrogen combined with the absence of macro (calcium, iron,

magnesium, phosphorous, potassium, sodium and zinc) and micro (copper and sulphur) minerals can influence the bacterial growth. Such nutritional burden can be partially overcome with the implementation of feed-batch cultures, in which fresh medium and supplements are added at specific time intervals after the initial induction (Shiloach and Fass, 2005; Riesenberg and Guthke, 1999). However, feed-batch cultures tend to be laborious and time consuming, and they tend to be applied for single cultures. These are limitations that have an effect on the overall throughput. Therefore, we decided to enrich the starting expression medium with specific host-target dependent nutrients via a feed-less expression protocol. We started adding a panel of supplements at their highest possible concentration depending on their solubility levels and avoiding any possible inhibitory effect (Riesenberg, 1991). The task was made difficult by the fact that the solubility of some of the additives was not only dependent on the concentration of the component itself but also by the presence of the other supplements, which could react with it (i.e. insoluble complexes of divalent metals, Dean (1979)) during the resolubilisation, the sterilisation or the fermentation processes. In the TM formulation this issue was successfully overcome by designing a core group of elements (glucose, NaCl, bacto-tryptone and bacto-yeast extract), which were then supplemented to the remaining additives (vitamins and AAs, which are thermosensitive together with salts and metals) prior use. A further element of bacterial growth inhibition is the acidification of the medium during the metabolisation/consumption of glucose as primary source of carbon, which is due to the realisation of acetic acid in the medium (Yup and Lee, 1996; Riesenberg, 1991; Luli and Strohl, 1990). Therefore we provided the TM with an isotonic, non-toxic buffer phosphate buffer system (i.e.  $\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$ ) capable of stabilising the pH of the medium at a neutral value (pH 7.2-7.4). If the selection of an ortho-phosphates-based buffer was based on the fact that bacterial cultures can reach high levels of biomass via the metabolisation of  $\text{PO}_4^{-2}$  anions (Paliy and Gunasekera, 2007), biotin and thiamine were instead chosen for their ability to improve protein expression (Riesenberg, 1991; Campbell and Williams, 1953). Indeed, if the former enhances the metabolism and avoids cellular stress, the latter favours the AA and sugar catabolism. Sulphur and nitrogen were also supplied as an energy source via the addition of ammonium sulphate. Positive mineral cations (e.g.  $\text{Ca}^{+2}$  and  $\text{Mg}^{+2}$ ), often utilised as enzymatic cofactors, were also added to the formulation with the addition of  $\text{CaCl}_2$  and  $\text{MgSO}_4$  (Yang et al., 2003). The role of bivalent metal ions like  $\text{Zn}^{+2}$  and  $\text{Fe}^{+2}$  was also taken into account for an optimal bacterial growth. Both metals are essential cofactors for a electron

transfer reactions and enzyme activation (e.g zinc-finger transcription factor enzymatic family). Finally, the optimisation of TM was completed with the addition of two AAs mixtures. The idea originated from the fact that although rare tRNAs were supplied through the use of the RGB *E. coli* strain, according to the preliminary results obtained with relatively rich medium (2xYT or AIM), it seemed not to be enough to overcome the difficulties that the RGB strain was having in producing soluble *Pf* proteins. This probably originated from the fact that the evolutionary distance between the two organisms led not only to the use of specific tRNAs but such bias was also translated into the use of certain AAs instead of others as shown in Chanda et al. (2005) and Lobry and Gautier (1994). A summary of the elements that form the TM formulation and their respective function is available in Table (3.19).

Due to time constrains we could not individually verify the relevance of each component of the formulation. However, when the performance of the this medium was compared to the 2xYT, which can be considered a TM deprived of its additives, the improvement was evident. The addition of specific AAs combined together with the ability of buffering the expression reaction to a neutral pH were found to be two critical features of this new formulations. In both experiments where the  $\Delta_{pH}$  of the medium and the AAs supplement were taken into account, the TM constantly outperformed the 2xYT. As for the role of the AA supplement, this was evident when the bacteria were cultured in TM with and without them. During the test, the addition of an AA mix improved both protein yield and solubility in all but one case, which was a protein already shown to be soluble in all the media tested.

Overall, the use of an optimised medium specifically designed for the expression of challenging *Pf* proteins had 3-fold effect. Firstly, it improved the bacterial growth and biomass, which in most of the cases implicates that more recombinant protein is produced. Secondly, the use of the TM medium more than doubled the number of soluble proteins (from less than 18-20% to more than 44%). In particular, the TM improved the solubility of those challenging *Pf* proteins, which were mainly insoluble in standard media. Last, but not least, the optimisation of the large-scale culturing conditions resulted not only in the soluble expression in the 500  $\mu\text{g}/\text{L}$  range of 76% candidate tested but also in a significant decrease of histidine rich stress-related bacterial proteins. This was crucial as the recombinant *Pf* proteins were eventually purified through use of their His-tags (see Section 3.6).

TM Core			
Element	Class	Known Role	Hypothetical Function in TM
Bacto-tryptone	Peptides	Source of AAs	Source of AAs
Bacto-yeast extract	Autolysed extract	Source of AAs, salts and carbohydrates	Source of AAs, salts and carbohydrates
Glucose	Sugar	Carbon and energy source	Carbon and energy source
NaCl	Salt	Sodium (Na <sup>+</sup> ) source and influence of water activity (e.g inverse of osmolarity) and availability	Stabilise bacterial growth
Buffer Solution			
Element	Class	Known Role	Hypothetical Function in TM
KH <sub>2</sub> PO <sub>4</sub> /K <sub>2</sub> HPO <sub>4</sub>	Buffer system	Isotonic, non-toxic buffer phosphate buffer system, pH 7.2-7.4	Stabilise TM pH and avoid medium acidification
Vitamins			
Element	Class	Known Role	Hypothetical Function in TM
Biotin	Vitamin	Also known as Vitamin B <sub>1</sub> , it acts as enzymatic cofactor and CO <sub>2</sub> transporter	Enhanced metabolism and decrease metabolic stress
Thiamine	Vitamin	Enzymatic cofactor and it enhance AAs and sugar catabolism	Enhanced metabolism and decrease metabolic stress
Minerals			
Element	Class	Known Role	Hypothetical Function in TM
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	Salt	Nitrogen (N), sulphur (S) and energy source	N, S and energy source
CaCl <sub>2</sub>	Salt	Enzymatic cofactor for heat shock and signalling proteins	Enhance protein folding capabilities by stabilisation of heat shock proteins
MgSO <sub>4</sub>	Salt	Magnesium (Mg <sup>++</sup> ) source, enzymatic cofactor, binds/ coordinates chaperonines and ATP-dependent enzymes	Enhance protein folding capabilities by stabilisation of chaperonines
Bivalent metal ions			
Element	Class	Known Role	Hypothetical Function in TM
FeSO <sub>4</sub>	Metal	Enzymatic cofactor mainly in the respiratory chain	Enhanced transfer of electron and reduce metabolic stress
ZnCl <sub>2</sub>	Metal	Enzymatic cofactor for zinc-finger transcription factor enzymes	Enhanced transcription enzymatic activity
Amino acids			
Element	Class	Known Role	Hypothetical Function in TM
Amino acids	AAs	Building blocks of proteins, source of energy	Enhance the protein synthesis and bacterial growth

Table 3.19: TM formulation.

### 3.5 *In vitro* expression of *Pf* candidates

The results of the heterologous *in vivo* expression in *E. coli*, highlighted that at least a third (31%, 59/189) of the *Pf* candidates is either lowly or not expressed despite the optimisation of the expression conditions. These findings further corroborated the idea that an alternative expression platform was needed as a back-up to rescue some of these challenging proteins. The use of an *in vitro* translation systems may overcome some of the issues observed so far in the *in vivo* expression (e.g. toxicity of the over-expressed product to the host cell, the segregation of insoluble products in the inclusion bodies or the rapid proteolytic degradation by intracellular proteases). Hence an *in vitro* expression system was set up in parallel to the *in vivo* expression system (see Section 3.4.1) and the purification of *Pf* antigens expressed *in vivo* (see Section 3.6). The purpose of this was two fold; firstly to be used as an alternative expression system for *Pf* proteins that were not successfully expressed *in vivo* and secondly to express *Pf* proteins to use in the microarray feasibility study.

#### 3.5.1 Prokaryotic cell-free system: *E. coli*

One of the initial tasks addressed in my PhD was the evaluation of a protein microarray platform for the multiplexed screening of *Pf* putative antigens. However, when this study was performed neither the *in vivo E. coli* expression platform nor the bioinformatic selection of the *Pf* candidates was available. Consequently, the analysis was carried out using a distinct set of 40 *Pf* genes *Pf* targets (see Table 3.20) provided by UNIPG (Dottorini, 2006) and expressed using an *E. coli* cell-free system, which I optimised during my MSc thesis.

N.	Gene ID	Known <i>Pf</i> antigens, info
1	<i>PF11_0344</i>	apical membrane antigen 1, AMA-1
2	<i>PF11_0224</i>	circumsporozoite-related antigen, CSP
3	<i>PFB0300c</i>	merozoite surface protein 2, MSP-2
4	<i>PFE1590w</i>	early transcribed membrane protein 5, ETRAMP-5
5	<i>PFB0305c</i>	merozoite surface protein 5, MSP-5
N.	Gene ID	Putative immuno-targets, info
6	<i>MAL7P1.170</i>	Plasmodium exported protein, unknown function
7	<i>MAL7P1.174</i>	Plasmodium exported protein (PHISTb), unknown function
8	<i>PF10_0013</i>	Plasmodium exported protein (hyp12), unknown function
9	<i>PF11_0040</i>	early transcribed membrane protein 11.2
10	<i>PF13_0197</i>	merozoite surface protein 7 precursor
11	<i>PF13_0275</i>	Plasmodium exported protein, unknown function
12	<i>PF14_0010</i>	glycophorin binding protein
13	<i>PF14_0290</i>	conserved Plasmodium protein, unknown function
14	<i>PF14_0358</i>	41-2 protein antigen precursor, transport protein particle (TRAPP) component, Bet3
15	<i>PF14_0678</i>	exported protein 2
16	<i>PF14_0742</i>	Plasmodium exported protein (hyp6), unknown function
17	<i>PFA0135w</i>	merozoite-associated tryptophan-rich antigen, putative
18	<i>PFA0670c</i>	Plasmodium exported protein (hyp8), unknown function
19	<i>PFA0680c</i>	Pfmc-2TM Maurer's cleft two transmembrane protein
20	<i>PFB0105c</i>	Plasmodium exported protein (PHISTc), unknown function
21	<i>PFB0106c</i>	Plasmodium exported protein, unknown function
22	<i>PFB0888w</i>	ribosome associated membrane protein RAMP4, putative
23	<i>PFB0926c</i>	Plasmodium exported protein (hyp2), unknown function
24	<i>PFC0085c</i>	Plasmodium exported protein, unknown function
25	<i>PFC0090w</i>	Plasmodium exported protein, unknown function
26	<i>PFC0555c</i>	conserved Plasmodium protein, unknown function
27	<i>PF1170c</i>	Plasmodium exported protein (PHISTb), unknown function
28	<i>PF1180w</i>	Plasmodium exported protein (PHISTb), unknown function
29	<i>PFE0050w</i>	Plasmodium exported protein, unknown function
30	<i>PFE0080c</i>	rhoptry-associated protein 2
31	<i>PFE0395c</i>	6-cysteine protein
32	<i>PFE1600w</i>	Plasmodium exported protein (PHISTb), unknown function
33	<i>PFF0615c</i>	6-cysteine protein
34	<i>PFF1060w</i>	conserved Plasmodium protein, unknown function
35	<i>PFF1525c</i>	Pfmc-2TM Maurer's cleft two transmembrane protein
36	<i>PFI1270w</i>	conserved Plasmodium protein, unknown function
37	<i>PFI1565w</i>	profilin, putative
38	<i>PFI1755c</i>	ring-exported protein 3
39	<i>PFI1770w</i>	Plasmodium exported protein (PHISTb), unknown function
40	<i>PFI1785w</i>	Plasmodium exported protein (PHISTb), unknown function

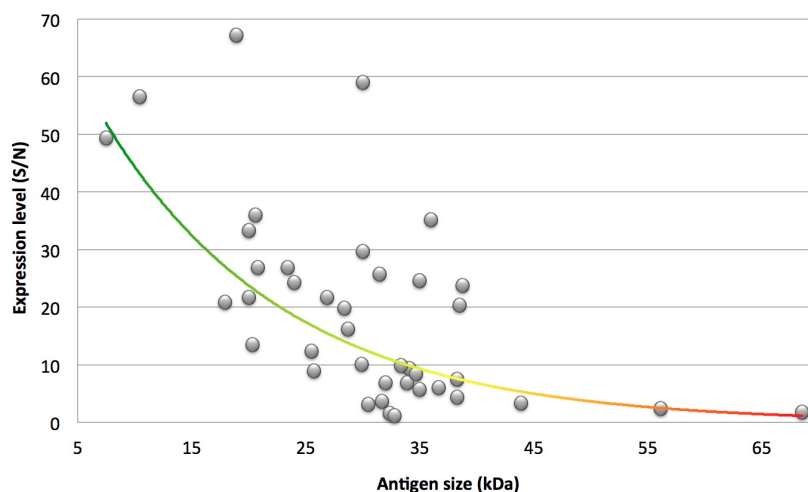
**Table 3.20: List of *Pf* genes utilised in the *E. coli in vitro* expression.**  
The selection of candidates was provided by UNIPG in the person of Dr. Dottorini.



At this stage of the project, the use of a different set of candidates was not critical as this proof-of-principle study would only provide essential information required for development of suitable protocols for the production and development of *Pf* antigen arrays (e.g. spotting condition, sample volume, working concentration etc). Nonetheless, the *in silico* analysis previously produced at UNIPG for the selection of the 40 *Pf* candidates, although smaller in terms of size, had similar selection criteria to the one eventually produced during the bioinformatic selection done for the FightMal project. The panel of candidates included: 5 well-documented malaria antigens *AMA* – 1, *MSP* – 2, *Exp* – 1, *MSP* – 5 and *ETRAP* – 5 (used as controls in the microarray); 11 characterised proteins (genes/proteins experimentally studied or with known homologies) and 24 hypothetical proteins (putative candidates selected during a previous internal bioinformatic study).

Gene constructs were created during my MSc project (2007) using pEXP5-NT/TOPO-TA vector (Invitrogen, UK) with a 6xHIS tag at the N-terminus. As no other expression method was yet available in our lab during the beginning of my PhD, and because of its simplicity of use and its scalability I used the *in vitro* *E. coli* cell-free expression system provided by Invitrogen: Expressway™ Plus Expression System. The Expressway™ Plus is a coupled transcription and translation T7-based expression platform that utilises an optimised *S30 E. coli slyD*-extract capable (according to the manufacturer) of producing up to  $\sim 300$  ng in a  $50\mu\text{L}$  reaction of recombinant active protein (i.e. chloramphenicol acetyltransferase protein, CAT, internal positive control) in less than 2 hours. Following the *in vitro* expression, the recombinant antigens were printed onto aldehyde-coated multi-well microscope slides (see Section 2.8.1) and used as an immunoassay substrate for assessing the presence of antigen-specific immunoglobulins classes and subclasses in African serum and European controls samples.

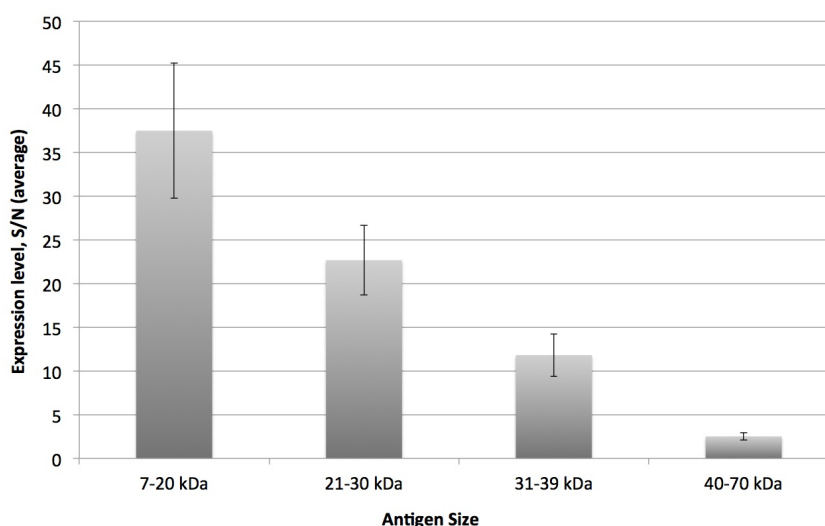
The arrays were run to assess the robustness of the proposed microarray platform and to assess the expression level of the various recombinant proteins. In particular, the expression levels were assessed by immune staining the chip with an anti-6xHis mouse IgG monoclonal antibody (4D11, see Section 2.8.2). The expression values, which were defined as signal to noise ratio (S/N), showed that expression levels (i.e. signal) were quite heterogeneous, ranging from 1.1 to 69-fold higher than the negative control (i.e. noise), Figure 3.22. A closer look at the data revealed a trend in protein yields that correlated with the protein size. In general, larger inserts (40-70 kDa) expressed less well than smaller inserts (7-30 kDa), with a critical limit of 31-40 kDa above which the



**Figure 3.22: Expression level plotted as a function of antigen size.**

The chart describes the variation of protein yields according to target size (kDa, x-axis). The y-axis shows the protein expression in terms of signal to noise ratio measured with an anti-6xHis tag antibody. The data suggests a correlation between the level of protein expression and the protein size. The trend line shows how smaller inserts tend to express better (U) than larger ones (U).

expression levels dropped to only a few fold higher ( $1 > S/N > 4$ ) than the negative control (*E. coli* extract lacking any expression plasmid), see Figure 3.23.



**Figure 3.23: Expression level plotted as a function of antigen size.**

The variation in expression yield as function of the protein size can be clustered in four main size-groups. Proteins ranging from 7 to 20 kDa possessed the high average yield, followed directly by the ones between 21-30 kDa. An increment of 10 kDa in the protein size resulted in a 50% drop in expression efficiency (31-39 kDa group). An even more sharp decrease was visible in the last group of proteins (40-70 kDa), where the drop was nearly 66%.

Overall, the analysis of the expression yields revealed that 82% (33 out of 40) of the targets generated expression signals ranging from 4 to 67 times higher than the noise. Only a minority (18%, 7 out of 40) showed low expression efficiency, which was close to the *E. coli* signal (1.1 to 4-fold higher than the negative signal), see Table 3.21.

Following the analysis of the expression levels the arrays were used in the microarray feasibility study (see Section 2.8.2 and 3.7) to assess which *Pf* antigens were recognised by sera from patients with malaria.

Expression level	Known antigens	Characterised protein	Hypothetical protein	Total
$4 > S/N > 67$	4	8	21	33
$1 > S/N > 4$	1	3	3	7
Negative	0	0	0	0

**Table 3.21: *In vitro* expression levels.**

Summary of the expression levels calculated as signal to noise ratio (S/N). Values were obtained probing the chips with an anti-6xHis monoclonal antibody. Noise value: average of background signal emitted by *E. coli* extract spots.

### 3.5.2 Eukaryotic cell-free system: wheat germ

During the initial cell-free expression carried out for the proof-of-principle microarray study, the prokaryotic *in vitro* platform showed several limitations (e.g. size-dependent expression, low yield etc.). Therefore, before investigating possible solutions to the size-dependent expression seen in the bacterial system (e.g. codon optimisation and gene synthesis), I decided to look into an alternative cell-free eukaryotic expression system. Between the different platforms taken into considerations (i.e. immortalised human cell lines, insect cells, rabbit reticulocyte, wheat germ), two were the ones most commonly used, the wheat germ and the rabbit reticulocyte lysate (see Table 3.22).

Info	S30 <i>E. coli</i>	Wheat germ extract	Rabbit reticulocyte
Classification	Prokaryote	Eukaryote	Eukaryote
Source	Expressway™ Plus (Invitrogen) or <i>E. coli</i> T7 S30 Extract System (Promega)	TNT® Coupled Wheat Germ Extract (Promega)	TNT® Coupled Reticulocyte Lysate(Promega)
Expression yield (50µL)	100-300ng	30-150ng	50-200ng
Codon Bias	Yes	No	No
Post-translational modification	No	No	Yes
Template source DNA	Circular>Linear	Circular<Linear	Circular<Linear
Presence of endogenous mRNAs	No	No	Yes
Sensitivity to additives	Low	Low	High
Cost	\$	\$\$	\$\$\$
Scalability	High	Medium	Low
Reproducibility	Medium	Medium	Low

**Table 3.22: Comparison of cell-free expression systems.**

Comparison of the features and advantages of each extract-based systems. Data was obtained from different cell-free expression system suppliers (i.e. Promega, Thermo Fisher Scientific and Roche).

Both platforms allow the simultaneous transcription and translation of the gene of interest, with no need for codon optimisation. To reach optimal expression level, however, a linearised T7-based plasmid containing both upstream and downstream regulatory sequences is required. According to the manufacturer's instructions, both systems can produce up to 150-200  $\mu$ g of protein in a 1.5 h reaction at 30°C.

Considering the biochemical nature of the *Pf* proteins (i.e. low glycosylation level), scalability (i.e. high-throughput expression), reproducibility and cost, the TNT® coupled wheat germ extract was chosen. The same plasmid as the one used in the *in vivo* system (pET160 backbone, see Figure 3.10) was ini-

tially used to save time and avoid subcloning. Indeed, although the system may require specific regulatory sequences upstream/downstream the gene of interest, the extract was capable of expressing a generic T7-based plasmid (according to the manufacturer's instructions). However, when I tested *PFE1605w* and *PFB0100c\_P2* constructs (two plasmids already successfully expressed in the *in vivo* platform), both resulted negative for expression. Moreover, the WB analysis showed a high level of background, with several unspecific bands, which were also present in the negative control (i.e. no plasmid added to the expression reaction). On the contrary, the luciferase gene, which was used as a control reaction, was successfully expressed in each occasion.

To avoid further delays and lengthy optimisation steps, we decided to evaluate the wheat germ platform using a new plasmid backbone (pIVEX, see Section 2.5.2), which was specifically created for the expression in eukaryotic systems, and in particular for the wheat germ. As the pIVEX vector did not have the same functional elements of our original pET160, I had to modify the pIVEX accordingly. This eventually facilitated the DNA subcloning between the two vectors (see Section 2.5.2). The plasmid modification was successfully accomplished by inserting a newly synthesised DNA cassette, containing some of the main features of the pET160 plasmid (e.g. 5'-6xHIS-Lumio<sup>®</sup>-TEV-attB<sub>1</sub>-CACC-attB<sub>2</sub>-3'), into the pIVEX vector via double enzymatic digestion. Digested products were then separated on agarose, purified and successfully ligated to form a new hybrid vector named pIVEX-PF.

One of the features of DNA cassette inserted was the presence of the *Pf* antigen gene coding for *MAL7P1.177*, a parasite protein of 26.0 kDa, which was successfully expressed *in vivo* (see Table 3.15). This allowed the direct expression of the newly ligated pIVEX-PF plasmid, without the need of a further double digestion to insert a *Pf* antigen.

Using this new construct, a series of small-scale expressions (50  $\mu$ L) were run to verify the performance of this cell-free system. Both the circular and linear form of the pIVEX-PF plasmid were used, as suggested in the manufacturing instructions. For each expression positive (luciferase gene) and negative (no plasmid) controls were successfully run. Unfortunately, in all three expression tests, the determination in WB of the expression level of pIVEX-PF was impaired by the same unspecific bands observed in the initial experiments, which ran at the same MW as the wheat. To identify the source of this unspecificity, the samples were probed with and without the primary anti-HIS tag m-Ab (only the secondary anti-mouse antibody used). This unmistakably identified the secondary antibody (anti-mouse-HRP) as the main cause of the

unspecific binding.

