# Expression and initial characterisation of the *Plasmodium falciparum* general transcription factors TFIIB and TLP

**By**

**Steven Bing**

**Dissertation presented for the degree of**

**Master of Science**

**In the Department of Molecular and Cell Biology**

**University of Cape Town**

**February 2015**



**Supervisor**

**Dr Thomas Oelgeschläger**

The financial assistance of the National Research Foundation (DAAD-NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the DAAD-NRF.

## Declaration

I know the meaning of Plagiarism and declare that all of the work in the document, save for that which is properly acknowledged, is my own.

Signed: Steven Bing

Signature: Signed by candidate

Signature removed

# Acknowledgements

I am deeply grateful to my family for the tireless love and support they have shown me, not only during my academic career, but in all things. Mom, Dad, and Graeme, thank you.

Thank you to my supervisor, Dr Thomas Oelgeschläger, for the guidance and advice, as well as the time and energy you have spent on my education.

To the members of my lab, Tom, Alma, Gertrud, and of course, Rob, I am so glad to have had the opportunity to work with you. You have all been such great lab mates and I appreciate the friendships we share. Thank you for the help, support, and buffers.

I have had the supreme pleasure of spending the last few years in the excellent company of the postgraduate and faculty members of the Department of Molecular and Cell Biology. I will continue our tradition of tea at 10:30, and think of you often.

A special thank you must of course be extended to the support staff of the Department. The research undertaken in the Department is truly a team effort, and I am grateful for the role you play in making sure research runs smoothly.

I would like to extend a general sense of gratitude to one and all who directly or indirectly, have lent their helping hand in this research project.

Finally, thank you Tyronne M$^{c}$Crindle. I truly do not know how I could have done this without you.

The following reagents were obtained through the MR4 as part of the BEI Resources Repository, NIAID, NIH: *Plasmodium falciparum P. falciparum* 3D7 library, MRA-296-299, deposited by D. Chakrabarti, and for that, we extend our gratitude.

# Abstract

Malaria is a leading cause of morbidity and mortality worldwide, and results in approximately 600,000 deaths annually. The life cycle of the parasite is complex, and has several distinct stages of development. The transitions between these stages are brought about through tightly controlled and highly synchronized changes in gene expression. *Plasmodium falciparum* causes the most lethal form of malaria in humans. The parasite is particularly virulent as it is able to evade immune detection by the infected host. This virulence is directly related to the expression of variable antigens on the surface of infected red blood cells. The control of gene expression is known to be largely regulated via RNA Polymerase II (RNAPII) transcription initiation, but in *P. falciparum* the underlying mechanisms have not been determined. This primarily because very little is known about both the key protein factors and DNA elements which guide the assembly of RNAPII components into the transcription initiation complex. Bioinformatics studies have shown that there is very little amino acid sequence conservation between human and *Plasmodium* RNAPII transcription initiation components. Together with the observation that the *Plasmodium* genome has an extremely high A+T content, this suggests that *Plasmodium* may have specific mechanisms to initiate transcription, which could be targeted by novel anti-malarials. The general transcription factor TFIIB and the TBP-like protein (TLP) are key proteins involved in the recognition of the core promoter, and the initiation of RNAPII transcription initiation complex assembly. TFIIB stabilises DNA binding of the primary promoter recognition factor, TATA-box binding protein (TBP), and is involved in promoter recognition through interactions with specific DNA sequences up- and downstream of the TBP DNA binding site. TBP-like protein is a member of the TBP protein family that has been implicated in life cycle stage specific gene transcription initiation in various eukaryotic model organisms. This research study reports the first successful expression and purification of recombinant epitope-tagged *Plasmodium falciparum* TFIIB and TLP proteins. Preliminary assays demonstrate DNA-binding activity for the recombinant *Plasmodium* TBP-like protein, and suggest DNA-binding activity in *Plasmodium* TFIIB protein, which has not been demonstrated before in eukaryotic TFIIB.

# Common Abbreviations

A,T,G,C - adenine, thymine, guanine, cytosine
Ad2ML - adenovirus 2 major late
ApiAP2 - Apicomplexan AP2
BLAST - Basic Local Alignment Search Tool
BLASTp - protein BLAST
bp – base pairs
BREu/d - upstream/downstream B-recognition element
BSA - bovine serum albumin
C-;N- terminus – carboxyl-; amino- terminus
CAP - catabolite activator protein
cDNA - DNA copy synthesized from mRNA
COBALT - Constraint-based Multiple Protein Alignment Tool
DNA - deoxyribonucleic acid
dNTP - deoxyribose nucleoside triphosphates
DPE - downstream promoter element
DTT - Dithiothreitol
*E. coli* - *Escherichia coli*
EDTA - Ethylenediaminetetraacetic acid
EMSA - electrophoretic mobility shift assay
EtBR - ethidium bromide
FastAP - FastAP Thermosensitive Alkaline Phosphatase
GBP - glycophorin-binding protein
GBP-130 - glycophorin binding protein 130
GSH - glutathione
GST - glutathione-S-transferase
GTF - general transcription factor
*H. sapiens* - *Homo sapiens*
HEPES - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRP - horseradish peroxidase
HTH- helix-turn-helix
Inr – initiator
IPTG - Isopropyl β-D-1-thiogalactopyranoside
iRBCs - infected red blood cells
KAHRP - knob-associated histidine rich protein
LA - lysogenic agar
LB - lysogenic broth
LCR - low-complexity region
mRNA - messenger RNA
MSA - Multiple sequence alignments
MTE - motif ten element
Mw - molecular weight
Ni-beads - PureProteome™ Nickel Magnetic Beads (*Merck Millipore*)
NP 40 - Nonidet™ P 40 Substitute (*Sigma-Aldrich*®)
OD - optical density
ORF - open reading frame
*P. falciparum* - *Plasmodium falciparum*
PCF - pooled column fraction
PCR - polymerase chain reaction
*Pf* - *Plasmodium falciparum*
*Pf*EMP - *P. falciparum* erythrocyte membrane protein
*Pf*TBP - *Plasmodium falciparum* TATA-box binding protein
*Pf*TFIIA - *Plasmodium falciparum* transcription factor IIA
*Pf*TFIIB - *Plasmodium falciparum* transcription factor IIB
*Pf*TLP - *Plasmodium falciparum* TBP-like protein
*Pf*TLPco - codon optimised *Pf*TLP
PIC - pre-initiation complex
PVDF - polyvinylidene difluoride
Q-, DEAE-, CM-, SP-Sepharose® - quaternary ammonium, diethylaminoethanol, carboxymethyl, sulphopropyl
RBC - red blood cells
RCF - relative centrifugal force
RNA - ribonucleic acid
RNAPII - RNA polymerase II
rRNA - ribosomal RNA
*S. cerevisiae* - *Saccharomyces cerevisiae*
SDS - sodium dodecyl sulphate
SDS-PAGE - SDS polyacrylamide gel electrophoresis
SELEX - Systematic evolution of ligands by exponential enrichment
SOC - Super Optimal broth with Catabolite repression
TAE - tris base, acetic acid and EDTA
TBE - tris base, boric acid and EDTA
TBP - TATA-box binding protein
TE - Tris-EDTA
TEV - Tobacco Etch Virus
TF - transcription factor
TFII- A,B,D - transcription factor II – A,B,D,
TAF - TBP-associated factor
TLP - TBP-like protein
TRF - TBP-related factor
tRNA - transfer RNA
TSS - transcription start site
UTR - untranslated region

# Contents

Table of Contents

Table of Contents

# Chapter 1

# Introduction

## 1.1 *Plasmodium falciparum*

### 1.1.1 The impact of *Plasmodium* infection

Malaria is a leading cause of morbidity and mortality across the globe (World Health Organisation 2014). Approximately 40% of the world population live in malaria endemic areas. In 2013 malaria resulted in about 200 million infections, and approximately six hundred thousand deaths. The vast majority of infections (~80%) occur in Africa, with ~90% of the deaths occurring on this continent, the majority of which are children under the age of five years. Malaria in humans is caused by four species of *Plasmodium*. These include *P. malariae, P. vivax, P. ovale,* and as the cause of cases with the highest rates of complications and death, *Plasmodium falciparum*.

### 1.1.2 Life cycle of *Plasmodium falciparum*

*Plasmodium falciparum* is an alveolate protist, of the group apicomplexa. Apicomplexans consist of single-celled eukaryotic organisms, many of which are parasitic to both invertebrates and vertebrates, leading to a host of diseases. As a group, the apicomplexans typically have multi-stage lifecycles, and with variable morphologies as they inhabit their multiple species-specific hosts (Neva & Brown 1994).

*Plasmodium falciparum* infects the female *Anopheles gambiae* mosquito and the human host. In taking up a blood meal from an infected human host, the mosquito takes up male and female gametocytes. Within the mosquito midgut, the male-derived micro- and female gametocyte-derived macrogametes of the parasite fuse to form a zygote (entering into the

sporogenic cycle), which proceeds to form the ookinete. The ookinete passes through the gut wall, and encysts on the exterior wall of the gut to form the oocyst. The oocyst eventually ruptures, and releases sporozoites into the mosquito body cavity. The sporozoites then travel to the mosquito salivary glands.

The sporozoites are released into bloodstream of the human host when the mosquito next feeds (the exo-erythrocytic cycle). The sporozoites then rapidly invade the hepatocytes of the liver. For *P. falciparum*, in the next 6-9 days the sporozoites develop into trophozoites, and then undergo multiple rounds of nuclear division forming a schizont, a single cell containing multiple nuclei without cell segmentation. After several rounds of nuclear division, the cell segments to form thousands of merozoites. Note that in *P. vivax,* the sporozoites may instead differentiate into hypnozoites, which may lie dormant in the liver, and are the cause of recurrent bouts of malaria (Karunaweera et al. 1992). Several tens of thousands of merozoites may be released from a bursting hepatocyte, and then individually invade red blood cells (erythrocytes; RBCs) entering the erythrocytic cycle (Cowman & Crabb 2006). Symptoms typically appear 7 – 14 days post sporozoite inoculation. In order to infect the erythrocytes, *P. falciparum* parasites recognise and bind to the erythrocyte surface receptors, for example, the sialoglycoproteins, glycophorins A and B. The parasites bind to glycophorin surface receptors with glycophorin-binding proteins (GBPs) via tandem binding repeats present on the GBP surface (Perkins 1984; Cowman & Crabb 2006). Having bound to the RBC, the parasite invades the target cell and resides within the parasitophorous vacuole.

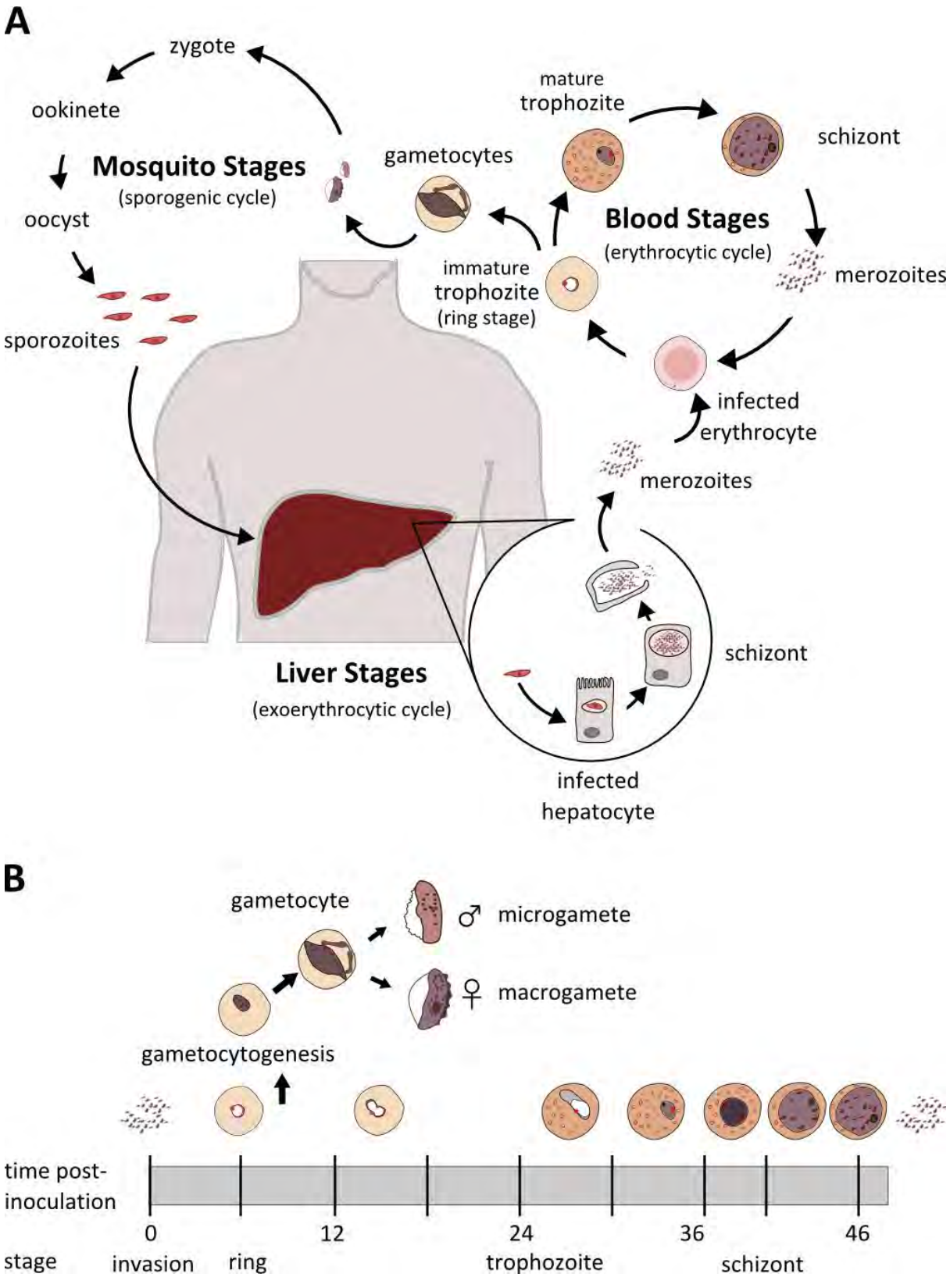



***Figure 1: Overview of the erythrocytic-stage life cycle of Plasmodium falciparum***

**A.** Life cycle of *P. falciparum*. Main stages of development are indicated.
**B.** Stages of the erythrocytic cycle of *P. falciparum.* Approximate time post-inoculation for the stages of development are indicated.
Illustrations are referenced from microscopy images of infected erythrocytes (White 2008; Alano 2007).

In the erythrocytes, the merozoites differentiate into the (immature) trophozoites. This is often called the 'ring' stage due to the morphology seen on a blood smear (Neva & Brown 1994). This is a highly metabolic stage, where the trophozoite enlarges for the next 12 hours, ingesting the erythrocyte cytoplasm and proteolysing the host haemoglobin (Francis et al. 1997). The mature trophozoites then differentiate further into a schizont. The schizonts produce a large number of merozoites (32 nuclei per schizont). The released merozoites then invade additional erythrocytes (Neva & Brown 1994).

Note that during the early erythrocytic stage (approximately 6 hours post-erythrocyte infection) some merozoites may differentiate into male and female gametocytes. These gametocytes reside within the erythrocytes until taken up again by the mosquito vector, and so complete the life cycle (Alano 2007). The mechanisms behind the commitment to sexual reproduction are unknown (Winzeler 2009).

The erythrocytic stage of the *Plasmodium falciparum* occurs over approximately 48 hours, from invasion to the bursting of the infected erythrocyte. The development into the ring stage occurs approximately 6 hours post-erythrocyte infection, the trophozoites mature at approximately 24 hours, the schizont forms at approximately 40 hours (Neva & Brown 1994).

### 1.1.3 Pathology of *Plasmodium falciparum*

The symptoms of malaria caused by *P. falciparum* differ from malaria caused by other *Plasmodium* species. As with all malaria infections, classic symptoms include paroxysm, a cyclical onset of sudden coldness followed by fever and sweating (Greenberg & Lobel 1990). Typically, paroxysms occur synchronously with the merozoite release in the intraerythrocytic stages. In *P. vivax/ovale* infections, this occurs every two days, and every

three days for *P. malariae*. *P. falciparum* infection may not have the classic paroxysms, and instead exhibit a continuous fever (Neva & Brown 1994).

In the determination of the severity of *falciparum* malarial infection, a primary determinant is the expression of proteins which are exported to the surface of infected erythrocytes, (Miller et al. 2013). The proteins expressed include the knob-associated histidine rich protein (KAHRP) and *P. falciparum* erythrocyte membrane protein 2 and 3 (*Pf*EMP2, *Pf*EMP3), which form a layer directly beneath the cell membrane, inducing a protruding 'knob' in the area by interacting with the host cellular structures and proteins (Lanzer et al. 1993). KAHRP proteins act as a platform for the presentation of the *P. falciparum* erythrocyte membrane protein 1 (*Pf*EMP1). This presentation of knobs and associated structures are a primary determinant of cytoadherence and severe malaria (Lanzer et al. 1993). The knobs cause infected red blood cells (iRBCs) to cytoadhere both to one another as well as other cells. Cytoadherence of iRBCs to non-infected RBCs as well as endothelial and other intravascular cells of the body prevents the clearance of the iRBCs by the spleen, and facilitates the positioning of infected cells to regions of optimum parasite growth (Buffet et al. 2010). These infected cells become sequestered in the micro-vascular structures of organs such as the heart, lungs and brain, and placenta, separate from peripheral blood circulation (Khoury et al. 2014). The consequence of this accumulation and sequestration of cells in organs is the onset of severe disease state of the human host, such as cerebral malaria, the outcome of which may be fatal comas and seizures (Buffet et al. 2010).

*Pf*EMP1 is encoded by the highly variable (*var*) genes. *P. falciparum* possess approximately 60 *var* genes, although only one is expressed at any one time (Guizetti & Scherf 2013). *P. falciparum* infections are detected by the host immune system when the parasites are in free peripheral circulation, and through the detection of iRBCs by the antigenic *Pf*EMP1 variants. The mechanism of switching the *var* genes allow the parasite to evade immune attack by the

host, preventing long-term immunity and so persist as a chronic infection (Bachmann et al. 2012).

The control of gene expression is not only required for the parasite to alter the erythrocyte displayed antigens rapidly, rather the complex life cycle of *P. falciparum* as a whole requires strictly controlled and co-ordinated changes in gene expression. Clearly, understanding of the molecular mechanisms underlying the control of gene expression will be a prerequisite to fully understand *Plasmodium* biology and the molecular basis of *Plasmodium* pathogenesis.

### 1.1.4 The *Plasmodium* genome

The genome of *P. falciparum* was published in 2002 (Gardner et al. 2002). This has allowed remarkable insight into the parasite, and it is hoped that insights from this resource will drive the knowledge of this parasite forward, not only in facilitating investigations into potential treatments and vaccines, but also into understanding the physiology and pathology of the organism. The *P. falciparum* genome contains up to 27 million bases, is separated into 14 chromosomes and expresses approximately 5500 genes (Florens et al. 2002).

An interesting feature of the *P. falciparum* genome is its adenine and thymine nucleotide (A + T) richness, with A + T contributing to an average of 80 -90% of the genome (Gardner et al. 2002). Other *Plasmodium* species do not have this degree of A+T richness, it being the highest A+T content of any genome sequenced. Several theories have been posited to explain the A+T content, including that the high levels of A+T richness may allow for rapid evolution and recombination, and so allow for increased immune system evasion. Due to the A+T richness it was found that protein sequences have diverged so extremely that they complicate the association of protein identity to hypothetical proteins. Of the ~5300 proteins predicted to be encoded by the genome, 60% do not have sufficient similarity to proteins in

other organisms to imply function/identity. This is likely due to the A+T-richness of the genome, and consequentially, the suite of codons available for use.

Proteins in the *P. falciparum* genome are heavily populated with low complexity regions (Zilversmit, Volkman, DePristo, et al. 2010). It has been found that the *P. falciparum* proteome is richer in low complexity regions (LCRs) than other organisms studied, with at least one LCR present in 87% of genes. Other organisms average between 65-70% of genes having LCRs. These LCRs also fall into different classes, which may have effects on their function and rate of evolution. There are multiple theories about the function of LCRs. LCRs may be an adaptive mechanism to adjust translation rates (Frugier et al. 2010), may be a method of regulating messenger ribonucleic acid (mRNA) stability, may be involved in generating antigenic diversity through recombination, or may simply be the result of increased recombination rates due to the A+T rich genome, discussed by Zilversmit et al. (2010).