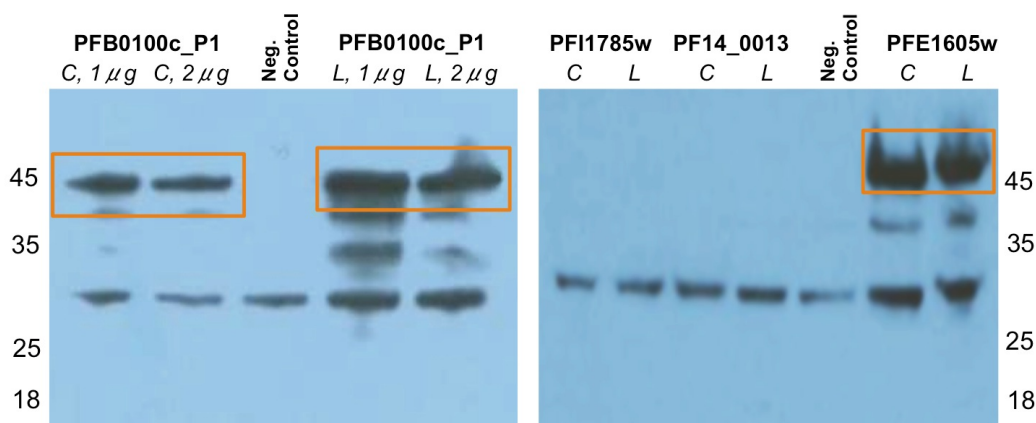
Eventually, once the cross-reactivity issues were overcome by using a different blocking solution (BSA/Milk over a proprietary super-block solution), we observed that *MAL7P1.177* was still not synthesised. To understand if the absence of expression was due to the kind of DNA target used or to an issue during the plasmid preparation, 4 additional genes were inserted into the newly developed plasmid (i.e. pIVEX-PF). *Pf* candidates were selected based on their expression rates observed during the AIM study, as this was the only information available at the time. One positive control, represented by a soluble protein (*PFB0100c\_P1*). One construct that was completely insoluble (*PFE1605w*) and two negatives (*PFI1785w*, *PF14\_0013*). As a temporary alternative to the Gateway<sup>®</sup> platform, the subcloning of these 4 genes was accomplished by F. Wolff via a double restriction enzyme digestion, which was enabled by the previous addition to the new hybrid plasmid of two octanucleotide-recognising restriction sites up- and down-stream of the gene of interest (NotI and AscI, respectively). Both restriction enzymes were chosen for their specificity towards two different 8bp GC palindromic sequences (NotI 5'-GCGGCCGC-3' and AscI 5'-GGCGCGCC-3'), which are not present in the *Pf* transcriptome. The double enzymatic digestion of both pIVEX-PF and pET160 plasmids followed by ligation of the open pIVEX-PF with the isolated *Pf* gene of interest offered a scalable and efficient method for a rapid vector backbone switch. For compatibility reasons over the digestion buffers the recombinant form of NotI, NotI-HF<sup>™</sup> was selected together with the AscI (NewEngland BioLabs).

Overall, 2 out of the 4 proteins were successfully expressed. The 2 newly synthesised targets were the positive control and the protein previously shown to be insoluble in the *in vivo* system, however, this time the expression product was soluble. Unfortunately, neither of the two previously not expressed proteins produced a visible band when the wheat germ expression extract was run on a WB (see Table 3.23).

Target	<i>In vivo</i> expression	<i>In vitro</i> expression
<i>MAL7P1.177</i>	Soluble	Negative
<i>PFB0100c_P1</i>	Soluble	Soluble
<i>PFE1605w</i>	Insoluble	Soluble
<i>PFI1785w</i>	Negative	Negative
<i>PF14_0013</i>	Negative	Negative

**Table 3.23: Wheat germ *in vitro* expression test.**

Summary of the solubility/expression levels obtained using the cell-free wheat germ extract.



**Figure 3.24: Wheat germ *in vitro* expression test.**

**Left panel:** *PFB0100c\_P1* was used as a control candidate during the wheat germ *in vitro* expression, as it resulted soluble during the *in vivo* expression. Two plasmid quantities were tested in the wheat germ expression (i.e. 1 or 2  $\mu\text{g}$ ), but no significant difference was observed between the two conditions. C: circular plasmid. L: linearised plasmids. **Right panel:** three additional plasmids were used to evaluate the performances of wheat germ extract. *PFI1785w* and *PF14\_0013* were two candidates that resulted negative during the *in vivo* expression, whereas *PFE1605w* was insoluble when expressed in AIM. Further analysis showed that this construct was still insoluble>soluble using the TM medium.

### 3.5.3 Discussion

The data obtained from the *in vitro* expression studies suggest that both prokaryotic and eukaryotic systems, although suitable for a mini-scale high-throughput protein expression, are not yet as efficient as expected. Being a relatively niche technology the cell-free expression system has yet to exploit its full potentials, which could bring the platform into the mainstream as a viable alternative to the *in vivo* systems for proteome wide protein analysis. In different occasions cell-free expression systems have been used for the production of large sets of *Pf* protein candidates. In particular, both *E. coli* (Crompton et al., 2010) and wheat germ (Tsuboi et al., 2010, 2008) *in vitro* systems have

been used in the literature to synthesised *Pf* proteins. As for the FightMal project, it could be utilised to rescue some of the candidates that failed to express *in vivo* due to either insolubility or toxicity.

Using the prokaryotic *in vitro* expression system, which was utilised for the feasibility study on the protein microarray platform, it was possible to expressed 95% (38/40 with a signal to noise ratio  $> 4$ ) of the 40 selected candidates. However, the levels of expression were quite heterogeneous, with an inverse correlation between protein expression and protein size (larger inserts generally expressed less than smaller inserts). An explanation for this trend can be found in the fact this prokaryotic *in vitro* system potentially suffers of the same limitations observed for the *E. coli in vivo* system, codon bias and AAs specific usage.

As for the eukaryotic *in vitro* platform, the results obtained in the initial tests were encouraging but limited to only a few plasmids, which were marginally evaluated due to both financial and time constrains. After the optimisation of the expression plasmid backbone, which was forced due to the unexpected lack of protein synthesis of the standard pET160 vector, 1 out of 3 of the challenging proteins was positive for soluble expression. The initial lack of expression of the pET160 plasmid can be attributed to the absence of some regulatory sequences needed for the optimal translation, a phenomenon already noted by Hino et al. (2008). In this work, the authors assessed the expression efficiency of different cell-free systems (i.e. *E. coli*, wheat germ, and rabbit reticulocytes) based on the synthesis of an eukaryotic enzyme (G3PDH). After attempting to express the protein using either a prokaryotic or an eukaryotic vector, they observed that the enhancer sequences present in the eukaryotic vector were essential for the correct synthesis of G3PDH in a *in vitro* eukaryotic expression system.

As for the inability of expressing *MAL7P1.177*, a small protein already soluble on the *in vivo* platform, it can potentially be explained by the fact that during the numerous subcloning steps needed for the creation of the pIVEX-PF vector the gene sequence could have been damaged. Indeed, as soon as the target sequence was replaced with a different gene, the vector was able to fully express the protein target, with no visible evidences of any errors/anomalies upstream or downstream the DNA of interest. In term of protein yield however, the system was not able to rescue any of the negative plasmids, which remained unexpressed even in this cell-free system. Whereas, for the two targets successfully expressed in a soluble form, the protein yield was limited and not sufficient to allow a downstream purification step. This initial results



indicate that the wheat germ expression system could overcome some of the solubility problems observed in the *in vivo* expression in *E. coli*. If this is confirmed and we manage to improve the protein yield, the use of this new expression platform may rescue some of the insoluble (20%) and the lowly expressed (11%) candidates potentially increasing the number of clones available for purification to over 70%, a 1.5-fold increase to the value obtained using the *in vivo* system alone. However, further experiments have to be carried out to investigate whether these findings can significantly improve the production of insoluble or toxic targets up to a level that can allow their purification.

If the consortium is willing to progress with this eukaryotic platform, it is envisaged that additional efforts have to be placed in increasing the overall protein yield. This possibility, although expensive, is already available in the market, in the form of continuous flow cell-free (CFCF) or continuous exchange cell-free (CECF) systems. The core of these technologies is based on the confinement of the protein synthesis in a expression chamber which is continuously filled (CFCF) or infused (via diffusional exchange, CECF) with a feeding solution that allows the simultaneous removal of by-products (Spirin et al., 1988). The methodology has been successfully applied to both prokaryotic and eukaryotic extracts (Shirokov et al., 2007; Ryabova et al., 1998; Spirin, 1992). The expression reaction can produce up to 5 mg/mL of protein in 24 h with reaction volumes ranging from 50  $\mu$ L to 10 mL. The continuous flow cell-free systems seem to overcome one of the biggest limitation of the standard *in vitro* expression. However, this comes at a high cost. In the last few years Roche Diagnostics GmbH has released a few CECF systems (i.e. RTS 100/500/9000, Hoffmann et al. (2004); Betton (2003)) in the market based on both *E. coli* and wheat germ. The price list per reaction can range from £25 (50 $\mu$ L, RTS 100) to £3,180 (10mL, RTS 9000, *E. coli* only) with the medium scale (1mL) system, RTS 500, selling for £326 per reaction.

With such price range the continuous flow cell-free systems are far from being applicable in our project. Even testing only the insoluble and lowly expressed candidates (31% of the total candidates) would cost around £2.500 only for the initial screening (n.100 50 $\mu$ L reaction of RTS 100 kit). To this figure we then have to add the number for medium scale reactions needed to produce enough protein for the purification step. Assuming that the system could rescue 1 every 3 protein tested, it would cost us over £10.700 (n.33 1 mL reaction of the RTS 500 kit). If this estimation is correct, the production cost of these challenging proteins could rise to nearly £450 per protein, without even including any pre-expression (e.g. vector exchange) or post ex-

pression (i.e. purification) manipulations. Such cost is nearly the full monthly allowance (i.e. €500) that each researcher was given for buying reagents under the FightMal project, and this cost does not include any pre-expression (e.g. vector exchange) or post expression (e.g. purification) manipulations.

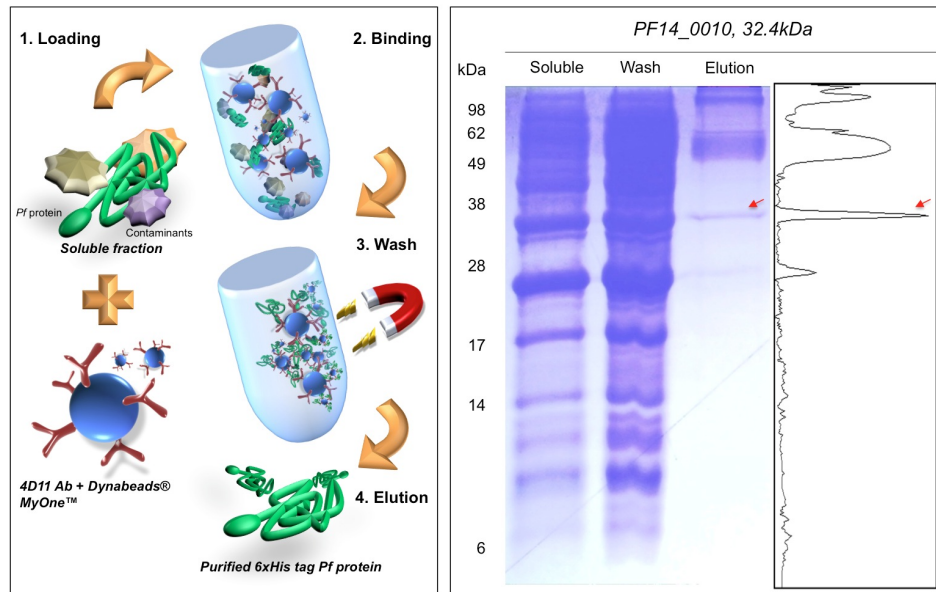
### 3.6 Purification of *Pf* recombinant proteins

A high purity grade of the spotted antigens is essential for a powerful and robust protein microarray platform. Hence, our aim was to identify and optimise a purification method that would enable us to obtain a few hundreds of  $\mu\text{g}$  of sufficiently pure ( $>90\%$ ) *Pf* proteins. Overall four purification procedures were evaluated:

1. Immunoaffinity purification via magnetic beads covalently bound to an anti-6xHis-tag antibody.
2. Immobilised metal ion affinity chromatography (IMAC) via gravity flow.
3. IMAC followed by size-exclusion chromatography (SEC) via fast protein liquid chromatography (FPLC).
4. IMAC based on high-speed, low pressure liquid chromatography (LPLC).

The first methodology to be assessed was the magnetic beads immunoaffinity purification. The evaluation was carried out at UNIPG, using one of the soluble recombinant proteins (i.e. *PF14\_0010*). The solid phase of the affinity chromatography system was composed of magnetic beads (Dynabeads<sup>®</sup> MyOne<sup>™</sup> Tosylactivated) covalently conjugated to our monoclonal anti-6xHis antibody. Generally, the antigen-antibody binding is pH dependent and can be reversed by adding an epitope competitor, by changing the ionic strength of the buffer or by dropping the pH of the solution. In our case, the binding to the capture antibody was promoted under a neutral pH. The endogenous bacterial proteins were removed by a series of PBS-T/BSA washing steps and subsequently the protein of interest was eluted by drastically dropping the pH from 7.4 to 2.5.

Purification efficacy was assessed by comparing the different amounts of 6xHis-tagged protein in both the flow-through and eluate against the initial quantity incubated with the beads (soluble fraction from the expression procedure). Unexpectedly, most of the protein of interest was lost during the washing step, with only a minor fraction being retained in the elution fraction (clear isolated band of the expected size, 32.4 kDa, indicating a low retention rate from the beads (see Figure 3.25). Additionally, at least three bands running at  $\sim 28$  kDa,  $\sim 60$  kDa and  $\sim 150$  were also present in the elution fraction, suggesting the co-elution of contaminants (e.g. 4D11 capturing antibody) during the release step. As a results of this contamination, our target band only



**Figure 3.25: Immunoaffinity purification via magnetic beads.**

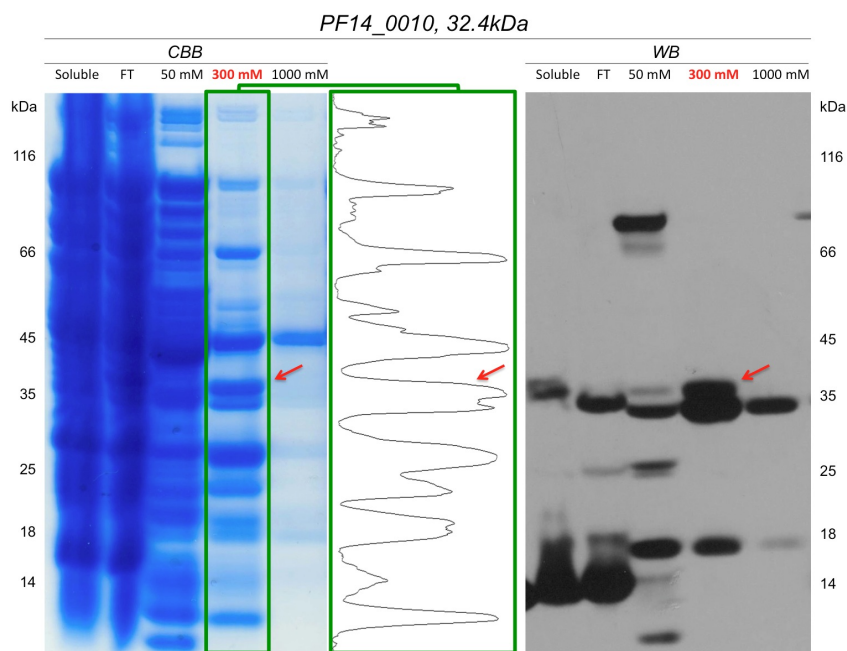
**Left panel:** Principles of the immunoaffinity purification based on the use of magnetic beads conjugated with the 4D11 monoclonal antibody. **Right panel:** image of the Coomassie stained gel (soluble, wash and elution fractions) relative to the purification of our reference protein candidate (i.e. *PF14\_0010*). A target-specific band is visible below the 38 kDa mark, in the fraction eluate. The intensity of the bands present in the elution fraction were quantified using ImageJ software (Abramoff et al., 2004), as illustrated in right side of the panel. The red arrows ( $\rightarrow$ ) indicate the protein of interest.

represented  $\sim 18\%$  of the protein content of the elution fraction, and a very low recovery rate was obtained (i.e. eluted fraction/soluble fraction:  $<10\text{-}20\%$ ).

In response to the sub-optimal performances (i.e. low recovery rate and low level of purity) observed with the magnetic beads purification, I started to look into an alternative purification system. With the help and know-how provided by the PX', I evaluated a new approach based on the IMAC technology. Via this technology, the purification is achieved by binding the 6xHis-tag of the protein to divalent metals (e.g  $\text{Ni}^{2+}$ , see Figure 2.5 and 2.7). The binding is pH dependent and can be reversed by dropping the pH, by increasing the ionic strength of the buffer or by adding EDTA or imidazole (as an histidine competitor, see Figure 2.6) to the elution buffer.

Initially, I evaluated a simple and scalable procedure, which consisted in purifying the soluble fraction (40 mL) of a large-scale bacterial culture (1 L) using a gravity flow IMAC method. A bed of 1 mL of  $\text{Ni}^{2+}$  sepharose resin was packed in a gravity chromatography column. Half of the soluble fraction (20 mL) was loaded and proteins binding weakly to the column were removed via an initial 50mM imidazole washing step. The protein of interest was eluted

with 300mM imidazole. This IMAC method was evaluated using *PF14\_0010* and five additional *Pf* proteins, which showed different levels of solubility (i.e. 2 over-expressed in a soluble form and 3 only partially soluble). A band specific for each protein was visible in all six eluted fractions. The general level of purity, however, was limited and mainly affected by a number of contaminant proteins (see Figure 3.26 and Figure 3.27). Around 8-10 unknown bands were observed when the eluted fractions were run and stained in Coomassie. In particular, two of the proteins co-eluted with our targets were also visible in WB (running at 35 kDa and 18 kDa), indicating either a C-terminal degradation of our recombinant protein or a possible presence of histine-rich sequences in the contaminants.

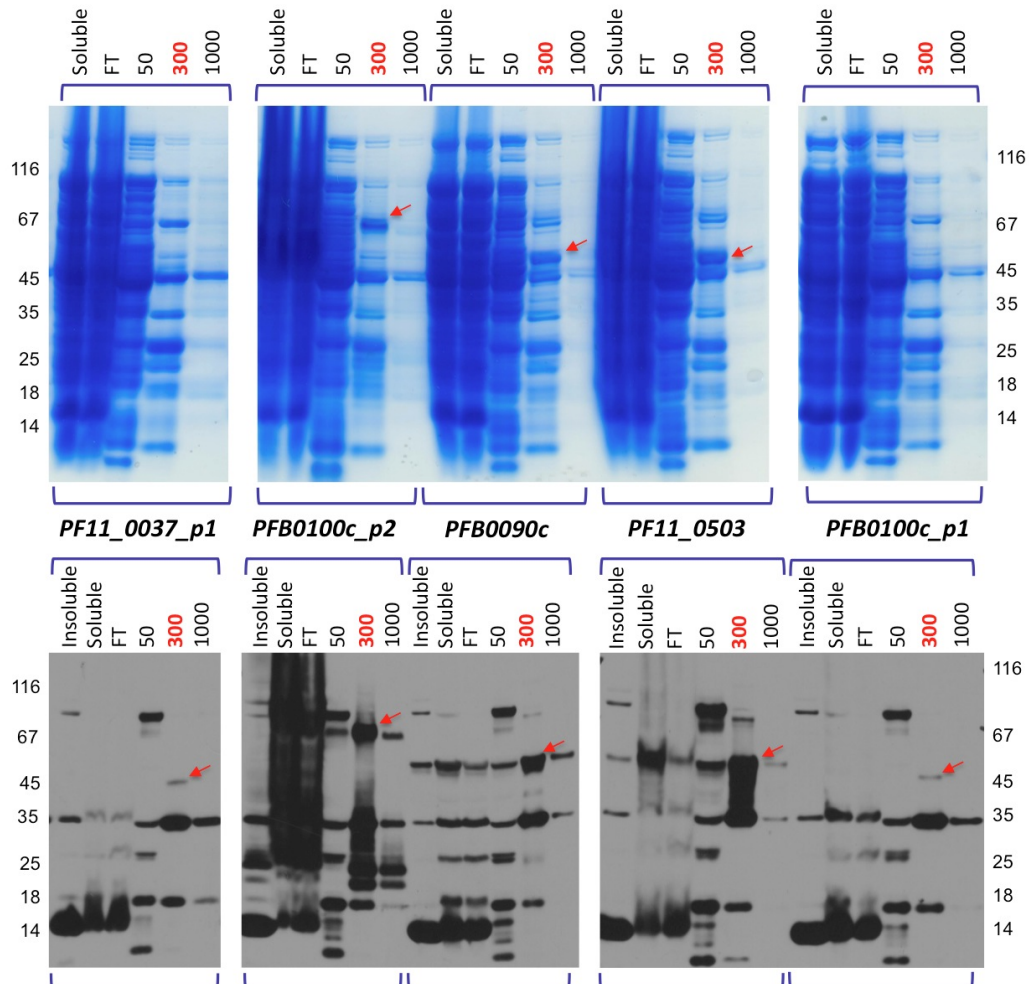


**Figure 3.26: Gravity flow IMAC, *PF14\_0010*.**

The left panel shows the Coomassie staining of the purification performed using our reference protein, *PF14\_0010* (soluble and over-expressed, 32.4 kDa size). The two green boxes highlight both the Coomassie and the chromatogram of the elution fraction. The quantification of the intensities relative to the bands present in the fraction eluted at 300 mM of imidazole were obtained using ImageJ software (Abramoff et al., 2004). The right panel contains the WB film of the *PF14\_0010* purification. In both Coomassie and WB a target-specific band is visible just above the 35 kDa mark ( $\rightarrow$ ), which is consistent with the migration pattern of *PF14\_0010* in non reducing conditions. Soluble: protein soluble fraction; FT: flow-through. 50 mM: washing step containing 50mM of imidazole; 300 mM: elution step performed using an elution buffer containing 300mM of imidazole; 1000 mM: column stripping step, performed at the end of the run to remove any protein still attached to the column.

The distribution of the unknown bands, however, was not random. Indeed, a common pattern was visible in the elution fractions from all 6 purifications.

Therefore we excluded the hypothesis that some of the extra bands observed in both Coomassie and WB were degradation products. Instead, these unknown contaminants are more likely to represent endogenous proteins rich in poly-histidine sequences. Consequently, they can bind the Ni<sup>2+</sup> sepharose resin and are also recognised by the anti-6xHis antibody.



**Figure 3.27: Gravity flow IMAC, 5 additional candidates.**

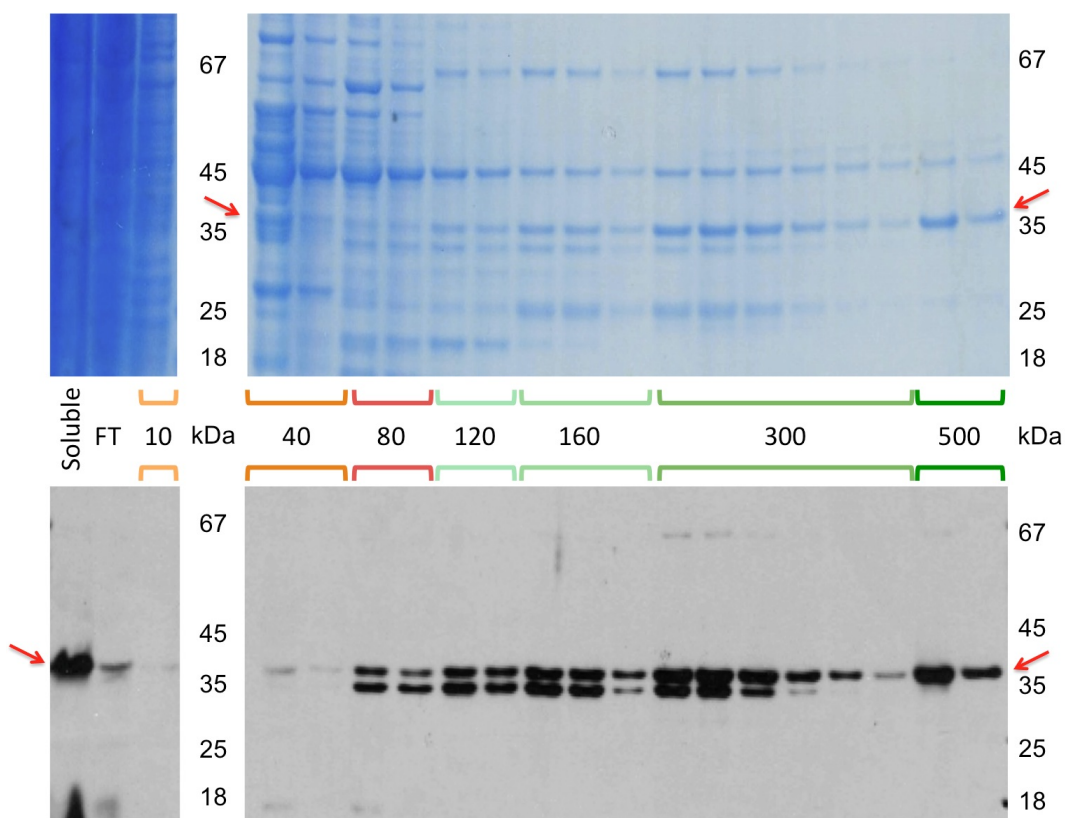
Five proteins, in addition to *PF14\_0010*, were expressed in 1 L bacterial culture. Twenty mL of the soluble fraction (**Soluble**) of each protein was loaded in a 1 mL chromatography column packed with Ni<sup>2+</sup> sepharose resin. The resulting flow through (**FT**) was collected and run together with fractions derived from the 50mM imidazole wash (**50**). The recombinant proteins were eluted using an elution buffer containing 300mM imidazole (**300**). Columns were eventually stripped of all remaining molecules bound to the solid phase via a final washing step with 1M imidazole (**1000**). Evaluation of the purification results was obtained by running each fraction in both Coomassie (*top panel*) and WB (*bottom panel*). The red arrows ( $\rightarrow$ ) indicate the specific band of each protein of interest. *PF11\_0037\_p1* over-expressed and soluble (37 kDa); *PFB0100c\_p2* over-expressed and soluble (45 kDa); *PFB0090c* partially soluble (50 kDa); *PF11\_0503* partially soluble (47 kDa); *PFB0100c\_p1* partially soluble (39 kDa).

Consequently, when I checked the level of purity of the *PF14\_0010*, it resulted close to  $\sim 12\%$ . As for the recovery level, no band was visible in the flow-through fraction (Figure 3.26, right panel), indicating a high rate of retention in the column. Such retention capability was confirmed by the absence of the specific-target band in most of the flow through fractions of the other 5 purifications (Figure 3.27, bottom panel). Noteworthy, is the fact that nearly all the candidates were efficiently eluted at 300mM of imidazole, as in 5 out of 6 cases no candidate specific band was visible in the stripping fraction. Overall, this purification procedure selectively removed the majority of the endogenous contaminants, while allowing elution of the *Pf* protein. However, the final purity level was limited. Hence, in an attempt to reduce the amount of contaminants in the elution fraction an alternative approach, based on similar principles, was evaluated.

The new purification procedure was carried out using the same methodology (IMAC) on 5 ml single-use pre-packed columns. This time, however, it was combined with the use of an ÄKTA FPLC system. To reduce the number of contaminants I increased the stringency of the binding condition by spiking the soluble fraction with 10mM imidazole prior loading it onto the column. For the same reason, the number of washes was increased from one at 50mM to three washes at 10, 40 and 80mM. To guarantee the widest spectrum of elution, four elution steps at 120, 160, 300 and 500mM imidazole were performed. The protocol was firstly assessed using the remaining 20 mL of the soluble fraction of *PF14\_0010*. Purification efficiency was assessed via both Coomassie and WB analysis, see Figure 3.28.

Under these new conditions, the majority of the protein of interest was successfully released during the four elution steps (see 120/160/300/500mM elution fractions, Figure 3.28). Importantly, only a limited fraction of *PF14\_0010* was lost in the flow-through and in the washes (FT and 40/80mM washing steps), which, on the other hand, successfully removed the majority of the contaminants. Despite the improved performances, however, three unspecific bands were still visible in Coomassie (120/160/300mM fractions) around 25, 45 and 67 kDa. It was also possible to detect a weak unspecific band close to 35 kDa in both Coomassie and WB. Interestingly, although the *PF14\_0010* band is visible in the Coomassie staining (i.e. 40mM imidazole washing fraction), the same band does not produce a proportional signal in WB. Hence, it could represent a possible N-terminal degradation product of our protein. Overall, the performance in terms of protein purity observed during the 300mM and



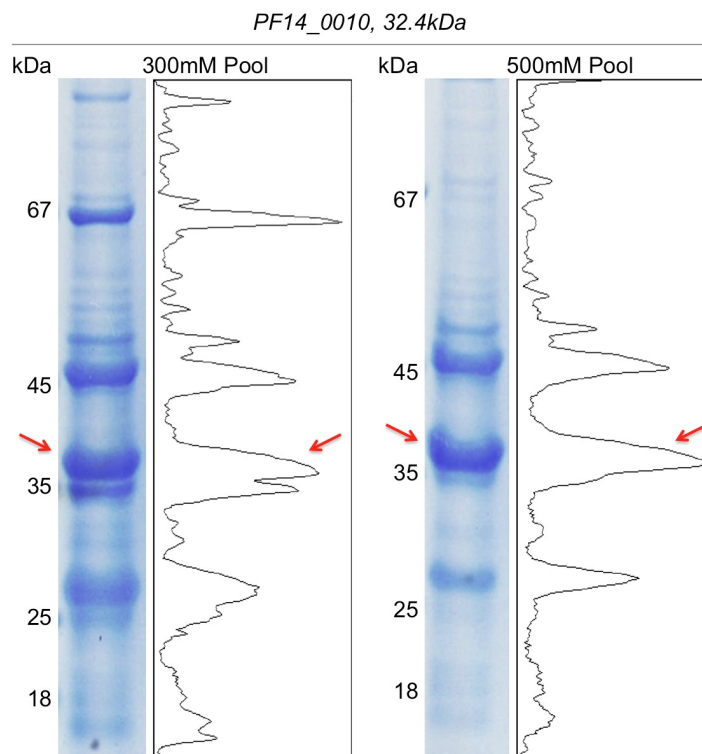


**Figure 3.28: Initial IMAC-FPLC purification test.**

500mM imidazole elution steps were considered interesting and their fraction pools (i.e. pool 300 and pool 500) were then concentrated and re-analysed in Coomassie for purity.

The Coomassie staining of the two mixes (see Figure 3.29) showed that both were still rich in bacterial contaminants. Indeed, if in one hand the *PF14\_0010* was the dominant band of the 500mM mixture, on the other hand two major contaminants were still present at 25 and 45 kDa. The same contaminants were present in the 300mM pool, which also contained an unknown protein running at 67 kDa. Overall, the purity grades (obtained by quantifying the pool's lanes with an imaging software), although higher than the one obtained with the previous two protocols, were still well below the 90% goal. In particular, 20% of the 300mM contained the recombinant protein. Whereas, the *PF14\_0010* represented 32% of the total protein content in the 500mM mixture. Therefore, although the new procedure produced nearly a 3-fold improvement in the purity level compared to both the magnetic beads and gravity-flow purification procedures (32% versus 18% and 12% respectively), the performance was still below the expectations. Consequently, I decided to attempt an additional purification step using a Superdex 75 gel filtration

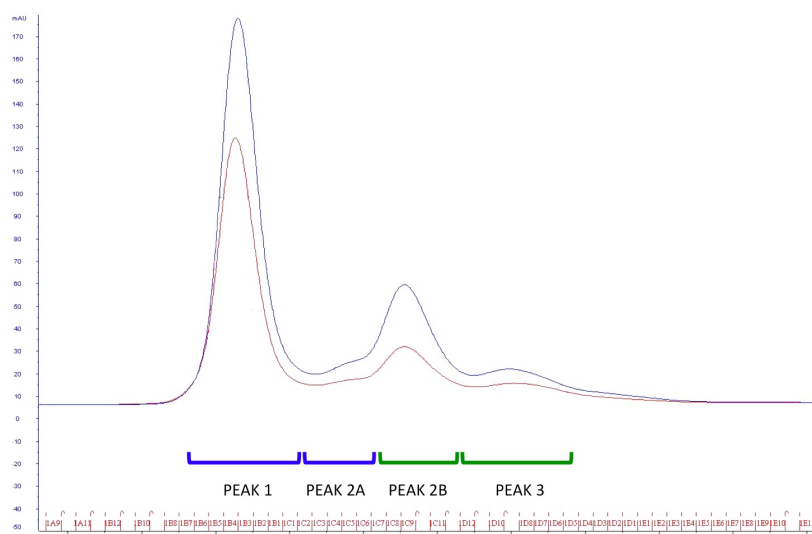




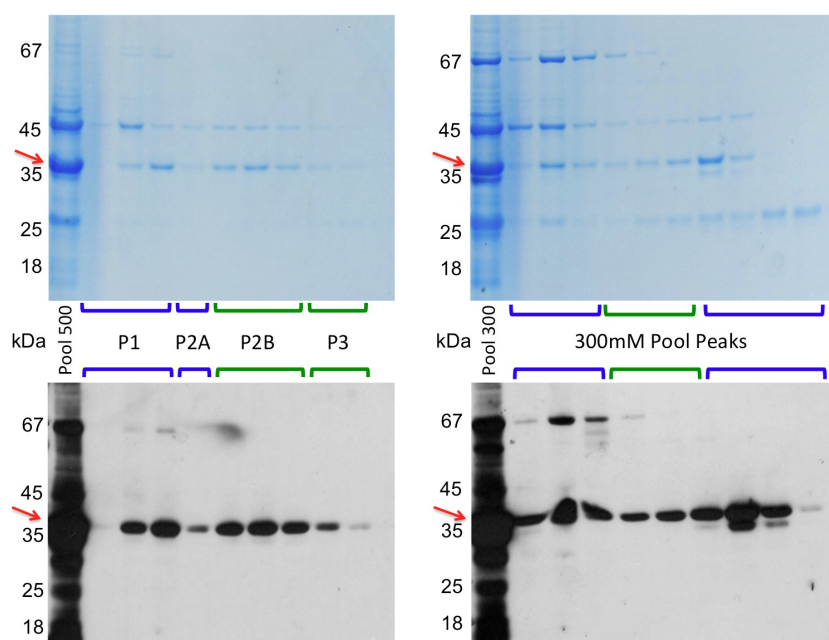
**Figure 3.29: IMAC-FPLC purification: results.**

column with optimal resolution between 3 and 70 kDa. Both pools were sequentially loaded and separated through the size-exclusion column, see Figure 3.30b.

The gel filtration performed on the 300mM pool efficiently separated most of the 67 kDa contaminant from the recombinant protein. However, the purity did not improve much, as both medium and low molecular weight contaminants (45 and 25 kDa respectively) were partially eluted in the elution fraction of the *Pf* protein, indicating that the resolution of the column was probably not sufficient to fully separate the contaminants from the target protein. The pool derived from the 500mM imidazole elution, however, produced a better result. Indeed, the majority of the 45 kDa contaminant was removed during the initial peak (P1, see Figure 3.30a) probably in the form of dimers. Whereas, the *PF14\_0010* was mainly eluted in the subsequent peaks, dimers first (P1 and P2A) and then monomers (mainly P2B). Such separation allowed me to collect, pool and concentrate the P2B and P3 fractions and evaluate both protein content and purity. Overall, 790 $\mu$ g of *PF14\_0010* were recovered (132 $\mu$ g/mL, for a total of 600 $\mu$ L), with a purity level of 65-70% (see Figure 3.31), which represent a 2-fold increase in purity from the previous IMAC-FPLC purification and up to 6-fold increase from the initial two purification protocols. To further



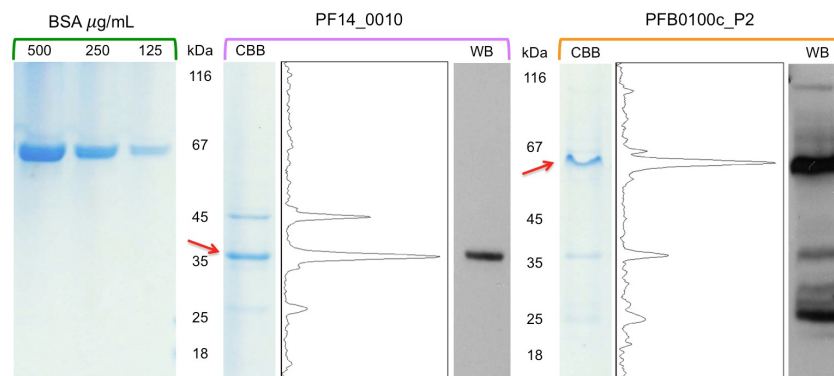
(a) 500mM pool purification chromatogram

(b) Coomassie (*top panel*) and WB (*bottom panel*) analysis**Figure 3.30: Second step FPLC purification, gel filtration.**

**Panel A** contains the chromatogram of the gel filtration. The three lines present in the chromatogram represent the  $A_{280}$ : □;  $A_{215}$ : □. **Panel B** shows the results of the gel filtration of both *PF14\_0010* pools (500mM left, 300mM right, target band is indicated with the red arrow). In particular for the 500mM pool, the fractions are named after the peak analysis visible in the panel A, P1: peak 1. P2A: peak 2A. P2B: peak 2B and P3: peak 3.

evaluate this new two-step purification procedure, I applied a similar protocol to the soluble fraction of *PFB0100c\_P2*. This time, however, the stringency

of the loading/binding step was increased by adding 20mM of imidazole to the soluble fraction. As for the *PF14\_0010*, the sample was first partially purified using  $\text{Ni}^{2+}$  sepharose resin and then further purified by gel filtration. The purest fractions were then pooled, concentrated and electrophoresed alongside a serial dilution of BSA to estimate the amount of *PF14\_0010*. Proteins were then analysed in Coomassie and WB (see Figure 3.31). The purity level obtained for *PFB0100c\_P2* pool was calculated at around 75-80%, with a final protein yield of 470 $\mu\text{g}$  (94 $\mu\text{g}/\text{mL}$ , for a total of 500 $\mu\text{L}$ ). With the help of F.



**Figure 3.31: IMAC-FPLC followed by SEC.**

Summary of the results obtained for the purification of two soluble over-expressed candidates *PF14\_0010* (32 kDa) and *PFB0100c\_P2* (45 kDa). To facilitate the protein quantification and a serial dilution of BSA was used as standard reference (0.5, 0.25 and 0.125  $\mu\text{g}/\text{mL}$ ). The running behaviour of the *PFB0100c\_P2* was not in complete accordance with its protein size (67 kDa vs 45 kDa). A different electrophoretic mobility can be caused by the amount of positive charges on the protein surface, which has been observed with a few other *Pf* proteins in this and in previous internal studies and it has also been reported in the literature (Adda et al., 2009).

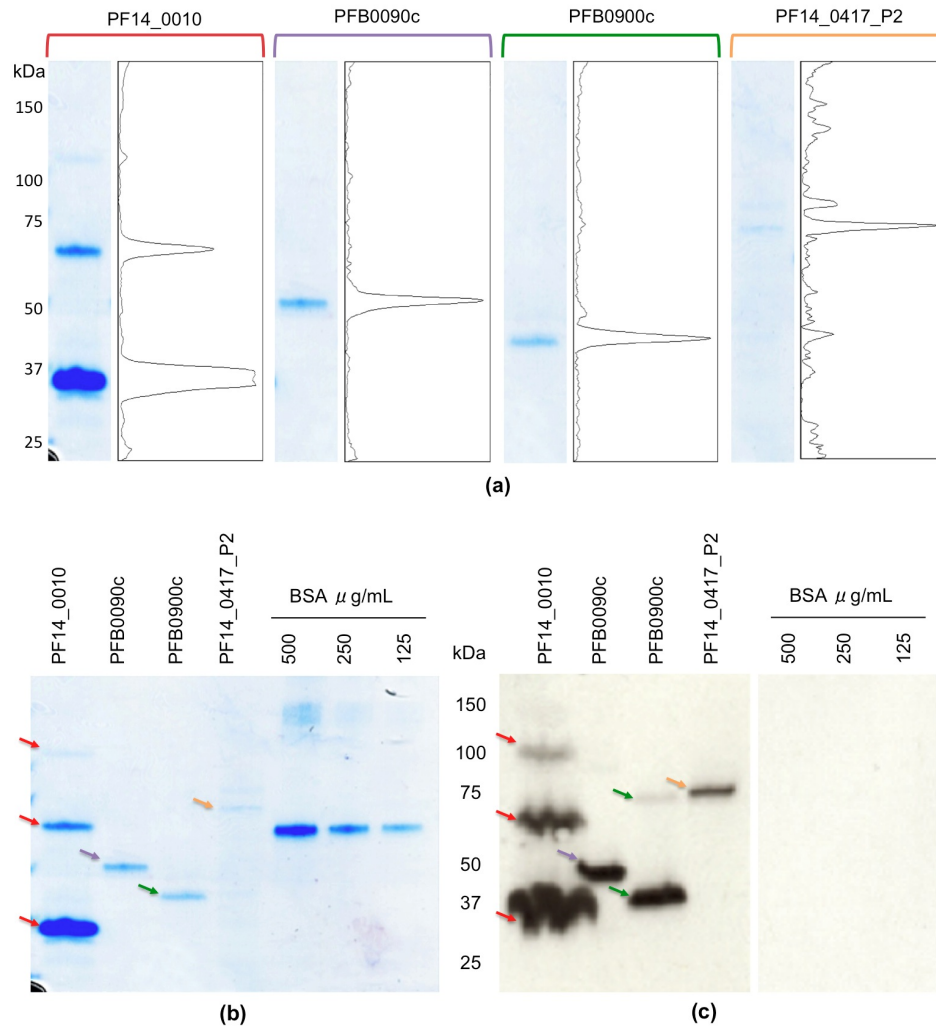
Wolff, eight *Pf* proteins were purified at PX', including *PF14\_0010*. The yield of the purified proteins ranged from 100 to 750  $\mu\text{g}$ . As for the purity level, proteins with higher expression rates (over-expressed) showed higher purity (n.4 candidates, purity average  $\sim 80\%$ ) than the ones with soluble expression visible only on WB and not on Coomassie (n.4 candidates, purity average  $\sim 65\%$ ). As these figures represented a substantial improvement compared to the initial magnetic-beads and gravity flow purifications, I decided to further evaluate the IMAC-FPLC platform. Indeed, if on one hand the protein yield and purity were promising; on the other hand the purity was still below the 90% threshold needed. Additionally, such performances were achieved only with highly expressed soluble proteins of similar sizes, which represent only a minority of the *Pf* candidate list. Hence, when I completed my secondment at PX' and moved back to MtM, I started evaluating the possibility to further improve and simplify the purification protocol.

Using a similar liquid chromatography machine (ÄKTAprimePlus<sup>TM</sup>, a low pressure liquid chromatography (LPLC) I first tried to increase the stringency during the binding to the column. This was initially achieved by decreasing the column bed volume (from 5 mL to 1 mL). Furthermore, the soluble fraction was spiked with twice the amount of imidazole (from 20mM to 40mM). Stringency of the washing steps was also increased from 10/40/80mM to 40/80/120mM. Finally, elution was achieved by 2 sequential steps of imidazole at 250mM and 500mM.

Using the *PF14\_0010* candidate, the new purification protocol yielded a protein purity of 91.3% (see Figure 3.32). This represents over 20% increase in purity level compared to the previous IMAC-SEC FPLC protocol (91% versus 65%). Noteworthy is the fact that the increase in the purification efficiency did not affect the final protein yield (761  $\mu$ g), which was very similar to the one observed after the double step purification (790 $\mu$ g).

Overall, 13 *Pf* candidates were tested using this new protocol out of the 25 tested in large-scale (see Table 3.18). The majority (77%, 10/13) of the candidates was successfully purified. For the remaining 3, the recombinant proteins co-eluted in the washing steps (40/80/120 mM imidazole). Interestingly, thanks to the combination of an optimised expression medium (see Section 3.4.1), and purification procedures, we were not only able to successfully purify 5 out of 6 (83%) of the soluble over-expressed proteins (including *PF14\_0010*), but we were also capable of purifying 5 out of 7 (71%) of those proteins that possessed only partial solubility and limited expression. The protein yield ranged from 72 to 760  $\mu$ g with a protein purity that was always above the 90% threshold (see Figure 3.32 top panel). Hence, the new purification procedure not only allowed us to obtain highly purified *Pf* proteins, but also to produce them in sufficient amount for the microarray analysis (minimum  $\sim$  50 $\mu$ g, see Figure 3.32 bottom panel).

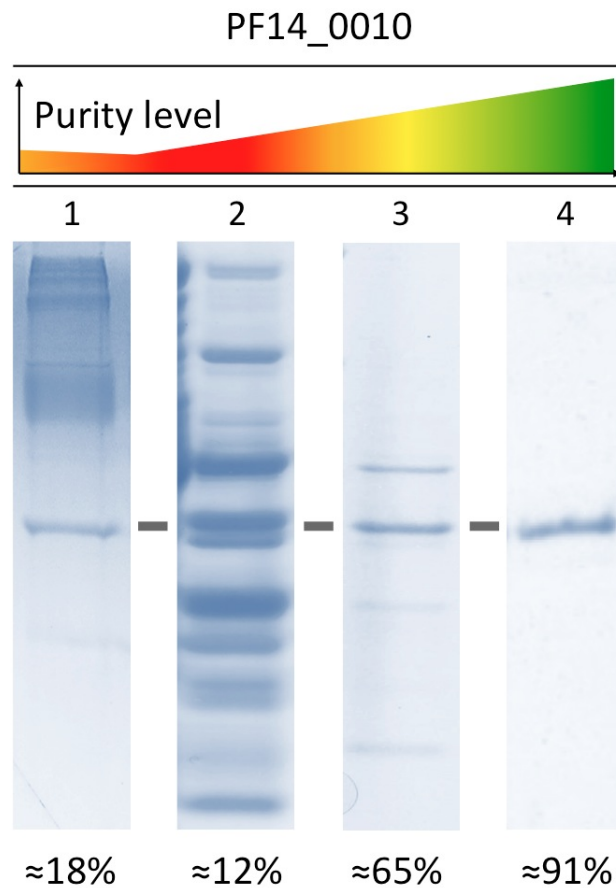
Figure 3.33 illustrates a comparison of the purification results obtained for the *PF14\_0010* candidate with the four different purification procedures. The analysis shows a nearly 8-fold improvement from the initial 18% purity of the magnetic-beads purification, to the 12% of the gravity-flow and the  $\sim$ 65% of the IMAC-SEC FPLC until the final 91% obtained via the IMAC-LPLC.



**Figure 3.32: Purification test FPLC, purity analysis MtM.**

**a)** Coomassie staining of four purified candidates expressed using the large-scale procedure and purified using the new single step protocol. *PF14\_0010*, over-expressed soluble protein, 32.4 kDa size. *PFB0090c*: partially soluble protein, 50.2 kDa. *PFB0900c*: partially soluble protein, 35.8 kDa. *PF14\_0417\_P2* soluble protein, 76.6 kDa. Sample purity was analysed by quantifying the intensity of bands using the ImageJ software (Abramoff et al., 2004).

**b)** Coomassie staining produced to quantify the purified products. The quantification was accomplished by interpolating the intensity of the product-specific band/s to a BSA standard curve, which was run alongside the samples. **c)** WB analysis of the same samples run for the protein quantification. As expected no bands were visible in the three lines containing the BSA serial dilutions. As visible in panel a, b and c both *PF14\_0010* and *PFB0900c* samples showed some degree of multimerisation with extra bands running at double and/or triple the expected size.



**Figure 3.33: Purification feasibility studies.**

Summary of the purification results (Coomassie) of the soluble *Pf* candidate *PF14\_0010*. The candidate was purified using four different procedures: **1)** Immuno-purification using magnetic beads conjugated with a anti-6xHis antibody; **2)** Gravity flow IMAC; **3)** FPLC-IMAC plus gel filtration; **4)** LPLC-IMAC, optimised single step purification. Purification efficiencies are expressed in purity percentage (e.g target/contaminants) and as a color scale with red being the lowest and green being the highest.

### 3.6.1 Discussion

One of the key elements of a sensitive and precise protein microarray immunoassay is the use of high quality purified putative targets, as the introduction of impurities may interfere with the assay at different levels. Indeed, if a plasmodial antigen is contaminated with traces of *E.coli* proteins, the contaminants may alter the ability of the target protein to bind the coated surface, which could result in an poor immobilisation or even the masking of some of the antigenic epitopes. More importantly, the contaminants along with the target molecule may have the ability to be recognised by the detection system. In our case, for instance human serum samples often possess high levels of immunoglobulin against *E. coli* endogenous proteins. Therefore, the use of *E. coli* contaminated antigens in immunoassays will produce strong cross-reactivity towards endogenous bacterial proteins, which will result in a very low level of specificity and sensitivity of the microarray.

As shown in Section 3.7.2, the humoral response against *E. coli* antigens can be partially overcome by pre-absorbing the human serum with bacterial extract. This technique has been already described in the literature (Davies et al., 2005) and its feasibility with malarial antigens was previously demonstrated during my proof-of-principle study that preceded this project (data not shown). However, it is not always the optimal solution. The efficiency of depletion of the *E. coli* specific antibodies has shown to be serum dependent and it cannot guarantee the complete elimination of the cross-reactivity. Thus one of the strategies of my PhD was to fine tune the *in vivo* expression and purification procedure to achieve a sufficient amount ( $\sim 250\text{-}500\mu\text{g}$ ) of high-quality and high-purity ( $\geq 90\%$ ) *Pf* candidates. Hence, the objective of this part of the project was to develop a simple, reproducible and highly efficient purification method capable of isolating *Pf* proteins, maintaining their protein structure, functionality and antigenic properties. The desired system had to be scalable, and easy to implement in a medium size research laboratory. Indeed, once a suitable procedure was developed, we planned to implement the methodology in the laboratories available within the FightMal consortium. A satisfactory protocol, capable of meeting the requirements initially set, was eventually achieved. The new procedure reaches levels of purity above 90%, after only a single step of purification, with a purified protein yield well above the  $\sim 50\mu\text{g}$  initially required for the protein microarray platform.