### 1.1.5 Gene expression in *Plasmodium falciparum*

There is evidence to support regulation of gene expression at various levels (reviewed in, Horrocks et al. 2009) including gene activation, post-transcriptional regulation, as well as epigenetic regulation.

Evidence for gene expression regulation arising from control of gene transcription has been shown in whole-genome/proteome studies (Florens et al. 2002; Bozdech, Llinás, et al. 2003). These studies making use of both microarray and proteomic data, and support a model of expression of genes in a life cycle stage-dependent manner, in which transcription of genes is tuned temporally for the physiological processes that are likely to occur at that stage of development. There is some evidence that this 'just in time model' needs revision (Horrocks et al. 2009). For example, it has been found that changes in the environment, such as the

stresses endured by temperature changes or glucose starvation, result in global expression changes. This suggests that the parasite is not locked into a specific pattern of gene expression (Horrocks et al. 2009). Further studies have also shown a lag between changes in mRNA transcript and protein levels, as well as a life cycle stage-specific rate of controlled mRNA decay, with increasing mRNA stability across the 48 hour erythrocytic cycle (Shock et al. 2007). Similarly, it has been found that the parasite also produces different sets of ribosomal RNAs (rRNAs) depending on life cycle stage, which is likely to be an additional level of translational control (Gardner et al. 2002). Altogether there is evidence now that there is a deeper level of transcriptional control than the 'just in time' model would suggest, and will remain the focus of future research.

The presence of LCRs maybe an additional mechanism used by the organism to regulate translation rate (Frugier et al. 2010). Other organisms may adjust translation rates by altering the codons in a gene to make use of more or less abundant transfer RNAs (tRNAs), and so either speed up or slow translation, but *P. falciparum* only has one copy of each tRNA-coding gene, suggesting equal numbers of the tRNAs, and so the loss of this method of translational regulation. Instead, by over-utilising a set of tRNAs in an LCR, the translational rate of the region is likely to reduce, and so allow more time for folding of complex secondary structures, such as the alignment of β-strands to form β-sheets (Frugier et al. 2010).

There has been a focus on regulation of transcription through epigenetics, due to the identification of multiple chromatin remodelling proteins (Coulson et al. 2004). Chromatin structure affects gene expression through either allowing the activation of genes through increasing the accessibility to them, or silencing gene expression through silencing of the region. A partial explanation for the stage-specific gene expression seen previously (Florens et al. 2002), is that genes of a shared process may be closely located on the chromosomes,

and so share activation by chromatin remodelling. Multiple studies investigating histone modifications have been performed, reviewed in (Hoeijmakers et al. 2012; Guizetti & Scherf 2013; Cui & Miao 2010).

Control of gene expression in *P. falciparum* at the transcriptional level is supported by the presence of features common to other eukaryotes, as found in early studies (Lanzer et al. 1992a; Lanzer et al. 1992b; Horrocks et al. 1998; Lanzer et al. 1993). Similar to model eukaryotes, transcription is regulated through a bipartite promoter system, consisting of (i) a core promoter region at which RNA polymerase II (RNAPII) forms the pre-initiation complex (PIC) with the general transcription factors (GTFs) and drives transcription from the transcription start site (TSS). Additional regulatory regions control the activity of the core promoter region, including (ii) *cis*-acting regulatory elements, to which additional specific transcription factors (TFs) may bind. Analogous to other model organisms, the binding of these *cis*-acting regulatory elements by positive or negative regulators may either support or inhibit the formation of the PIC, and so are correlated with the activity of the promoter and the accumulation of mRNA (Horrocks et al. 2009). Transcription of *P. falciparum* genes typically results in monocistronic transcripts which contain intronic and exonic regions, conserved splicing sites (and machinery) as well as 5' and 3' untranslated regions (UTRs), which are themselves modified by capping and polyadenylation, respectively (Horrocks et al. 2009).

The formation of the PIC as the key step in transcriptional gene control made it a target for finding the *Plasmodium* protein orthologues involved. The key components of the PIC machinery were identified by hidden Markov model profile searching in the *P. falciparum* genome (Bischoff & Vaquero 2010). This search uncovered orthologues of RNAPII and most, but not all, GTFs. These include the TATA binding protein (TBP) and some of the other components of the transcription factor IID (TFIID) system, the TBP-associated factors

(TAFs). So far, only TAFs 1, 2, and 7 could be identified in *P. falciparum*. By contrast metazoans possess at least 13 TAFs. The search also uncovered orthologues for the transcription factors (TFII-) A, B, E, F, and H, and the TBP homologue TBP-like protein (Bischoff & Vaquero 2010). Interestingly, the TAFs that have not been found (excepting TAF5) all have histone-fold domains. Similarly, the histone H1, which also contains this motif, is absent in *P. falciparum*. Identification of other transcription-associated proteins resulted in only approximately 1.3% of the proteins uncovered, about one third of the number found in other free-living eukaryotes (Coulson et al. 2004). This has led to the suggestion that control of gene expression may arise predominantly from post-transcriptional and epigenetic regulatory mechanisms.

The A+T richness of the *Plasmodium* genome may be partly responsible for the paucity of transcription factors found by bioinformatics approaches, due to low sequence homology of *Plasmodium* genes to other protists. Of the proteins which have been putatively identified as being involved in transcriptional regulation, very few have been fully characterised, either *in vitro* or *in vivo*. The *Plasmodium* Apicomplexan AP2 (ApiAP2) factors are a group of transcription factors common to the apicomplexan group, which have been studied *in vitro* through EMSA and protein-binding microarray assays. Five of the ApiAP2 factors have also been characterised *in vivo*. These include factors involved in the function of sporozoites (Yuda et al. 2010), the expression of invasion related genes, *var* gene silencing, and two factors important for gametocytogenesis (De Silva et al. 2008; Painter et al. 2011; Sinha et al. 2014). Two other transcription factors have been partly characterised, the *Pf*Myb1 protein which binds to several promoters and controls the expression of the associated genes (Gissot et al. 2005), and the PREB protein, which has been shown to bind and activate gene expression at the *Pf1-cys-prx cis*-element (Komaki-Yasuda et al. 2013).

Locating *cis*-acting recognition elements, as well as the sequences for the core promoter has been particularly challenging, in part due to the A+T content of the genome (~90% intergenic), but also due to technical challenges in the mapping of TSSs (Wakaguri et al. 2009; Horrocks et al. 2009; Brick et al. 2008). Several studies have attempted to identify these regions through bioinformatics (Young et al. 2008; Jurgelenaite et al. 2009; Brick et al. 2008), however, the sequences obtained sometimes conflict, and without biochemical investigation cannot be confirmed.

Although several of the general transcription factors and RNAPII have been identified *in silico*, there is still very little information available regarding the function of this crucial transcription apparatus. Without this information the understanding of transcriptional regulation in *Plasmodium,* as well as understanding of the characteristics of the unique genome cannot be uncovered.

### 1.1.6 The RNA polymerase II pre-initiation complex

The following description of the formation of the PIC and the components thereof are thoroughly reviewed in Thomas & Chiang, (2006). The formation of the PIC occurs at the core promoter region surrounding the TSS (designated as +1 on the sense strand) Figure 2B. These core promoter elements have been identified in other eukaryotes, such as humans and yeast (Decker & Hinton 2013). They include the TATA box, a conserved region of sequence TATAWAAR (in metazoans), which is occasionally present in the core promoter, 25-30 nucleotides upstream of the TSS, and to which TBP binds. In metazoans and yeast, the PIC is nucleated by the binding of TBP, as part of the TFIID complex, to the core promoter region. The binding of the TFIID protein-complex is an induced fit mechanism, where the binding of DNA by TBP and insertion of the conserved phenylalanine residues leads to bending of the DNA, causing torsional stress on the DNA as well as a conformational shift in the TFIID complex itself (Cianfrocco & Nogales 2013). The PIC then forms in a step-

wise fashion. The binding of TFIIA stabilises TBP-DNA binding, and competes with proteins which inhibit TBP DNA-binding and/or function (Kokubo et al. 1998; Grünberg & Hahn 2013). TFIIB binds next, and enhances TBP binding to the TATA box (Zhao & Herr 2002; Imbalzano et al. 1994). TFIIB then assists in the recruitment of TFIIF and RNAPII (Chen & Hampsey 2004; Kostrewa et al. 2009; Bushnell et al. 2004; Tubon et al. 2004), and

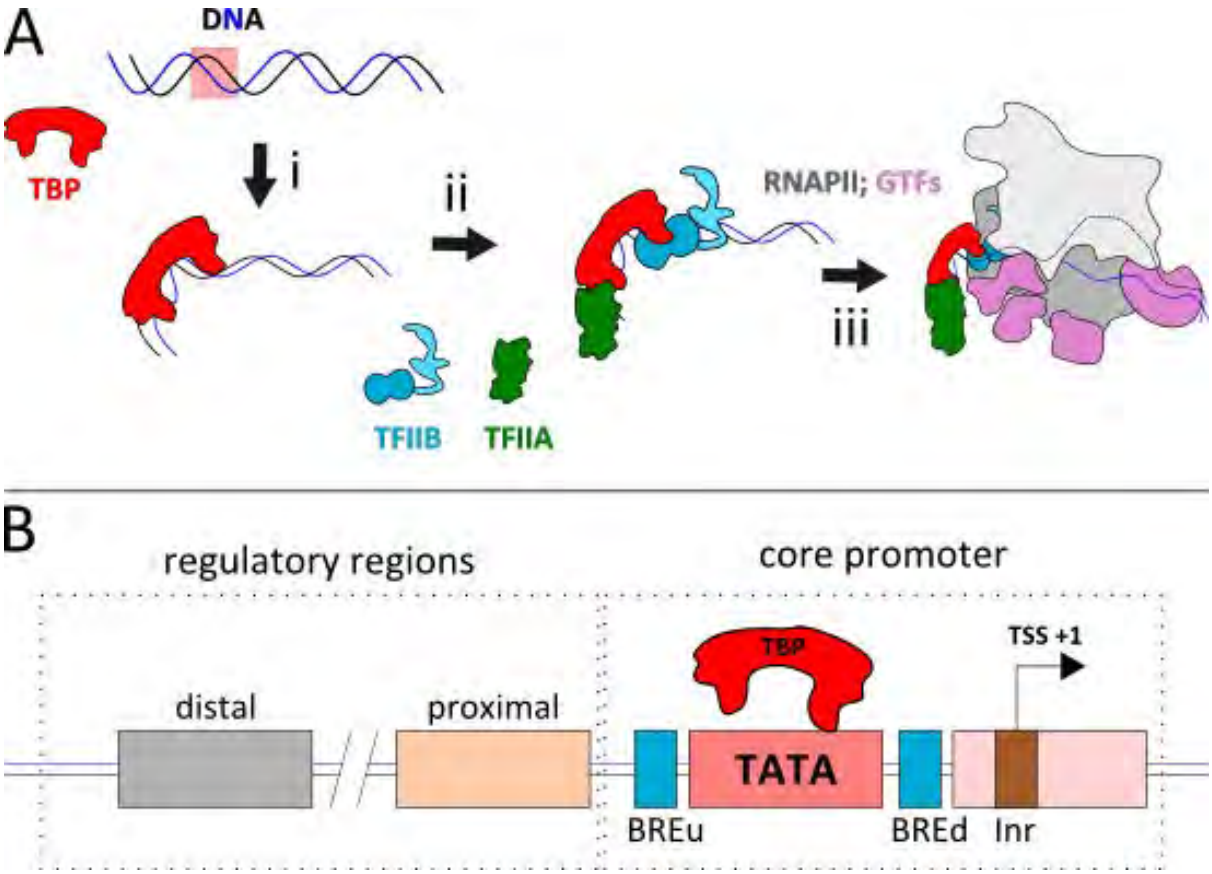



***Figure 2: Simplified model of the core promoter and RNA polymerase II pre-initiation complex (PIC) assembly.***

**A.** The PIC is initiated by (i) the binding of TBP as part of the TFIID complex (not shown) to the TATA-box (pink region) which causes a bend in the DNA. (ii) TFIIA and TFIIB bind to TBP and stabilise the association to DNA, and to recruit (iii) RNAPII and the rest of the GTFs. The DNA around the transcription start site is melted and transcription is initiated.

**B.** A simplified illustration of gene regulatory regions present in eukaryotes. The core promoter contains the transcription start site (TSS) and core promoter elements, such as the TATA box (recognised by TBP) and B-recognition elements up- (u) and downstream (d) of TATA, recognised by TFIIB. PIC assembly at the core promoter is regulated by proximal and distal (enhancer) promoter regions that contain binding sites for sequence-specific DNA-binding transcription activator or repressor proteins.

finally TFIIE and TFIIH. The binding of additional GTFs causes the DNA to melt, promoter opening, and transcription to occur (Figure 2A). The TATA box is not ubiquitous in the core promoter, and other core promoter elements may be present, both in tandem with TATA, or without (Thomas & Chiang 2006). The combination of core promoter elements differ between genes, and none have been found to be indispensable, although some combinations do occur more frequently than others.

The core promoter may contain an initiator (Inr) sequence, which surrounds the TSS, and is bound by TAFs 1/2, a downstream promoter element (DPE) 28-34 nucleotides downstream of the TSS (in *Drosophila*) bound by TAFs 6/9 (Thomas & Chiang 2006). There may also be a motif ten element (MTE, at +18 to +29) which functions in tandem with the Inr region, and may substitute for a missing TATA box and/or DPE (in conjunction with the Inr region). The downstream core element (DCE) is present (as three sub-elements at +6 to +11, +16 to +21, and +30 to +34) but mutually exclusive with the DPE (located at +23 to +34) interact with TAFs 1 and 6 respectively (Thomas & Chiang 2006). TFIIB stabilises the TBP-DNA interaction through the recognition of elements in the core promoter, the B-recognition elements up and downstream of the TATA box (BREu and BREd, -38 to -32 and -23 to -17 respectively, in yeast) (Lagrange et al. 1998; Deng & Roberts 2005). The presence or absence of these BREs also selects for the mutually exclusive recruitment of either TFIIA or NC2 (respectively) (Deng et al. 2009).

### 1.1.1 The general transcription factors

TFIID

TFIID is a multi-protein complex, composed of TBP and the TAFs. TBP is the predominant DNA-binding component of TFIID and recognises the TATA-box (in core promoters with

this element) while the TAFs recognise other core promoter elements, (Thomas & Chiang 2006; Juven-Gershon et al. 2008).

TBP (and its homologues/orthologues) consist of a very highly-conserved carboxyl- (C-) terminal domain, with a large and variable amino- (N-) terminal domain, to which no structure has been prescribed (Akhtar & Veenstra 2011). The C-terminal domain, or core domain, is responsible for DNA binding. The core is saddle-shaped, with a concave and positively charged surface which makes contact with the DNA, while the convex surface makes contact with other proteins/factors (Figure 3). The functional role of the variable N-terminal is not well characterised. Early *in vitro* studies suggested that the domain is required for TATA-dependent transcription, (Lescure et al. 1994), although others have shown no loss of cell-growth or transcription in mice with homozygous deletions of the N-terminus (Schmidt et al. 2003). In mice, the loss of the N-terminus has only been shown to lead to rejection of implanted embryos in females, while males were seemingly unaffected (Hobbs

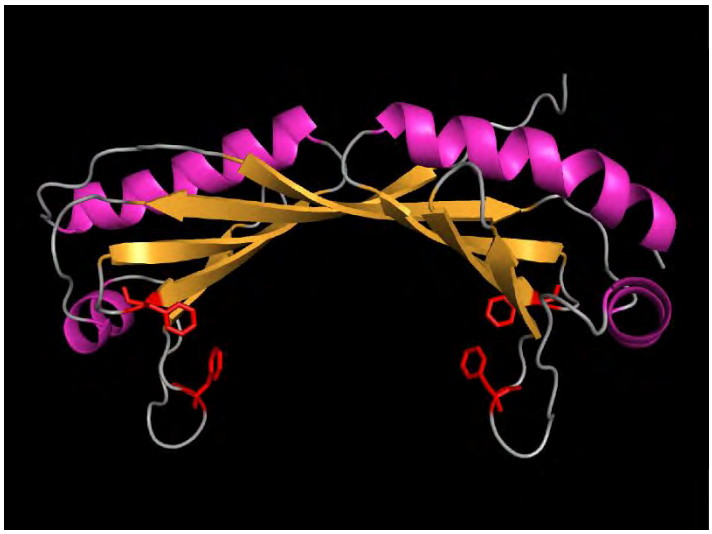


***Figure 3: Tertiary structure of human TATA-box binding protein (core region).***

Concave region formed by β-sheets (orange) which interact with DNA. Phenylalanine residues highlighted in red. Convex region formed by α-helices in purple. Structure model shown was prepared using PyMol based on the PDB entry 1CDW (Nikolov et al. 1996).

et al. 2002). The expansion of the glutamine repeats in the region leads to spino-cerebellar ataxia in humans (van Roon-Mom et al. 2005), although the mechanism is not understood.

*P. falciparum* TBP (*Pf*TBP) was cloned before the bioinformatics identification of the general transcription factors (McAndrew et al. 1993), but has only been partially characterised thus far in a single study (Ruvalcaba-Salazar et al. 2005). In this study, the core region of *Pf*TBP was recombinantly expressed and N-terminally tagged with glutathione-S-transferase (GST). The DNA-binding sites of the protein was examined by electrophoretic mobility shift assay (EMSA) and DNase I foot printing assay to the *kahrp* and *gbp-130* genes. This study was successful in mapping the *Pf*TBP binding to TATA-like sequences upstream of the TSS, at -81 and -186 nucleotides upstream from the TSS of the *kharp* and *gbp-130* genes respectively. This conservation of the TATA-locality of TBP binding is remarkable, given the A+T rich genome, and the prevalence of cryptic TATA-like elements. Interestingly, Brick et al. (2008) suggest the presence of physiochemical signals in the region of the TSS, identified computationally, may be used to identify the core-promoter elements.

TBP-related factors.

Although TBP is the most studied of the proteins in the family, there exist several TBP-related factors (TRFs), reviewed extensively by Akhtar & Veenstra (2011).

The discovery of TRFs at different times and in different organisms, with different numbers of TRFs, and combined with difficulty in establishing homology of these TRFs, has led a fair amount of confusion in the naming of the TRFs. The first TRF was discovered in *Drosophila*, and was named TRF1. This TRF1 is not found in vertebrates. Metazoans may have another TRF, in *Drosophila* named TRF2, as it is in humans, but may alternately be named the TBP-related protein (TRP) or TBP-like factor (TLF) or the TBP-like protein

(TLP) or TBP-like protein 1 (TBPL1). TLP is the name of the orthologue that will be referred to in this thesis. Another TRF is present in vertebrates, named alternately TBP2 or TRF3, or TBPL2 (Zhang et al. 2001).

The functional roles of the TRFs are less well characterised than TBP. It has been suggested that the role of the TRFs is to provide an additional method of gene transcriptional regulation. In this model TBP-like protein of TFIID is switched for TBP and allows expression from tissue specific promoters (Hochheimer & Tjian 2003).

TBP2 is the most similar of the TRFs, with ~90% similarity between the human paralogs over the core domain. TBP2 is known to be able to associate to the TATA box, and may replace TBP at certain genes (Bartfai et al. 2004; Jallow et al. 2004). For example, TBP2 of *Xenopus laevis* has been shown to be expressed predominantly in the oocytes, as an alternative to TBP (Akhtar & Veenstra 2009). TBP2 in mice gametogenesis appears to complement TBP, seemingly activating alternative genes to TBP (Gazdag et al. 2007). TRF1 from *Drosophila* appears to function predominantly with RNA polymerase III transcription (Akhtar & Veenstra 2011).

TLP is a more distant paralogue to TBP and TBP2 (Akhtar & Veenstra 2011), with lower genetic similarity than the relationship between TBP2 and TBP (over the core domain). It has not been demonstrated to have affinity for the TATA box (Moore et al. 1999), however TLP does appear to be able to bind both TFIIA and TFIIB, as well as the TFIIA-like factor (ALF). TLP has been shown to act as both a transcriptional inhibitor as well as an activator, (Moore et al. 1999). Interestingly, in a study of the TATA-less promoter of deoxynucleotidyl transferase, an insertion of a TATA-element removed the ability of TLP to activate transcription, (T. Ohbayashi et al. 2003). A dependence on TLP by some embryogenesis genes, complementary to TBP, has been seen in TLP knockouts of *X. laevis*, (Veenstra et al.

2000). Similar results have been seen for embryogenesis in *C. elegans*, (Dantonel et al. 2000). Knockdown studies in mice have shown that TLP does not appear to have a role in embryogenesis, but is important for spermatogenesis (Martianov et al. 2001). In these knockout mice, females had no phenotype and were fertile, while males were unable to complete sperm production, due to inhibition of the elongation of the spermatids.

In the identification of GTFs in *P. falciparum*, a candidate gene named TFIID-like was found, (Bischoff & Vaquero 2010) (accession number *PF*3D7_1428800.2 on PlasmoDB, www.plasmodb.org). In *Plasmodium* there appears to be only a single TBP-like protein (hereafter referred to as *Plasmodium falciparum* TLP or *Pf*TLP). This protein has not been characterised thus far, but may hold an intriguing avenue of exploration for the characterisation of transcription initiation. Notably, and similarly to other TLPs described thus far, proteomic and transcriptomic data suggests upregulation of the *Pf*TLP gene/protein during the sexual development stages, such as gametocytogenesis (Aurrecoechea et al. 2009). This is in contrast to *Pf*TBP expression which appears to be either downregulated or expressed at lower levels than *Pf*TLP during these stages. Should *Pf*TLP be involved in the transcription of specific genes, this may provide interesting and possibly useful insights into transcription initiation, and potentially future drug targets.

TFIIB

TFIIB is has been implicated in many aspects of transcription, both in the initiation as well as in the elongation of the RNA strand (Kostrewa et al. 2009). As TBP induces a dramatic bend in the DNA, TFIIB binds to TBP and strengthens the association of the complex to DNA through its interactions with the BREu and BREd regions up- and downstream of the TATA box. TFIIB also recruits RNAPII, and assists in the identification of the transcription start site (Imbalzano et al. 1994; Zhao & Herr 2002; Nikolov et al. 1995; Lagrange et al. 1998; Chen & Hahn 2004; Chen & Hahn 2003; Kostrewa et al. 2009).

Structurally, TFIIB is composed of an N-terminal region which binds and directs RNAPII to the TSS (Figure 4; Tubon et al. 2004; Kostrewa et al. 2009). The region in the N-terminus, which binds to RNAPII has been termed the B-ribbon, and consists of a zinc-finger domain which is conserved in eukaryotic and archaeal bacterial TFIIB proteins (Qureshi & Jackson 1998; Gietl et al. 2014). Adjacent to the B-ribbon is the B-reader which contains a helix-loop structure, thought to assist in the localisation of RNAPII to the TSS, through the identification of Inr sequence motifs.

The N-terminal region also regulates transcription through a highly conserved region consisting of multiple charged residues, called the charged cluster domain. TFIIB may exist in either a closed state, where intramolecular interactions between the N and C terminals appear to prevent transcription initiation (but not the formation of the PIC), or the open state where initiation is possible (Elsby & Roberts 2004; Glossop et al. 2004; Zhang et al. 2000).

The C-terminal or core domain of the TFIIB molecule is composed of two cyclin-like repeats, made up of 5 α-helices, and connected by a linker region which contains an

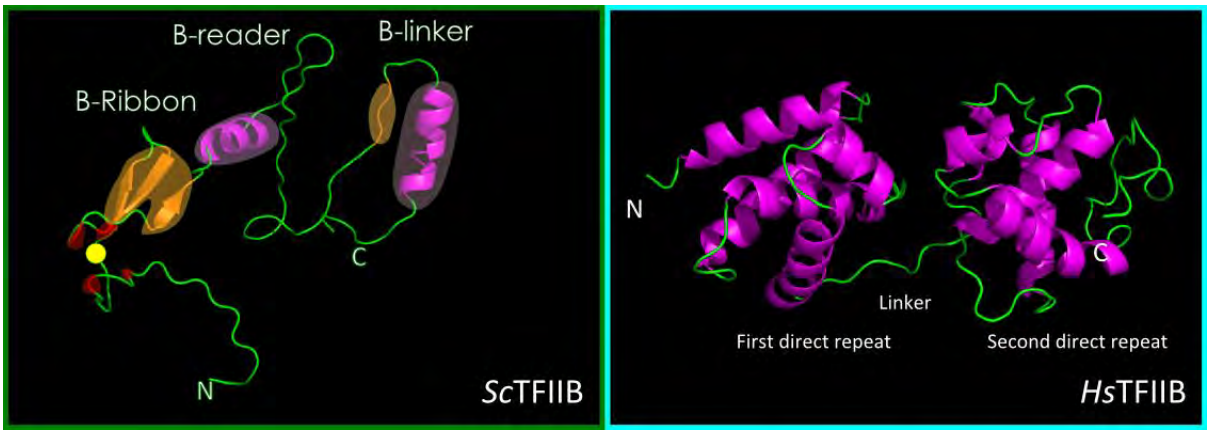



***Figure 4: Tertiary structure of TFIIB.***

The structure of the N-terminus of *Sc*TFIIB (left, (Kostrewa et al. 2009)) is composed of a zinc-finger domain, a B-ribbon made up of anti-parallel β-sheets (orange), the B-reader composed of an α-helix (purple) and a loop region, and the B-linker composed of a β-strand and α-helix. The C-terminal cyclin-like repeats of human TFIIB (right, (Nikolov et al. 1995)) are each composed of four amphipathic α-helices surrounding a core α-helix, separated by a linker region.

Structure models shown were prepared using PyMol based on PDB entries 3K1F.M (Kostreva et al. 2009) and 1VOL (Nicolov et al. 1995).

additional α-helix (Figure 4). This linker region has been shown to form a cleft, into which TBP binds to TFIIB, while the α-helix region interacts with and stimulates the activity of RNAPII (Xin Liu, David A. Bushnell, Dong Wang, Guillermo Calero 2010; Tsai & Sigler 2000; Nikolov et al. 1995).

TFIIB is theorised to be involved in the transition between initiation and elongation of transcription (Kostrewa et al. 2009). TFIIB assists TBP in the melting of the DNA, opening a bubble in the DNA and exposing this DNA to RNAPII. Whilst bound to RNAPII after the formation of the DNA bubble, the B-linker is then located in a position to stabilise the transcription bubble, while the B-reader is able to locate the Inr, and localise RNAPII to the TSS. The RNA-strand grows as a DNA-RNA hybrid, and after growth beyond seven nucleotides, TFIIB is released and the elongation complex is formed (promoter escape).

The identification of GTFs in *P. falciparum* led to identification of a TFIIB orthologue (Bischoff & Vaquero 2010). Given the importance of TFIIB in the formation of the PIC, through interactions with various GTFs and RNAPII, the characterisation of the molecule is predicted to be fundamentally important for the characterisation of transcriptional initiation in the parasite.

## 1.2 The aims of this study

Currently, a major focus of the work in the research group is to identify the DNA sequences which direct the formation of the pre-initiation complex, and so direct transcription in *Plasmodium falciparum.* Studies which have utilised bioinformatics approaches to identify TSSs to map and predict promoter regions and regulatory sequences have yielded largely inconclusive results due to the A+T-richness of the *P. falciparum* genome.

In the formation of the PIC, the *P. falciparum* GTFs must localise to the gene promoter regions. It may, therefore, be assumed that by characterising the binding specificity of GTFs

specifically that of *Pf*TBP/*Pf*TLP in conjunction with *Pf*TFIIB and *Pf*TFIIA, will provide an important starting point to identifying *P. falciparum* promoter regions, and subsequently the mechanisms behind *P. falciparum* transcriptional regulation.

Currently, the expression and characterisation of *Pf*TBP and *Pf*TFIIA are being investigated by another student in the research group (Robert Milton). TFIIB is known to be important for promoter recognition, discussed above, through the identification of the BREs. The possibility that *Pf*TFIIB performs a similar role in *P. falciparum* promoters must be investigated. Additionally, the potential of *Pf*TLP to regulate gene expression as an alternative to *Pf*TBP at certain genes, and hence, recognise alternative promoter sequences to *Pf*TBP is a possibility that must be explored. Canonically, TBP is known to bind to the TATA-box with high affinity. It is the focus of this research group to answer the question of how the *P. falciparum* TBP orthologue is able to recognise a TATA-box, in a veritable sea of A+T.

The specific aims of this research project were to investigate whether epitope-tagged *Pf*TFIIB and *Pf*TLP can be expressed in a soluble form in *E. coli* BL21-CodonPlus RIL cells and, if so, optimise expression conditions for maximal yield. Establish the affinity-purification of *Pf*TFIIB and *Pf*TLP. Finally to establish an investigation into the potential DNA-binding properties of *Pf*TFIIB and *Pf*TLP.

The aims of this research project form part of current research activities in the lab, that aim to make use of recombinant *P. falciparum* GTFs to determine the sites of PIC assembly, both in previously characterised putative *P. falciparum* promoters, and through a systematic evolution of ligands by exponential enrichment (SELEX) strategy (Ogawa & Biggin 2012).

# Chapter 2

# Materials and Methods

## 2.1 Bioinformatics analysis of protein structure and function

### 2.1.1 Multiple sequence alignments

Gene sequences used for multiple sequence alignments:

The TBP/TRF and TFIIB protein sequences used in the multiple sequence alignments are listed in Table 6 in the Appendix 5.1, and were retrieved from the National Centre for Biotechnology Information (NCBI) online database (www.ncbi.nlm.nih.gov/). *Plasmodium* protein sequences were retrieved from the PlasmoDB database (Aurrecoechea et al. 2009) with the exception of *P. reichenowi*, retrieved from NCBI.

Multiple sequence alignments (MSA) were performed using Clustal Omega, hosted on the Analysis Tool Web Services from the EMBL-EBI (Larkin et al. 2007; McWilliam et al. 2013; Sievers et al. 2011), on the default settings. As well as the Constraint-based Multiple Protein Alignment Tool (COBALT, www.ncbi.nlm.nih.gov/blast; Papadopoulos & Agarwala 2007), on the default settings.

Phylogenetic analysis of proteins performed in MEGA6 program (Tamura et al. 2013), using the Minimum Evolution method on the default settings: Interior-branch test (bootstrap) with 500 iterations; Substitution model: Poisson model; Uniform rate among sites. Tree Inference Options: Close-Neighbour-interchange; initial tree by Neighbour-joining.

### 2.1.2 Domain identification in protein sequences

BLASTp as well as Domain Enhanced Lookup Time Accelerated (DELTA)-BLASTp (Altschul et al. 1997; States & Gish 1994) were performed on the full-length *Pf*TLP and *Pf*TFIIB translated protein sequences (Appendix 5.1) with the default parameters. Conserved

domain information was determined from the results of the Conserved Domain Database search (Marchler-Bauer et al. 2009; Marchler-Bauer & Bryant 2004; Marchler-Bauer et al. 2011).

### 2.1.3 Secondary and tertiary structural prediction of proteins

Phyre (version 2.0; Phyre2) was used to predict the secondary and tertiary structures in the *Pf*TLP and *Pf*TFIIB, using the Intensive modelling mode (Kelley & Sternberg 2009).

Visualisation of tertiary protein structures

Visualisation and analysis of protein structures were performed in PyMol v1.6.00 (Schrödinger, LLC 2010) and Chimera v1.8.1 (Pettersen et al. 2004) programs. Solved protein structures used for comparison purposes include: *Hs*TFIIB C-terminus: PDB entry 1VOL (Nikolov et al. 1995); *Sc*TFIIB N-terminus: PDB entry 3K1F.M, (Kostrewa et al. 2009); *Hs*TBP core: PDB entry 1CDW (Nikolov et al. 1996).

## 2.2 Vectors and primers used for gene cloning and sequencing

### 2.2.1 Vectors used for gene cloning

The plasmid vectors for the cloning and expression of the *Plasmodium* proteins *Pf*TLP and *Pf*TFIIB were derived from the commercially available pET11d vectors (*Novagen*). The pET11d expression vectors contain the β-lactamase gene which confers ampicillin resistance. The vectors used in this study are summarised in Table 9 in the Appendix 5.4, as are the vector diagrams. The *Nde*I and *BamH*I restriction enzyme sites of the multiple cloning sites in the vectors were used for cloning in this study. The pET11d-6His-*Hs*TBP vector was obtained from Dr T. Oelgeschläger, and this served as the vector frame for the cloning of pET11d-6His-*Pf*TLP. The pET11d-GST-6His vector was constructed by Chenjerai Muchapirei, and served for the cloning of *PfTLPco.*

### 2.2.2 Primers used for gene cloning and sequencing

The primers used for cloning and sequencing of the *PfTLP* open reading frame (ORF) are listed in Table 1. Primers for cloning were designed in order to overlap the N-terminus and C-terminus of the target gene. Restriction enzyme (endonuclease) sites were incorporated in the primers for insertion of the PCR product into digested vector frame. Primers for cloning were designed by hand in SerialCloner (Version 2.6.1, *SerialBasics*), melting temperatures and self-complementarity of the primers was checked using OligoCalc (www.basic.northwestern.edu/biotools/oligocalc.html).

***Table 1: List of primers for cloning and sequencing***

Endonuclease restriction sites are shown underlined, start codons are **bold**, and the stop codons are in *italics*.

| Number | Primer Name | 5' to 3' sequence | Restriction Site | Gene |
|---|---|---|---|---|
| 1 | *Pf*TLP-FWD | GGCAGCCAT**ATG**TATCCCCCTT GTAAAAAGAAAAAAC | *Nde*I | *PfTLP* |
| 2 | *Pf*TLP-REV | GCAGCCAGATCT*CTATTA*ATGTT GCGATTTACTTTTAATTAAATAT GG | *BglI*I | *PfTLP* |
| 3 | *Pf*TLP-IntFWD | GTCCCCGTCACTTTAAGTAC | N/A | *PfTLP* |
| 4 | *Pf*TLP-IntREV | ATGTCCACTTTGCTTTTATCCT | N/A | *PfTLP* |
| 5 | *Pf*TLPco-FWD | TTATAAGGCAT**ATG**TATCCGCC GTGTAAA | *Nde*I | *PfTLPco* |
| 6 | *Pf*TLPco-REV | TATGGATCC*TTA*TCAGTGTTGGC TTTTAC | *BamH*I | *PfTLPco* |
| 7 | *Pf*TLPco-IntFWD | GTGCCGGTTACCCTGAGTACG | N/A | *PfTLPco* |
| 8 | *Pf*TLPco-IntREV | CTTTGTTGTCGTTAGATTTGTTTT CATCGTTC | | *PfTLPco* |
| 9 | T7-FWD | TAATACGACTCACTATAGG | N/A | |
| 10 | pET11d-REV | GTCAGGCACCGTGTATGAAA | N/A | |
| 11 | GST-FWD | ACAAATTGATAAGTACTTGAAA TCCA | | |

## 2.3 Cloning and sequencing of *Plasmodium* TLP

### 2.3.1 Agarose gel electrophoresis

Electrophoresis of DNA was performed using 0.8-2% (w/v) agarose containing 0.02µg/ml ethidium bromide (EtBR). Gels were prepared using either 0.5×TBE (40mM Tris-Cl pH 8.3, 45mM boric acid, 1mM Ethylenediaminetetraacetic acid [EDTA]) or 1×TAE (40mM Tris-Cl pH 8.3, 0.11% (v/v) glacial acetic acid, 1mM EDTA) buffer. Electrophoresis was conducted at 80-100volts until the DNA was sufficiently separated. DNA bands were visualised by employing a long wavelength (365 nm) UV trans-illuminator.

### 2.3.2 Polymerase chain reactions (PCRs)

PCR reactions (Sambrook et al. 1989) were performed using KAPA Taq PCR Kits making use of Buffer A (*Kapa Biosystems*), following the manufacturer's instructions, in 50µl reaction volumes. dNTPs for the reactions were supplied by *Thermo Scientific.* Amplification of the *PfTLP* and *PfTLPco* genes, with their respective primers, made use of the same reaction cycling conditions (Table 9). The PCR reaction profile was adjusted to include a long and lower elongation temperature due to the A+T richness of the gene sequences (Su et al. 1996).

PCR amplification of the open reading frame for *Plasmodium TLP* was performed using 2µl of each of the cDNA libraries MRA-296 and 297 (MR4, ATCC® Manassas Virginia; contributed by D. Chakrabarti).

A synthesised *PfTLP* gene, optimised for expression in *E. coli* was ordered from *GenScript*, (www.genscript.com/) termed '*PfTLP* codon optimised' or *PfTLPco* gene, and was supplied in a pUC57 vector as a lyophilised powder. The vector was reconstituted in reverse osmosis purified water ($_{Ro}H_2O$; Milli-Q purified, *Millipore*) to a final concentration of 200ng/µl. 5.5fmol of the vector was used per PCR reaction.

All PCR products were analysed by agarose gel electrophoresis, and then purified using either the MinElute™ Reaction Cleanup Kit (*Qiagen*) or QIAquick PCR Purification Kit (*Qiagen*), following the manufacturer's instructions. Purified DNA concentrations were determined with a NanoDrop 2000 spectrophotometer ('Nanodrop', *Thermo Scientific*).

### 2.3.3 Restriction enzyme digestion and ligation of PCR amplified inserts

The original *PfTLP* ORF purified PCR product was simultaneously digested with *Nde*I and *Bgl*II restriction enzymes in 2×tango buffer (*Thermo Scientific*) to cleave the N and C-termini respectively, leaving 'sticky' ends compatible with the expression vector. The digested PCR product was then gel isolated after agarose gel electrophoresis, and purified with the QIAEX II Gel Extraction Kit (*Qiagen*) following the manufacturer's protocol, and the concentration determined by Nanodrop and confirmed by agarose gel electrophoresis with a calibrated molecular weight marker.

***Table 2: PCR cycling conditions for PfTLP gene amplification***

| Duration | Temperature (°C) | Number of Cycles |
|---|---|---|
| | | |
| 10 min | 95 | 1 |
| | | |
| 30 sec | 90 | 25 |
| 30 sec | 58 | |
| 90 sec | 60 | |
| | | |
| 10 min | 60 | 1 |
| 10 min | 72 | 1 |

The *PfTLPco* purified PCR product was digested with *Nde*I restriction enzyme in *Nde*I buffer (*New England Biolabs*) and purified with the MinElute Reaction Cleanup Kit (*Qiagen*). The purified DNA was then digested with *BamH*I restriction enzyme in *BamH*I buffer (*Thermo Scientific*), and purified with the MinElute Reaction Cleanup Kit. All procedures following the manufacturer's instructions. The concentration of the purified insert DNA was determined by Nanodrop and confirmed by agarose gel electrophoresis with a calibrated molecular weight marker.

### 2.3.4 Preparation of pET11d-vectors for cloning

Plasmid isolation

In order to generate sufficient vector frame for the cloning of the *PfTLP* and *PfTLPco* genes, either One Shot® Stbl3™ or TOP10 Chemically Competent *E. coli* cells (Stbl3, TOP10, *Invitrogen*) were transformed with the appropriate vector (pET11d-6His-*Hs*TBP or pET11d-GST-6His), according to the manufacturer's instructions. Transformants were spread-plated on lysogenic agar plates (LA; 1% w/v tryptone powder, 0.5% w/v yeast extract, 0.5% w/v NaCl, 1.5% w/v agar, pH adjusted to 7.5) containing 100µg/ml ampicillin, and grown overnight (8-12 hours) at 37°C. The transformed cells were then used to inoculate 10ml lysogenic broth (LB; 1% w/v tryptone powder, 0.5% w/v yeast extract, 0.5% w/v NaCl, pH adjusted to 7.5) cultures containing 100µg/ml ampicillin, and grown overnight at 37°C with shaking. Plasmid DNA was isolated from the cultures making use of the GeneJET Plasmid Miniprep Kit (*Thermo Scientific*), following the manufacturer's instructions. The concentration of purified plasmid DNA (pDNA) was determined by Nanodrop, and confirmed by linearizing the DNA by restriction enzyme digestion, and agarose gel electrophoresis with a calibrated molecular weight marker.

Vector restriction enzyme digestion

Purified plasmid DNA was double digested with *Bam*HI and *Nde*I in 2×Tango buffer (*Thermo Scientific*) and analysed by agarose gel electrophoresis. The cleaved vector frame was dephosphorylated with FastAP Thermosensitive Alkaline Phosphatase (FastAP, *Thermo Scientific*) and the sample heat inactivated at 75°C for 10 minutes. The cleaved vector frame was then isolated after agarose gel electrophoresis, and purified with the QIAEX II Gel Extraction Kit. The concentration of purified plasmid vector frame was determined by Nanodrop, and confirmed by agarose gel electrophoresis with a calibrated molecular weight marker. All procedures were performed following the manufacturer's instructions.