The final goal of a purity level equal or above 90% has been very challenging

to achieved and required a long "try and fail" approach that started initially with the evaluation of an immuno-purification via magnetic beads conjugated with a monoclonal anti-6xHis tag antibody. The limited purification degree observed with this initial procedure can be justified by the fact that none of the assay conditions were experimentally optimised. In particular, the stringency of the elution step (pH 2.5) probably caused the co-elution of fragments of the capture antibody (4D11) together with antigen. Co-elution of the capturing antibody can be circumvented either by crosslinking the antibody to the magnetic-beads or by optimising the elution conditions. The direct coupling of the capture antibody to the beads can be achieved at least in two ways. The first one is by binding the primary amino-groups of the 4D11 to an aldehyde-activated beads (AminoLink Plus Coupling Resin and Immobilisation Kit, Thermo Scientific n.44894). The second one is by covalently binding the *Fc*-portion of the antibody to protein A or protein G resins (Pierce Crosslink IP, Thermo Scientific n.26147). As for the optimal conditions for the release of the target protein, they can be optimised by testing a variety of elution buffers until a gentle but effective antigen release is obtained. The most common elution buffer may contain an alkaline ( $\sim 10$  pH) or acidic ( $\sim 3$  pH) solution, a high ionic strength or a competing ligand/analog that can displace the antigen (i.e. 6xHis epitope), (Pierce, 2009). Even if both approaches can potentially improve the efficiency by which the antigen-antibody binding interaction is disrupted, they can not guarantee the certainty that some fragments of the capture antibody are not co-eluted together with the target protein. This represents a major drawback of the approach. Even a minimal contamination of our recombinant proteins by the 4D11 could hamper the possibility to evaluate/normalise the quantity of the antigens immobilised onto the microarray slide. Indeed, this process will use the same anti-6xHis mouse antibody as a primary detection agent, followed by an anti-mouse IgG antibody used as a secondary detector. In these circumstances, the signal coming from the spotted targets will have a double source: the 4D11 correctly bound to the recombinant proteins and the undefined quantity of 4D11 fragments co-eluted with them. If the partially purified *Pf* proteins contained traces of the antibody that co-eluted during the purification step, this would invalidate any attempt to determine the amount of antigen in the immunoassay.

On these bases, although the immunoaffinity method showed high specificity (e.g no or little bacterial contaminants present) and limited running cost (the capture antibody was produced in-house), we decided to abandon it in favour of an alternative procedure not relying on the 4D11 antibody. Conse-



quently, I shifted my attention towards the use of an IMAC platform, which has been shown in the literature to produce high purity grade proteins (Graslund et al., 2008; Block et al., 2009).

I started by evaluating both the gravity-flow and semi-automated FPLC IMAC procedures. The gravity-flow platform, although it is the cheapest option of the two, showed many intrinsic limitations that could not justify its implementation on a large-scale. The procedure resulted slow ( $<0.5$  mL/min flow rate), laborious and unreliable. Most of the eluted *Pf* proteins were contaminated by a number of *E. coli* endogenous proteins, some of which cross-reacted with our anti-6xHis antibody (bands visible in WB, Figure 3.27). Additionally, I observed that relevant portions of recombinant proteins were lost due to a low column retention rates (protein present in the flow-through and the 50mM washing step), which was unexpected, as we decided to set the initial purification conditions to a low-mild stringency. The volume of resin was enough to capture over 15 mg of 6xHis-tag protein. The loading buffer was not spiked with imidazole to avoid any competition with the protein target. Nonetheless, the recovery was sub-optimal.

Being a low-cost platform, the gravity flow purification has only a few parameters that can be controlled during the procedure. The operator, for instance, has only a limited control on the flow rate, which is essential for an efficient binding, or on the quality of the self-packed column (i.e. resin density and presence of air-bubbles). Therefore, this approach is better suited for the purification of highly over-expressed soluble proteins (e.g. 5-10 mg per L of bacterial culture), where the recovery rate is not an issue. It is relatively easy to set up and it does not require expensive chromatography equipment. However, it is not optimal for the purification of proteins, which tend to express at a low level (e.g. *Pf* proteins), as protein recovery is limited.

Nonetheless, the information gained from the gravity-flow trial (i.e. behaviour of both nickel-charged resins and our 6xHis-tag proteins), helped to set up a revised purification procedure. The new protocol was based on the use of a semi-automated FPLC system, which allowed me to fully regulate both antigen-resin interaction (e.g. by controlling the flow-rate) and the stringency of the elution (e.g. by using either a gradient or step elution).

The use of the FPLC improved the purification yield but the purity (although better than the previous procedure) was still below the 90% needed. However, thanks to the relatively high yield of purified proteins obtained, I was able to implement an additional purification step based on size exclusion chromatography. Unfortunately, if on one hand the gel filtration helped to

achieve a 2-fold increase in purity, on the other hand it increased both the the complexity/cost of the procedure and the amount of protein lost.

An improvement was observed when the resin bed volume was decreased from 5 mL to 1 mL. This can be explained by the fact that the excess of nickel-charged resin available in the larger format (over 75 mg for the 5 mL column) favoured the binding of unspecific histidine-rich bacterial proteins. A second improvement was obtained by increasing the stringency of protein binding to the column via the addition of imidazole. Nonetheless, a number of endogenous *E. coli* contaminants were still co-eluted with our protein of interest. In particular, a 45 kDa bacterial protein seemed to be unaffected by the increased stringency (see band in the 500mM imidazole elution, Figure 3.29). Even more importantly, the size difference between the two proteins was not sufficient to allow a complete separation between the two via the size exclusion chromatography.

Overall, the improvements observed with the IMAC-SEC FPLC procedure demonstrated that the methodology was still a promising tool. However, the use of a two-step purification was not ideal for our project. To fully exploit the purification power of this procedure, a large quantity of high-quality soluble protein is needed, which it is rare when *Pf* are expressed in *E. coli*. Additionally, the purification is not only time consuming (1 day longer than the standard single step) but it requires costly chromatography equipment. Finally, due to the limited resolution of the size-exclusion power of the SEC, this approach does not guarantee an efficient depletion of those bacterial contaminants with a similar size to the protein of interest.

Regarding the bacterial contaminants, unfortunately, due to time limitations it was not possible to fully characterised them by N-terminal sequencing. However, according to the literature, some of these bacterial proteins, and in particular the 25 and 45 kDa bands, could represent two stress-related proteins known to be co-purified in nickel based IMAC systems (Howell et al., 2006; Bolanos-Garcia and Davies, 2006).

The discovery of stress induced histidine-rich contaminants led us to further investigate whether it was possible to reduced their presence by decreasing the level of stress under which the RGB were synthesising our proteins. Such investigation started as soon as I was back to the MtM facilities at Imperial College London. Under the supervision of Dr. Dottorini and with the help of Dr. Trivelli we developed the TM medium which was capable of reducing the burden sustained by the RGB during the heterologous expression of plasmodial proteins. As described in the previous Section (3.4.1) the new expression

conditions improved not only the protein yield and solubility, but did also promote a more sustainable bacterial growth.

The combination of a highly efficient expression medium together with an optimised IMAC-LPLC proved to be a success. When I tried to purify the same *PF14\_0010* candidate under the new conditions, the amount of contaminants co-eluting with the protein target were significantly reduced, if not completely eliminated. The system produced a nearly 30% increase in the candidate purity. Such performances were eventually confirmed when a total of 15 recombinant *Pf* candidates were expressed and purified. For the majority of the candidates a degree of purity  $\geq 90\%$  and a final protein yield ranging from 72  $\mu\text{g}$  to over 760  $\mu\text{g}$  was observed. Additionally, we demonstrated that this new procedure is capable of purifying not only soluble proteins that are over-expressed, but also proteins that resulted only partially soluble with a limited level of expression rate (only visible on WB).

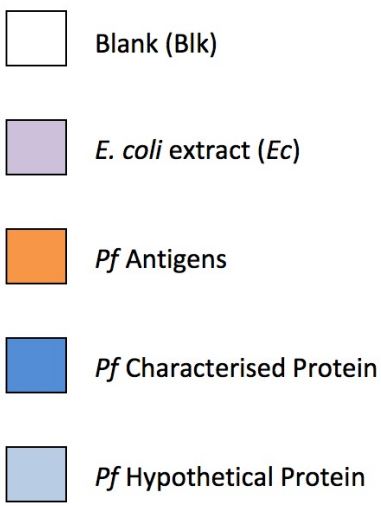
The combination of an optimised expression system together with a more efficient single step purification resulted in high-quality and high-purity *Pf* putative antigens. The main improvements were driven by the decreased number of stress related histidine-rich contaminants and by fine tuning the purification parameters, which overall increased both the yield and the purity of the purified antigens. Overall, protein production efficiency was around 66% (10/15), which is more than twice the percentage (30.2%) obtained by Vedadi et al. (2007) during their attempt to express and purify over 1000 apicomplexan genes, and more than three times higher than the one (18.7%) observed by Mehlin et al. (2006). The authors of the latter study, however, obtained a higher protein yield (i.e. between 0.9 and 406 mg per L of culture) for 63 out of the 337 *Pf* genes that they successfully expressed in *E. coli*.

## 3.7 Protein microarray feasibility study

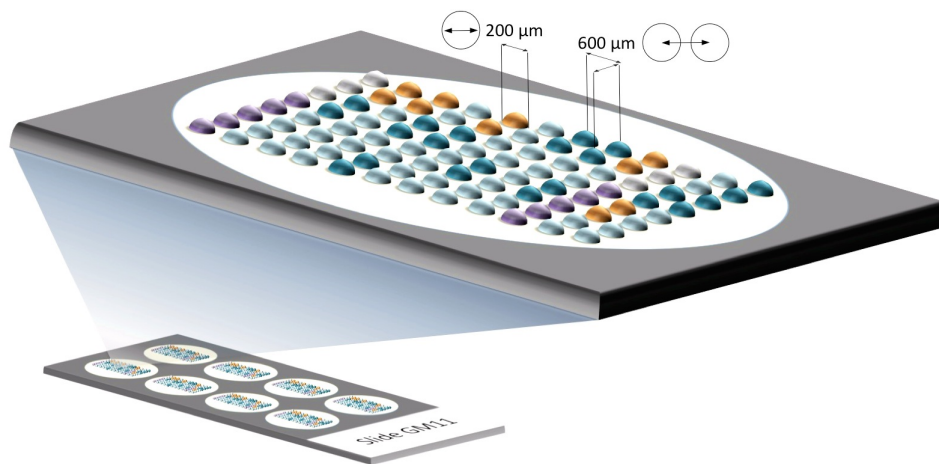
At the beginning of the my PhD, I performed a proof-of-principle study aimed at evaluating the capabilities of the protein microarray platform and the cell-free expression system. The selected candidates were chosen from a list of putative antigens produced by UNIPG during a bioinformatic study done in 2006. The plasmids harbouring the 40 candidates were created during my MSc project, which I carried out in the Crisanti's laboratory using a pEXP5-NT/TOPO-TA vector with a 6xHis tag at the N-terminus (but not c-MYC at the C-terminus). During my PhD I expressed the *Pf* candidates using an *in vitro E. coli* system (see Section 2.5.1). The recombinant antigens were printed onto aldehyde-coated multi-well microscope slides creating a microarray chip of 40 *Pf* antigens, which included: 5 well-documented *Pf* malaria antigens AMA-1, MSP-2, Exp-1, MSP-5 and ETRAMP-5 (as controls); 11 characterised proteins (genes/proteins experimentally studied or with known homologies) and 24 proteins derived from hypothetical genes. The chip was used as an immunoassay substrate for assessing both the expression levels (see Section 3.5.1) and the candidate antigenicity towards the human immunoglobulins G present in either African or European (negative control) serum samples.

### 3.7.1 Production of protein microarray chips

During this proof-of-principle study, 112 protein arrays were produced by spotting the expression reactions onto 8-well aldehyde-activated glass microscope slides. The microarrays used were designed as 8 x12 and 8 x13 matrices (96-104 discrete spots with an array density of  $\sim 280$  spots/cm<sup>2</sup> and a spot diameter  $\sim 200\mu\text{m}$ , Figure 3.34) allowing the incorporation of the 40 candidate antigens in duplicate, *E. coli* extract as a negative control and replicates of the printing buffer as a blank (Figure 3.34).

	1	2	3	4	5	6	7	8	
1	<i>Ec</i>	<i>Ec</i>	<i>Ec</i>	<i>Ec</i>	<i>Ec</i>	Blk	Blk	Blk	
2	1	1	11	11	21	21	31	31	
3	2	2	12	12	22	22	32	32	
4	3	3	13	13	23	23	33	33	
5	4	4	14	14	24	24	34	34	
6	5	5	15	15	25	25	35	35	
7	6	6	16	16	26	26	36	36	
8	7	7	17	17	27	27	37	37	
9	8	8	18	18	28	28	38	38	
10	<i>Ec</i>	<i>Ec</i>	<i>Ec</i>	<i>Ec</i>	<i>Ec</i>	Blk	Blk	Blk	
11	9	9	19	19	29	29	39	39	
12	10	10	20	20	30	30	40	40	

(a)



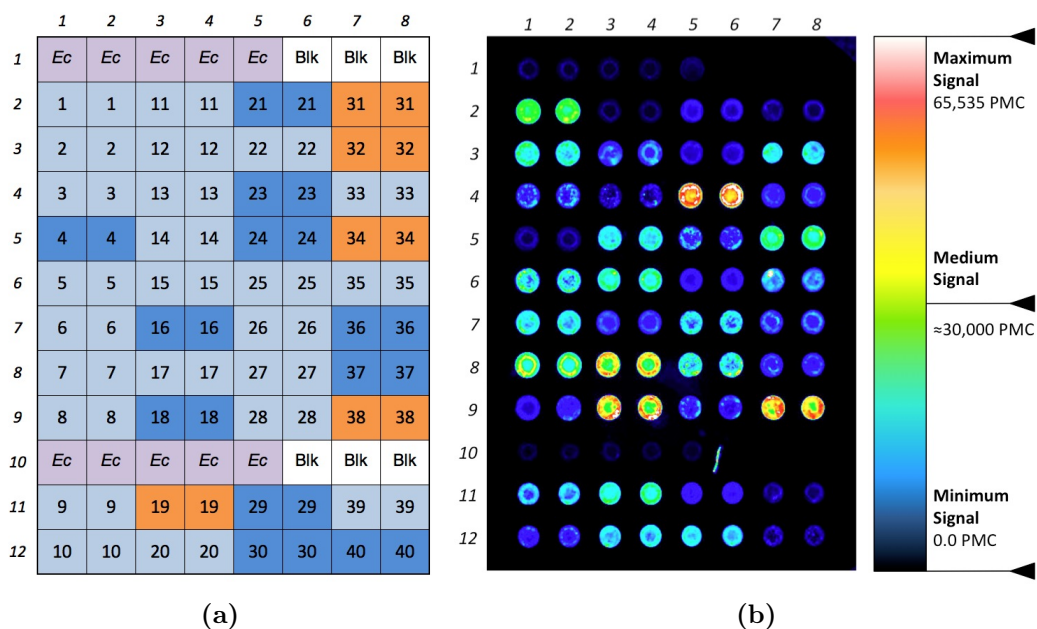
(b)

**Figure 3.34: Microarray printing scheme and features.**

**a)** The panel represents the printing scheme of the first generation chip (8x12 matrix). The scheme describes the position of each protein solution in the array. The numbers indicate the printing location of the various antigens as described in Table 3.20. **b)** The panel shows a magnified representation of an 8x12 array printed on an 8-well slide (average spot diameter = 200 μm; pitch = 600 μm).

### 3.7.2 Processing of protein microarray chips

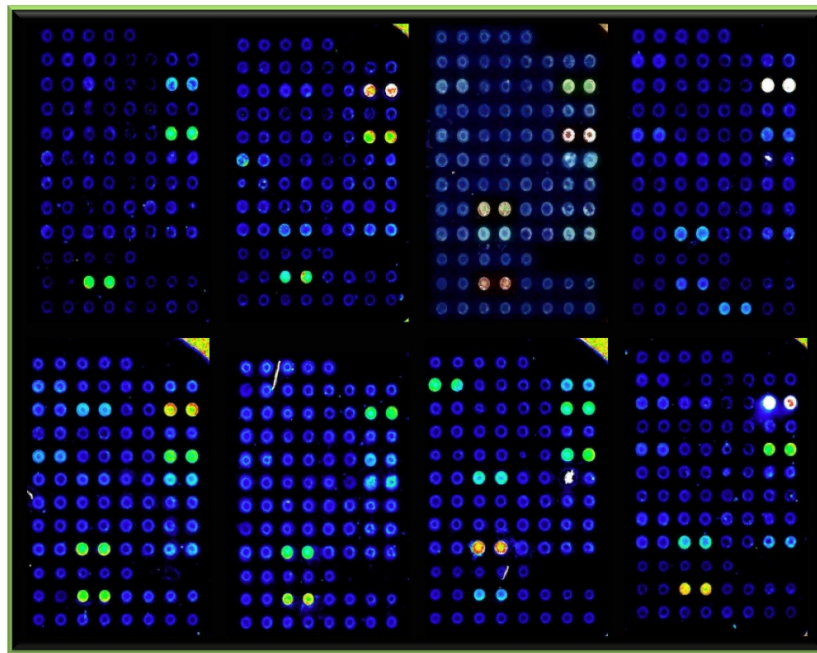
The arrays were first used to assess the expression level of the various recombinant proteins. This was performed by incubating the chip with an anti-6xHis mouse IgG monoclonal antibody. As shown in Section 3.5.1, the spotted antigens showed a quite heterogeneous reactivity against the 4D11 antibody, indicating highly variable expression levels (see Figure 3.35). The candidate's signals were from 1.1 to 67-fold higher than the negative control (i.e. *E. coli* extract-no expression plasmid). Both blank and *E. coli* extract spots emitted either a low or no signal, indicating a high level of specificity of the system.



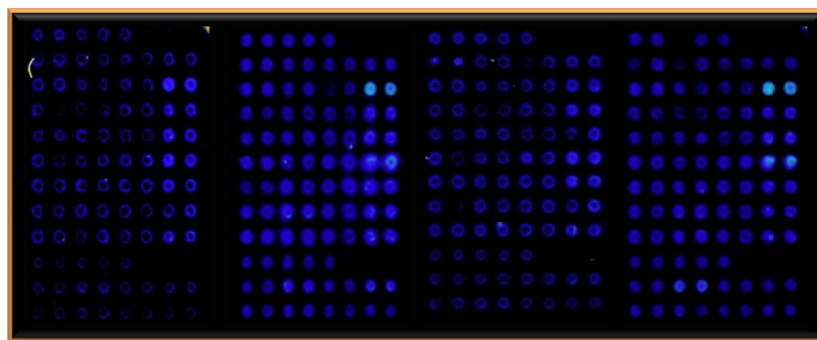
**Figure 3.35: Microarray printing scheme and features.**

**a)** The panel shows the printing scheme of the 8x12 chip with a corresponding image (panel b) of the array having been developed with the 4D11 anti-6xHis-tag antibody. The candidate expression rate was evaluated using the 4D11 antibody to detect the amount of expressed target protein present in spots of the array. **b)** As visible from the reactivity in panel b the candidate expression level resulted quite heterogeneous. In particular, the expression levels, which were as a signal:noise ratio (S/N) (i.e. the signal of the candidate divided by the signal of the negative control), ranged from 1.1 (antigen 11, intensity= dark blue) to 67-fold (antigen 23, intensity=red-white) higher than the negative control (*Ec*, intensity=dark-blue/black). PMC: Photo Multiplier Count, minimum signal= 0.0 PMC (black); maximum signal= 65,535 PMC (white).

The protein arrays were then used to reveal the presence of specific IgG antibodies in human sera against the panel of printed candidate antigens. In order to achieve this, the arrays were probed with 33 malaria iper-immune African serum samples (see Figure 3.36a) and 16 European “negative” serum samples to assess the specificity and sensitivity of the system (see Figure 3.36b).



(a) African sera



(b) European sera

**Figure 3.36: Reactivity of serum samples against *Pf* antigens.**

**a)** Sera from 33 adult individuals (males and females ranging in age between 16 and 65) living in Gambia were exposed to 8x12 arrays containing the arrayed malaria antigens. A representative sample of array images from 8 sera is shown. **b)** As a control, 16 European sera (images from 4 are shown) were utilised to confirm the specificity of the assay and to determine positivity cut-offs. The mean + 2SD of the S/N ratios measured for the 16 European sera against each antigen were used to define cut-off values for each candidate.

It is important to note that all the serum samples were initially pre-absorbed against an *E. coli* extract preparation to overcome issues caused by printing the crude expression reaction mixtures. Using this approach, antigen-specific IgG were detected for the majority of selected candidate antigens. 27% (11 out of 40) of antigens were frequently recognised, with reactivity recorded from at least a third of the adult African sera tested (e.g. 11 sera or more); 45% (18 out of 40) were recognised by between 3 and 10 serum samples; 20% (8 out of 40) were recognised by 1 or 2 sera, leaving their antigenicity interpretation uncertain and , finally,  $\sim 8\%$  (3 out of 40) of antigens were recognised by none of the sera (see Table 3.24).

Candidate Positivity	Hits	<i>Pf</i> Malaria antigens	Characterised proteins	Hypothetical proteins	Total
Highly recognised	11 to 33	5	2	4	11
Medium recognised	3 to 10	0	5	13	18
Lowly recognised	1-2	0	3	5	8
Not recognised	0	0	1	2	3

**Table 3.24: Serum sample reactivity data.**

Summary of the reactivity of the selected candidates expressed in terms of hits (e.g. number of times that the selected candidate was recognised by different sera. The first group (11-33 hits) contains all those antigens recognised by at least 11 African sera or more (max 33); the second group (3-10 hits) is formed by those antigens that were reactive with a minimum of 3 and a maximum of 10 African sera. The third group (1-2 hits) contains those proteins with uncertain interpretation, as they were reactive only once or twice. Finally the fourth and last group (0 hits) is formed by the *Pf* that did not react with any of the African sera.

As a strategy to overcome the heterogeneous nature of the expression levels and its influence in the antibody recognition, the results generated for discrete antigens were further analysed by combining serum sample reactivity data (i.e. number of times the antigen is recognised by an African serum, n. hits) with expression level data (i.e. S/N values obtained with the anti-6xHis mouse IgG monoclonal antibody) generated upon probing chips with the monoclonal mouse anti-6xHis antibody.

As a result of this signal normalisation I obtained a new reactivity value, “reactivity per unit of expression” (e.g. PFB0300c, MSP2: 28 serum samples positive and S/N expression level 29.6; thus the reactivity per unit of expression was equal to  $28/29.6=0.95$ ). Using this approach the reactivity of the antigens were clustered in three groups: highly reactive (“+++”) value  $> 0.9$ ; medium reactive value (“++”) ranging from 0.3 to 0.9; low reactive value (“+”) ranging from 0.1 to 0.3 (see Table 3.25). Notably, the reactivity for 2 of these antigens was extremely high: AMA-1 Hits/Exp =7.3 and MAL7P1.170



Reactivity per unit of expression	<i>Pf</i> Malaria antigens	Characterised proteins	Hypothetical proteins	Total
High (+++)	4	2	4	10
Medium (++)	1	5	7	13
Low (+)	0	3	11	14
Negative	0	1	2	3

**Table 3.25: Antigen reactivity per unit of expression.**

Summary of the reactivity of the selected candidates in terms of number of reactivity (hits, n. of positive serum) per unit of expression (S/N). The normalisation of the antigen signal towards the expression level should guarantee the same probability of antigenicity for those proteins expressed at a lower level (less protein, less epitopes) as for the ones highly expressed (more protein, more epitopes). Categorisation criteria based on the hits/expression (Hits/Exp) ratio: High= Hits/Exp >0.9; Medium= 0.3>Hits/Exp<0.9; Low= 0.1>Hits/Exp<0.3; Negative= Hits/Exp <0.1.

Hits/Exp= 8.7 (the reactivity values of the remaining candidates ranged from 0 to 1.93). Medium-high reactivity (Hits/Exp >0.3) was observed in 57.5% (23 out of 40) of the printed antigens. In particular, such categorisation grouped four out of five *Pf* malaria antigens in the most reactive group, with values ranging from 0.9 (ETRAP 5) to 7.27 (AMA-1). Remarkably, two *Pf* characterised proteins (*MAL7P1.170* and *PFF0615c*) and four hypothetical proteins (*PFC0085c*, *PFE1600w*, *PFB0926c*, *PFE0050w*) were clustered in the same highly reactive group. Low reactivity (Hits/Exp <0.3, “+”) was instead observed for 35% (14 out of 40) of the printed targets, most of which were hypothetical proteins (11/14). A detailed summary table showing the reactivity per unit of expression for each antigen is showed in Table 3.26.