### 2.3.5 Isolation of *Pf*TLP expressing pET11d-vectors

Ligation, plasmid isolation identification of succefuly cloned vectors

Digested insert and vector frame were ligated by incubation with T4 DNA Ligase (*Thermo Scientific*), insert:vector ratios of 1:1 and 3:1 respectively, in 20µl volumes. The ligation mixes were incubated for 2 hours at room temperature. Of the ligation mixes, 1/10 volume of each ligation mix was then transformed into 25µl-100µl of either Stbl3 or TOP10 Chemically Competent *E. coli* cells, following the manufacturer's instructions. The transformed cells were then spread-plated on ampicillin containing lysogenic agar (LA) plates, and incubated overnight at 37°C.

Colonies successfully transformed with insert-containing vectors were identified by colony PCR. In this assay, a colony is lightly touched with a sterile pipette tip. The tip is dipped into 2µl of $_{Ro}H_2O$, and heated to 90°C for 5 minutes. This is then used in the PCR reaction (described above) with the appropriate forward and reverse primers, and the cycling conditions in Table 2. *Pf*TLP-FWD and *Pf*TLP-REV primers were used to identify transformants containing pET11d-6His-*Pf*TLP and pET-11d-GST-6His-*Pf*TLP vector, and *Pf*TLPco-FWD and *Pf*TLPco-REV primers for transformants containing the pET-11d-GST-

6His-*Pf*TLPco vector (Table 1). 1/5 volume of the PCR reactions were analysed by agarose gel electrophoresis, and vector DNA isolated from positively identified colonies, as described above. The sequences of the vectors were then confirmed by DNA sequence analysis (Stellenbosch Central Analytical Facility, Stellenbosch, South Africa) on an ABI3730xl DNA analyser (Applied Biosystems, Foster City, USA). The pET11d-6His-*Pf*TLP and pET11d-GST-6His-*Pf*TLP vectors were sequenced with the primers 3, 4, 9, and 10; pET11d-GST-6His-*Pf*TLPco vector was sequenced with the primers 7, 8, 10 and 11 (Table 1) to ensure full sequence coverage. Sequence data was analysed using Chromas software (Version 2.01, Technelysium Pty Ltd, Queensland, Australia), and SerialCloner.

### 2.3.6 Transformation of protein expression vectors into protein expressing cells

Note that the cloning of the pET11d-GST-6His-*Pf*TLPco was initially unsuccessful, and a vector was isolated with an additional two codons in between the 6His-tag and *Pf*TLPco ORF. This vector, termed pET11d-GST-6His-*Pf*TLPco(aa), was used in expression trials until the correct pET11d-GST-6His-*Pf*TLPco vector could be cloned.

The pET11d-6His-*Pf*TLP, pET11d-6His-*Pf*TFIIB (cloned in the lab, and provided by Dr Thomas Oelgeschläger), pET11d-GST-6His-*Pf*TLP, pET11d-GST-6His-*Pf*TLPco and pET11d-GST-6His-*Pf*TLPco(aa) vectors were transformed into BL21-CodonPlus® (DE3)-RIL competent cells (*Agilent Technologies,* Stratagene 2005) for expression, following the manufacturer's instructions. Transformants were spread plated onto either LA plates or Super Optimal Broth with Catabolite Repression (SOC; 2% w/v tryptone powder, 0.5% w/v yeast extract, 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 20mM glucose, 1.5% w/v agar) plates, containing 50µg/ml chloramphenicol and 100µg/ml ampicillin, and incubated overnight at 37°C.

## 2.4 Expression of recombinant *Plasmodium* proteins in *E. coli*

Note that in all of the following experiments, the LB media used contained both ampicillin and chloramphenicol antibiotics at 100µg/ml and 50µg/ml.

Starter culture procedure

Either 10 or 20ml of LB, with or without 1% (w/v) glucose, was inoculated from either a single colony from a freshly transformed plate, or from a scraping of the top of a frozen glycerol culture (maintained at -80°C). The inoculated culture was incubated with shaking at 37°C overnight (typically for 16 hours). The overnight culture was pelleted by centrifugation for 5 minutes at 2822 relative centrifugal force (RCF) in a Beckman Ultra Centrifuge for 10 minutes at 4°C. The cell pellet was resuspended in 10ml of fresh LB in order to remove secreted β-lactamases in the cell culture, and used to inoculate 200ml of LB. If the expression culture was 100ml of LB, 5ml of starter culture was used (1/20 vol.). The optical density of the cell culture was measured at wavelength of 600nm ($OD_{600nm}$) making use of a light spectrophotometer.

### 2.4.1 Expression of recombinant *Pf*TFIIB protein in *E. coli* BL21-CodonPlus® (DE3)-RIL cells

Original expression of *Pf*TFIIB protein protocol

Two 200ml flasks of cell culture containing *E. coli* BL21-CodonPlus® (DE3)-RIL cells transformed with pET11d-6His-*Pf*TFIIB were incubated at 37°C, and $OD_{600nm}$ of the cell cultures recorded. Protein expression of one culture was induced with 1mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) at cell densities which ranged from 0.4 - 0.8 $OD_{600nm}$ units. The culture densities of both the induced and non-induced were monitored, and approximately 2-2.5 hours post-induction the cell cultures were pelleted through centrifugation at 3468 RCF for 10 minutes. The cell pellets were frozen and stored at -20°C.

Expression of recombinant *Pf*TFIIB through auto-induction

Overnight cell cultures of *E. coli* BL21-CodonPlus® (DE3)-RIL cells transformed with pET11d-6His-*Pf*TFIIB were grown in LB containing 1% glucose, and pelleted by centrifugation as above. Pelleted overnight cultures were resuspended in fresh media and used to inoculate two 200ml flasks of LB (containing 1% glucose). The flasks were incubated at 37°C for several hours, and the densities of the cell cultures monitored. The cell cultures were pelleted through centrifugation at 3468 RCF for 10 minutes. The cell pellets were frozen and stored at -20°C.

### 2.4.1 Expression of recombinant *Pf*TLP protein in *E. coli* BL21-CodonPlus® (DE3)-RIL cells

Expression of recombinant *Pf*TLP protein

*E. coli* BL21-CodonPlus® (DE3)-RIL competent cells transformed with pET11d-6His-*Pf*TLP, pET11d-GST-6His-*Pf*TLP or pET11d-GST-6His-*Pf*TLPco/(aa) were used to inoculate 10-20ml cultures and incubated overnight (8-10 hours) at 37°C. Overnight cultures were used to inoculate expression cultures as described above. The expression culture was incubated at 37°C and the $OD_{600nm}$ of the culture monitored. Protein expression was induced with 0.1-1mM IPTG, final concentration. Cultures may or may not have been cold-shocked to 20-25°C in an ice-bath (indicated in the Results section), and the temperature monitored with a sterile thermometer. Cultures were then incubated for up to 20 hours between 20-25°C (as indicated). The $OD_{600nm}$ of the cultures were recorded, and then pelleted through centrifugation at 3468 RCF for 10 minutes. The cell pellets were frozen and stored at -20°C.

### 2.4.2 Preparation of *E. coli* glycerol stocks

Bacterial glycerol stocks were prepared by taking 0.75ml of mid-exponential phase (~0.4-0.6 $OD_{600nm}$) bacterial culture, and adding 0.25ml sterile 60% ice-cold glycerol in a

1.5ml microcentrifuge tube. The cell solution was vortexed to ensure even dispersal, and then snap-frozen in liquid nitrogen and stored at -80°C.

### 2.4.3 Determination of recombinant protein expression by pilot-scale affinity purification

Isolation of the total soluble *E. coli* cellular protein fraction

Frozen cell pellets were resuspended in sonication buffer, at 5ml/g of cell pellet. Sonication buffer is composed of 50mM Tris-base (pH 7.2 at 21°C), 300mM KCl, 20mM imidazole, 0.01% Triton™ X-100 (*Sigma-Aldrich*®), and 10mM β-mercaptoethanol. 50µl of Protease Inhibitor Cocktail for use in purification of Histidine-tagged proteins (protease inhibitor; *Sigma-Aldrich*®) and either 0.5 or 1 mg/ml of Lysozyme from chicken egg white (lysozyme; *Sigma-Aldrich*®) was added to the resuspended cell solution. The resuspended cell solution was then incubated on ice for an hour, and then the cells were sheared through physical disruption by sonication. Sonication was performed with an Ultrasonics Inc. Cell Disruptor Model W-225R for 6 rounds of 10 seconds, with 20 second refractory period, on ice at power level four. The cell lysate was then centrifuged at 15000 RCF for 30 minutes at 4°C. The cleared cell lysate was then siphoned from the cell debris pellet, and kept on ice until purification, or a stored at -20°C.

Isolation of insoluble *E. coli* cellular protein fraction

After the soluble protein was removed from the lysed cell pellet, the resultant pellet would be resuspended in sonication buffer containing 8M urea, through sonication. The insoluble protein would then be then centrifuged at 15000 RCF for 30 minutes at room temperature, and the supernatant collected and kept on ice until purification, or a stored at -20°C.

Pilot-scale glutathione affinity purification

Pilot-scale purification of GST-tagged proteins was performed making use of glutathione-affinity purification from the cleared cell lysate using glutathione-agarose (*Sigma-Aldrich*®). In brief, 50µl of 40% (v/v) glutathione-agarose resin slurry in 2M NaCl and 1mM sodium azide, was equilibrated in 1ml of sonication buffer containing 0.1% Nonidet™ P 40 Substitute (NP 40; *Sigma-Aldrich*®). The resin was settled by brief centrifugation at 350 RCF for 1 minute, and the sonication buffer siphoned from the settled resin. 1ml of cleared cell lysate with 0.1% NP 40 was added to the resin and then mixed for 1 hour at 4°C. The glutathione-agarose was then recovered by brief centrifugation, and the unbound protein fraction siphoned off and stored on ice. The resin was washed once in 1ml of sonication buffer, and moved to a sterile microcentrifuge tube. The resin would then be suspended in 2×SDS loading dye (typically 45µl in volume) and the proteins heat denatured at 95°C for 5 minutes. The resultant denatured protein would then be moved to a fresh tube for SDS-PAGE analysis, or a stored at -20°C.

1) Pilot-scale nickel affinity purification

Pilot-scale purification of the hexa-histidine (6His)-tagged proteins was performed making use of nickel-affinity purification from the cleared cell lysate using PureProteome™ Nickel Magnetic Beads (*Merck Millipore;* Ni-beads). In brief, 5 or 10µl of suspended beads would be pipetted into a sterile microcentrifuge tube. The beads were separated from the storage buffer using a magnetic separator stand. The beads were equilibrated in 1ml of sonication buffer (with or without 0.1% NP 40), and recovered using the magnetic separator. 1ml of cleared cell lysate was added to the beads, and mixed for 1 hour at 4°C. The nickel magnetic beads were recovered with the magnetic separator, and the unbound protein fraction removed and stored on ice. The beads were washed either 3 times in 50µl of sonication buffer, or once in 1ml of sonication buffer. The resuspended beads were moved to a sterile microcentrifuge

tube. The beads were suspended in 2×SDS loading dye (typically 30µl in volume) and heated at 95°C for 5 minutes to denature the proteins. The resultant denatured protein would then be moved to a fresh tube for SDS-PAGE analysis, or a stored at -20°C.

SDS-PAGE analysis

SDS-PAGE gels were either poured in custom SDS-PAGE casting plates, and electrophoresis performed with custom gel apparatus, or the *Bio-Rad* mini-gel system was used. The volumes of stacking and running gel were adjusted accordingly. Typically the custom gels required 5ml of the 5% stacking gel and 10ml of the 12 resolving gel. The *Bio-Rad* mini-gels required approximately 1.25ml 5% stacking gel and 4.5ml 12% resolving gel. Combs of 1mm width were used for all the gels.

***Table 3: SDS-Page solutions***

| **12% Resolving gel: Reagent** | **Volume** |
|---|---|
| Water | 4.2ml |
| 40% 37.5:1 bis-acrylamide/acrylamide mix (*Sigma-Aldrich*®) | 3ml |
| 1.5 M Tris-HCl pH 8.8 | 2.5ml |
| 10% SDS solution | 0.1 ml |
| Ammonium persulphate (100mg/ml) | 0.2ml |
| Tetramethylethylenediamine (TEMED) | 10µl |
| Total | 10ml |

| **5% Stacking gel: Reagent** | **Volume** |
|---|---|
| Water | 3 ml |
| 40% 37.5:1 bis-acrylamide/acrylamide mix (*Sigma-Aldrich*®) | 0.625 ml |
| 1.0 M Tris-HCl pH 6.8 | 1.25 ml |
| 10% SDS solution | 50µl |

| | |
|---|---|
| Ammonium persulphate (10% w/v in $H_2O$) | 0.1 ml |
| Tetramethylethylenediamine (TEMED) | 5µl |
| Total | 5ml |

Samples were mixed to be in either 1 or 2×SDS loading dye (100 mM Tris-Cl (pH 6.8), 4% (w/v) SDS, 0.2% (w/v) bromophenol blue, 20% (v/v) glycerol, 200mM dithiothreitol [DTT]) and heated to 90°C for 2 minutes. Samples were then briefly centrifuged to allow the condensate to mix into the sample. The samples were then loaded into the SDS-PAGE and electrophoresis begun, with the ampere kept constant at typically 35-45mA until completion. The gels were removed from the glass plates, and washed in deionised water 3 times for 10 minutes, and then stained with Bio-Safe™ G-250 Coomassie premixed stain (Bio-Rad) for 30 minutes to an hour. The gels were de-stained by washing repeatedly with water. Gels were visually inspected to determine the presence of protein, and photographed.

### 2.4.4 Identification of 6His-*Pf*TLP and 6His-*Pf*TFIIB proteins by mass spectrometry based analysis

SDS-PAGE gels loaded with the 2×SDS loading dye denatured proteins (6His-*Pf*TLP and 6His-*Pf*TFIIB loaded onto separate gels) was performed. The proteins on the SDS-PAGE were visualised with Coomassie staining, described above, and the protein bands to be investigated excised. An additional control band of visually blank gel was also excised. The gel slices were sent for peptide mass fingerprinting (PMF) analysis to the Centre for Proteomic and Genomic Research (Faculty of Health Science, University of Cape Town, South Africa). MALDI/MS analysis was performed using a 4800 MALDI ToF/ToF system (AB SCIEX) with instrument control through Series Explorer. A summary of the identified proteins, and PMF spectra was returned.

## 2.5 Accumulation and purification of recombinant *Plasmodium* TFIIB protein

### 2.5.1 Accumulation of recombinant 6His-*Pf*TFIIB expressed protein *E. coli* cell mass

Overnight cultures of 5×10ml, inoculated from a glycerol stock of *E. coli* BL21-CodonPlus® (DE3)-RIL cells transformed with pET11d-6His-*Pf*TFIIB (0.6 $OD_{600nm}$ at the point of making the glycerol stock), were allowed to grow for approximately 12 hours at 37°C. The overnight cultures were pelleted at 2000 RCF for 10 minutes at room temperature, and the supernatant removed. The cells were resuspended in media (containing 1% glucose) and used to inoculate 400ml cultures. The $OD_{600nm}$ was measured to be 0.03-0.05 $OD_{600nm}$ at this time. The cell culture was then grown to $OD_{600nm}$ of approximately 1.6 - 1.8, approximately 7 hours. The cell culture would then be pelleted by centrifugation at 2000 RCF and the cell pellet weighed and small scraping of cell pellet taken for protein expression analysis, as described above, prior to freezing the pellet in liquid nitrogen and storage at -80°C. This was repeated until sufficient cell mass was generated.

### 2.5.2 Bulk purification of 6His-*Pf*TFIIB protein by Nickel-affinity purification

Production of soluble protein in cleared cell lysate

The accumulated cell pellets were resuspended in 3-5ml of sonication buffer containing 5µM $ZnCl_2$, per gram of cell pellet, up to 20ml per cell pellet. The resuspended cell pellets were incubated with 1mg/ml final concentration of lysozyme for 30 minutes on ice. The cell solutions were sonicated until the cell pellets were suspended, 9 rounds of 10 second bursts with 20 second refractory periods. The cell lysate was centrifuged at 15300 relative centrifugal force for 40 minutes. The cleared cell lysate was then decanted into chilled and sterile plastic tubes and the remaining cell debris/insoluble cell fraction was frozen and stored at -80°C.

Purification of 6His-*Pf*TFIIB with Ni-NTA affinity purification resin

2.1ml packed resin volume of Sigma HIS-Select® HF-Nickel Affinity Gel agarose resin was equilibrated in 20× resin volume sonication buffer. The buffer was removed, and the resin incubated with half of the cleared cell lysate in sterile plastic 50ml tubes for 1 hour with slow mixing at 4°C. The tubes were centrifuged in a clinical centrifuge at approximately 300RCF, and the protein fraction not bound to the resin was removed. The resin was then incubated with the remainder of the cleared cell lysate as before. Finally, the samples were centrifuged, and the unbound sample removed. The resin was washed twice in 20×packed resin volume of wash buffer (20mM Tris-Cl pH 7.2, 0.2mM EDTA, 100mM KCl, 0.1% NP 40, 20mM imidazole, 5µM $ZnCl_2$) by mixing the resin in the buffer, slowly centrifuging the slurry, and removing the wash buffer. The resin was resuspended in 10×packed resin volume of wash buffer, and the resin loaded equally onto two Poly-Prep® Chromatography columns (*BioRad*). The wash buffer was allowed to flow through the columns by gravity-flow, and collected. Each column was then loaded with 5×packed resin volume of elution buffer (20mM Tris-Cl pH 7.2, 0.2mM EDTA, 100mM KCl, 0.1% NP 40, 250mM imidazole, 5µM $ZnCl_2$). Fractions of the elution of approximately 200µl were collected into sterile micro-centrifuge tubes.

Determination of protein containing elution fractions:

10µl of the BioRad Protein assay solution was incubated with 1µl of each fraction, to test for protein content. The highest protein containing fractions (PCF) were then pooled; as well as the first 4 fractions and first 5 fractions of columns 1 and 2 respectively. Fractions 9-11 of both columns were pooled together. Aliquots of the PCF were then made, and the crude protein preparation flash-frozen in liquid nitrogen and stored at -80°C. Finally, titrations of bovine serum albumin (BSA) protein (*Roche*), and the PCF were loaded onto an SDS-PAGE gel to estimate recombinant *Pf*TFIIB protein quantity and purity.

### 2.5.3 Further purification of 6His-*Pf*TFIIB by Sepharose® resins

6His-*Pf*TFIIB was further purified making use of the Sepharose® Fast-Flow resins Q, DEAE, CM, or SP as indicated. In each case 20µl of settled Sepharose® resin volume was equilibrated in BC-100 buffer (20 mM Tris, pH 7.2 at RT, 100mM KCl, 0.02mM EDTA, 10mM β-mercaptoethanol, 0.1% NP 40). Approximately 200ng of recombinant *Pf*TFIIB protein preparation was incubated with the resin on ice for up to an hour. The resin was centrifuged at 380RCF for 2 minutes, and the unbound fraction removed. The resin was washed in 100µl of BC-100 buffer, centrifuged again and the wash fraction removed. 30µl of BC-× buffer was added, where '×' is the concentration of KCl in the buffer required to give the desired final KCl concentration (taking void-volume of the resin into account) The resin was incubated for 2-5 minutes on ice, and the eluted protein removed by centrifugation and aspiration. The next buffer of increased KCl concentration was added. This was repeated in step-wise fashion. Finally, after aspiration of the final buffer, the resin was suspended in 2×SDS-loading buffer. The elution buffer from each step-elution was mixed with 2×SDS-loading buffer, and then the maximum amount of sample capable of being loaded was loaded (10µl of the buffer eluted, effectively 1/3$^{rd}$ of the eluted protein).

The experimentation was analysed by SDS-PAGE analysis.

Nuclease treatment of protein prior to use

From initial experiments, it was determined that nucleic acids may be contaminating the *Pf*TFIIB preparation. In order to investigate the effect of DNA on the binding of *Pf*TFIIB, Benzonase® endonuclease (nuclease; *Sigma-Aldrich*®) was used to treat the protein preparation. Approximately ~0.2µl of nuclease (~50 units of enzyme) and incubated at 37ºC for 15 minutes. The purification on SP-Sepharose® resins continued as above.

## 2.6 Accumulation and purification of recombinant *Plasmodium* TLP

### 2.6.1 Accumulation of recombinant GST-6His-*Pf*TLP expressed protein *E. coli* cell mass

Overnight cultures of 150ml LB containing 1% glucose were inoculated from a glycerol stock of *E. coli* BL21-CodonPlus® (DE3)-RIL cells transformed with pET11d-GST-6His-*Pf*TLPco (0.6 $OD_{600nm}$ at the point of making the glycerol stock), and allowed to grow for 12 hours. The overnight cultures were pelleted at 2000 RCF for 10 minutes at 4°C, and the supernatant removed. The cells were resuspended in LB and used to inoculate 500ml cultures. The $OD_{600nm}$ was measured to be typically 0.03-0.05 $OD_{600nm}$ at this time. The cell culture was then grown at 37°C to $OD_{600nm}$ of approximately 1.5 - 1.6, approximately 5 hours, and then cold-shocked to 21°C and maintained at 21°C with shaking for half an hour prior to induction with 0.5mM IPTG. The cells were incubated for approximately 23-25 hours at 21°C. The cell cultures were pelleted by centrifugation at 2000 RCF and the cell pellets weighed. A small scraping of cell pellet was taken for protein expression analysis by pilot-scale affinity purification, prior to freezing the pellet in liquid nitrogen and storage at -80°C. This was repeated until sufficient cell mass was generated.

### 2.6.2 Preparation of soluble protein

The accumulated cell pellets were resuspended in sonication buffer (50mM Tris-HCl pH7.2, 300mM KCl, 20mM imidazole, 0.01% triton-x100, 10mM β-mercaptoethanol) 3ml of sonication buffer per gram of cell pellet. The resuspended cell pellet solution contained 0.5mg/ml of lysozyme enzyme, 40µl/g cell pellet of protein protease inhibitor, and 1250 units of Bensonaze®. The resuspended cell solution was incubated on ice for one hour. The resuspended cell pellets were sonicated until the cell pellets were suspended, 5 bursts of 10 seconds each on power level 5 with 20 second refractory period. The resuspended cell pellets were then incubated on ice for 2 hours. The sample were then sonicated at power level 4 for

6 bursts of 10 seconds, with 20 seconds of refractory period after each burst. This occurred while chilled on ice. Additional sonication buffer was then added to the tubes for centrifugation purposes. The cell lysate was then centrifuged at 10,000 RCF for 40 minutes, and the cleared cell lysate then decanted into a sterile plastic tube. This cell lysate was then centrifuged once more for 20 minutes at 10,000 RCF, and the supernatant decanted into a sterile plastic tube.

### 2.6.3 Binding to Glutathione-Agarose (Sigma-Aldrich®):

500µl of packed resin volume glutathione-agarose was loaded onto a Poly-Prep® Chromatography columns. The resin was equilibrated with 20×resin volume of sonication buffer. The cleared cell lysate was loaded onto the column. The unbound flow-through was collected. The column was washed with 20×resin of wash buffer (20mM Tris-Cl pH 7.2, 0.2mM EDTA, 100mM KCl, 0.1% NP 40). The protein was eluted from the column with 4ml glutathione-elution buffer (20mM Tris-Cl pH 7.2, 0.2mM EDTA, 100mM KCl, 10mM β-mercaptoethanol, 0.1% NP 40, 20mM glutathione, solution adjusted to pH 8 with 5N NaOH), collecting fractions of approximately 250µl. 5µl of fractions 1, 3, 5, 7, 9, and 11 were analysed for protein content by SDS-PAGE analysis. Protein containing fractions were pooled, and titrated with BSA on an SDS-PAGE gel to estimate the quantity of protein obtained.

### 2.6.4 Thrombin cleavage of GST-6His-*Pf*TLP

2000 units of thrombin enzyme (Thrombin Factor IIa, *Sigma-Aldrich*®) were incubated with the pooled protein preparation, and incubated at 4°C for several hours. Thrombin cleavage of the GST-6His-*Pf*TLP protein preparation analysed making use of 200µl of the protein preparation and pilot-scale nickel affinity batch purification and SDS-PAGE analysis, as above.

### 2.6.5 Nickel affinity purification of cleaved 6His-*Pf*TLP protein

A settled resin volume of 0.5ml of HIS-Select® Nickel Affinity Gel was loaded onto a Poly-Prep® Chromatography Column and equilibrated with the glutathione-elution buffer used previously. The remaining protein preparation was loaded to the column (~3.7ml), and the unbound flow-through stored on ice. The column was washed with 20×resin volume wash buffer (20mM Tris-Cl pH 7.2, 0.2mM EDTA, 100mM KCl, 0.1% NP 40, 20mM imidazole). The protein was to be mixed with glycerol for freeze protection (40% glycerol solution) after elution. The elution buffer was, therefore, twice as concentrated (40mM Tris-Cl pH 7.2, 0.4mM EDTA, 200mM KCl, 20mM β-mercaptoethanol, 0.1% NP 40), so that it may be diluted down to 'BC-100' with 40% glycerol. The imidazole concentration was kept at 250mM for the elution. The protein was eluted in 250µl fractions, and the fractions analysed by SDS-PAGE for protein content. The protein containing fractions were then pooled and the 6His-*Pf*TLP protein quantified by SDS-PAGE with a BSA titration.

It was found that the ‘unbound’ flow-through still contained 6His-*Pf*TLP. This was most likely due to the presence of competing 6His-peptides cleaved from the G6Hp. The unbound fraction was then passed through a Ni-column a second time, and the protein eluted as above.

### 2.6.6 Investigation of SP-Sepharose® purification of 6His-*Pf*TLP

Made use of SP-Sepharose® Fast Flow resin (*Sigma-Aldrich*®). The protocol used was similar to that described in 2.5.3, albeit without making use of the nuclease enzymes in these assays.

### 2.6.7 Depletion of GST from *Pf*TLP protein preparation

Glutathione-agarose (0.5ml settled resin volume) was equilibrated in 20×resin volume of BC-100 buffer (20 mM Tris, pH 7.2 at RT, 100mM KCl, 0.02mM EDTA, 10mM β-mercaptoethanol, 0.1% NP 40). The eluted fractions from the nickel-affinity purification

were loaded onto the column. The protein was incubated on the column for 30 minutes at 4°C. The protein was then allowed to flow through and collected. A small amount of resin (~10µl) was taken for analysis. Analysis of the experiment was performed by SDS-PAGE.

## 2.7 Protocol for immunoblot analysis

Immunoblots were performed making use of the Mini Trans-Blot® Electrophoretic Transfer Cell (*Bio-Rad*), following the manufacturers suggestions. An overview of the methodology is used is below:

Transfer of protein to PVDF membrane

Millipore Immobilon-P transfer membrane (PVDF) was activated in 100% methanol for 15 seconds, then Membrane and filter pads were equilibrated in ice-cold transfer buffer. The gel for transfer was first washed briefly in deionized water then in ice-cold transfer buffer (25mM Tris, 192mM glycine). The 'sandwich' was then assembled. The gel was transferred at between 70-100volts for at least 1 hour (dependent on voltage achieved by power supply used). After transfer, the membrane was washed in 1×TBST buffer (137mM NaCl, 2.7mM KCl, 19mM Tris, 1% Tween-20 [*Sigma-Aldrich*®]) 3 times for 5 minutes. The excess TBST buffer was drained. The membrane was then blocked with blocking buffer (5%fat-free milk powder made up in 1×TBST buffer) either over night or for at least 1 hour at 4ºC with shaking. Excess blocking buffer was washed off with three washes in TBST.

Visualisation of immunoblots

The membrane was then incubated with the appropriate primary antibody for 1 hour at room temperature. The primary antibody was removed and the membrane washed 3× for at least 10 minutes with TBST. The secondary antibody was diluted to either 1:5000 or 1:3000 (10ml total) in 1×TBST buffer, and incubated with the membrane for 30 minutes to 1 hour at room

temperature. The secondary antibody was poured off and the membrane washed 3 x 10 min with TBST.

Chemiluminescent detection was performed with the ThermoScientific Pierce ECL Western Blotting Substrate kit, following the manufacturer's instructions, enough substrate was used to cover the membrane. Excess substrate was removed from the membrane, and the membrane wrapped in plastic film. X-ray films were then exposed to the membrane, treated with developer and fixer.

## 2.8 Characterisation of *Pf*TLP antibody

The production of polyclonal antibodies was performed by *BioGenes (BioGenes*, Berlin, Germany). Peptide-antibodies were raised in rabbits against the peptide sequence:

294-307: C-NDENKSNDNKEQND

Cleared cell lysate from cells in which expression of G6H-*Pf*TLP was induced was mixed with 30ng of purified 6His-*Pf*TLP:

| Name | Enriched | MW | Purified | Enriched |
|---|---|---|---|---|
| Lane Number | 1 | 2 | 3 | 4 |
| Protein Plus Kaleidoscope Mw Marker (*BioRad*) | | 4 | | |
| 6His*Pf*TLP (10ng/µl) | 3 (30ng) | | 3 | 3 |
| Cleared Lysate (1.8µg/µl) | 3 (5.4µg) | | | 3 |
| BC-100 | 4 | | 7 | 4 |
| Total (µl) | 10 | 4 | 10 | 10 |

The samples were mixed with equal volume 2×SDS-loading buffer, separated on a 5/12% SDS-PAGE gel, and transferred to a PVDF membrane and visualised as described above. The strips were probed with decreasing concentrations of anti-serum, shown is 1/1000 and

1/3000 dilutions of the anti-serum in TBST buffer. The secondary antibody used was horseradish peroxidase-conjugated anti-rabbit IgG; produced in goat (*Sigma-Aldrich*®).

## 2.9 DNA binding assays.

### 2.9.1 Preparation of DNA probes

Promoter sequence overview

The vectors containing the *Plasmodium* probes were produced by Tomas Hessler (unpublished results, 2014). In brief, the region from -179 to +57 of the *gbp-130* transcription start site were amplified (region includes positions 654647 – 654882 on chromosome 10; TSS at 654703), from a gDNA library of *Plasmodium* 3D7, and cloned into a pGEM-72f (*Promega*) derived vector. Similarly, the region from -171 to +65 of the *kahrp* TSS were also cloned (106605 to 106840 on chromosome 2; TSS at 106669). See amplified probe below.

EMSAs

The EMSA probes were produced by PCR reactions making use of both 5' and 3' biotinylated primers. These probes were amplified, gel extracted and purified prior to use.

Grey regions are primer amplification sites, the underlined and bold nucleotide indicate the TSS. Note that the probes have been constructed with the upstream region of the TSS on the 3' end, and the downstream region on the 5' end.

*gbp-130* – 282bp

5′CGAGCGGAGACTCTAGAATTCtttgaagtacactcaaaataagttatataccatatgtt
tttaacatatattatatatatatat**a**tatatataatataatataatattataattatttttt
taatattattaaattgaacataattattttatatcttactattatttttagaaaatttatt
atatatacatgcaatcataaataatgttttccctgaacctttttttcaatgaaataagttaa
cacaccattccttttGGATCCGGAGAGCTCCCAACGCGTT-3′

*kahrp* – 282bp

5′CGAGCGGAGACTCTAGAATTCtgatgatatgttttatatttcttgttttaaattattag
aataaaaagaaataattattatttc**a**tggaaataatataatattattatttttttttttaa
ttatttagtagttatgttttgtcgtttttctcatttatttattataatttacctatagta
tatactatgatgtatatttactctagtatgaagaataaagttaatgtaaaatattactaca
ctacatgcagtttaGGATCCGGAGAGCTCCCAACGCGTT-3′

The Ad2ML EMSA probe was amplified from the pTOG5TdT(-41TATAto+33) vector, provided by Dr T. Oelgeschläger. The 'TATAWAAR' consensus sequence of the canonical TATA-box is underlined in bold. The probe was produced by PCR reactions making use of both 5' and 3' biotinylated primers.

Ad2ML – 120bp

5′CGAGCGGAGACTCTAGAATTCggtacctatg**tataaaat**ggtgagaggacatcagagcc
ctcattctggagacaccacctgatggcacagacagAGGATCCGGAGAGCTCCCAACGCGTT
-3′

Immobilised template assays

The immobilised template assay probes were longer than the EMSA probes, as they were generated by a different primer set. Additionally, to ensure that the probes bound only on one end to the streptavidin beads, only one end (5') was biotinylated. Additionally, the immobilised template assay probes contained 5 GAL4 sites, although these were not expected to effect the binding of transcription factors to the probes. The additional length of the probes was used to ensure that the TSS of the probes was of sufficient distance away from the beads, and would allow the proteins to bind unhindered. Lower case regions indicate the sequence of the gene of interest. Note that the probes have been constructed with the upstream region of the TSS on the 3' end, and the downstream region on the 5' end.

The template DNA for PCRs were the *Plasmodium* promoter containing vectors and the pTOG5TdT(-51TATAto+62) vector for Ad2ML.

The forward primer was biotinylated, resulting in 5'-end biotin labelled probes.

Biotin-pTOPCRUS-Fwd 5'- AAT TGG GCC CGA CGT CGC A -3'
pTO(G5)TdTPCR-Rev 5'- AAC GCG TTG GGA GCT CTC C -3'

*gbp-130* – 422bp

5′AATTGGGCCCGACGTCGCATGCTCCTCTAGAGCTTGCATGCCTGCAGGTCGGAGTACTG
TCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCT
CCGAGCGGAGTACTGTCCTCCGAGCGGAGACTCTAGAATTCtttgaagtacactcaaaata
agttatataccatatgtttttaacatatattatatatatatat**a**tatatataatataatat
aatattataattatttttaatattattaaattgaacataattattttatatcttactatt
atttttagaaaatttattatatatacatgcaatcataaataatgttttccctgaacctttt
ttcaatgaaataagttaacacaccattccttttGGATCCGGAGAGCTCCCAACGCGTT-3′

*kahrp* – 422bp

5′AATTGGGCCCGACGTCGCATGCTCCTCTAGAGCTTGCATGCCTGCAGGTCGGAGTACTG
TCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCT
CCGAGCGGAGTACTGTCCTCCGAGCGGAGACTCTAGAATTCtgatgatatgttttatattt
cttgttttaaattattagaataaaaagaaataattattatttc**a**tggaaataatataatat
tattatttttttttttaattatttagtagttatgttttgtcgtttttttctcatttatttat
tataatttacctatagtatatactatgatgtatatttactctagtatgaagaataaagtta
atgtaaaatattactacactacatgcagtttaGGATCCGGAGAGCTCCCAACGCGTT-3′

The Ad2ML probe – 260bp

5′AATTGGGCCCGACGTCGCAtgctcctctagagcttgcatgcctgcaggtcggagtactg
tcctccgagcggagtactgtcctccgagcggagtactgtcctccgagcggagtactgtcct
ccgagcggagtactgtcctccgagcggagactctagaattcggtacctatgtataaaatgg
tgagaggacatcagagccctcattctggagacaccacctgatggcacagacagaGGATCCG
GAGAGCTCCCAACGCGTT 3′

Generating immobilised template assay probes by PCR

Immobilised template assay probes were generated by PCR (as above) with the following cycling conditions:

| Duration | Temperature (°C) | Number of Cycles |
|---|---|---|
| 5 min | 95 | 1 |
| 30 sec | 95 | 30 |
| 30 sec | 56 | |
| 90 sec | 72 | |
| 10 min | 60 | 1 |
| 10 min | 72 | 1 |

The probes were analysed on 0.5×TBE/1.5% agarose/EtBr gel. After checking the correct sizes, the remainder of the probes were separated on an agarose gel, the bands excised, and purified using the Thermo Scientific GeneJET Gel Extraction Kit to purify the DNA from the gel, following the manufacturer's instructions. The purified probe concentrations were then determined by Nanodrop and confirmed by an agarose/EtBr gel.

Binding of probes to Streptavidin magnetic beads

Dynabeads® M-280 Streptavidin (*Invitrogen*™) were used. Biotinylated probes were bound to the beads following the manufacturer's instructions, albeit making use of twice the quantity of beads recommended. Probe-bound beads were resuspended in the 1×wash buffer (10mM Tris-Cl pH 7.5 at RT; 0.5mM EDTA; 1M NaCl) containing 100ng/µl BSA protein (in BC-100 buffer). BSA was included to pre-block the beads with protein prior to immobilised template assays, to reduce non-specific protein-bead associations.

The binding efficiency of the beads was tested by washing an aliquot of the beads in 1×wash buffer without NaCl (10mM Tris-Cl pH 7.5 at RT; 0.5mM EDTA) or BSA. The beads were then boiled in this buffer with 0.2% SDS for 10 minutes at 100°C, to disrupt the biotin-

streptavidin binding. The beads were then separated from the supernatant in a magnetic separator, and loaded on a 1% agarose gel, and quantitated from a molecular weight marker with known DNA concentrations.

### 2.9.2 Polyacrylamide gel electrophoresis mobility shift assay

Native polyacrylamide gels were poured (8cm, made use of the *Bio-Rad* 4-gel Mini-PROTEAN Tetra cell system, with 10-well 1mm casting combs) of either 5 or 3.5% 49:1 *bis*-acrylamide/acrylamide solutions (*Sigma-Aldrich*®), containing 0.5×TBE, 0.5mM dithiothreitol (DTT), 2mM $MgCl_2$. Gels were then allowed to set overnight, before pre-running for one hour at 4°C with 0.5×TBE containing 2mM $MgCl_2$.

For each reaction mix, two solutions, a DNA containing buffer (8µl; 33mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid pH 8.4 [HEPES], 8.3mM $MgCl_2$, 167ng/µl BSA, 0.17% NP 40) and the tested protein(s) in BC-100 buffer (12µl); 20mM Tris-Cl pH 7.2, 0.2mM EDTA, 100mM KCl, 10mM β-mercaptoethanol, 0.1% NP 40) were combined to form a DNA-protein reaction mix buffer. The reaction mix was then incubated at 30°C for 45 minutes. For loading purposes, 15% Ficoll® PM 400 (*Sigma-Aldrich*®, made up in 1M Tris-Cl pH 7.2) was then added to the reaction mixes and mixed.

The reaction mixes were then loaded onto the gel, along with a lane containing only free-probe and dye (0.2% Bromophenol blue dye. 0.2%, Xylene cyanol dye, 0.2% Orange G) and run at 400volts for 5 minutes, and then 100 volts until the dye had migrated sufficiently.

The gel was then removed from the glass plates, and transferred to a (pre-equilibrated) positively charged Immobilon-Ny+ nylon membrane (nylon membrane; *Merck Millipore*), making use of the *Bio-Rad* Mini Trans-Blot® Electrophoretic Transfer Cell and ice-cold 0.5×TBE (no $MgCl_2$). Transfer was conducted at 100volts for 2 hours at 4°C. After transfer, the DNA was crosslinked to the membrane at 120mJ/cm2 using a commercial UV-light cross

linker. The DNA probe was visualised on X-ray film using the Chemiluminescent Nucleic Acid Detection Module (*Thermo Scientific*).

Human transcription factors EMSAs

The EMSA with human transcription factors made use of 6His-*Hs*TBP, 6His-*Hs*TFIIB in BC-100 and recombinant human TFIIA (p55/p12), obtained from Dr T. Oelgeschläger.

Probe concentrations

It was found that as little as 2fmol of double-biotin tagged probe was able to be visualised in this method, 5fmol was more commonly used.

### 2.9.3 Agarose gel electrophoresis mobility shift assay

For the agarose EMSAs, a 1.4% agarose/0.5×TBE gel was used, containing 5mM 5mM $MgCl_2$. The gel was pre-run for 1.5 hours at 100volts, in 0.5×TBE (5mM $MgCl_2$) at 4°C.

Protein-DNA reactions were performed as above, making use of half the total volume (10µl). For these assays, 10fmols of the probes were used. The gels were loaded as above, and electrophoresis occurred for approximately 2 hours at 100 volts at 4°C, 0.5×TBE used as running buffer.

Probe transfer

A 240mbar diaphragm pump (Charles Austen DA7C) was utilised in conjunction with Southern blotting apparatus (ATTO, USA) to construct horizontal vacuum-assisted transfer apparatus for agarose gels. The wells of the gel were filled with molten but cooled agarose, and the gel placed on top of the nylon membrane. A vacuum-proof mask covering the gel and nylon membrane sandwich, and subjected to vacuum. Vacuum was adjusted manually through the usage of a screw-valve. The vacuum was set at a negative pressure from which from visual observation <50% of the gel was contracted. Transfer was allowed to occur for two hours room temperature, using ice-cold 0.5×TBE. After transfer, the DNA was

crosslinked to the membrane at 120mJ/cm2 using a commercial UV-light cross linker. The DNA probe was visualised on X-ray film using the Chemiluminescent Nucleic Acid Detection Module (*Thermo Scientific*).