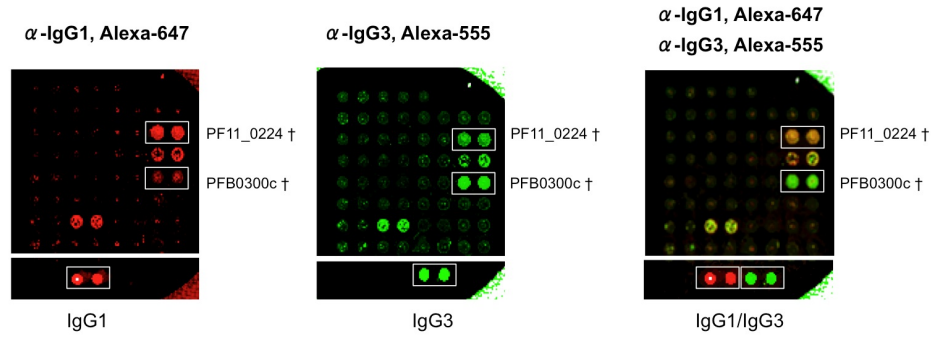
As the final step of the microarray feasibility study we decided to optimise a protocol that would also allow for the detection of immunoglobulin subtypes (see Figure 3.37a). Indeed, as shown in the literature, the presence of cytophilic antibodies (IgG1 and and IgG3) in combination with white cells correlates with an increase capability of parasite clearance. On this basis I decided to assess a protocol capable of evaluating the immune profiles of IgG1 or IgG3 cytophilic subclasses to the reactive antigens in the same cohort of sera used in the previously described study (n.33 African and n.13 European pre-absorbed serum samples).

Candidate	Protein	ORF Size	N. Hits	Expression Level	Reactivity per unit of expression	
ID	Info, <i>Plasmo DB</i> database	kDa	n.33	S/N	Hits/Exp	Level
<i>PF11_0344</i> †	apical membrane antigen 1 precursor, AMA-1	68.5	13	1.8	7.27	+++
<i>PF11_0224</i> †	circumsporozoite related antigen, Exp-1	17.9	28	20.8	1.34	+++
<i>PFB0300c</i> †	merozoite surface protein 2 precursor, MSP-2	30.0	28	29.6	0.95	+++
<i>PFE1590w</i> †	early transcribed membrane protein, ETRAMP 5	20.0	30	33.3	0.90	+++
<i>PFB0305c</i> †	merozoite surface protein 5, MSP5	30.0	21	59.0	0.36	++
<i>MAL7P1.170</i> ◊	ring stage expressed protein	32.3	14	1.6	8.66	+++
<i>PFF0615c</i> ◊	<i>Plasmodium falciparum</i> membrane protein <i>Pf</i> 12 precursor	38.3	7	7.5	0.93	+++
<i>PFE0080c</i> ◊	rhoptry-associated protein 2, RAP2	43.9	2	3.3	0.60	++
<i>PF14_0010</i> ◊	glycophorin binding protein-related antigen	33.9	4	6.8	0.58	++
<i>PF11_0040</i> ◊	early transcribed membrane protein, ETRAMP 11.2	10.5	25	56.5	0.44	++
<i>PF13_0197</i> ◊	merozoite surface protein 7, MSP7	38.7	9	23.8	0.38	++
<i>PFD1170c</i> ◊	RESA-like protein, truncated	34.1	3	9.4	0.32	++
<i>PFD1180w</i> ◊	trophozoite antigen r45-like protein, truncated	34.7	2	8.4	0.24	+
<i>PF14_0358</i> ◊	41-2 protein antigen precursor	20.4	3	13.5	0.22	+
<i>PFI1565w</i> ◊	conserved protein	18.9	2	67.2	0.03	+
<i>PF14_0678</i> ◊	exported protein 2, Exp-2	31.7	0	3.6	0.00	-
<i>PFC0085c</i> *	hypothetical protein	35.0	11	5.7	1.93	+++
<i>PFE1600w</i> *	hypothetical protein	56.1	4	2.4	1.67	+++
<i>PFB0926c</i> *	hypothetical protein	29.9	12	10.1	1.19	+++
<i>PFE0050w</i> *	hypothetical protein	28.7	15	16.2	0.93	+++
<i>PF13_0275</i> *	hypothetical protein	31.5	20	25.8	0.78	++
<i>PFB0106c</i> *	hypothetical protein	32.0	5	6.9	0.73	++
<i>PF14_0290</i> *	hypothetical protein	38.3	3	4.3	0.69	++
<i>PFA0135w</i> *	hypothetical protein	30.5	2	3.1	0.64	++
<i>PF10_0013</i> *	hypothetical protein	25.5	6	12.4	0.48	++
<i>MAL6P1.15</i> *	hypothetical protein, conserved in <i>Pf</i> , PFF1525c	20.8	9	26.8	0.34	++
<i>PFI1785w</i> *	hypothetical protein	36.6	2	6.0	0.33	++
<i>PFI1755c</i> *	hypothetical protein	36.0	10	35.1	0.28	+
<i>PFA0680c</i> *	hypothetical protein, conserved in <i>Pf</i>	20.6	8	36.0	0.22	+
<i>PFC0090w</i> *	hypothetical protein conserved	28.4	4	19.8	0.20	+
<i>PFF1060w</i> *	hypothetical protein	23.4	5	26.8	0.19	+
<i>PFI1270w</i> *	hypothetical protein	24.0	4	24.3	0.16	+
<i>MAL7P1.174</i> *	hypothetical protein	35.0	4	24.5	0.16	+
<i>PFA0670c</i> *	hypothetical protein	26.8	3	21.7	0.14	+
<i>PFB0105c</i> *	hypothetical protein	33.3	1	9.8	0.10	+
<i>PFB0888w</i> *	hypothetical protein	7.5	5	49.3	0.10	+
<i>PF14_0742</i> *	hypothetical protein	20.0	2	21.6	0.09	+
<i>PFE0395c</i> *	hypothetical protein	38.5	1	20.3	0.05	+
<i>PFC0555c</i> *	hypothetical protein	25.7	0	8.9	0.00	-
<i>PFI1770w</i> *	hypothetical protein	32.8	0	1.1	0.00	-

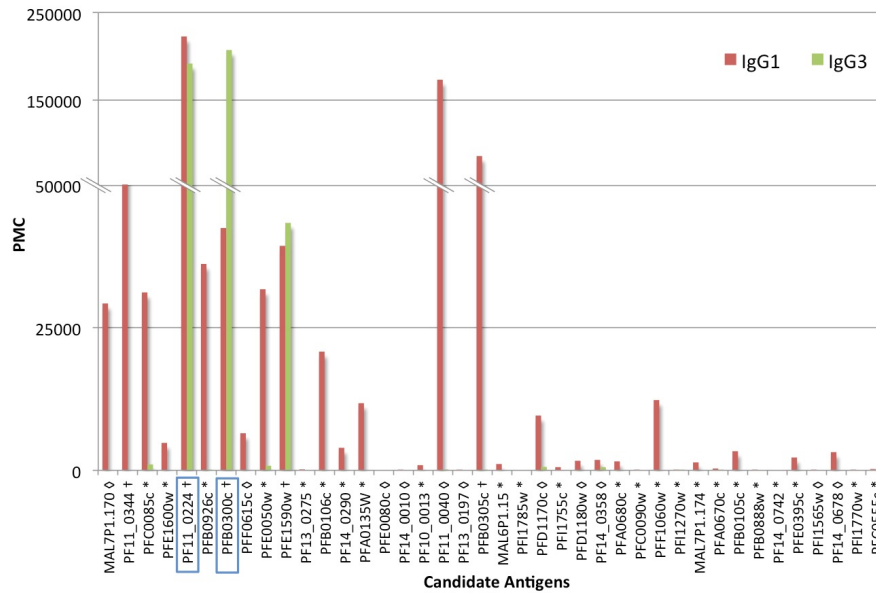
**Table 3.26: Antigen expression and reactivity levels.**

The table shows a summary of the reactivity of the selected candidate antigens. From the left: 1<sup>st</sup> column, gene ID of the candidates; 2<sup>nd</sup> column, gene information in GeneDB/PlasmoDB databases (accessed May 2008); 3<sup>rd</sup> column, protein size expressed in kDa; 4<sup>th</sup> column, number of hits (max n. 33); 5<sup>th</sup> column, expression level expressed in S/N ratio produced probing the chips with an anti-HIS antibody; 6<sup>th</sup> and 7<sup>th</sup> columns reactivity level (x) expressed in hits per unit of expression, High:  $x > 0.9$  (“+++”, ■); Medium:  $0.3 > x < 0.9$  (“++”, ■); Low:  $0.1 > x < 0.3$  (“+”, ■); Neg:  $x < 0.1$  (“-”, ■). (†) *Pf* Malaria antigen; (◊) Characterised protein; (\*) Hypothetical protein.

The response visualised was mainly IgG1 driven, as expected (see Figure



(a) Detection of cytophilic IgG1/IgG3 subclasses

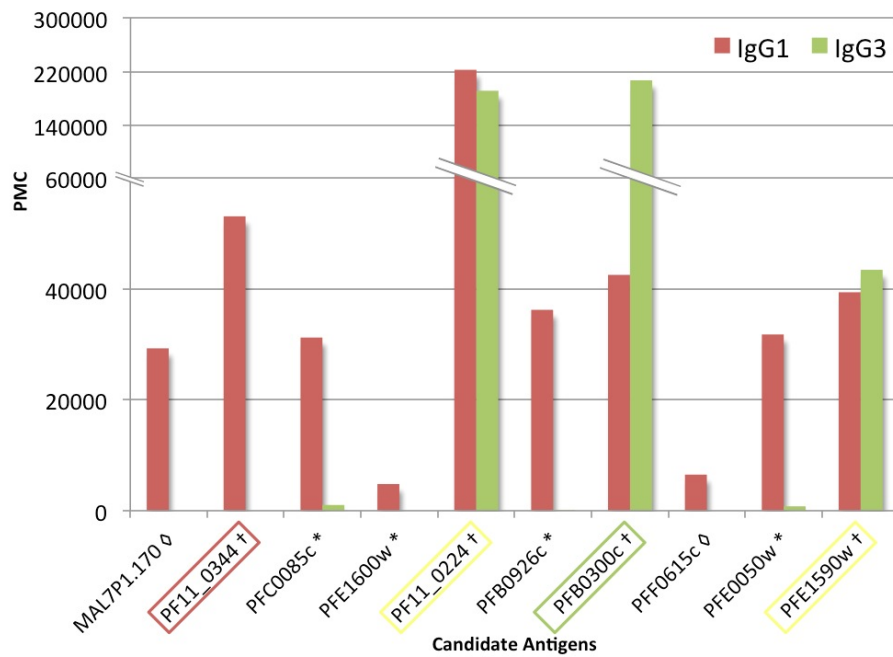


(b) Immune profile of cytophilic IgG subclasses

**Figure 3.37: Serum antibody reactivities of sera from heavily exposed Gambian adults.**

**a)** The figure shows three partial views of a chip probed with an African serum and processed using an optimised protocol for the simultaneous detection of cytophilic IgG1 and IgG3 subclasses. The core of the detection system was an optimised mixture of two different fluorophore-labelled antibodies,  $\alpha$ -IgG1-Alexa-647 and  $\alpha$ -IgG3-Alexa-555. At completion of the processing the slide was read at two different wavelengths 532/575 and 635/670 nm. The right panel represent the composite image of two readings. Highlighted in the white boxes are two major *Pf* malaria antigens (*PF11\_0224*: Exp-1, *PFB0300c*: MSP-2) and the positive controls (purified IgG1 or IgG3). **b)** The figure shows the average IgG1 (■) vs IgG3 (■) reactivity of antigens in terms of raw signal. Exp-1 and MSP-2 reactivity histograms are boxed in bright blue (□). PMC: Photo Multiplier Count. Due to the necessity of using the two laser sources, the slides were read and analysed using a different scanner instrument, hence the reactivity scale is different from the one describe in Figure 3.35).

3.37b). However, the immune response observed against MSP-2 was clearly IgG3 dominant, whereas, targets such as Exp-1 or ETRAMP-5 showed an approximately even split between antigen-specific IgG1 and IgG3 reactivities (see Figure 3.38).



**Figure 3.38: Immune reactivity, IgG1 vs IgG3 subtypes.**

Summary of the reactivity towards the cytophilic IgG1/IgG3 subclasses for a selected panel of antigens. The histograms describes the average IgG1 (■) vs IgG3 (■) reactivity of a selected panel of antigens in terms of raw signal (PMC: Photo Multiplier Counts). Reactivity for AMA-1 (*PF11\_0344*) was clearly IgG1 driven (□), whereas the Exp-1 (*PF11\_0224*) and ETRAMP-5 (*PFE1590w*) immune responses were split even between IgG1 and IgG3 (□). MSP-2, represented the best example of IgG3 driven response (□). Reactivity values were not normalised for expression level, however, the IgG subclasses reactivity bias should not be affected by the expression value, as the number of epitopes available for both IgGs were identical. For the same reasons no assumptions were made on the basis of absolute reactivity values for this set of data.

### 3.7.3 Discussion

As an initial feasibility study of the microarray platform, I investigated the protein microarray production and detection. The study consisted of expressing (via *in vitro* cell-free *E. coli* expression) and spotting recombinant proteins as crude *E. coli* extracts to aldehyde-activated glass microscope slides. These arrays then formed the substrate of an immunoassay and were used to profile serum reactivities in adult African serum samples pre-absorbed against *E. coli* proteins. The data obtained highlighted the power of the overall approach, and in particular the power of the protein microarray, in rapidly assessing a large number of candidates in parallel. Indeed, in a matter of a few days, 33 African serum samples from adults heavily exposed to malaria and 16 negative European sera were successfully tested for antigen-specific IgG (total, IgG1 and IgG3) against 40 *Pf* candidate antigens. Using a relatively short development process ( $< 2$  h) and minute amounts of serum ( $2 \mu\text{L}$ ), we were able to develop a powerful immunoassay capable of revealing the specific reactivity profile of distinct sera against candidates in a truly parallel, analyte-to-analyte comparison that is both sensitive and specific. It is worth noting that if a larger number of proteins had been available (e.g. 100s or 1000s of parasite antigens) the time taken to assess the 33 serum samples against this increased number of antigens would have remained the same. Importantly, the pre-absorption step very successfully overcomes the issues caused by printing the crude expression reaction mixtures. The use of the various background and negative control spots allowed for very effective discrimination between unspecific and specific signal. Remarkably, of the 40 candidate antigens printed, the majority (37/40) were detected by antigen-specific IgGs, leaving the antigenicity interpretation uncertain for only a minority of low reactive antigens (8/37, recognised only by 1 or 2 serum samples). Even more interestingly is the fact that 11 of these 37 antigens were commonly recognised by more than a third of the African samples, suggesting that they could represent immune dominant antigens. Furthermore, the system was able to reveal not only the existence of 24 new proteins (annotated in the database as hypothetical) but also confirmed antigenicity in adults for at least 22 of them. However, it is important to consider that 5 of these 22 proteins showed a level of reactivity (detected by only 1 or 2 sera) that requires more detailed analysis to reach a definitive decision on their antigenicity.

These findings together with the recent publications of a similar research carried out by Crompton et al. (2010) and Trieu et al. (2011) further corrob-

rate the power of the microarray format for the identification of novel antigens as new possible vaccine candidates. The Crompton et al. (2010) and Trieu et al. (2011) studies, in particular, made use of the *in vitro* expressed protein microarray platform developed by Felgner P. L. (see Section 1.4.2), in which around 23% (~1200 proteins) of the *Pf* proteome was expressed in a *E. coli* cell-free system and spotted as crude extract. In both studies the teams managed to identify a *Pf* antigenic signature which correlates with the malaria immunity. In particular, Trieu et al. (2011) found that the reactivity towards a panel of 19 parasite antigens strongly correlates with malaria sterile protection induced by irradiated sporozoites. As for the study of Crompton et al. (2010), they found that the reactivity towards 49 *Pf* proteins was significantly higher in children (age 8–10 years) with no malaria symptoms.

As far as the normalisation step is concerned, although it is not critical for investigations based on the assessment of binary antibody responses (positive vs negative), a normalisation analysis will guarantee that all the candidates are spotted in sufficient amounts to be detectable. This will also improve the ability to declare antigens that are not detected as being “non-antigenic”, something that is not possible with a system in which expression is highly heterogeneous (e.g. *in vitro* expressed products) and recombinant proteins possess just one N-terminal tag. This is why, one of the key requirements set for the FightMal project was to produce high-quality purified recombinant *Pf* proteins containing both a N- and C-terminal tag. Tags that will eventually be used to normalise the protein content of each spot. Indeed, the ability to normalise the amount of printed antigen will open up the opportunity to perform quantitative inter-analyte comparisons with constructed arrays. Such quantitative analyses may be necessary for a deeper understanding of the protective immune response, allowing, for instance, mapping of the progression of antibody reactivity profiles (and especially titre levels) in vaccine trials over time.

Although these results do not help identifying the antigens involved in the development of NAI (as the samples are from exposed adults it is impossible to tell which responses are a cause of natural immunity), they certainly highlight the eligibility of the system for proteome-wide antigenicity screening and serum antibody profiling to narrow down the search for those proteins capable of eliciting an immune response in humans.

## Chapter 4

# CONCLUSIONS

The work documented in this thesis demonstrates that the objectives and deliverables defined at the beginning of the project have been successfully validated and implemented. To date, the following objectives have been met: i) three sets of *Pf* antigens have been selected and annotated through an *in silico* analysis aimed at identifying the relevant candidates to spot in CHIP 1, 2 and 3; ii) an initial feasibility study aimed at assessing the suitability of selected medium-throughput cloning, expression and purification procedures has been successfully carried out; iii) the optimised procedures have been efficiently implemented using 189 CHIP 2 candidate targets; iv) a microarray feasibility study aimed at evaluating protocols for the spotting, detection and analysis of *Pf* candidates has been completed. For the microarray optimisation, suitable procedures for the immobilisation of *in vitro* expressed antigens and the detection of antigen-specific human IgGs (total and sub-types) from patients chronically exposed to malaria have been successfully established.

### 4.1 The *Pf* genome and proteome analysis

During the initial phase of my PhD a bioinformatic analysis was performed to identify both known and potential *Pf* antigens likely to be involved in the establishment of the NAI towards the parasite, which is commonly observed in the field. Particular attention was given to identify those antigens predicted to be exposed during the blood stage of the parasite. This allowed not only the selection of ~360 vaccine candidates (n.25 genes for CHIP 1, n.260 genes / n.448 gene portions for CHIP 2 and n.74 genes for CHIP 3) for the FightMal project but also highlighted the complexity of the *Pf* genome (highly AT-rich, highly polymorphic, complex gene structure and expression pattern based on

the different life cycles of *Pf*, etc). In particular, several incongruences were found in the nomenclature and the annotations of the genes. Discrepancies in the sequences were recorded and they represent valuable information for future database updates and revisions. Most of the parasite peculiarities were addressed one by one to develop a suitable cloning platform capable of producing hundreds of expression plasmids containing the DNA of interest in a limited amount of time and without the use of automation. It is envisaged that both the know-how and the information gathered through this study will generate a wealth of novel genomic and proteomic information, which may reveal important insights into the parasite pathogenesis. The bioinformatic analysis, the subsequent cloning and expression of the selected *Pf* putative antigens add valuable scientific data to that in the literature by identifying genes that are likely to be expressed specifically during the blood stage of *Pf*, by establishing suitable strategies for the medium-throughput production of high quality *Pf* and by identifying novel *Pf* genes as functional genes encoding proteins. It is expected that once all the candidates have been expressed and purified, they will represent one of the largest *Pf* protein datasets available within the scientific community. A database containing all the information regarding the selected candidates for CHIP 1, 2 and 3 will be made available, first to the members of the consortium and later to the wider scientific community. To date, the database only contains the candidate information retrieved during the bioinformatic selection, cloning, expression and purification of each specific target. However, with the progress of the project the information regarding the candidate antigenicity derived from the microarray immune profiling activity will be made available via the FightMal database.

## 4.2 Expression of selected *Pf* candidates in heterologous hosts

*Pf* genes and proteins are known from the literature to be difficult to clone and express due to some of the organism peculiarities. The parasite not only has a highly polymorphic AT-rich genome, which encodes for proteins from 1.5 to 2-fold longer than the yeast or *E. coli* ones, but it is also influenced by codon bias and its proteins are rich in low-complexity and highly repetitive regions. All these elements make the *Pf* a challenging organism to study and characterise. Indeed, given the fact that only in rare occasions plasmodial proteins can be isolated from the organism in levels sufficient for structural



and functional studies, many attempts have been carried out in the past to establish a suitable platform capable of producing soluble and functional *Pf* proteins. Of the different heterologous expression systems available, *E. coli* is the most common. Indeed, the use of this bacterial host for the expression of recombinant protein has many advantages (e.g. limited cost of implementation, ease of use, throughput and genetic manipulation). However, the literature has so far presented *E. coli* as a sub-optimal host for the production of soluble plasmidial proteins, which if expressed tend to be accumulated in the inclusion bodies in the form of protein aggregates (e.g. the study done by Vedadi et al. (2007) in which only 30% of the 1008 genes produced soluble/partially soluble proteins). On the contrary, both yeast (i.e. *S. cerevisiae* and *Pichia pastoris*) and baculovirus-infected insect cells are regarded as suitable choices for an efficient production of *Pf* proteins, given their high yield of expression (10-75 mg/L) and ability to produce correctly folded proteins. However, both systems require laborious procedures and costly apparatus (e.g. bio-reactors). Additionally, to express a gene in either of the two platforms, plasmidial sequences have to be modified to allow the correct synthesis of the protein target (e.g. codon-optimisation, depletion of N- and O-glycosylation sites etc.). Hence, neither the yeast nor the baculovirus expression platforms match the simplicity, flexibility, throughput and costs offered by the *E. coli* host, which is arguably considered the workhorse for proteome wide studies. It is on this basis that for the FightMal project we decided to investigate whether these shortfalls of the *E. coli* host could be somehow overcome. Remarkably, I demonstrated that if the conditions of growth are properly optimised, the *E. coli* host was able to express over 75% of the *Pf* candidates tested. In particular, ~60% of these proteins can be produced in soluble or partially soluble form and in an amount sufficient for a single step IMAC purification. Yields of 100-500  $\mu\text{g}$  of purified proteins allow the use of these recombinant targets for the production of thousands of microarray chips. The system developed within my PhD is a low-cost, rapid and easy procedure, which enables the RGB bacterial strain followed by the IMAC purification to produce of high quality highly purified *Pf* proteins. The optimisation of the expression media and temperature of the culture enabled us to minimise the expression of histidine-rich endogenous stress-related bacterial proteins allowing the successful purification of recombinant *Pf* antigens. While the amounts of protein achieved may not be enough for an extensive immune profiling study using the ELISA system they are more than sufficient for the production of thousands of protein microarrays, as each chip requires only 1/1000 of what is used in

the ELISA (75  $\mu$ g of purified protein, 50  $\mu$ g/ml spotting concentration, 1.5  $\eta$ L spotted in each chip Gray et al. (2007)).

### 4.3 The *Pf* microarray platform as a platform to identify malaria vaccine candidates and as an *in vitro* diagnostic tool

The goal to eradicate the malaria parasite is still far from being achieved. Steady progress and recent breakthroughs in both parasite and vector control have, however, made malaria elimination a more feasible, but still difficult objective. Indeed, if on one hand we are now more capable of monitoring and managing the disease transmission, on the other hand the genetic variability of the parasite enables it to quickly develop resistances to antimalarial drugs. All these factors strengthen the idea that an effective vaccine is needed, as, in the past, vaccination has proven to be one of the most effective tools for the control and eradication of infectious diseases. Despite the efforts of many scientific teams, however, an effective vaccine against malaria is still not available. Arguably the main cause of this failure is the lack of available vaccine candidates. In the last 30 years, scientists have been trying to develop new formulations based only on less than 0.6% of the *Pf* proteome. For the same reason, until now only, less than 30 purified antigens have been used in the literature for the serum characterisation of the immunity towards the parasite acquired naturally. Moreover, only a limited subset of these antigens has been tested in parallel to study a possible protective serum profile. Therefore, the availability of  $\sim$ 300 (25 from CHIP 1, 100-150 CHIP 2 and 74 CHIP 3) purified *Pf* putative and characterised antigens represent a remarkable 10-fold improvement in the resolution power of a NAI immune profiling assay. Given the recent finding on this field Trieu et al. (2011); Crompton et al. (2010); Gray et al. (2007), it is now clear that if we want to resolve the jigsaw of the malaria NAI and find new correlates of protection it is essential to have a powerful tool of investigation capable of screening hundreds of protected and unprotected malaria patients against a wide panel of *Pf* antigens. On this basis, the results obtained during my PhD play an essential role in developing a new powerful immune profiling tool capable of characterising specific antibody protective signatures that can help to select the next generation of malaria vaccine candidates. Indeed, as previously demonstrated by Gray et al.

(2007), our platform was already capable of highlighting possible correlates of protection, but with the improvements brought by my PhD it is now possible not only to increase by 100-fold the number of *Pf* candidates available for testing but it is also possible to quantitatively characterise the immunoglobulin response down to the sub-class level (e.g. IgG1, IgG3). I believe that the application of our microarray platform, however, is not limited to the NAI immune profiling, but may be extended to the diagnostic field. Indeed, as demonstrated in the recent report of the WHO (see Section 1.3.2) the demand for more accurate and informative malaria rapid diagnostic tests (RDTs) is growing exponentially. On these regards, our microarray chip could serve not only as an efficient serum-based multiplex assay platform that will identify vaccine candidates but it could also help in the determination of vaccine-induced immune responses and finally it could be used as a potent diagnostic tool for epidemiological and clinical studies. The use of the microarray technology in the malaria RDT field, however, can only be achieved if the newly developed platform represents a cost-effective alternative to systems already available in the market. So far, not a single protein microarray based in vitro diagnostic (IVD) has been able to enter this market. This is mainly due to the difficulties that the IVD industry is facing when working with such complex but powerful technology. The complex optimisation of the chemistry, which is essential for an accurate and a reliable assay, the limited supplier availability (e.g. glass-coated slides, which are produced only in small batches), the high running cost for both production and use are only a few of the issues that have prevented the microarray technology to play a major role in the IVD field. However, the demand for multiplex assay platforms is surging and the recent progress made in the assembly and automation of the protein microarray assay (e.g. MtM ADAM platform) may help the technology to break into the IVD market. In particular, the use of lateral flow devices, which simplify the handling of the miniaturised assay chemistry, together with the use of an automated LED-based microarray processor and analyser may offer an affordable alternative to the more expensive and labour demanding ELISA immunoassays, which currently represent the most common platform in epidemiological and clinical studies.

## 4.4 FightMal project: future work

There are a few points that remain to be addressed in order to guarantee the final success of the FightMal project. In particular, it will be crucial to fine-tune the antigen production procedure to guarantee an efficient large-scale *in vivo* expression and purification. Indeed, to increase the throughput of the antigen production it will be of pivotal importance to efficiently transfer the know-how derived from my PhD to the laboratories of the other members of the FightMal project (i.e. UNIPG and PX'). Contrary to an *in vitro* expression system, our *in vivo* platform is far from having a high-throughput capability. However, from my experience in developing quantitative microarray immunoassays (e.g. for allergy, influenza and malaria) the quality of the antigen used in the assay is of pivotal importance. Hence, I believe that, although our antigen production platform has a limited throughput, its production capabilities are still higher and less expensive than any alternative *in vivo* system (e.g. yeast or baculovirus) and more importantly the quality of the sample produced is superior to any of the *in vitro* systems that I have personally tested so far. In the coming months, all the partners of the FightMal consortium will collaborate to guarantee the completion of the predefined deliverables. Notably, cloning activities will ensue until the entire set of CHIP 2 candidates have been successfully inserted into the expression plasmid. Meanwhile, the protein expression will continue until all the available candidates have been characterised via a mini-scale analysis (5 mL, TM) and the ones resulting positive for soluble or partially soluble expression have been expressed and purified. It is estimated that our platform will allow the expression and purification of around 1-3 proteins per week per laboratory, for a total of 3-9 proteins per week considering the MtM, UNIPG and PX' laboratories are working side by side. If this estimation is correct it will take between 3 to 6 months to test and produce over 100-150 high quality *Pf* candidates. As for CHIP 1 and 3 the members responsible for sourcing the antigens (LSHTM and UNIPG) are actively working on obtaining either the purified proteins or the plasmid constructs needed for the expression. As soon as a sufficient number of purified antigens (i.e. 20-30) are available for spotting the partner responsible for the production of the microarrays (MtM) will start a small feasibility study to establish the optimal spotting conditions for these purified proteins. Validation of the microarray immunoassays will be carried out using serum samples from subjects with known reactivity to malaria antigens as determined by ELISA. The serum samples will be investigated in parallel in both the microarray and

the ELISA assays to establish how the two systems correlate in terms of prevalence and concentration of the antibodies. Once optimal procedures have been established, the robustness of the microarray assay will be established to guarantee that the test is stable under conditions similar to the ones observed in the field (i.e. high temperature and high/low humidity). The final *Pf* microarray will be used to test samples from children enrolled in a longitudinal, case-control study in a malaria endemic area of Uganda (i.e. Apach district). To look for correlates of protection, divisive clustering algorithms and detailed statistical analyses will be utilised to identify individual antigens or patterns of antigens within immune reactivity profiles that allow discrimination between children protected/non-protected against the clinical manifestations of malaria.

# Bibliography

- Abramoff M.D., Magalhaes P.J., and Ram S.J. Image Processing with ImageJ. *Biophotonics International*, 11(7):36–42, 2004.
- Adda C. G., Murphy V. J., Sunde M., Waddington L. J., Schloegel J., Talbo G. H., Vingas K., Kienzle V., Masciantonio R., Howlett G. J., Hodder A. N., Foley M., and Anders R. F. Plasmodium falciparum merozoite surface protein 2 is unstructured and forms amyloid-like fibrils. *Molecular and Biochemical Parasitology*, 166(2):159–71, 2009.
- Afonso A., Hunt P., Cheesman S., Alves A. C., Cunha C. V., do Rosario V., and Cravo P. Malaria parasites can develop stable resistance to artemisinin but lack mutations in candidate genes atp6 (encoding the sarcoplasmic and endoplasmic reticulum  $\text{Ca}^{2+}$  ATPase), tctp, mdr1, and cg10. *Antimicrob Agents Chemother.*, 50(2):480–9, 2006.
- Aguiar J. C., LaBaer J., Blair P. L., Shamailova V. Y., Koundinya M., Russell J. A., Huang F., Mar W., Anthony R. M., Witney A., Caruana S. R., Brizuela L., Sacci Jr. J. B., Hoffman S. L., and Carucci D. J. High-throughput generation of *P. falciparum* functional molecules by recombinational cloning. *Genome Research*, 14(10B):2076–82, 2004.
- Aide P., Bassat Q., and Alonso P. L. Towards an effective malaria vaccine. *Archives of Disease in Childhood*, 92(6):476–479, 2007.
- Akashi H. and Gojobori T. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America*, 99(6):3695–700, 2002.
- Alonso Pedro L, Sacarlal Jahit, Aponte John J, Leach Amanda, Macete Eusebio, Milman Jessica, Mandomando Inacio, Spiessens Bart, Guinovart Caterina, Espasa Mateu, Bassat Quique, Aide Pedro, Ofori-Anyinam Opokua, Navia Margarita M, Corachan Sabine, Ceuppens Marc, Dubois Marie-Claude, Demoitié Marie-Ange, Dubovsky Filip, Menéndez Clara, Tornieporth Nadia, Ballou W Ripley, Thompson Ricardo, and Cohen Joe. Efficacy of the RTS,S/AS02A vaccine against Plasmodium falciparum infection and disease in young African children: randomised controlled trial. *The Lancet*, 364(9443):1411 – 1420, 2004.

- Alonso Pedro L, Sacarlal Jahit, Aponte John J, Leach Amanda, Macete Eusebio, Aide Pedro, Sigauque Betuel, Milman Jessica, Mandomando Inacio, Bassat Quique, Guinovart Caterina, Espasa Mateu, Corachan Sabine, Lievens Marc, Navia Margarita M, Dubois Marie-Claude, Menendez Clara, Dubovsky Filip, Cohen Joe, Thompson Ricardo, and Ballou W Ripley. Duration of protection with RTS,S/AS02A malaria vaccine in prevention of *Plasmodium falciparum* disease in Mozambican children: single-blind extended follow-up of a randomised controlled trial. *The Lancet*, 366(9502): 2012 – 2018, 2005.
- Amino R., Thiberge S., Martin B., Celli S., Shorte S., Frischknecht F., and Menard R. Quantitative imaging of *Plasmodium* transmission from mosquito to mammal. *Nature medicine*, 12(2):220–4, 2006.
- Aponte John J, Aide Pedro, Renom Montse, Mandomando Inacio, Bassat Quique, Sacarlal Jahit, Manaca M Nelia, Lafuente Sarah, Barbosa Arnoldo, Leach Amanda, Lievens Marc, Vekemans Johan, Sigauque Betuel, Dubois Marie-Claude, Demoitié Marie-Ange, Sillman Marla, Savarese Barbara, McNeil John G, Macete Eusebio, Ballou W Ripley, Cohen Joe, and Alonso Pedro L. Safety of the RTS,S/AS02D candidate malaria vaccine in infants living in a highly endemic area of Mozambique: a double blind randomised controlled phase I/IIb trial. *The Lancet*, 370(9598):1543 – 1551, 2007.
- Archer C. T., Kim J. F., Jeong H., Park J. H., Vickers C. E., Lee S. Y., and Nielsen L. K. The genome sequence of *E. coli* W (ATCC 9637): comparative genome analysis and an improved genome-scale reconstruction of *E. coli*. *BMC genomics*, 12:9, 2011.
- Arrow K. J., Panosian C., and Gelband H. *Saving Lives, Buying Time: Economics of Malaria Drugs in an Age of Resistance*. Committee on the Economics of Antimalarial Drugs. The National Academies Press, 2004.
- Atkinson C. T. and Aikawa M. Ultrastructure of malaria-infected erythrocytes. *Blood cells*, 16(2-3):351–68, 1990.
- Audran R., Cachat M., Lurati F., Soe S., Leroy O., Corradin G., Druilhe P., and Spertini F. Phase I malaria vaccine trial with a long synthetic peptide derived from the merozoite surface protein 3 antigen. *Infection and immunity*, 73(12):8017–26, 2005.
- Baca A. M. and Hol W. G. Overcoming codon bias: a method for high-level overexpression of *Plasmodium* and other AT-rich parasite genes in *Escherichia coli*. *International journal for parasitology*, 30(2):113–8, 2000.
- Bacarese-Hamilton T., Bistoni F., and Crisanti A. Protein microarrays: from serodiagnosis to whole proteome scale analysis of the immune response against pathogenic microorganisms. *Biotechniques*, Suppl:24–9, 2002.
- Bacarese-Hamilton T., Ardizzoni A., Gray J., and Crisanti A. Protein arrays for serodiagnosis of disease. *Methods Mol Biol.*, 264:271–83, 2004.

- Ballou W. R., Arevalo-Herrera M., Carucci D., Richie T. L., Corradin G., Diggs C., Druilhe P., Giersing B. K., Saul A., Heppner D. G., Kester K. E., Lanar D. E., Lyon J., Hill A. V., Pan W., and Cohen J. D. Update on the clinical development of candidate malaria vaccines. *Am J Trop Med Hyg.*, 71(2 Suppl):239–47, 2004.
- Baum J., Gilberger T. W., Frischknecht F., and Meissner M. Host-cell invasion by malaria parasites: insights from *Plasmodium* and *Toxoplasma*. *Trends in parasitology*, 24(12):557–63, 2008.
- Beare P. A., Chen C., Bouman T., Pablo J., Unal B., Cockrell D. C., Brown W. C., Barbian K. D., Porcella S. F., Samuel J. E., Felgner P. L., and Heinzen R. A. Candidate antigens for Q fever serodiagnosis revealed by immunoscreening of a *Coxiella burnetii* protein microarray. *Clinical and vaccine immunology : CVI*, 15(12):1771–9, 2008.
- Bell D., Wongsrichanalai C., and Barnwell J. W. Ensuring quality and access for malaria diagnosis: how can it be achieved? *Nature reviews. Microbiology*, 4(9 Suppl):S7–20, 2006.
- Bendtsen J. D., Nielsen H., von Heijne G., and Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol*, 340(4):783–95, 2004.
- Benedict M. Q. and Robinson A. S. The first releases of transgenic mosquitoes: an argument for the sterile insect technique. *Trends in parasitology*, 19(8): 349–55, 2003.
- Betton J. M. Rapid translation system (RTS): a promising alternative for recombinant protein production. *Current protein & peptide science*, 4(1): 73–80, 2003.
- Blattner F. R., Plunkett 3rd G., Bloch C. A., Perna N. T., Burland V., Riley M., Collado-Vides J., Glasner J. D., Rode C. K., Mayhew G. F., Gregor J., Davis N. W., Kirkpatrick H. A., Goeden M. A., Rose D. J., Mau B., and Shao Y. The complete genome sequence of *Escherichia coli* k-12. *Science*, 277(5331):1453–62, 1997.
- Block H., Maertens B., Spriestersbach A., Brinker N., Kubicek J., Fabis R., Labahn J., and Schafer F. Immobilized-metal affinity chromatography (IMAC): a review. *Methods in enzymology*, 463:439–73, 2009.
- Bolanos-Garcia V. M. and Davies O. R. Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography. *Biochimica et biophysica acta*, 1760(9):1304–13, 2006.
- Bonn D. Filling the vaccine gap. *The Lancet Infectious Diseases*, 5(1):7, 2005.
- Bull P. C., Lowe B. S., Kortok M., Molyneux C. S., Newbold C. I., and Marsh K. Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. *Nature medicine*, 4(3):358–60, 1998.



- Campbell Jr. L. L. and Williams O. B. Observations on the biotin requirement of thermophilic bacteria. *Journal of bacteriology*, 65(2):146–7, 1953.
- Carlton J. M., Escalante A. A., Neafsey D., and Volkman S. K. Comparative evolutionary genomics of human malaria parasites. *Trends in parasitology*, 24(12):545–50, 2008.
- Carter J. A., Ross A. J., Neville B. G., Obiero E., Katana K., Mung’ala-Odera V., Lees J. A., and Newton C. R. Developmental impairments following severe falciparum malaria in children. *Trop Med Int Health.*, 10(1):3–10, 2005.
- Carter R., Mendis K. N., Miller L. H., Molineaux L., and Saul A. Malaria transmission-blocking vaccines-how can their development be supported? *Nature medicine*, 6(3):241–4, 2000.
- Catteruccia F., Benton J. P., and Crisanti A. An *Anopheles* transgenic sexing strain for vector control. *Nat Biotechnol.*, 23(11):1414–7 Epub 2005 Oct 9, 2005.
- Chanda I., Pan A., and Dutta C. Proteome composition in *Plasmodium falciparum*: higher usage of GC-rich nonsynonymous codons in highly expressed genes. *Journal of molecular evolution*, 61(4):513–23, 2005.
- Chen P. Q., Li G. Q., Guo X. B., He K. R., Fu Y. X., Fu L. C., and Song Y. Z. The infectivity of gametocytes of *Plasmodium falciparum* from patients treated with artemisinin. *Chinese medical journal*, 107(9):709–11, 1994.
- Cheng J., Randall A. Z., Sweredoski M. J., and Baldi P. SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Research*, 33(suppl 2):W72–W76, 2005. doi: 10.1093/nar/gki396.
- Chilengi R. and Gitaka J. Is vaccine the magic bullet for malaria elimination? a reality check. *Malaria Journal*, 9 Suppl 3:S1, 2010.
- Chima R. I., Goodman C. A., and Mills A. The economic impact of malaria in Africa: a critical review of the evidence. *Health Policy.*, 63(1):17–36, 2003.
- Chin C. D., Laksanasopin T., Cheung Y. K., Steinmiller D., Linder V., Parsa H., Wang J., Moore H., Rouse R., Umviligihozo G., Karita E., Mwambarangwe L., Braunstein S. L., van de Wiggert J., Sahabo R., Justman J. E., El-Sadr W., and Sia S. K. Microfluidics-based diagnostics of infectious diseases in the developing world. *Nature medicine*, 17(8):1015–9, 2011.
- Chowdhury D. R., Angov E., Kariuki T., and Kumar N. A potent malaria transmission blocking vaccine based on codon harmonized full length Pfs48/45 expressed in *Escherichia coli*. *PloS one*, 4(7):e6352, 2009.

- Clyde D. F., McCarthy V. C., Miller R. M., and Hornick R. B. Specificity of protection of man immunized against sporozoite-induced falciparum malaria. *The American journal of the medical sciences*, 266(6):398–403, 1973.
- Cohen S., Mc Gregor Ia, and Carrington S. Gamma-globulin and acquired immunity to human malaria. *Nature*, 192:733–7, 1961.
- Collins F. H. and Paskewitz S. M. Malaria: current and future prospects for control. *Annu Rev Entomol.*, 40:195–219, 1995.
- Cortese J. F., Caraballo A., Contreras C. E., and Plowe C. V. Origin and dissemination of *Plasmodium falciparum* drug-resistance mutations in South America. *The Journal of infectious diseases*, 186(7):999–1006, 2002.
- Cowman A. F. and Crabb B. S. Invasion of red blood cells by malaria parasites. *Cell*, 124(4):755–66, 2006.
- Cox-Singh Janet, Davis Timothy M. E., Lee Kim-Sung, Shamsul Sunita S. G., Matusop Asmad, Ratnam Shanmuga, Rahman Hasan A., Conway David J., and Singh Balbir. Plasmodium knowlesi malaria in humans is widely distributed and potentially life threatening. *Clin Infect Dis.*, 46(2):165–171, JAN 15 2008.
- CRGQ Collaboration Research Group for Qinghaosu. Antimalaria studies on Qinghaosu. *Chinese medical journal*, 92(12):811–6, 1979.
- Crompton P. D., Kayala M. A., Traore B., Kayentao K., Ongoiba A., Weiss G. E., Molina D. M., Burk C. R., Waisberg M., Jasinskas A., Tan X., Doumbo S., Doumtabe D., Kone Y., Narum D. L., Liang X., Doumbo O. K., Miller L. H., Doolan D. L., Baldi P., Felgner P. L., and Pierce S. K. A prospective analysis of the Ab response to *Plasmodium falciparum* before and after a malaria season by protein microarray. *Proceedings of the National Academy of Sciences of the United States of America*, 107(15):6958–63, 2010.
- D’Alessandro U., Leach A., Drakeley C. J., Bennett S., Olaleye B. O., Fegan G. W., Jawara M., Langerock P., George M. O., Targett G. A., and et al. Efficacy trial of malaria vaccine SPf66 in Gambian infants. *Lancet*, 346 (8973):462–7, 1995.
- Davies D. H., Liang X., Hernandez J. E., Randall A., Hirst S., Mu Y., Romero K. M., Nguyen T. T., Kalantari-Dehaghi M., Crotty S., Baldi P., Villarreal L. P., and Felgner P. L. Profiling the humoral immune response to infection by using proteome microarrays: high-throughput vaccine and diagnostic antigen discovery. *Proc Natl Acad Sci U S A.*, 102(3):547–52 Epub 2005 Jan 12, 2005.
- Dean J.A. *Lang’s handbook of chemistry*. McGraw-Hill, New York, 12th edition, 1979.
- Delehanty J. B. Printing functional protein microarrays using piezoelectric capillaries. *Methods in molecular biology*, 264:135–43, 2004.

- Di Cristina M., Nunziangeli L., Giubilei M. A., Capuccini B., d'Episcopo L., Mazzoleni G., Baldracchini F., Spaccapelo R., and Crisanti A. An antigen microarray immunoassay for multiplex screening of mouse monoclonal antibodies. *Nature protocols*, 5(12):1932–44, 2010.
- Dondorp A. M., Yeung S., White L., Nguon C., Day N. P., Socheat D., and von Seidlein L. Artemisinin resistance: current status and scenarios for containment. *Nature reviews. Microbiology*, 8(4):272–80, 2010.
- Doolan D. L. and Hoffman S. L. The complexity of protective immunity against liver-stage malaria. *Journal of immunology*, 165(3):1453–62, 2000.
- Doolan D. L., Southwood S., Freilich D. A., Sidney J., Graber N. L., Shatney L., Bebris L., Florens L., Dobano C., Witney A. A., Appella E., Hoffman S. L., Yates 3rd J. R., Carucci D. J., and Sette A. Identification of *Plasmodium falciparum* antigens by antigenic analysis of genomic and proteomic data. *Proceedings of the National Academy of Sciences of the United States of America*, 100(17):9952–7, 2003.
- Doolan D. L., Mu Y., Unal B., Sundaresh S., Hirst S., Valdez C., Randall A., Molina D., Liang X., Freilich D. A., Oloo J. A., Blair P. L., Aguiar J. C., Baldi P., Davies D. H., and Felgner P. L. Profiling humoral immune responses to *P. falciparum* infection with protein microarrays. *Proteomics*, 8(22):4680–94, 2008.
- Doolan Denise L and Stewart V Ann. Status of malaria vaccine R&D in 2007. *Expert Review of Vaccines*, 6(6):903–905, 2007.
- Dottorini T., Sole G., Nunziangeli L., Baldracchini F., Senin N., Mazzoleni G., Proietti C., Balaci L., and Crisanti A. Serum IgE Reactivity Profiling in an Asthma Affected Cohort. *PloS one*, 6(8):e22319, 2011.
- Druilhe P. and Perignon J. L. A hypothesis about the chronicity of malaria infection. *Parasitol Today.*, 13(9):353–7, 1997.
- Duffy P. E., Craig A. G., and Baruch D. I. Variant proteins on the surface of malaria-infected erythrocytes—developing vaccines. *Trends in parasitology*, 17(8):354–6, 2001.
- Dzikowski R., Frank M., and Deitsch K. Mutually exclusive expression of virulence genes by malaria parasites is regulated independently of antigen production. *PLoS pathogens*, 2(3):e22, 2006.
- Eisenhaber B., Bork P., and Eisenhaber F. [prediction of potential gpi-modification sites in proprotein sequences]. *Journal of molecular biology*, 292(3):741–58, 1999.
- Eyles J. E., Unal B., Hartley M. G., Newstead S. L., Flick-Smith H., Prior J. L., Oyston P. C., Randall A., Mu Y., Hirst S., Molina D. M., Davies D. H., Milne T., Griffin K. F., Baldi P., Titball R. W., and Felgner P. L.

- Immunodominant *Francisella tularensis* antigens identified using proteome microarray. *Proteomics*, 7(13):2172–83, 2007.
- Faber B. W., Remarque E. J., Morgan W. D., Kocken C. H., Holder A. A., and Thomas A. W. Malaria vaccine-related benefits of a single protein comprising *Plasmodium falciparum* apical membrane antigen 1 domains I and II fused to a modified form of the 19-kilodalton C-terminal fragment of merozoite surface protein 1. *Infection and immunity*, 75(12):5947–55, 2007.
- Fankhauser N. and Maser P. Identification of GPI anchor attachment signals by a Kohonen self-organizing map. *Bioinformatics*, 21(9):1846–52, 2005.
- Felgner P. L., Kayala M. A., Vigil A., Burk C., Nakajima-Sasaki R., Pablo J., Molina D. M., Hirst S., Chew J. S., Wang D., Tan G., Duffield M., Yang R., Neel J., Chantratita N., Bancroft G., Lertmemongkolchai G., Davies D. H., Baldi P., Peacock S., and Titball R. W. A *Burkholderia pseudomallei* protein microarray reveals serodiagnostic and cross-reactive antigens. *Proceedings of the National Academy of Sciences of the United States of America*, 106(32):13499–504, 2009.
- Fernando S. D., Gunawardena D. M., Bandara M. R., De Silva D., Carter R., Mendis K. N., Wickremasinghe A. R., De Silva N. R., Pathmeswaran A., Weerasinghe C. R., Selvaratnam R. R., Padmasiri E. A., and Montresor A. The impact of repeated malaria attacks on the school performance of children. Impact of mass chemotherapy for the control of filariasis on geohelminth infections in Sri Lanka. *Am J Trop Med Hyg.*, 69(6):582–8, 2003.
- Filip Dubovsky M.D. Creating a Vaccine against Malaria. *The Malaria Vaccine Initiative PATH*, 2001.
- Florens L., Washburn M. P., Raine J. D., Anthony R. M., Grainger M., Haynes J. D., Moch J. K., Muster N., Sacci J. B., Tabb D. L., Witney A. A., Wolters D., Wu Y., Gardner M. J., Holder A. A., Sinden R. E., Yates J. R., and Carucci D. J. A proteomic view of the *Plasmodium falciparum* life cycle. *Nature*, 419(6906):520–6, 2002.
- Florens L., Liu X., Wang Y., Yang S., Schwartz O., Peglar M., Carucci D. J., Yates 3rd J. R., and Wub Y. Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes. *Molecular and Biochemical Parasitology*, 135(1):1–11, 2004.
- Fluck C., Schopflin S., Smith T., Genton B., Alpers M. P., Beck H. P., and Felger I. Effect of the malaria vaccine Combination B on merozoite surface antigen 2 diversity. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 7(1):44–51, 2007.
- Foth B. J., Ralph S. A., Tonkin C. J., Struck N. S., Fraunholz M., Roos D. S., Cowman A. F., and McFadden G. I. Dissecting apicoplast targeting in the malaria parasite *Plasmodium falciparum*. *Science.*, 299(5607):705–8, 2003.