### 2.9.1 Immobilised template assay

Reaction conditions

The reaction conditions for the immobilised template assays were performed in 40µl reaction volumes. The reaction mix was the same as that described for the EMSA assay above, with the exception that magnetic beads bound to probe were used in place of the free-probe. In these assays, the equivalent of 1000fmol of DNA-probe were used in the assays. For the negative, 'No DNA' control, the equivalent amount of streptavidin magnetic beads were used. These beads were treated and prepared in the same manner as the DNA-bound beads, albeit without DNA probe bound.

In this assay, the same reaction was repeated with beads bound to probe, and beads without probe. The beads were equilibrated in the DNA-protein reaction mix without protein prior to use. After removal of this equilibration buffer, the protein-reaction mix with the appropriate protein was added. The reactions were then incubated at 30°C for 45 minutes. Afterwards, the beads were separated from the reaction mix (containing the unbound protein fraction), making use of a magnetic stand. The beads were then very briefly washed in 100µl of the equilibration buffer, and this wash buffer removed making use of the magnetic stand. The beads were then resuspended in 20µl of 2×SDS-loading buffer. Of the unbound fraction, 10µl was taken and 10µl of 2×SDS-loading buffer added (this results in all of the unbound fractions visualised on the immunoblot as being ¼ of the input protein). The samples were then heated at 95°C for 5 minutes.

An SDS-PAGE gel was then performed. A titration curve of the input protein was included, of quantity equal to the input protein, half the input, a quarter of the input and an eighth of the input. The unbound fraction (1/4 of total) was loaded, was the supernatant of the SDS-denatured beads. Molecular weight markers were included (4µl of Precision Plus Kaleidoscope, *Bio-Rad*).

The SDS-PAGE was then transferred to a PVDF membrane and visualised as described above.

Proteins used for the immobilised template assays

For each reaction, 1000fmol of each protein was used. In the results Chapter I report on immobilised template assays performed with 6His-*Pf*TLP; 6His-*Hs*TBP and recombinant 6His-*Hs*TFIIB (both provided by Dr T. Oelgeschläger).

Antibodies used for immunoblot visualisation of the immobilised template assays

A commercially available antibody was used for the detection of 6His-*Hs*TFIIB (1/2000 dilution of anti- *Hs*TFIIB rabbit antibody (SC-225; *Santa Cruz biotechnology*). The anti-*Hs*TBP and anti-*Hs*TFIIA antibodies were provided by Dr T. Oelgeschläger, and are not commercially available, they were used at dilutions of 1/2000 and 1/3000 respectively.

The secondary antibody used was the Anti-Rabbit IgG peroxidase; produced in goat *(Sigma-Aldrich*®) at 1/3000 dilution.

# Chapter 3

# Results and Discussion

## 3.1 Bioinformatics analysis of putative *Pf*TFIIB – structure and function

### 3.1.1 Amino acid sequence alignment analysis

In order to assess the evolutionary conservation of *Pf*TFIIB with well-studied TFIIB proteins, amino acid sequence analysis was performed. Sequence alignments provide information about the evolutionary relationship between proteins, and this information may be useful to infer structural and functional information.

Pairwise alignments with human (*Hs*TFIIB) and *S. cerevisiae* TFIIB (*Sc*TFIIB; SAU7) was performed with BLASTp on the default settings.

The amino acid sequence of *Pf*TFIIB shares 28% sequence identity and 45% similar residues (amino acid substitutions where the amino acids share similar physiochemical properties) with *Sc*TFIIB and 24% sequence identity and 45% similarity with *H. sapiens* TFIIB. Conservation of these proteins to *Pf*TFIIB was especially low over the second cyclin repeat, with the presence of several 'gaps' in the alignment (not shown).

The low sequence homology of *Pf*TFIIB to *Hs*TFIIB and *Sc*TFIIB prompted the use of a multiple sequence alignment. TFIIB molecules from evolutionarily separate eukaryotes were used for the alignment in order to gain insight into the sequence diversity of TFIIB, as well as to highlight regions of greater conservation, often an indication of residues which are functionally important. Shared conservation of these regions in *Pf*TFIIB may indicate conserved functions of the molecule.

The multiple sequence alignment of *Pf*TFIIB and TFIIB molecules from eukaryotes including fission (S*chizosaccharomyces pombe*) and budding yeast (*Saccharomyces*

*cerevisiae), Homo sapiens*, *Drosophila melanogaster,* the related protist *Toxoplasma gondii*, and the archaea bacterium *P. furiosus* is presented in Figure 5. Regions and residues which are known to be functionally important for *Hs*TFIIB or *Sc*TFIIB function are indicated (by colour) in the alignment (Tsai & Sigler 2000; Kostrewa et al. 2009; Lagrange et al. 1998; Deng & Roberts 2005).

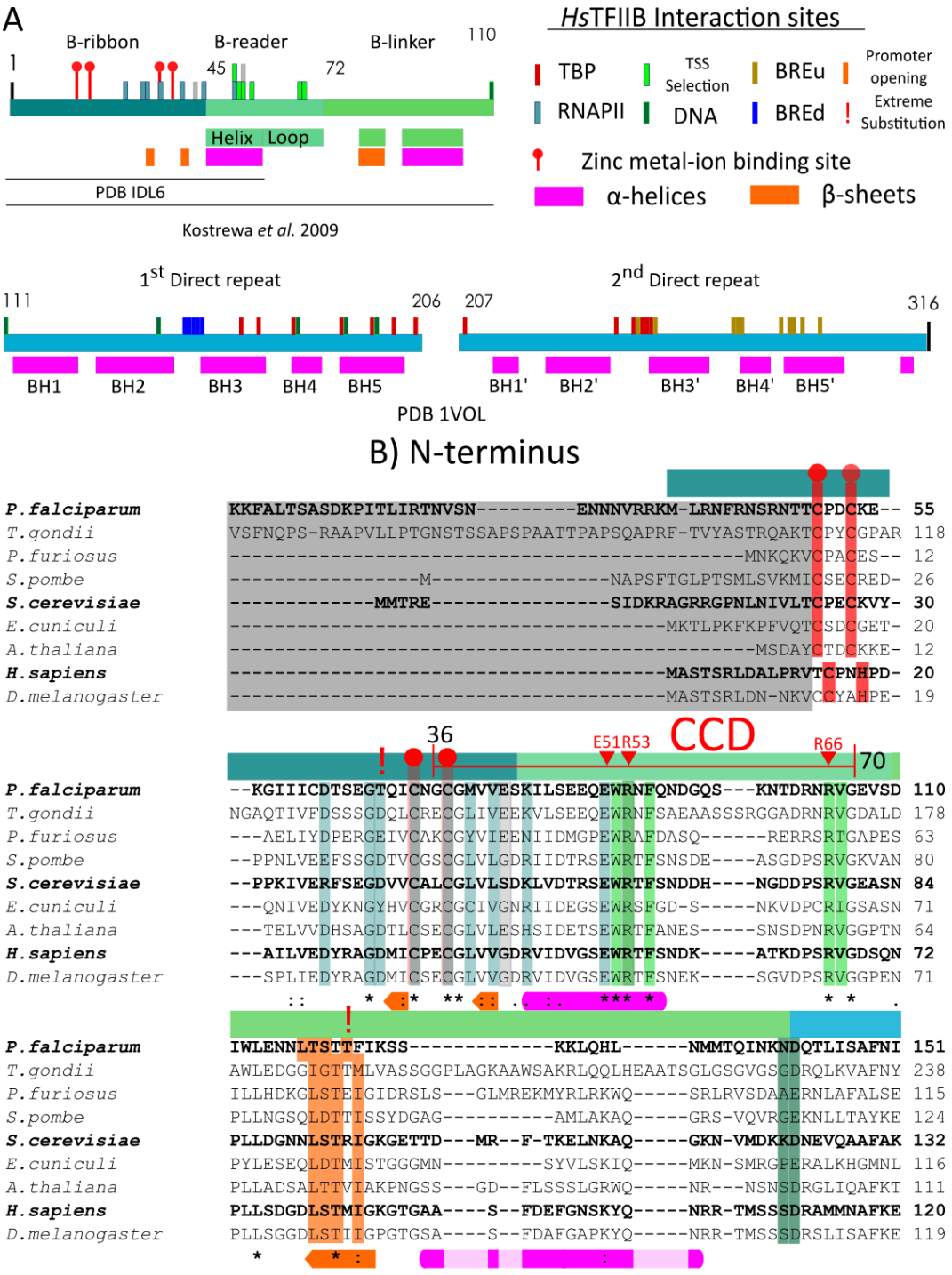



***Figure 5: Multiple sequence alignment of TFIIB proteins.***

**A** Schematic of HsTFIIB, with functional regions identified in colours. Secondary structures are indicated by purple and orange bars (α-helices and β-sheets respectively).

**B** Multiple sequence alignment performed with Clustal Omega of diversely related eukaryote TFIIB sequences. Structural regions are indicated by colour bars above sequences: B-ribbon, B-reader and B-linker regions are shown in shades of green. Cyclin repeats are shown in blue. Residues with known functional interactions are highlighted with coloured bars according to colour code in A. Interaction partners are referred to. Extreme amino acid substitutions in *Pf*TFIIB as compared to *Hs*TFIIB/*Sc*TFIIB indicated with red exclamation marks.

Abbreviations used: BH: β-helices; TSS: Transcription start site; BREu/d:up-/downstream B-recognition element .

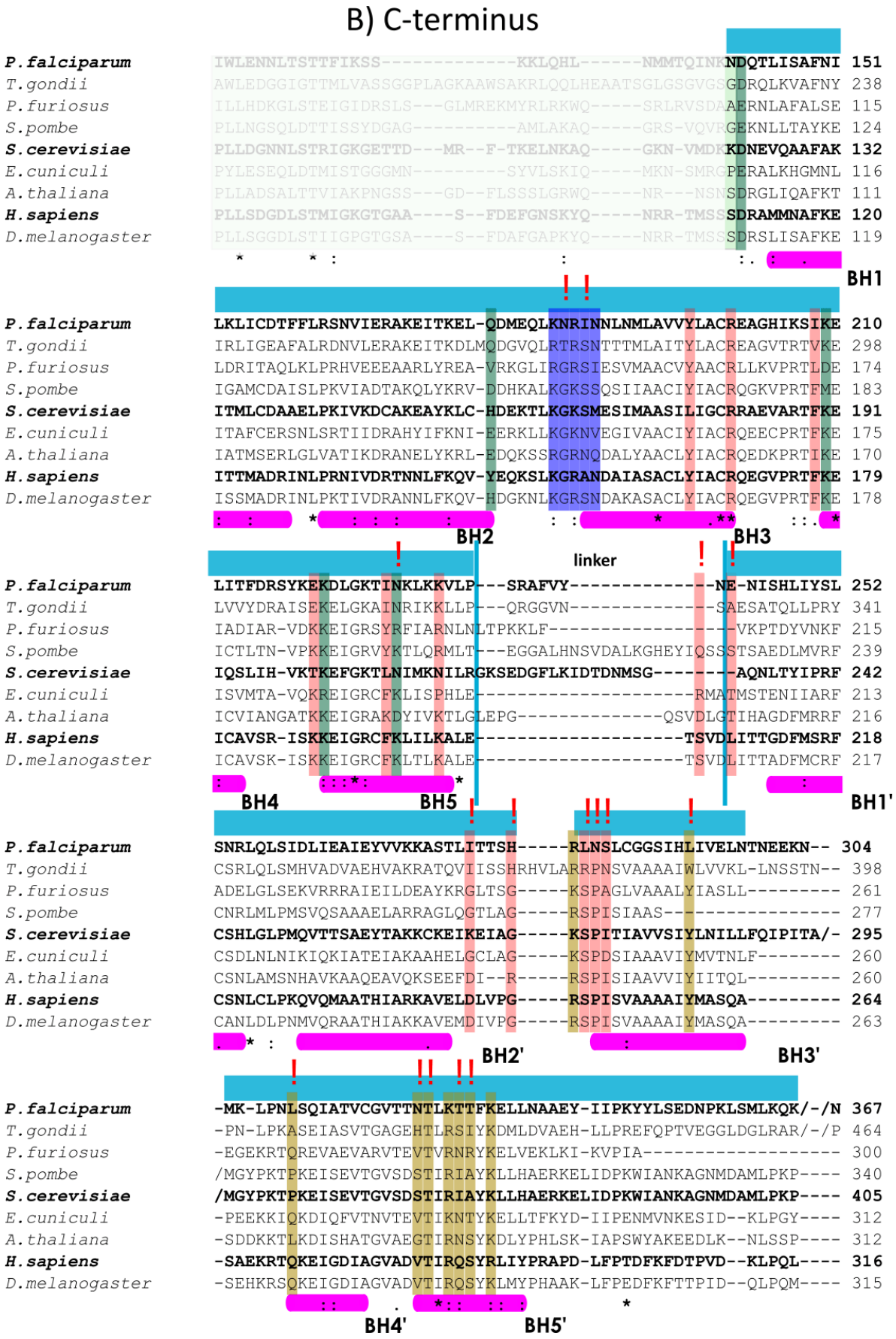



***Figure 5 continued – Comparison of the C-terminus of eukaryotic TFIIB proteins.***

*Pf*TFIIB shares conservation of many functionally important residues

Eukaryotic TFIIB (including *Pf*TFIIB) molecules do not share a great deal of conservation across the N-terminal region. This is especially true for the N-most terminal region (blocked in grey, up to residue 49 in *Pf*TFIIB) which is variable in both length and composition. This region is also the least conserved between putative *Plasmodium* species TFIIB proteins with less than 50% similarity between *P. falciparum* and other *Plasmodium* TFIIB proteins. Beyond the 49th amino acid residue (relative to *P. falciparum*) there is close to 90% sequence identity across *P. vivax*, *P. cynomolgi, P. knowlesi, P. chabaudi (chabaudi), P. berghei*, and *P. yoelli (yoelli),* as compared to *P. falciparum* (data not shown).

Beyond the 49$^{th}$ residue (relative to *Pf*TFIIB) there is increased conservation of residues with known function, both in position as well as the identity of the residues. These residues are highlighted in colours associated to their function in Figure 5.

The N-terminus of *Pf*TFIIB maintains conserved functionally important residues

The function of the N-terminus in eukaryotic TFIIB includes the formation of the zinc-finger containing B-ribbon domain which interacts with the "dock" domain of RNAPII in the formation of the PIC (*Sc*TFIIB - residues C45, C48, L50, L52 and E62 (Kostrewa et al. 2009)). Adjacent to the zinc-ribbon motif is the B-reader region. The B-reader region (when associated to RNAPII) follows the RNA exit channel from which synthesised RNA extends. The B-reader contains helix and loop structures which are thought to locate RNAPII to the +1 position (TSS) of the template strand. The mutation of these residues leads to an alteration of the TSS (Deng & Roberts 2007). TFIIB localises RNAPII to the TSS through recognition of Inr sequence motifs (Fairley et al. 2002; Butler & Kadonaga 2002) (YYANWYY in mammals, TCAKTY in *Drosophila* and TYAYWWW in the apicomplexan *Babesia bovis*, (Yamagishi et al. 2014) with specific residues (in yeast; *Sc*TFIIB – E62, W63, R64, F66, R78, V79) (Kostrewa et al. 2009). Interestingly these residues are conserved in *Pf*TFIIB,

although so far an Inr motif has not been identified or confirmed in *P. falciparum* gene core promoters.

Between the B-reader region and the first cyclin lies the B-linker region. This region is required to bind to the clamp coiled-coil of RNAPII and the residues contained are required for promoter opening (*Sc*TFIIB – L71, S72, T73, I75; Grünberg et al. 2012). Although the linker region is shorter in *Pf*TFIIB as compared to the human and yeast homologues (by 10 residues), these residues are largely conserved in *Pf*TFIIB.

Additionally, a highly conserved region, called the charged cluster domain (CCD; indicated by a red bar), is important in regulating the overall TFIIB conformation. TFIIB may be in either a closed state, where intramolecular interactions between the N and C terminals appear to prevent transcription initiation (but not the formation of the PIC), or the open state where initiation is possible (Glossop et al. 2004). The residues implicated in this are highlighted in Figure *5*B, (*Hs*TFIIB - R53 and R66) and are conserved in *Pf*TFIIB.

The C-terminal cyclin domains of *Pf*TFIIB differ in conservation to other eukaryotic TFIIBs

In *Hs*TFIIB, The C-terminal domain (B-core or $TFIIB_C$) is composed of two imperfect direct repeats, each consisting of 5 α-helices, termed BH1-5 in the first repeat (amino acids 124-200), and BH1'-BH5' in the second repeat (218-294) Figure 5A. *Hs*TFIIBc interacts with TBP in the cleft formed by residues linking the two cyclin-like domains. The amino acid sequences over the linker region between the cyclin-like repeats differ considerably between species, both in length and composition (Figure 5). This region has been shown to form another α-helix (BH0), absent in Figure 5, which makes contact with RNAPII (Liu et al. 2010). The variable length of this region has been suggested partly explain the inability for human TFIIB to substitute for yeast TFIIB in yeast cells (Shaw et al. 1997; Shaw et al. 1996; Liu et al. 2010).

In the stabilisation of TBP-DNA binding, *Hs*TFIIB$_c$ interacts with DNA and makes sequence-specific contact with DNA elements located directly upstream and downstream of the TATA box. This interaction brings additional stability to the TFIIB-TBP-DNA ternary structure by increasing the total number of DNA residues contacted in the ternary TBP/TFIIB/TATA complex (Tsai & Sigler 2000). These sites are termed the B-recognition elements, upstream BRE (BREu) and downstream BRE (BREd) respectively, with the consensus sequences of SSRCGCC and RTDKKKK (Lagrange et al. 1998; Deng & Roberts 2005). The amino acid sequence over the first cyclin-like repeat domain shows conservation of residues thought to interact with the TBP molecules, as well as moderate conservation at residues thought to interact with BREd (Figure 5).

The sequence over the region of the second cyclin-like domain in *Hs*TFIIB shows less conservation of *Pf*TFIIB to other eukaryotic TFIIB proteins. It is thought that the BREu is contacted by the helix-turn-helix motif composed of evolutionarily conserved residues exposed on the recognition loops between the BH4'-BH5' helices of *Hs*TFIIB$_c$ (Lagrange et al. 1998). This interaction is thought to be TBP-independent, and to help direct the direction of the forming PIC. As in yeast, there is very low conservation of the residues at this site in *Pf*TFIIB (Tsai & Sigler 2000). However, there is conservation of positively charged R/K residues in the region of the HTH motif which does suggest some conservation of function.

In addition, there is low conservation of the residues which in the human system interact with TBP (and TLP) (Tsai & Sigler 2000). The interaction between *Pf*TBP and *Pf*TFIIBc was modelled computationally in 2005 by Buendía-Orozco and co-workers (Buendía-Orozco et al. 2005), who note that certain residues in the supposed interaction surfaces are conserved, such as the highly conserved residues which form the salt-bridge between *Hs*TBP and *Hs*TFIIB, but that *Pf*TFIIB also contains several unique amino acid substitutions which

do not support interactions between the molecules. These predictions need to be tested directly in protein-protein interaction and mutation analysis studies

### 3.1.2 Prediction of *Pf*TFIIB secondary and tertiary structures

The low-conservation of the overall sequence of *Pf*TFIIB to the well-described *Hs*TFIIB and *Sc*TFIIB prompted the determination of a predicted secondary and tertiary structure for the molecule. The program Phyre was used for this analysis.

Phyre makes use of the PsiPred, SSPro and JNet programs in order to generate secondary structure predictions (Kelley & Sternberg 2009), based on the comparison of the input sequence to homologous sequences with solved tertiary structures.

Phyre provides a confidence score from 0-9 on the prediction of secondary structures or disordered regions. The scores which are assigned to the secondary structures are illustrated in Figure 6.

*Pf*TFIIB has similar secondary structure arrangement to solved eukaryotic TFIIB molecules

The secondary structures of *Hs*TFIIB, along with the structural interaction sites/residues important for TFIIB function described in the literature, are displayed in Figure 6A. The secondary structure arrangement predicted for *Pf*TFIIB, along with the functionally important residues homologous to *Hs*TFIIB, are in Figure 6B. Table 7 in the Appendix 5.3 lists the positions of predicted secondary structures.