- Fried M. and Duffy P. E. Adherence of *Plasmodium falciparum* to chondroitin sulfate A in the human placenta. *Science*, 272(5267):1502–4, 1996.
- Gardner M. J., Hall N., Fung E., White O., Berriman M., Hyman R. W., Carlton J. M., Pain A., Nelson K. E., Bowman S., Paulsen I. T., James K., Eisen J. A., Rutherford K., Salzberg S. L., Craig A., Kyes S., Chan M. S., Nene V., Shallom S. J., Suh B., Peterson J., Angiuoli S., Pertea M., Allen J., Selengut J., Haft D., Mather M. W., Vaidya A. B., Martin D. M., Fairlamb A. H., Fraunholz M. J., Roos D. S., Ralph S. A., McFadden G. I., Cummings L. M., Subramanian G. M., Mungall C., Venter J. C., Carucci D. J., Hoffman S. L., Newbold C., Davis R. W., Fraser C. M., and Barrell B. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature.*, 419(6906):498–511, 2002.
- Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., and Bairoch A. Protein Identification and Analysis Tools on the ExPASy Server. Technical report, The Proteomics Protocols Handbook, Humana Press, 2005. Available at: <http://web.expasy.org/protparam/>.
- Gatewood Brown Marena and Rousseau Ronald W. Effect of Sodium Hydroxide on the Solubilities of L-Isoleucine, L-Leucine, and L-Valine. *Biotechnology Progress*, 10(3):253–257, 1994.
- Genton B., Betuela I., Felger I., Al-Yaman F., Anders R. F., Saul A., Rare L., Baisor M., Lorry K., Brown G. V., Pye D., Irving D. O., Smith T. A., Beck H. P., and Alpers M. P. A recombinant blood-stage malaria vaccine reduces *Plasmodium falciparum* density and exerts selective pressure on parasite populations in a phase 1-2b trial in Papua New Guinea. *The Journal of infectious diseases*, 185(6):820–7, 2002.
- Gilabert M., Audebert S., Viens P., Borg J. P., Bertucci F., and Goncalves A. [proteomics and breast cancer: a search for novel diagnostic and theragnostic biomarkers]. *Bulletin du cancer*, 97(3):321–39, 2010.
- Gilson P. R., Nebl T., Vukcevic D., Moritz R. L., Sargeant T., Speed T. P., Schofield L., and Crabb B. S. Identification and stoichiometry of glycosylphosphatidylinositol-anchored membrane proteins of the human malaria parasite *Plasmodium falciparum*. *Molecular and cellular proteomics : MCP*, 5(7):1286–99, 2006.
- Girard M. P., Reed Z. H., Friede M., and Kieny M. P. A review of human vaccine research and development: malaria. *Vaccine*, 25(9):1567–80, 2007.
- Goldberg D. E. and Cowman A. F. Moving in and renovating: exporting proteins from *Plasmodium* into host erythrocytes. *Nature reviews. Microbiology*, 8(9):617–21, 2010.
- Good M. F., Kaslow D. C., and Miller L. H. Pathways and strategies for developing a malaria blood-stage vaccine. *Annual review of immunology*, 16:57–87, 1998.

- Goodenbour J. M. and Pan T. Diversity of tRNA genes in eukaryotes. *Nucleic acids research*, 34(21):6137–46, 2006.
- Graslund S., Nordlund P., Weigelt J., Hallberg B. M., Bray J., Gileadi O., Knapp S., Oppermann U., Arrowsmith C., Hui R., Ming J., dhe Paganon S., Park H. W., Savchenko A., Yee A., Edwards A., Vincentelli R., Cambillau C., Kim R., Kim S. H., Rao Z., Shi Y., Terwilliger T. C., Kim C. Y., Hung L. W., Waldo G. S., Peleg Y., Albeck S., Unger T., Dym O., Prilusky J., Sussman J. L., Stevens R. C., Lesley S. A., Wilson I. A., Joachimiak A., Collart F., Dementieva I., Donnelly M. I., Eschenfeldt W. H., Kim Y., Stols L., Wu R., Zhou M., Burley S. K., Emtage J. S., Sauder J. M., Thompson D., Bain K., Luz J., Gheyi T., Zhang F., Atwell S., Almo S. C., Bonanno J. B., Fiser A., Swaminathan S., Studier F. W., Chance M. R., Sali A., Acton T. B., Xiao R., Zhao L., Ma L. C., Hunt J. F., Tong L., Cunningham K., Inouye M., Anderson S., Janjua H., Shastry R., Ho C. K., Wang D., Wang H., Jiang M., Montelione G. T., Stuart D. I., Owens R. J., Daenke S., Schutz A., Heinemann U., Yokoyama S., Bussow K., and Gunsalus K. C. Protein production and purification. *Nature methods*, 5(2):135–46, 2008.
- Gray J. C., Corran P. H., Mangia E., Gaunt M. W., Li Q., Tetteh K. K., Polley S. D., Conway D. J., Holder A. A., Bacarese-Hamilton T., Riley E. M., and Crisanti A. Profiling the Antibody Immune Response against Blood Stage Malaria Vaccine Candidates. *Clin Chem.*, 53(7):1244–53 Epub 2007 May 17, 2007.
- Greenwood B. and Targett G. Do we still need a malaria vaccine? *Parasite immunology*, 31(9):582–6, 2009.
- Greenwood B. M., Bojang K., Whitty C. J., and Targett G. A. Malaria. *Lancet.*, 365(9469):1487–98, 2005.
- Greenwood Brian M., Fidock David A., Kyle Dennis E., Kappe Stefan H.I., Alonso Pedro L., Collins Frank H., and Duffy Patrick E. Malaria: progress, perils, and prospects for eradication. *The Journal of Clinical Investigation*, 118(4):1266–1276, 4 2008.
- Gunsalus I.C. and Stanier R.Y. *The bacteria*. Academic Press, Inc., New York, 1960.
- Gupta S., Snow R. W., Donnelly C., and Newbold C. Acquired immunity and postnatal clinical protection in childhood cerebral malaria. *Proceedings. Biological sciences / The Royal Society*, 266(1414):33–8, 1999.
- Guyatt H. L. and Snow R. W. Malaria in pregnancy as an indirect cause of infant mortality in sub-Saharan Africa. *Trans R Soc Trop Med Hyg.*, 95(6): 569–76, 2001.
- Hall D. A., Zhu H., Zhu X., Royce T., Gerstein M., and Snyder M. Regulation of gene expression by a metabolic enzyme. *Science*, 306(5695):482–4, 2004.

- Hall D. A., Ptacek J., and Snyder M. Protein microarray technology. *Mechanisms of Ageing and Development*, 128(1):161–7, 2007.
- Hartmann M., Roeraade J., Stoll D., Templin M. F., and Joos T. O. Protein microarrays for diagnostic assays. *Analytical and bioanalytical chemistry*, 393(5):1407–16, 2009.
- Henderson D. A. Principles and lessons from the smallpox eradication programme. *Bulletin of the World Health Organization*, 65(4):535–46, 1987.
- Hertz-Fowler C., Peacock C. S., Wood V., Aslett M., Kerhornou A., Mooney P., Tivey A., Berriman M., Hall N., Rutherford K., Parkhill J., Ivens A. C., Rajandream M. A., and Barrell B. GeneDB: a resource for prokaryotic and eukaryotic organisms. *Nucleic Acids Res*, 32(Database issue):D339–43, 2004.
- Hiller N. L., Bhattacharjee S., van Ooij C., Liolios K., Harrison T., Lopez-Estrano C., and Haldar K. A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science*, 306(5703):1934–7, 2004.
- Hino M., Kataoka M., Kajimoto K., Yamamoto T., Kido J., Shinohara Y., and Baba Y. Efficiency of cell-free protein synthesis based on a crude cell extract from *Escherichia coli*, wheat germ, and rabbit reticulocytes. *Journal of biotechnology*, 133(2):183–9, 2008.
- Hoffman S. L., Billingsley P. F., James E., Richman A., Loyevsky M., Li T., Chakravarty S., Gunasekera A., Chattopadhyay R., Li M., Stafford R., Ahumada A., Epstein J. E., Sedegah M., Reyes S., Richie T. L., Lyke K. E., Edelman R., Laurens M. B., Plowe C. V., and Sim B. K. Development of a metabolically active, non-replicating sporozoite vaccine to prevent *Plasmodium falciparum* malaria. *Human vaccines*, 6(1):97–106, 2010.
- Hoffmann M., Nemetz C., Madin K., and Buchberger B. Rapid translation system: a novel cell-free way from gene to protein. *Biotechnology annual review*, 10:1–30, 2004.
- Holt R. A., Subramanian G. M., Halpern A., Sutton G. G., Charlab R., Nuskern D. R., Wincker P., Clark A. G., Ribeiro J. M., Wides R., Salzberg S. L., Loftus B., Yandell M., Majoros W. H., Rusch D. B., Lai Z., Kraft C. L., Abril J. F., Anthouard V., Arensburger P., Atkinson P. W., Baden H., de Berardinis V., Baldwin D., Benes V., Biedler J., Blass C., Bolanos R., Boscus D., Barnstead M., Cai S., Center A., Chaturverdi K., Christophides G. K., Chrystal M. A., Clamp M., Cravchik A., Curwen V., Dana A., Delcher A., Dew I., Evans C. A., Flanagan M., Grundschober-Freimoser A., Friedli L., Gu Z., Guan P., Guigo R., Hillenmeyer M. E., Hladun S. L., Hogan J. R., Hong Y. S., Hoover J., Jaillon O., Ke Z., Kodira C., Kokoza E., Koutsos A., Letunic I., Levitsky A., Liang Y., Lin J. J., Lobo N. F., Lopez J. R., Malek J. A., McIntosh T. C., Meister S., Miller J., Mobarry C., Mongin E., Murphy S. D., O’Brochta D. A., Pfannkoch C., Qi R., Regier M. A., Remington K., Shao H., Sharakhova M. V., Sitter C. D., Shetty J., Smith

- T. J., Strong R., Sun J., Thomasova D., Ton L. Q., Topalis P., Tu Z., Unger M. F., Walenz B., Wang A., Wang J., Wang M., Wang X., Woodford K. J., Wortman J. R., Wu M., Yao A., Zdobnov E. M., Zhang H., Zhao Q., et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science*, 298(5591):129–49, 2002.
- Howell J. M., Winstone T. L., Coorsen J. R., and Turner R. J. An evaluation of in vitro protein-protein interaction techniques: assessing contaminating background proteins. *Proteomics*, 6(7):2050–69, 2006.
- Ingram R. J., Metan G., Maillere B., Doganay M., Ozkul Y., Kim L. U., Baillie L., Dyson H., Williamson E. D., Chu K. K., Ascough S., Moore S., Huwar T. B., Robinson J. H., Sriskandan S., and Altmann D. M. Natural exposure to cutaneous anthrax gives long-lasting T cell immunity encompassing infection-specific epitopes. *Journal of immunology*, 184(7):3814–21, 2010.
- Jambou R., Legrand E., Niang M., Khim N., Lim P., Volney B., Ekala M. T., Bouchier C., Esterre P., Fandeur T., and Mercereau-Puijalon O. Resistance of *Plasmodium falciparum* field isolates to in-vitro artemether and point mutations of the SERCA-type PfATPase6. *Lancet*, 366(9501):1960–3, 2005.
- John J. Role of injectable and oral polio vaccines in polio eradication. *Expert review of vaccines*, 8(1):5–8, 2009.
- Jones R. B., Gordus A., Krall J. A., and MacBeath G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature*, 439(7073):168–74, 2006.
- Kall L., Krogh A., and Sonnhammer E. L. Advantages of combined transmembrane topology and signal peptide prediction—the phobius web server. *Nucleic acids research*, 35(Web Server issue):W429–32, 2007.
- Keesey J. Epitope Tagging Basic Laboratory Methods. Technical report, Roche Applied Bioscience, 1996. Available at: [www.roche-applied-science.com/PROD\\_INF/MANUALS/epitope/epi\\_toc.htm](http://www.roche-applied-science.com/PROD_INF/MANUALS/epitope/epi_toc.htm).
- Kiraga J., Mackiewicz P., Mackiewicz D., Kowalczyk M., Biecek P., Polak N., Smolarczyk K., Dudek M. R., and Cebrat S. The relationships between the isoelectric point and: length of proteins, taxonomy and ecology of organisms. *BMC genomics*, 8:163, 2007.
- Kissinger J. C., Brunk B. P., Crabtree J., Fraunholz M. J., Gajria B., Milgram A. J., Pearson D. S., Schug J., Bahl A., Diskin S. J., Ginsburg H., Grant G. R., Gupta D., Labo P., Li L., Mailman M. D., McWeeney S. K., Whetzel P., Stoeckert C. J., and Roos D. S. The *Plasmodium* genome database. *Nature*, 419(6906):490–2, 2002.
- Kocken C. H., Withers-Martinez C., Dubbeld M. A., van der Wel A., Hackett F., Valderrama A., Blackman M. J., and Thomas A. W. High-level



- expression of the malaria blood-stage vaccine candidate *Plasmodium falciparum* apical membrane antigen 1 and induction of antibodies that inhibit erythrocyte invasion. *Infection and immunity*, 70(8):4471–6, 2002.
- Kricka L. J., Imai K., and Fortina P. Analytical ancestry: evolution of the array in analysis. *Clinical chemistry*, 56(12):1797–803, 2010.
- Krogh A., Larsson B., von Heijne G., and Sonnhammer E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*, 305(3):567–80, 2001.
- Kunnath-Velayudhan S., Salamon H., Wang H. Y., Davidow A. L., Molina D. M., Huynh V. T., Cirillo D. M., Michel G., Talbot E. A., Perkins M. D., Felgner P. L., Liang X., and Gennaro M. L. Dynamic antibody responses to the *Mycobacterium tuberculosis* proteome. *Proceedings of the National Academy of Sciences of the United States of America*, 107(33):14703–8, 2010.
- Kyte J. and Doolittle R. F. A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology*, 157(1):105–32, 1982.
- LaCount D. J., Vignali M., Chettier R., Phansalkar A., Bell R., Hesselberth J. R., Schoenfeld L. W., Ota I., Sahasrabudhe S., Kurschner C., Fields S., and Hughes R. E. A protein interaction network of the malaria parasite *Plasmodium falciparum*. *Nature.*, 438(7064):103–7, 2005.
- Langhorne J., Ndungu F. M., Sponaas A. M., and Marsh K. Immunity to malaria: more questions than answers. *Nature immunology*, 9(7):725–32, 2008.
- Le Roch K. G., Zhou Y., Blair P. L., Grainger M., Moch J. K., Haynes J. D., De La Vega P., Holder A. A., Batalov S., Carucci D. J., and Winzeler E. A. Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science.*, 301(5639):1503–8 Epub 2003 Jul 31, 2003.
- Lee C. L., Ow D. S., and Oh S. K. Quantitative real-time polymerase chain reaction for determination of plasmid copy number in bacteria. *Journal of microbiological methods*, 65(2):258–67, 2006.
- Leiriao P., Mota M. M., and Rodriguez A. Apoptotic *Plasmodium*-infected hepatocytes provide antigens to liver dendritic cells. *The Journal of infectious diseases*, 191(10):1576–81, 2005.
- Lengeler C. Insecticide-treated bed nets and curtains for preventing malaria. *Cochrane database of systematic reviews*, (2):CD000363, 2004.
- Liang L., Leng D., Burk C., Nakajima-Sasaki R., Kayala M. A., Atluri V. L., Pablo J., Unal B., Ficht T. A., Gotuzzo E., Saito M., Morrow W. J., Liang X., Baldi P., Gilman R. H., Vinetz J. M., Tsois R. M., and Felgner P. L. Large scale immune profiling of infected humans and goats reveals differential recognition of *Brucella melitensis* antigens. *PLoS neglected tropical diseases*, 4(5):e673, 2010.

- Lobry J. R. and Gautier C. Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 *Escherichia coli* chromosome-encoded genes. *Nucleic acids research*, 22(15):3174–80, 1994.
- Logie D. E., McGregor I. A., Rowe D. S., and Billewicz W. Z. Plasma immunoglobulin concentrations in mothers and newborn children with special reference to placental malaria: Studies in the Gambia, Nigeria, and Switzerland. *Bulletin of the World Health Organization*, 49(6):547–54, 1973.
- Luli G. W. and Strohl W. R. Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. *Applied and environmental microbiology*, 56(4):1004–11, 1990.
- MacBeath G. and Schreiber S. L. Printing proteins as microarrays for high-throughput function determination. *Science.*, 289(5485):1760–3, 2000.
- Magnan C. N., Zeller M., Kayala M. A., Vigil A., Randall A., Felgner P. L., and Baldi P. High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics*, 26(23):2936–43, 2010.
- Mahanty S., Saul A., and Miller L. H. Progress in the development of recombinant and synthetic blood-stage malaria vaccines. *The Journal of experimental biology*, 206(Pt 21):3781–8, 2003.
- malERA Consultative Group. A Research Agenda for Malaria Eradication: Diagnoses and Diagnostics. *PLoS medicine*, 8(1), 2011a.
- malERA Consultative Group. A research agenda for malaria eradication: drugs. *PLoS medicine*, 8(1), 2011b.
- malERA Consultative Group. A research agenda for malaria eradication: vaccines. *PLoS medicine*, 8(1), 2011c.
- Malkin E. M., Diemert D. J., McArthur J. H., Perreault J. R., Miles A. P., Giersing B. K., Mullen G. E., Orcutt A., Muratova O., Awkal M., Zhou H., Wang J., Stowers A., Long C. A., Mahanty S., Miller L. H., Saul A., and Durbin A. P. Phase 1 clinical trial of apical membrane antigen 1: an asexual blood-stage vaccine for *Plasmodium falciparum* malaria. *Infection and immunity*, 73(6):3677–85, 2005.
- Marti M., Good R. T., Rug M., Knuepfer E., and Cowman A. F. Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science*, 306(5703):1930–3, 2004.
- McGregor I. A., Rowe D. S., Wilson M. E., and Billewicz W. Z. Plasma immunoglobulin concentrations in an African (Gambian) community in relation to season, malaria and other infections and pregnancy, journal = *Clinical and experimental immunology*. 7(1):51–74, 1970.

- McMurry J., Sbai H., Gennaro M. L., Carter E. J., Martin W., and De Groot A. S. Analyzing *Mycobacterium tuberculosis* proteomes for candidate vaccine epitopes. *Tuberculosis*, 85(1-2):95–105, 2005.
- Mehlin C., Boni E., Buckner F. S., Engel L., Feist T., Gelb M. H., Haji L., Kim D., Liu C., Mueller N., Myler P. J., Reddy J. T., Sampson J. N., Subramanian E., Van Voorhis W. C., Worthey E., Zucker F., and Hol W. G. Heterologous expression of proteins from *Plasmodium falciparum*: results from 1000 genes. *Molecular and Biochemical Parasitology*, 148(2):144–60, 2006.
- Mezzasoma L., Bacarese-Hamilton T., Di Cristina M., Rossi R., Bistoni F., and Crisanti A. Antigen microarrays for serodiagnosis of infectious diseases. *Clin Chem*, 48(1):121–30, 2002.
- Miller L. H. and Greenwood B. Malaria—a shadow over Africa. *Science*, 298(5591):121–2, 2002.
- Miller L. H., Good M. F., and Milon G. Malaria pathogenesis. *Science*, 264(5167):1878–83, 1994.
- Mochon A. B., Jin Y., Kayala M. A., Wingard J. R., Clancy C. J., Nguyen M. H., Felgner P., Baldi P., and Liu H. Serological profiling of a *Candida albicans* protein microarray reveals permanent host-pathogen interplay and stage-specific responses during candidemia. *PLoS pathogens*, 6(3):e1000827, 2010.
- Molina D. M., Pal S., Kayala M. A., Teng A., Kim P. J., Baldi P., Felgner P. L., Liang X., and de la Maza L. M. Identification of immunodominant antigens of *Chlamydia trachomatis* using proteome microarrays. *Vaccine*, 28(17):3014–24, 2010.
- Moll K., Ljungström I., Perlmann H., Scherf A., and Wahlgren M. Methods in malaria research, fifth edition. Technical report, MR4/ATCC and BioMalPar, 2008. Available at: [www.mr4.org/Portals/3/Methods\\_In\\_Malaria\\_Research\\_5theditionv5-2.pdf](http://www.mr4.org/Portals/3/Methods_In_Malaria_Research_5theditionv5-2.pdf).
- Moorthy V.S., Diggs C., Ferro S., Good M.F., Herrera S., Hill A.V., Imoukhuede E.B., Kumar S., Loucq C., Marsh K., Ockenhouse C.F., Richie T.L., and Sauerwein R.W. Report of a Consultation on the Optimization of Clinical Challenge Trials for Evaluation of Candidate Blood Stage Malaria Vaccines, 18-19 March 2009, Bethesda, MD, USA. *Vaccine*, 27(42):5719 – 5725, 2009.
- Moss W. J. and Griffin D. E. Global measles elimination. *Nature reviews. Microbiology*, 4(12):900–8, 2006.
- Mu J., Awadalla P., Duan J., McGee K. M., Keebler J., Seydel K., McVean G. A., and Su X. Z. Genome-wide variation and identification of vaccine

- targets in the *Plasmodium falciparum* genome. *Nature genetics*, 39(1):126–30, 2007.
- Mueller A. K., Labaied M., Kappe S. H., and Matuschewski K. Genetically modified *Plasmodium* parasites as a protective experimental malaria vaccine. *Nature*, 433(7022), 2005.
- MVI-PATH Malaria Vaccine Initiative. Accelerating Progress Toward Malaria Vaccines. *PATH Publications*, 2007.
- Nakamura Yasukazu, Gojobori Takashi, and Ikemura Toshimichi. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Research*, 28(1):292, 2000. doi: 10.1093/nar/28.1.292.
- Nolan T., Papathanos P., Windbichler N., Magnusson K., Benton J., Catteruccia F., and Crisanti A. Developing transgenic *Anopheles* mosquitoes for the sterile insect technique. *Genetica*, 139(1):33–9, 2011.
- Nosten F., Luxemburger C., Kyle D. E., Ballou W. R., Wittes J., Wah E., Chongsuphajaisiddhi T., Gordon D. M., White N. J., Sadoff J. C., and Heppner D. G. Randomised double-blind placebo-controlled trial of SPf66 malaria vaccine in children in northwestern Thailand. Shoklo SPf66 Malaria Vaccine Trial Group. *Lancet*, 348(9029):701–7, 1996.
- Ockenhouse C. F., Sun P. F., Lanar D. E., Welde B. T., Hall B. T., Kester K., Stoute J. A., Magill A., Krzych U., Farley L., Wirtz R. A., Sadoff J. C., Kaslow D. C., Kumar S., Church L. W., Crutcher J. M., Wizel B., Hoffman S., Lalvani A., Hill A. V., Tine J. A., Guito K. P., de Taisne C., Anders R., Ballou W. R., and et al. Phase I/IIa safety, immunogenicity, and efficacy trial of NYVAC-Pf7, a pox-vectored, multiantigen, multistage vaccine candidate for *Plasmodium falciparum* malaria. *The Journal of infectious diseases*, 177(6):1664–73, 1998.
- Oeuvray C., Bouharoun-Tayoun H., Gras-Masse H., Bottius E., Kaidoh T., Aikawa M., Filgueira M. C., Tartar A., and Druilhe P. Merozoite surface protein-3: a malaria protein inducing antibodies that promote *Plasmodium falciparum* killing by cooperation with blood monocytes. *Blood*, 84(5):1594–602, 1994.
- Olotu Ally, Lusingu John, Leach Amanda, Lievens Marc, Vekemans Johan, Msham Salum, Lang Trudie, Gould Jayne, Dubois Marie-Claude, Jongert Erik, Vansadia Preeti, Carter Terrell, Njuguna Patricia, Awuondo Ken O., Malabeja Anangisye, Abdul Omar, Gesase Samwel, Mturi Neema, Drakeley Chris J., Savarese Barbara, Villafana Tonya, Lapierre Didier, Ballou W. Ripley, Cohen Joe, Lemnge Martha M., Peshu Norbert, Marsh Kevin, Riley Eleanor M., von Seidlein Lorenz, and Bejon Philip. Efficacy of RTS,S/AS01E malaria vaccine and exploratory analysis on anti-circumsporozoite antibody titres and protection in children aged 5-17 months in Kenya and Tanzania: a randomised controlled trial. *The Lancet Infectious Diseases*, 11(2):102–109, 2011.