Overall it appears as though the secondary structures of *Pf*TFIIB are highly conserved with those present within the known structures of *Hs*TFIIB and *Sc*TFIIB. The predicted structure of putative *Pf*TFIIB is predicted to have the secondary structure arrangement necessary to form the zinc-finger, supported by the presence of conserved cysteine residues (in red), and most probably the rest of the secondary structures of the canonical N-terminus. Although Phyre only predicts these structures with low confidence, this is most probably due to the

lack of structures of the N-terminus present in the protein databases (Kostrewa et al. 2009). The length of the region, and the conservation of key interaction residues would suggest the formation of the same secondary structure arrangement. The Phyre analysis also predicts the presence of conserved cyclin repeats, with a similar number of B-helices in each repeat and at locations comparable to the *Hs*TFIIB core domain. There does appear to be a longer C-terminal region beyond this core domain in *Pf*TFIIB, conserved in *Plasmodium spp.* which may be a feature of *Plasmodium* TFIIB proteins.

Overall, Phyre predicts that *Pf*TFIIB has a similar structure to other studied TFIIB proteins. Along with general conservation of functionally described residues, this suggests similar functionality of *Pf*TFIIB to known eukaryotic TFIIB proteins. Additionally, given the high degree of sequence homology, these functions are conserved in other *Plasmodium* orthologous, Figure 6C.

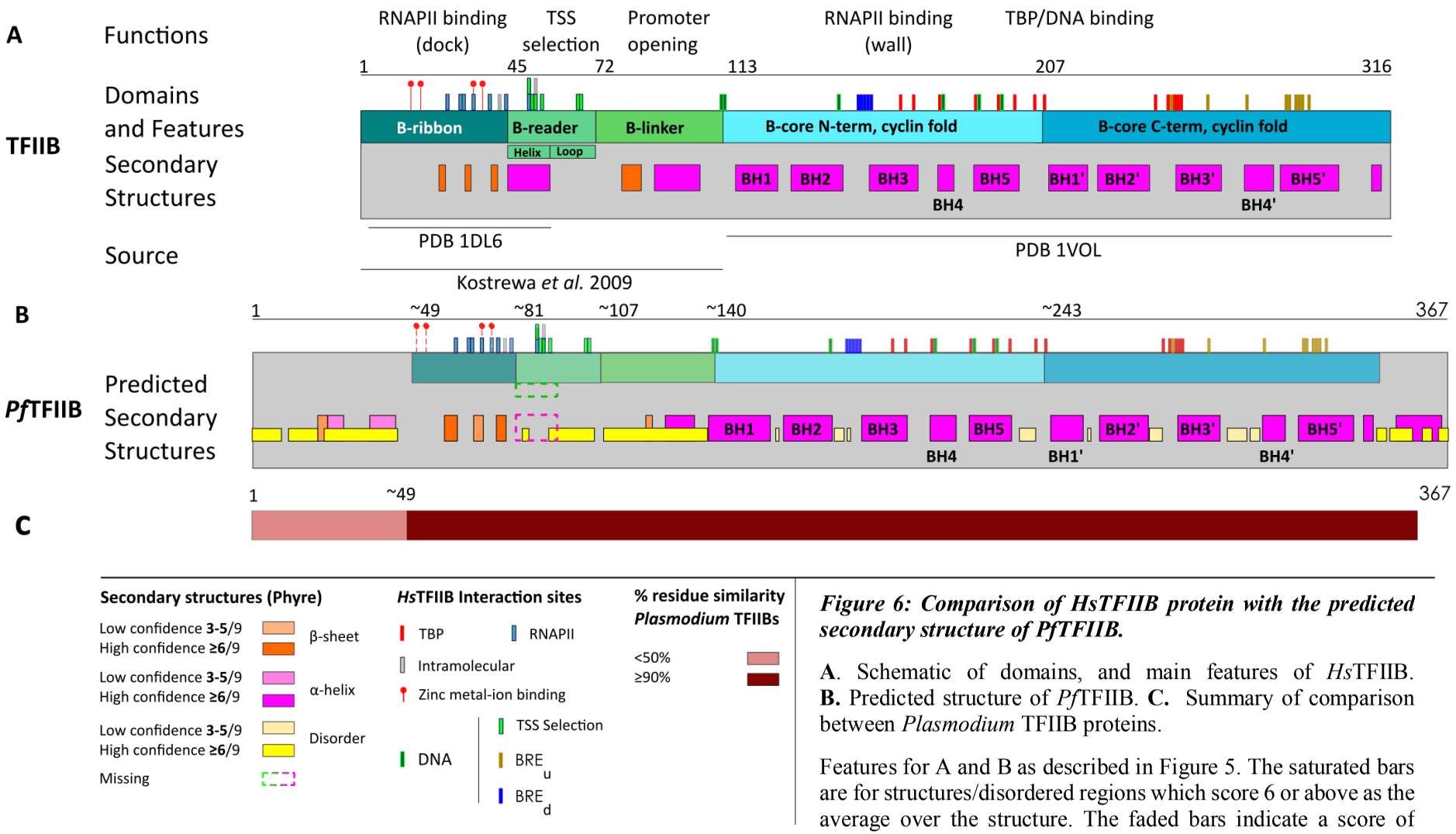



***Figure 6: Comparison of HsTFIIB protein with the predicted secondary structure of PfTFIIB.***

**A**. Schematic of domains, and main features of *Hs*TFIIB. **B.** Predicted structure of *Pf*TFIIB. **C.** Summary of comparison between *Plasmodium* TFIIB proteins.

Features for A and B as described in Figure 5. The saturated bars are for structures/disordered regions which score 6 or above as the average over the structure. The faded bars indicate a score of between 3 and 6. Scores of three or less are not illustrated. Dark maroon bar in C describes over 90% similarity of *Plasmodium* TFIIB pairwise aligned to *Pf*TFIIB. Light maroon describes less than 50% similarity.

*Pf*TFIIB protein tertiary structure is similar to eukaryotic TFIIB proteins

Phyre generates a predicted tertiary structure derived from solved structures previously deposited in molecular databanks. The predicted secondary structures of the putative *Pf*TFIIB were compared to solved TFIIB structures, shown in Figure 7A and B. The N-terminal domain description is derived from a solved crystal structure of *Sc*TFIIB, PDB entry 3K1F.M, solved by Kostrewa et al. (2009). The C-terminal or TFIIB-core domains structural information are inferred from the solved structure PDB entry 1VOL (Nikolov et al. 1995).

Phyre predicts that *Pf*TFIIB contains a C-terminus with two cyclin-like direct repeats composed of 5 alpha-helices each, and separated by a linker region (Nikolov et al. 1995), as described in other eukaryotes.

A comparison of the residues present in the cyclin domains was performed as a simple, non-quantitative test of the integrity of the predicted tertiary structure of *Pf*TFIIB$_C$. Cyclin domains are characterised by the presence of 5 alpha-helices, 4 of which surround a central alpha-helix. In order to achieve this structure, the central helix (H3/H3') is composed of predominantly hydrophobic residues, while the 4 external helices (H1, H2, H4, H5/H1', H2', H4', H5') are amphipathic, with hydrophobic residues facing inwards towards the central helix. The second cyclin shows the lowest sequence homology of *Pf*TFIIB to other eukaryotic TFIIBs, as is shown here. The residues in the second cyclins of *Hs*TFIIB and *Pf*TFIIB were coloured according to their broad physiochemical properties at pH 7 (Monera et al. 1995) (Figure 8). From a comparison of the two structures, it is revealed that the predicted cyclin regions of *Pf*TFIIB contain amino acid residues that would contribute

towards the folding of the cyclin domains. This supports the Phyre prediction that the *Pf*TFIIB C-terminal region folds into the canonical cyclin domains.

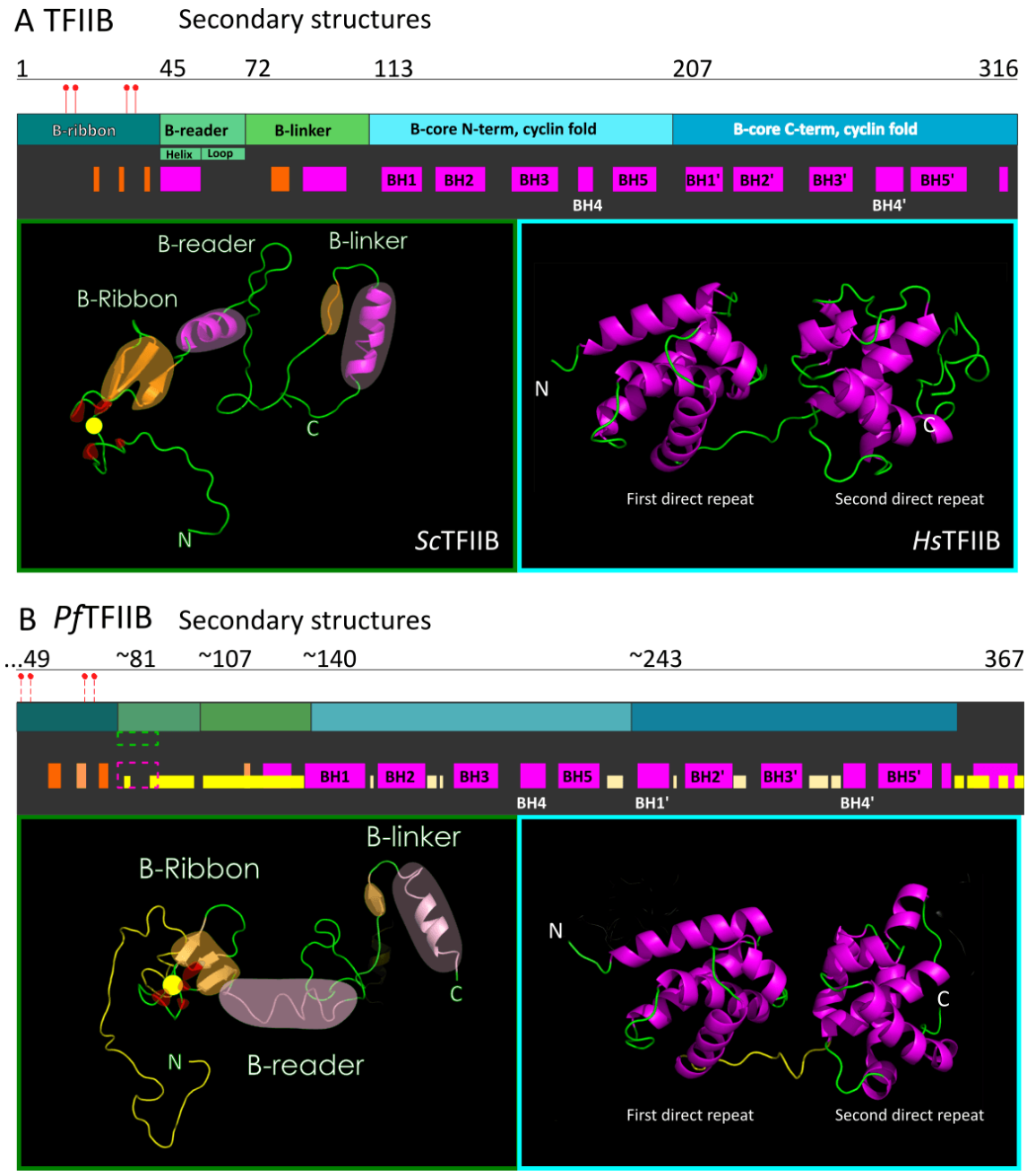



***Figure 7: Comparison of TFIIB solved crystal structure with predicted PfTFIIB tertiary structure.***

Solved and predicted structures of known TFIIB (*Sc* and *Hs*TFIIB) and Phyre predicted *Pf*TFIIB are shown in box insets, with a schematic of the broad secondary structure arrangements shown above. α-helices are coloured in purple, β-sheets are coloured orange. Disordered regions are coloured in yellow. The (cys/his) residues responsible for the formation of the zinc-finger are coloured red. The α -helices of the $TFIIB_c$ are labelled as B-helices or BH 1-5 and 1'-5', respectively. *Hs*TFIIB C-terminus: PDB entry 1VOL (Nikolov et al. 1995); *Sc*TFIIB N-terminus: PDB entry 3K1F.M, (Kostrewa et al. 2009).

### 3.1.3 Summary of *in silico* analysis of *Pf*TFIIB

The amino acid sequence alignments and the *in silico* analysis of the putative *Pf*TFIIB structure suggest an overall conservation of the main structural and functional features of classical TFIIB. Although there is a low-sequence homology at the amino acid level of the protein, there is strong support to the classification of the protein as a TFIIB molecule, albeit with some notable *Plasmodium*-specific features.

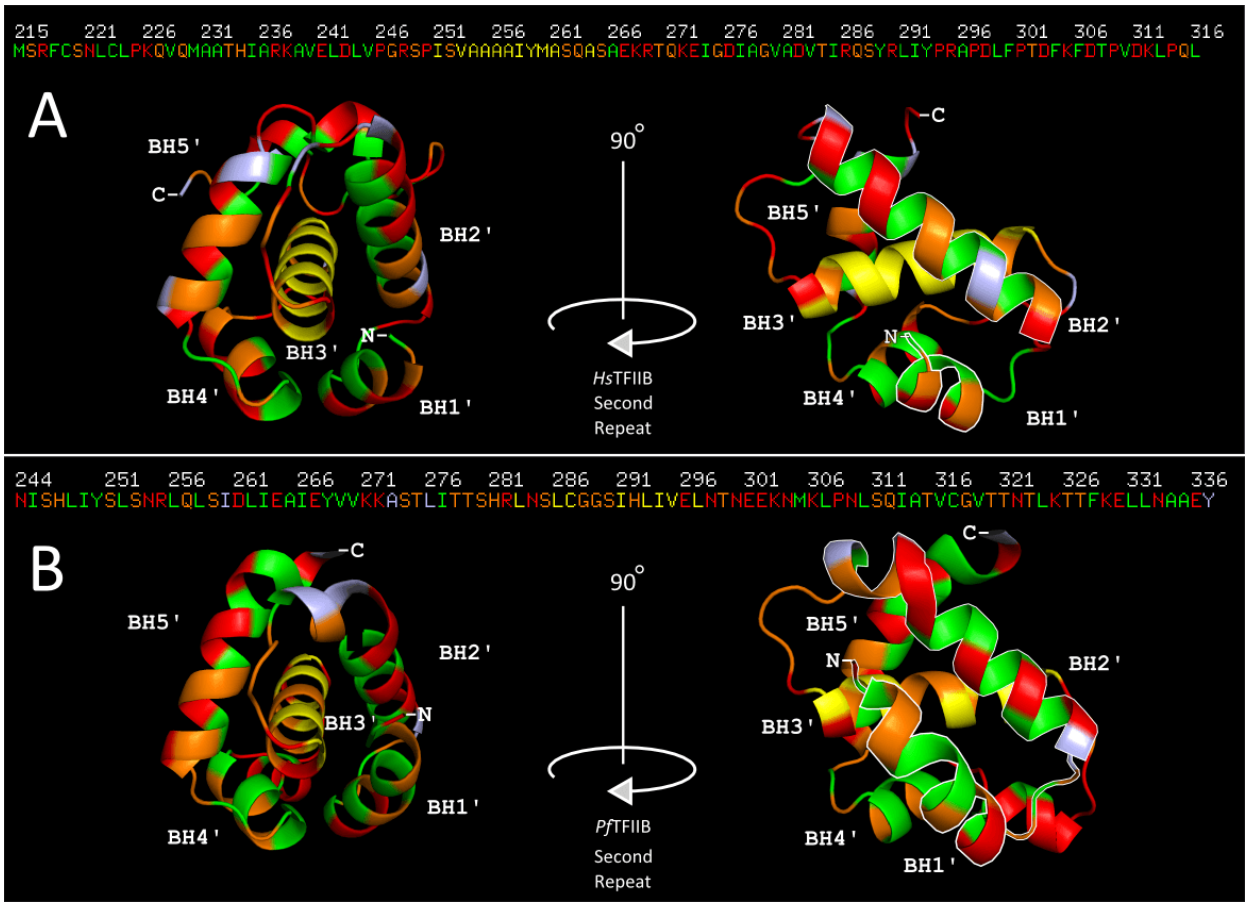



***Figure 8: Comparison of the second cyclin domain of HsTFIIB and PfTFIIB.***

Physiochemical properties of the residues of *Hs*TFIIB (**A**) and predicted *Pf*TFIIB (**B**) proteins are shaded according to the residue physiochemical properties at pH 7:

Yellow - hydrophobic residues of H3' core
Green - hydrophobic residues interacting with H3' core
Pale blue - hydrophobic residues that do appear to interact with H3' core
Red – Charged (acidic and basic) residues (K,R,D,E)
Orange - non-charged residues

Although the majority of the molecule is very highly conserved between *Plasmodium spp.*, the N-most terminal region is longer in *P. falciparum* (49 residues) and is extremely basic. This region has an average calculated isoelectric point of 12.53, higher than that calculated in human TFIIB, Table 4. Similarly, the C-most terminal region is longer than other eukaryotic TFIIB molecules, and is more positively charged at pI 10.19 as compared to *Hs*TFIIB with a pI of 4.45. Phyre predicts the possibility of helices in both of these regions. This may suggest the presence of additional DNA-binding motifs in *Pf*TFIIB not seen in other eukaryotic organisms. Future work into visualising this region may be important to the understanding of this putative TFIIB orthologue.

There is conservation of canonical N-terminal structures and residues important for TFIIB inter- and intramolecular interactions in *Pf*TFIIB. These interactions, such as those involved in RNAPII interactions, are supported by the deep conservation of these TFIIB and TFB molecules in the eukaryote and archaebacterial orthologues studied (Kostrewa et al. 2009; Liu et al. 2010). It seems likely, therefore, that *Pf*TFIIB maintains its role in the formation of the PIC. Interestingly, considering the A+T richness of the *P. falciparum* genome, there is also conservation of the residues involved in both the localisation of RNAPII to TSS, as well as the residues used to identify the Inr motif in other eukaryotes, although the Inr motif has not been identified in *P. falciparum*.

***Table 4: Comparison of calculated charge of human and putative Plasmodium TFIIB proteins, separated by region***

| Region | *Pf*TFIIB | | *Hs*TFIIB | |
|---|---|---|---|---|
| | **Residues** | **pI** | **Residues** | **pI** |
| N-terminus | 1 - 48 | 12,53 | 1 - 13 | 10,4 |
| B-ribbon | 49 - 81 | 4,26 | 14 - 45 | 4,04 |
| B-reader | 82 - 108 | 4,38 | 46 - 71 | 4,59 |
| B-linker (sheet) | 109 - 121 | 3.55 | 72 - 85 | 3.49 |
| B-linker (helix) | 127 - 140 | 11.28 | 86 - 113 | 10.24 |
| Core (BH1-BH5') | 141 - 338 | 9,69 | 114 - 293 | 9,88 |
| C-terminus | 339 - 367 | 10,19 | 294 - 316 | 4,45 |
| | | | | |
| **Average** | | **9.73** | | **8.48** |

There appears to be greater conservation of the first cyclin domain than that of the second cyclin domain. There is greater conservation of classical TBP-interaction sites in the first repeat than in the linker region and second cyclin. The reciprocal residues in TBP, and potential interactions between the molecules, should be investigated.

The increased conservation of the first cyclin compared to the second is also true for the residues implicated in locating the BREd sequence motif, as compared to the less conserved BH3'-BH4'. It is possible that *P. falciparum* promoters contain a BREd element similar to the classical BREd motifs described. The BREu recognition HTH motif in *Pf*TFIIB is far less conserved, with the exception of the basic residues in the HTH and supporting H3' helix. This is similar to plants and yeast, where the HTH motif appears to be largely missing (Tsai & Sigler 2000). This suggests either a divergence of the BREu motif in *Plasmodium* promoters, or perhaps significant alteration to the consensus sequence.

## 3.2 Bioinformatics analysis of putative *Pf*TLP – structure and function

### 3.2.1 Amino acid sequence alignment analysis

In order to assess the evolutionary conservation of *Pf*TLP as a member of the TBP-family, amino acid sequence analysis was performed.

Pairwise alignments of *Pf*TLP with *Hs*TBP (the core region, amino acids 163-339) or *Hs*TLP were attempted. However, it was not possible to generate a meaningful result from these alignments. The low sequence homology of *Pf*TLP to *Hs*TBP or *Hs*TLP resulted in large gaps and low-confidence alignments (data not shown).