- O'Meara W. P., Mangeni J. N., Steketee R., and Greenwood B. Changes in the burden of malaria in sub-Saharan Africa. *The Lancet Infectious Diseases*, 10(8):545–55, 2010.
- Owusu-Agyei S., Koram K. A., Baird J. K., Utz G. C., Binka F. N., Nkrumah F. K., Fryauff D. J., and Hoffman S. L. Incidence of symptomatic and asymptomatic *Plasmodium falciparum* infection following curative therapy in adult residents of northern Ghana. *The American journal of tropical medicine and hygiene*, 65(3):197–203, 2001.
- Paliy O. and Gunasekera T. S. Growth of *E. coli* BL21 in minimal media with different gluconeogenic carbon sources and salt contents. *Applied microbiology and biotechnology*, 73(5):1169–72, 2007.
- Pan W., Huang D., Zhang Q., Qu L., Zhang D., Zhang X., Xue X., and Qian F. Fusion of two malaria vaccine candidate antigens enhances product yield, immunogenicity, and antibody-mediated inhibition of parasite growth in vitro. *Journal of immunology*, 172(10):6167–74, 2004.
- Parekh F. K. and Richie T. L. Characterization of immune reactivity profiles using microarray technology may expedite identification of candidate antigens for next generation malaria vaccines. *Clin Chem.*, 53(7):1183–5, 2007.
- Perlmann P. and Troye-Blomberg M. Malaria and the immune system in humans. *Chemical immunology*, 80:229–42, 2002.
- Pierce. Optimize elution conditions for immunoaffinity purification. Technical report, Thermo Fisher Scientific Inc, 2009. Available at: [www.piercenet.com/files/TR0027 - Elution - conditions.pdf](http://www.piercenet.com/files/TR0027 - Elution - conditions.pdf).
- Pierce Susan K. and Miller Louis H. World Malaria Day 2009: What Malaria Knows about the Immune System That Immunologists Still Do Not. *The Journal of Immunology*, 182(9):5171–5177, 2009. doi: 10.4049/jimmunol.0804153.
- Pinder M., Moorthy V. S., Akanmori B. D., Genton B., and Brown G. V. MALVAC 2009: progress and challenges in development of whole organism malaria vaccines for endemic countries, 3-4 June 2009, Dakar, Senegal. *Vaccine*, 28(30):4695–702, 2010.
- Pluess B., Tanser F. C., Lengeler C., and Sharp B. L. Indoor residual spraying for preventing malaria. *Cochrane database of systematic reviews*, (4): CD006657, 2010.
- Pollastri G., Przybylski D., Rost B., and Baldi P. Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins*, 47(2):228–35, 2002.
- Rappuoli Rino. Reverse vaccinology. *Current Opinion in Microbiology*, 3(5): 445 – 450, 2000.

- RBM Roll Back Malaria Partnership. Global Malaria Action Plan - GMAP objectives June 2011. Technical report, RBM, Roll Back Malaria Partnership, 2011.
- Remarque E. J., Faber B. W., Kocken C. H., and Thomas A. W. Apical membrane antigen 1: a malaria vaccine candidate in review. *Trends in parasitology*, 24(2):74–84, 2008.
- Renberg B., Nordin J., Merca A., Uhlen M., Feldwisch J., Nygren P. A., and Karlstrom A. E. Affibody molecules in protein capture microarrays: evaluation of multidomain ligands and different detection formats. *Journal of proteome research*, 6(1):171–9, 2007.
- Richie T. L. and Saul A. Progress and challenges for malaria vaccines. *Nature.*, 415(6872):694–701, 2002.
- Rieckmann K. H., Beaudoin R. L., Cassells J. S., and Sell K. W. Use of attenuated sporozoites in the immunization of human volunteers against falciparum malaria. *Bulletin of the World Health Organization*, 57 Suppl 1: 261–5, 1979.
- Riesenbergs D. High-cell-density cultivation of *Escherichia coli*. *Current opinion in biotechnology*, 2(3):380–4, 1991.
- Riesenbergs D. and Guthke R. High-cell-density cultivation of microorganisms. *Applied microbiology and biotechnology*, 51(4):422–30, 1999.
- Riley E. M., Wagner G. E., Akanmori B. D., and Koram K. A. Do maternally acquired antibodies protect infants from malaria infection? *Parasite immunology*, 23(2):51–9, 2001.
- Roca-Feltrer A., Carneiro I., Smith L., Schellenberg J. R., Greenwood B., and Schellenberg D. The age patterns of severe malaria syndromes in sub-Saharan Africa across a range of transmission intensities and seasonality settings. *Malaria Journal*, 9:282, 2010.
- Rodriguez-Ortega M. J., Norais N., Bensi G., Liberatori S., Capo S., Mora M., Scarselli M., Doro F., Ferrari G., Garaguso I., Maggi T., Neumann A., Covre A., Telford J. L., and Grandi G. Characterization and identification of vaccine candidate proteins through analysis of the group A Streptococcus surface proteome. *Nature biotechnology*, 24(2):191–7, 2006.
- Roussilhon C., Oeuvray C., Muller-Graf C., Tall A., Rogier C., Trape J. F., Theisen M., Balde A., Perignon J. L., and Druilhe P. Long-term clinical protection from falciparum malaria is strongly associated with IgG3 antibodies to merozoite surface protein 3. *PLoS medicine*, 4(11):e320, 2007.
- Russell T. L., Govella N. J., Azizi S., Drakeley C. J., Kachur S. P., and Killeen G. F. Increased proportions of outdoor feeding among residual malaria vector populations following increased use of insecticide-treated nets in rural Tanzania. *Malaria Journal*, 10:80, 2011.

- Ryabova L. A., Morozov IYu, and Spirin A. S. Continuous-flow cell-free translation, transcription-translation, and replication-translation systems. *Methods in molecular biology*, 77:179–93, 1998.
- Sabchareon A., Burnouf T., Ouattara D., Attanath P., Bouharoun-Tayoun H., Chantavanich P., Foucault C., Chongsuphajaisiddhi T., and Druilhe P. Parasitologic and clinical human response to immunoglobulin administration in falciparum malaria. *Am J Trop Med Hyg*, 45(3):297–308, 1991.
- Sabot Oliver, Cohen Justin M, Hsiang Michelle S, Kahn James G, Basu Suprotik, Tang Linhua, Zheng Bin, Gao Qi, Zou Linda, Tatarsky Allison, Aboobakar Shahina, Usas Jennifer, Barrett Scott, Cohen Jessica L, Jamison Dean T, and Feachem Richard GA. Costs and financial feasibility of malaria elimination. *The Lancet*, 376(9752):1604 – 1615, 2010. ISSN 0140-6736.
- Sachs J. and Malaney P. The economic and social burden of malaria. *Nature.*, 415(6872):680–5, 2002.
- Sachs J. D. A new global effort to control malaria. *Science.*, 298(5591):122–4, 2002.
- Sakharkar M. K., Chow V. T., and Kanguene P. Distributions of exons and introns in the human genome. *In silico biology*, 4(4):387–93, 2004.
- Sargeant T. J., Marti M., Caler E., Carlton J. M., Simpson K., Speed T. P., and Cowman A. F. Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome biology*, 7(2):R12, 2006.
- Saul A., Lawrence G., Allworth A., Elliott S., Anderson K., Rzepczyk C., Martin L. B., Taylor D., Eisen D. P., Irving D. O., Pye D., Crewther P. E., Hodder A. N., Murphy V. J., and Anders R. F. A human phase 1 vaccine clinical trial of the *Plasmodium falciparum* malaria vaccine candidate apical membrane antigen 1 in Montanide ISA720 adjuvant. *Vaccine*, 23(23):3076–83, 2005.
- Saul Allan and Battistutta Diana. Codon usage in *Plasmodium falciparum*. *Molecular and Biochemical Parasitology*, 27(1):35 – 42, 1988. ISSN 0166-6851. doi: 10.1016/0166-6851(88)90022-9.
- Schena M., Shalon D., Davis R. W., and Brown P. O. Quantitative monitoring of gene expression patterns with a complementary dna microarray. *Science.*, 270(5235):467–70, 1995.
- Scherf A., Lopez-Rubio J. J., and Riviere L. Antigenic variation in *Plasmodium falciparum*. *Annual review of microbiology*, 62:445–70, 2008.
- Sette Alessandro and Rappuoli Rino. Reverse Vaccinology: Developing Vaccines in the Era of Genomics. *Immunity*, 33(4):530–541, 2010.

- Shiloach J. and Fass R. Growing *E. coli* to high cell density—a historical perspective on method development. *Biotechnology advances*, 23(5):345–57, 2005.
- Shirokov V. A., Kommer A., Kolb V. A., and Spirin A. S. Continuous-exchange protein-synthesizing systems. *Methods in molecular biology*, 375:19–55, 2007.
- Shuman S. Recombination mediated by vaccinia virus DNA topoisomerase I in *Escherichia coli* is sequence specific. *Proc Natl Acad Sci U S A.*, 88(22):10104–8, 1991.
- Singh Balbir, Sung Lee Kim, Matusop Asmad, Radhakrishnan Anand, Shamsul Sunita SG, Cox-Singh Janet, Thomas Alan, and Conway David J. A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *The Lancet*, 363(9414):1017 – 1024, 2004.
- Smith J. D., Chitnis C. E., Craig A. G., Roberts D. J., Hudson-Taylor D. E., Peterson D. S., Pinches R., Newbold C. I., and Miller L. H. Switches in expression of *Plasmodium falciparum* var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell*, 82(1):101–10, 1995.
- Spirin A. S. [gene expression in cell-free systems on a preparative scale]. *Bioorganicheskaya khimiya*, 18(10-11):1394–402, 1992.
- Spirin A. S., Baranov V. I., Ryabova L. A., Ovodov S. Y., and Alakhov Y. B. A continuous cell-free translation system capable of producing polypeptides in high yield. *Science*, 242(4882):1162–4, 1988.
- Steller S., Angenendt P., Cahill D. J., Heuberger S., Lehrach H., and Kreutzberger J. Bacterial protein microarrays for identification of new potential diagnostic markers for *Neisseria meningitidis* infections. *Proteomics*, 5(8):2048–55, 2005.
- Stevenson M. M. and Riley E. M. Innate immunity to malaria. *Nature reviews. Immunology*, 4(3):169–80, 2004.
- Stoute J. A., Gombe J., Withers M. R., Siangla J., McKinney D., Onyango M., Cummings J. F., Milman J., Tucker K., Soisson L., Stewart V. A., Lyon J. A., Angov E., Leach A., Cohen J., Kester K. E., Ockenhouse C. F., Holland C. A., Diggs C. L., Wittes J., and Gray Heppner Jr. D. Phase 1 randomized double-blind safety and immunogenicity trial of Plasmodium falciparum malaria merozoite surface protein FMP1 vaccine, adjuvanted with AS02A, in adults in western Kenya. *Vaccine*, 7:7, 2005.
- Struik S. S. and Riley E. M. Does malaria suffer from lack of memory? *Immunol Rev.*, 201:268–90, 2004.
- Subbarao S. K. and Sharma V. P. *Anopheline* species complexes and malaria control. *The Indian journal of medical research*, 106:164–73, 1997.



- ter Kuile F., White N. J., Holloway P., Pasvol G., and Krishna S. *Plasmodium falciparum*: in vitro studies of the pharmacodynamic properties of drugs used for the treatment of severe malaria, journal = Experimental parasitology. 76(1):85–95, 1993.
- Terpe K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied microbiology and biotechnology*, 60(5):523–33, 2003.
- Thailayil J., Magnusson K., Godfray H. C., Crisanti A., and Catteruccia F. Spermless males elicit large-scale female responses to mating in the malaria mosquito *Anopheles gambiae*. *Proceedings of the National Academy of Sciences of the United States of America*, 108(33):13677–81, 2011.
- Theisen M., Soe S., Oeuvray C., Thomas A. W., Vuust J., Danielsen S., Jepsen S., and Druilhe P. The glutamate-rich protein (GLURP) of *Plasmodium falciparum* is a target for antibody-dependent monocyte-mediated inhibition of parasite growth *in vitro*. *Infection and immunity*, 66(1):11–7, 1998.
- Trieu A., Kayala M. A., Burk C., Molina D. M., Freilich D. A., Richie T. L., Baldi P., Felgner P. L., and Doolan D. L. Sterile protective immunity to malaria is associated with a panel of novel *P. falciparum* antigens. *Molecular and cellular proteomics : MCP*, 2011.
- Trouiller P. and Olliaro P. L. Drug development output from 1975 to 1996: what proportion for tropical diseases? *Int J Infect Dis.*, 3(2):61–3, 1998.
- Tsuboi T., Takeo S., Iriko H., Jin L., Tsuchimochi M., Matsuda S., Han E. T., Otsuki H., Kaneko O., Sattabongkot J., Udomsangpetch R., Sawasaki T., Torii M., and Endo Y. Wheat germ cell-free system-based production of malaria proteins for discovery of novel vaccine candidates. *Infection and immunity*, 76(4):1702–8, 2008.
- Tsuboi T., Takeo S., Arumugam T. U., Otsuki H., and Torii M. The wheat germ cell-free protein synthesis system: a key tool for novel malaria vaccine candidate discovery. *Acta tropica*, 114(3):171–6, 2010.
- Vacic V., Uversky V. N., Dunker A. K., and Lonardi S. Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *BMC bioinformatics*, 8:211, 2007.
- Valero M. V., Amador L. R., Galindo C., Figueroa J., Bello M. S., Murillo L. A., Mora A. L., Patarroyo G., Rocha C. L., Rojas M., and et al. Vaccination with SPf66, a chemically synthesised vaccine, against *Plasmodium falciparum* malaria in Colombia. *Lancet*, 341(8847):705–10, 1993.
- van Ooij C., Tamez P., Bhattacharjee S., Hiller N. L., Harrison T., Liolios K., Kooij T., Ramesar J., Balu B., Adams J., Waters A. P., Janse C. J., and Haldar K. The malaria secretome: from algorithms to essential function in blood stage infection. *PLoS pathogens*, 4(6):e1000084, 2008.

- Vedadi M., Lew J., Artz J., Amani M., Zhao Y., Dong A., Wasney G. A., Gao M., Hills T., Brokx S., Qiu W., Sharma S., Diassiti A., Alam Z., Melone M., Mulichak A., Wernimont A., Bray J., Loppnau P., Plotnikova O., Newberry K., Sundararajan E., Houston S., Walker J., Tempel W., Bochkarev A., Kozieradzki I., Edwards A., Arrowsmith C., Roos D., Kain K., and Hui R. Genome-scale protein expression and structural biology of *Plasmodium falciparum* and related Apicomplexan organisms. *Molecular and Biochemical Parasitology*, 151(1):100–10, 2007.
- Venter J. C., Adams M. D., Myers E. W., Li P. W., Mural R. J., Sutton G. G., Smith H. O., Yandell M., Evans C. A., Holt R. A., Gocayne J. D., Amanatides P., Ballew R. M., Huson D. H., Wortman J. R., Zhang Q., Kodira C. D., Zheng X. H., Chen L., Skupski M., Subramanian G., Thomas P. D., Zhang J., Gabor Miklos G. L., Nelson C., Broder S., Clark A. G., Nadeau J., McKusick V. A., Zinder N., Levine A. J., Roberts R. J., Simon M., Slayman C., Hunkapiller M., Bolanos R., Delcher A., Dew I., Fasulo D., Flanigan M., Florea L., Halpern A., Hannenhalli S., Kravitz S., Levy S., Mobarry C., Reinert K., Remington K., Abu-Threideh J., Beasley E., Biddick K., Bonazzi V., Brandon R., Cargill M., Chandramouliswaran I., Charlab R., Chaturvedi K., Deng Z., Di Francesco V., Dunn P., Eilbeck K., Evangelista C., Gabrielian A. E., Gan W., Ge W., Gong F., Gu Z., Guan P., Heiman T. J., Higgins M. E., Ji R. R., Ke Z., Ketchum K. A., Lai Z., Lei Y., Li Z., Li J., Liang Y., Lin X., Lu F., Merkulov G. V., Milshina N., Moore H. M., Naik A. K., Narayan V. A., Neelam B., Nusskern D., Rusch D. B., Salzberg S., Shao W., Shue B., Sun J., Wang Z., Wang A., Wang X., Wang J., Wei M., Wides R., Xiao C., Yan C., et al. The sequence of the human genome. *Science*, 291(5507):1304–51, 2001.
- Vogel Gretchen. The ‘Do Unto Others’ Malaria Vaccine. *Science*, 328(5980): 847–848, 2010.
- Volkman S. K., Sabeti P. C., DeCaprio D., Neafsey D. E., Schaffner S. F., Milner Jr. D. A., Daily J. P., Sarr O., Ndiaye D., Ndir O., Mboup S., Duraisingh M. T., Lukens A., Derr A., Stange-Thomann N., Waggoner S., Onofrio R., Ziaugra L., Mauceli E., Gnerre S., Jaffe D. B., Zainoun J., Wiegand R. C., Birren B. W., Hartl D. L., Galagan J. E., Lander E. S., and Wirth D. F. A genome-wide map of diversity in *Plasmodium falciparum*. *Nature genetics*, 39(1):113–9, 2007.
- Wang D., Liu S., Trummer B. J., Deng C., and Wang A. Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells. *Nat Biotechnol.*, 20(3):275–81, 2002.
- White N. J. The Treatment of Malaria. *New England Journal of Medicine*, 335(11):800–806, 1996.
- White N. J. Qinghaosu (Artemisinin): The Price of Success. *Science*, 320(5874):330–334, 2008.

- WHO World Health Organisation. The World Malaria Report 2010. Technical report, The World Health Organisation, 2010a.
- WHO World Health Organisation. Global report on antimalarial drug efficacy and drug resistance: 2000-2010. Technical report, The World Health Organisation, 2010b.
- WHO World Health Organisation. Guidelines for the treatment of malaria - 2nd edition. Technical report, The World Health Organisation, 2010c.
- WHO World Health Organization. Indoor residual spraying, Global Malaria Programme. Technical report, The World Health Organisation, 2006.
- Wipasa Jiraprapa and Riley Eleanor M. The immunological challenges of malaria vaccine development. *Expert Opinion on Biological Therapy*, 7(12): 1841–1852, 2007.
- Wongsrichanalai C., Pickard A. L., Wernsdorfer W. H., and Meshnick S. R. Epidemiology of drug-resistant malaria. *The Lancet Infectious Diseases*, 2 (4):209–18, 2002.
- Wyss J. H. Screwworm eradication in the Americas. *Ann N Y Acad Sci.*, 916: 186–93, 2000.
- Yang L., Guo S., Li Y., Zhou S., and Tao S. Protein microarrays for systems biology. *Acta biochimica et biophysica Sinica*, 43(3):161–71, 2011.
- Yang Q., Xu J., Li M., Lei X., and An L. High-level expression of a soluble snake venom enzyme, glosedobin, in *E. coli* in the presence of metal ions. *Biotechnology letters*, 25(8):607–10, 2003.
- Yu X., Schneiderhan-Marra N., and Joos T. O. Protein microarrays for personalized medicine. *Clinical chemistry*, 56(3):376–87, 2010.
- Yup Sang and Lee. High cell-density culture of *Escherichia coli*. *Trends in Biotechnology*, 14(3):98 – 105, 1996.
- Zahn K. Overexpression of an mRNA dependent on rare codons inhibits protein synthesis and cell growth. *Journal of bacteriology*, 178(10):2926–33, 1996.
- Zhou Z., Schnake P., Xiao L., and Lal A. A. Enhanced expression of a recombinant malaria candidate vaccine in *Escherichia coli* by codon optimization. *Protein Expr Purif.*, 34(1):87–94, 2004.
- Zhu H., Bilgin M., Bangham R., Hall D., Casamayor A., Bertone P., Lan N., Jansen R., Bidlingmaier S., Houfek T., Mitchell T., Miller P., Dean R. A., Gerstein M., and Snyder M. Global analysis of protein activities using proteome chips. *Science.*, 293(5537):2101–5 Epub 2001 Jul 26, 2001.

Zhu H., Hu S., Jona G., Zhu X., Kreiswirth N., Willey B. M., Mazzulli T., Liu G., Song Q., Chen P., Cameron M., Tyler A., Wang J., Wen J., Chen W., Compton S., and Snyder M. Severe acute respiratory syndrome diagnostics using a coronavirus protein microarray. *Proc Natl Acad Sci U S A.*, 103(11): 4011–6 Epub 2006 Mar 7, 2006.

Zubay G. In vitro synthesis of protein in microbial systems. *Annu Rev Genet.*, 7:267–87, 1973.

Zumstein Ronald C. and Rousseau Ronald W. Solubility of L-isoleucine in and recovery of L-isoleucine from neutral and acidic aqueous solutions. *Industrial & Engineering Chemistry Research*, 28(8):1226–1231, 1989.

# Appendix

## International conferences (posters)

### **10<sup>th</sup> National Biotechnology Conference (CNB X)**

17/09/08 - 19/09/08, Perugia, Italy.

GENOME SCALE ANALYSIS OF THE IMMUNE RESPONSE AGAINST  
MALARIA FOR VACCINE CANDIDATES IDENTIFICATION

Tania Dottorini, **Giorgio Mazzoleni**, Carla Proietti, Davide Domenico  
Pettinato, Roberta Spaccapelo, Manlio Di Cristina, Julian Gray and Andrea  
Crisanti

### **5<sup>th</sup> Annual BioMalPar. Conference on the Biology and Pathology of the Malaria Parasite**

18/05/09 - 20/05/09, Heidelberg, Germany.

FIGHTMAL: CORRELATING PROTECTION AGAINST MALARIA  
WITH SERUM PROFILES AGAINST *PLASMODIUM FALCIPARUM*  
ANTIGEN REPERTOIRES

Francesca Baldracchini, Patrick Corran, Tania Dottorini, Chris Drakeley,  
Marilyne Faily, Julian Gray, Bernard Kanoi, Walter Low, Christophe  
Martin, **Giorgio Mazzoleni**, Edward Hosea Ntege, Brenda Okech, Davide  
Pettinato, Carla Proietti, Kevin Tetteh, Florence Wolff, Tristan Rouselle,  
Thomas Egwang, Eleanor Riley and Andrea Crisanti

### **6<sup>th</sup> European Congress on Tropical Medicine and International Health and 1<sup>st</sup> Mediterranean Conference on Migration and Travel Health**

06/09/09 - 10/09/09, Verona, Italy.

FIGHTMAL: CORRELATING PROTECTION AGAINST MALARIA  
WITH SERUM PROFILES AGAINST *PLASMODIUM FALCIPARUM*  
ANTIGENS

Patrick Corran, Chris Drakeley , Carla Proietti , Kevin Tetteh , Eleanor  
Riley, Bernard Kanoi , Edward Hosea Ntege, Thomas Egwang, Francesca  
Baldracchini, Julian Gray, Davide Pettinato, Walter Low, Christophe Martin,  
Florence Wolff, Tristan Rouselle, Marilyne Faily, Tania Dottorini, **Giorgio  
Mazzoleni** and Andrea Crisanti

**5<sup>th</sup> MIM Pan-African Malaria Conference**

02/11/09 - 06/11/09, Nairobi, Kenya.

FIGHTMAL: CORRELATING PROTECTION AGAINST MALARIA WITH SERUM PROFILES AGAINST *PLASMODIUM FALCIPARUM* ANTIGENS

Tania Dottorini, Patrick Corran, Teun Bousema, Chris Drakeley, Edward Hosea Ntege, Walter Low, Julian Gray, Brenda Okech, Kevin Tetteh, Francesca Baldracchini, Marilyn Faily, Florence Wolff, **Giorgio Mazzoleni**, Francesco Trivelli, Bernard Kanoi, Carla Proietti, Davide Pettinato, Christophe Martin, Tristan Rouselle, Thomas Egwang, Eleanor Riley and Andrea Crisanti

## Seminars and presentations

**Hospital for Tropical Diseases, Seminar**

29/05/08, London, UK

PROTEIN MICROARRAYS: APPLICATIONS IN SERODIAGNOSTICS AND VACCINE DEVELOPMENT

**Giorgio Mazzoleni**

**XXXVII AMCLI National Congress. Application of protein microarrays in microbiology: state of the art and prospects**

05/10/08-08/10/08, Stresa, Italy

IDENTIFICATION OF CORRELATES OF PROTECTION AGAINST THE *PLASMODIUM FALCIPARUM*

**Giorgio Mazzoleni**

**Imperial College London, Final Year PhD Student Poster Presentation**

12/05/10, London, UK

MICROARRAY-BASED ANTIGENICITY SCREENING FOR THE IDENTIFICATION OF NOVEL VACCINE CANDIDATES FOR *PLASMODIUM FALCIPARUM* MALARIA

**Giorgio Mazzoleni**, Tania Dottorini, Patrick Corran, Teun Bousema, Chris Drakeley, Edward Hosea Ntege, Walter Low, Mauro Maccari, Brenda Okech, Kevin Tetteh, Francesca Baldracchini, Marilyn Faily, Florence Wolff, Francesco Trivelli, Bernard Kanoi, Carla Proietti, Davide Pettinato, Christophe Martin, Tristan Rouselle, Thomas Egwang, Eleanor Riley and Andrea Crisanti

## Publications

Gepoclu: a software tool for identifying and analyzing gene positional clusters in large-scale gene expression analysis.

Dottorini T, Senin N, **Mazzoleni G**, Magnusson K, Crisanti A.

*BMC Bioinformatics*. 2011 Jan 26;12:34.

PubMed PMID: 21269436; PubMed Central PMCID: PMC3040130.

Serum IgE reactivity profiling in an asthma affected cohort.

Dottorini T, Sole G, Nunziangeli L, Baldracchini F, Senin N, **Mazzoleni G**, Proietti C, Balaci L, Crisanti A.

*PLoS One*. 2011;6(8):e22319. Epub 2011 Aug 4.

PubMed PMID:21829614; PubMed Central PMCID: PMC3150333.

An antigen microarray immunoassay for multiplex screening of mouse monoclonal antibodies.

Di Cristina M, Nunziangeli L, Giubilei MA, Capuccini B, d'Episcopo L, **Mazzoleni G**, Baldracchini F, Spaccapelo R, Crisanti A.

*Nat Protoc*. 2010 Dec;5(12):1932-44.

Epub 2010 Nov 10. PubMed Central PMCID: 21127487.