*Pf*TLP is 368 residues in length, *Hs*TLP is 186 residues in length, and the full-length *Hs*TBP molecule is 339 amino acids in length, with a core region of about 180 residues. Whereas *Hs*TBP and many of its homologues, have a long and variable glutamine-rich N-terminus which extends the length of the protein, *Pf*TLP instead has lengthy low-complexity regions (LCRs) in the sequence. The presence of LCR regions combined with the sequence divergence of *Plasmodium* protein, prevents pairwise alignments with the protein orthologues in other eukaryotes.

In view of the inability to perform direct one-on-one alignments of *Pf*TLP with other TBPs, several (simultaneous) approaches were made to define the regions which contain TBP-like functional regions of *Pf*TLP.

Firstly, a multiple-sequence alignment was performed with TBP and TRF molecules from various organisms making use of the constraint-based alignment tool (COBALT) for multiple protein sequences (Papadopoulos & Agarwala 2007), a summary of which is displayed in Figure 9. *Pf*TLP has not been described at all in the literature, and the relationship it holds with other TBP or TBP-related factors is unknown. *Plasmodium* is an apicomplexan protist, and is evolutionarily distant from organisms which have well-

characterised TBP/TRF molecules. For this reason, a selection of TBP/TRFs were chosen for the alignment, and included animal (*Homo sapiens,* the insect *Drosophila melanogaster, Danio rerio, Xenopus laevis*), plant (*Arabidopsis thaliana)*, yeast (*Saccharomyces cerevisiae)*, archaebacterial (*Pyrococcus furiosus, Methanococcus maripaludis*), as well as other putative apicomplexan (*Plasmodium falciparum, Theileria annulata, Toxoplasma gondii*) TBP and TRFs. For ease of interpretation only a selection of the proteins from the alignment are shown in Figure 9.

Appended to Figure 9 are residues which have been identified in *Hs*TBP to have functional roles, such as interactions with TFIIB, TFIIA, and DNA, as entered in the NCBI entry for *Hs*TBP (Nikolov et al. 1995; Kim & Burley 1994; Geiger et al. 1996). Additionally, for the ease of interpretation, the LCR regions of *Pf*TLP and the variable N-terminal regions of the TBP molecules are not fully displayed.

*Pf*TLP shares conservation of functionally important residues

Eukaryotic (and prokaryotic) TBP and TRFs differ substantially in the N-terminal region. The core region shows a great deal of conservation across the proteins, especially at the functionally important residues (residues which interact with TFIIB, TFIIA and DNA). *Pf*TLP does not appear to have many of the residues conserved for the association to TFIIA (green in Figure 9). There are extreme amino acid substitutions at these residues. Additionally, the presence of the large LCR adjacent to the canonical TFIIA binding site does suggest that *Pf*TFIIA may not bind to the *Plasmodium* orthologue of TFIIA. This stands in contrast to the conservation of many homologous residues required for the binding of TFIIB.

Finally, when TBP binds to the TATA box a conserved set of phenylalanine residues are projected into the minor groove of the DNA, and cause the DNA to bend. These phenylalanine resides are highly conserved in TBP molecules, but in TLP molecules one or more of these resides are frequently missing. *Hs*TLP contains only a single conserved

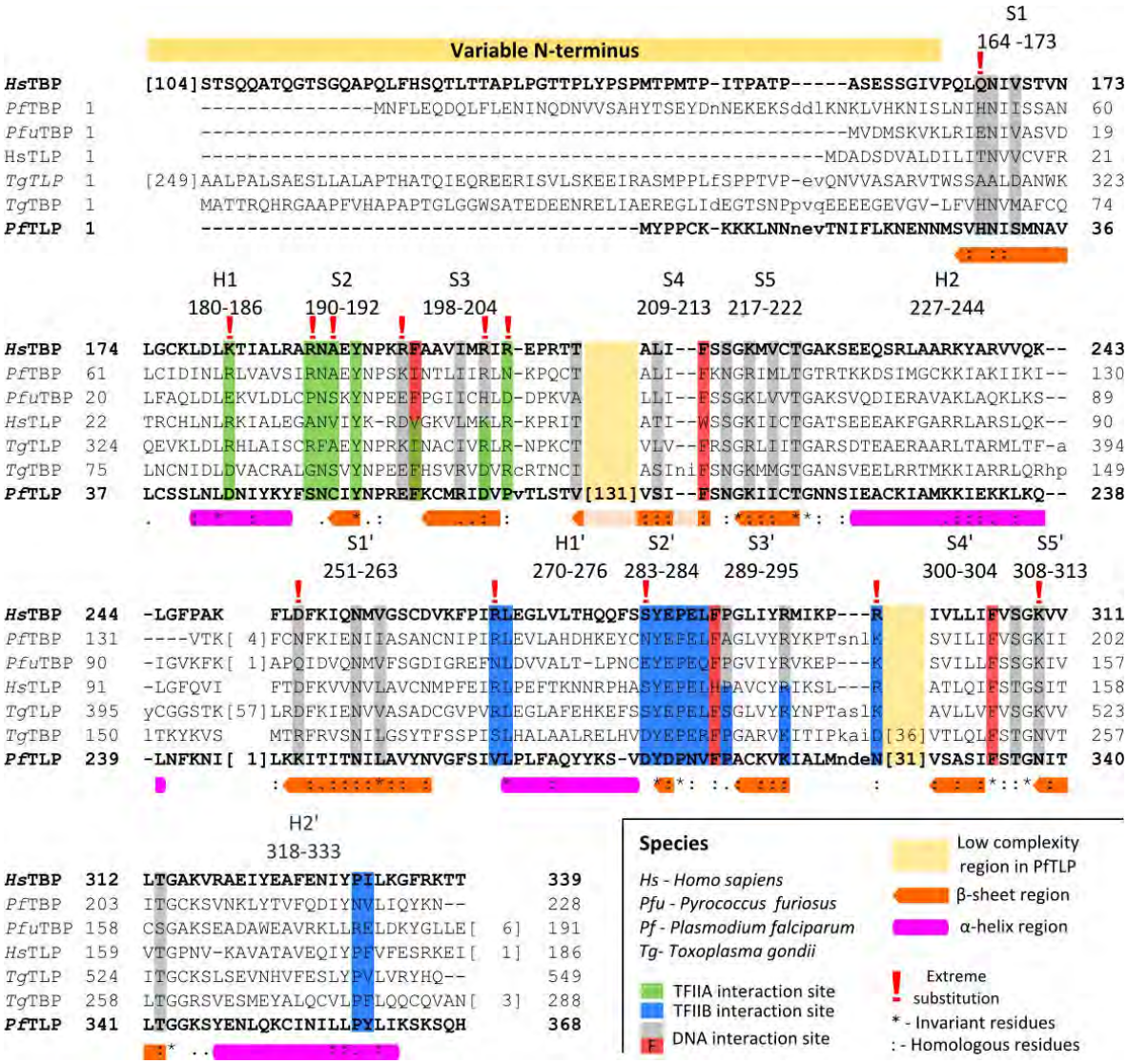



***Figure 9: Multiple sequence alignment of TBP-family proteins.***

COBALT alignment of selected sequences from a multiple sequence alignment of TBP and TBP-related factors from a range of eukaryotic and prokaryotic organisms. Inset in the box are the full names of the species displayed.

Secondary structures of *Hs*TBP are shown below the sequences. H and S refer to α-helices (purple) and β-sheets (orange) respectively. Low complexity regions in *Pf*TLP are indicated in pale yellow. *Hs*TBP sites which interact with *Hs*TFIIA, *Hs*TFIIB and DNA are coloured in green, blue and grey respectively. The conserved DNA-interacting phenylalanine residues are indicated in red. Invariant residues are indicated by an asterisk. Homologous residues are indicated by a colon, and in cases where the majority of the residues are conserved between species, a period is used.

phenylalanine residue, and the DNA binding of this molecule is not sequence specific for TATA box (Thomas & Chiang 2006). This is not the case for *Pf*TLP, where all four phenylalanine residues are conserved. This is interesting given the A+T rich content of the *P. falciparum* genome. Also intriguingly, *Pf*TBP is missing the first phenylalanine residue.

*Pf*TLP may be a distinct member of the TBP-related factors

Apicomplexan TLPs have not been described in the literature. In order to gain an understanding of their relatedness to other TRF proteins, the multiple-sequence alignment generated for Figure 9 was used to generate a minimal-evolution phylogenetic tree (Figure 10). The phylogenetic analysis places the apicomplexan TBPs and TLPs into separate clades. *Pf*TBP is placed in a large clade which encompasses both eukaryotic and archaebacterial TBP proteins. *Pf*TLP and the other putative apicomplexan TLPs are placed distinctly separately from the rest of the TLP molecules. Note that *Toxoplasma gondii* TLP and TBP molecules are placed in alternate clades. The sequence alignment data combined with the phylogenetic information suggest that the proteins have been incorrectly annotated; *Tg*TBP is most likely a TLP and *vice versa*.

The separation of the putative apicomplexan TLPs suggest that they may have arisen from an alternative gene expansion event from those which gave rise to the other eukaryotic TLPs. However, several caveats to this analysis must be mentioned. Firstly, it was not possible to take into account the very varied rates of evolution between such diverse organisms. In the formation of the dataset, gaps and missing data were removed from the aligned sequences, (157 positions were used in the final dataset) this would have removed the LCR regions as well as affected any other variable regions between the molecules. The effect of these variable length regions have not been characterised, and the impact they have on the rate of evolution is unknown. This result, therefore, is determined from the minimal evolution that would be required to generate the core-regions of the TBP and TBP-like proteins compared.

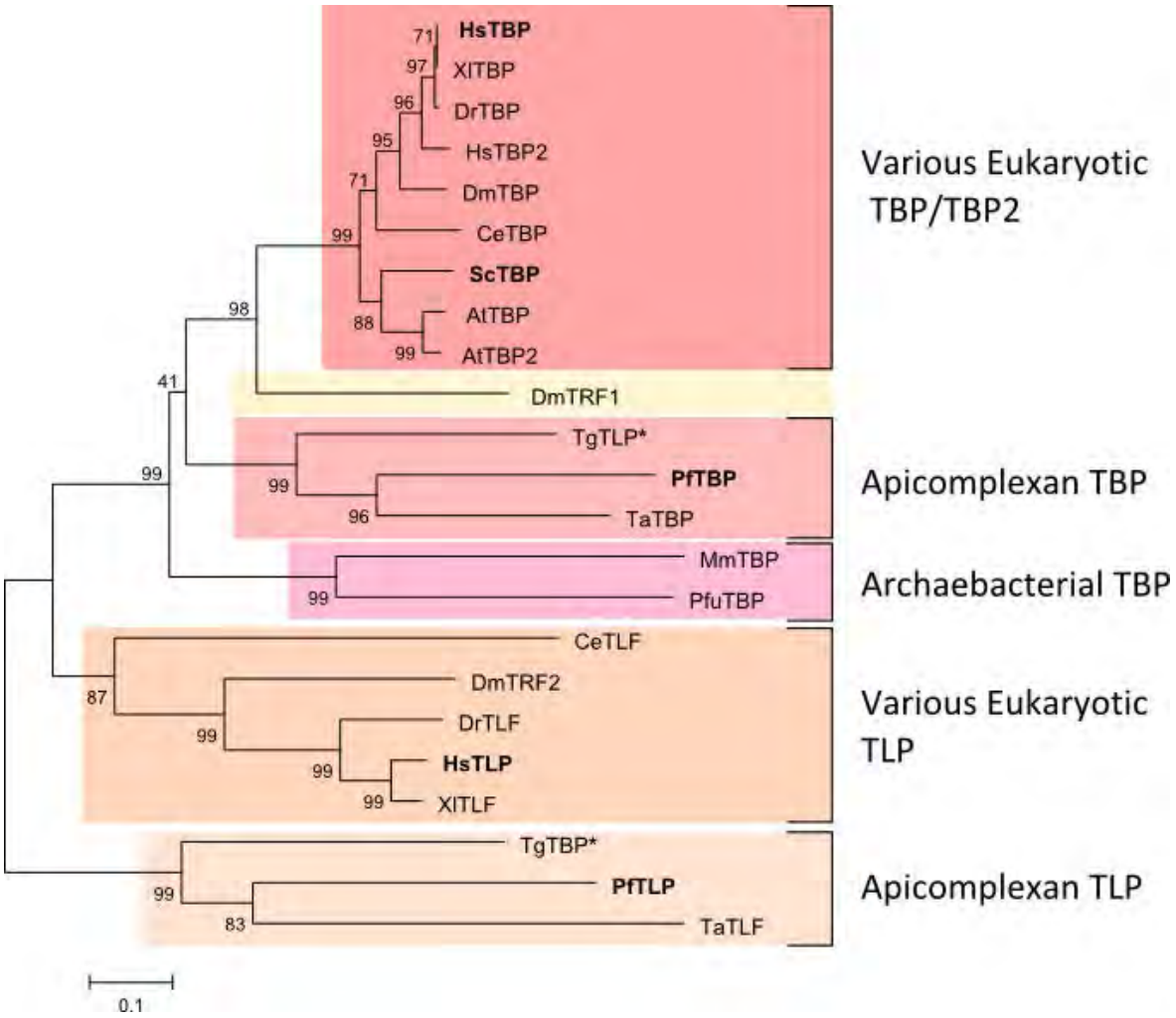



***Figure 10: Phylogenetic analysis of TBP-family proteins.***

Minimum-Evolution (CNI algorithm) tree of a sequence alignment of members of the TBP-family of proteins aligned in COBALT (Figure 9). Branch length is proportional to evolutionary distances computed by Poisson correction. Bootstrap test (500 replicates) was performed.

Colours are used to make the clades distinct. Red/pink indicate TBP/TBP2; orange indicates TLP (TRF2/TLF); yellow indicates TRF1 (*Drosophila* specific).

### 3.2.2 Prediction of *Pf*TLP secondary and tertiary structures

The high degree of conservation of *Pf*TLP to other TBP/TLP molecules suggested in Figure 9 suggests that *Pf*TLP has many features that are common to TBP/TLPs. However, the presence of the low-complexity regions does call into question if the molecule can form the appropriate secondary structures ascribed to TBP molecules, and if so, whether this structure is able to fold into an appropriate tertiary structure.

TBP (and its homologues/orthologues) core region/ C-terminal domain is a semi-symmetrical saddle-shaped structure, with a concave and positively charged surface which makes contact with the DNA, while the convex surface makes contact with other proteins/factors (Nikolov et al. 1995). The concave surface is made up of anti-parallel β-sheets, S1-S5 and S1'-S5'. Two long α-helices run along and on top of this beta-barrel motif, with two short α-helices at either end.

Phyre secondary structure prediction was performed with the *Pf*TLP sequence. The result of this prediction is presented in Figure 11, where it is compared to the known secondary structure of *Hs*TBP.

In Figure 11, the regions boxed together are the regions with the greatest sequence homology. There appears to be general conservation of predicted secondary structures of *Pf*TLP to *Hs*TBP in these boxed regions. Additionally, the general arrangement of secondary structures in this region are similar to those in *Hs*TBP. The large LCR region between the presumptive S3 and S4 beta-sheets also has several predicted structures. These structures are likely the result of repetitive sequences in the region, which may make partial alignments to regions in the scaffold sequences used to generate the secondary structure prediction. Interestingly, there is symmetry in the location of the LCR regions, with one LCR placed between the S3 and S4 sheets, while the other is placed between the S3' and S4'.

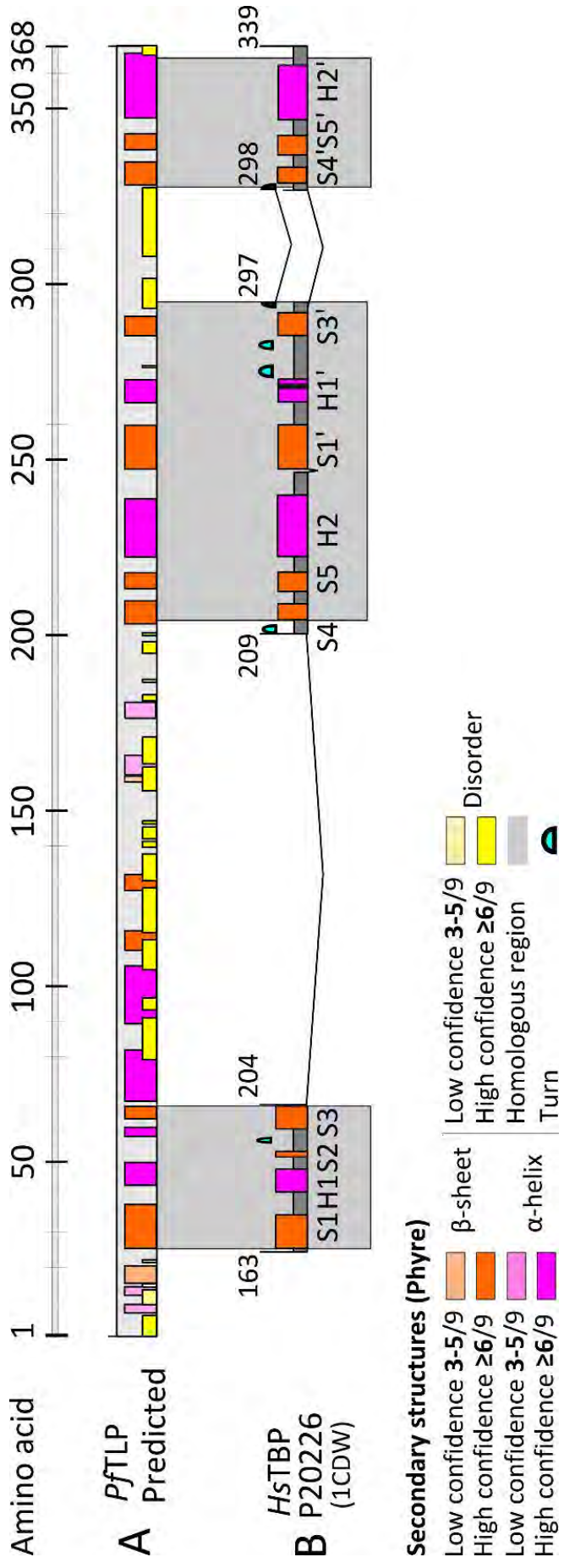



***Figure 11: PfTLP has conserved secondary structures as present in* Hs*TBP.***

**A.** Secondary structure arrangement of *Pf*TLP predicted by Phyre. **B.** Secondary structures of *Hs*TBP that aligns to *Pf*TLP regions based on sequence alignment. Note that LCR regions (indicated in yellow) are present in the regions where turns are present in the β-barrel motif. A confidence score from 0-9 on the prediction of secondary structures or disordered regions is indicated by the saturation of the bars in the figure. Saturated bars are for structures/disordered regions which score 6 or above as the average over the structure. The faded bars indicate a score of between 3 and 6. Scores of three or less are not illustrated. Disordered regions are similarly indicated in yellow. Turns in *Hs*TBP are indicated with a blue semi-circle.

*Plasmodium* TLP multiple sequence alignment

In order to investigate whether the predicted LCRs are likely to be correct, and not a feature of TLPs in *Plasmodium* TLP proteins, a multiple sequence alignment of putative, computationally identified, *Plasmodium* TLPs were compared. This alignment is presented in Figure 12, and the core region of *Hs*TBP has been appended to the graphical alignment to give positioning to the described TBP structural arrangement. From this figure it becomes clear that the conservation between the *Plasmodia* TLPs is greatest in the regions which are most conserved with other organisms, and least in the first LCR. The ~70-200 amino acid residue regions of *Pf*TLP show very little homology to regions in the other *Plasmodium* spp. barring that *P. reichenowi*, a very close relative. Increased susceptibility to mutations through less stringent evolutionary pressure on the residues to remain the same may account for both the variability in amino acids as and the variation in length of this region in the *Plasmodium* TLPs. There is moderate homology in the second LCR, between S3' and S4', but not the degree that is present in the presumed functional regions.

Interestingly, in *Pf*TLP the first LCR (from residue ~74 to residue ~222, or nucleotides ~222-666) has a guanine + cytosine (G, C) content of 18.3% (a Heterogeneous LCR) while the second LCR (from residues ~293-326, or nucleotides ~879-978) has a G+C content of 33.3% and appears to be a HighGC class LCR (Zilversmit, Volkman, Depristo, et al. 2010). Note that it is difficult without confirmed structural and functional data to assign rigid boundaries to the LCR positions in the *Pf*TLP sequence. These locations have been chosen with consideration given to their alignments against other TBP/TLP molecules, as well as the predicted secondary structure locations. They are open to further definition. Although both LCRs are present within symmetrically

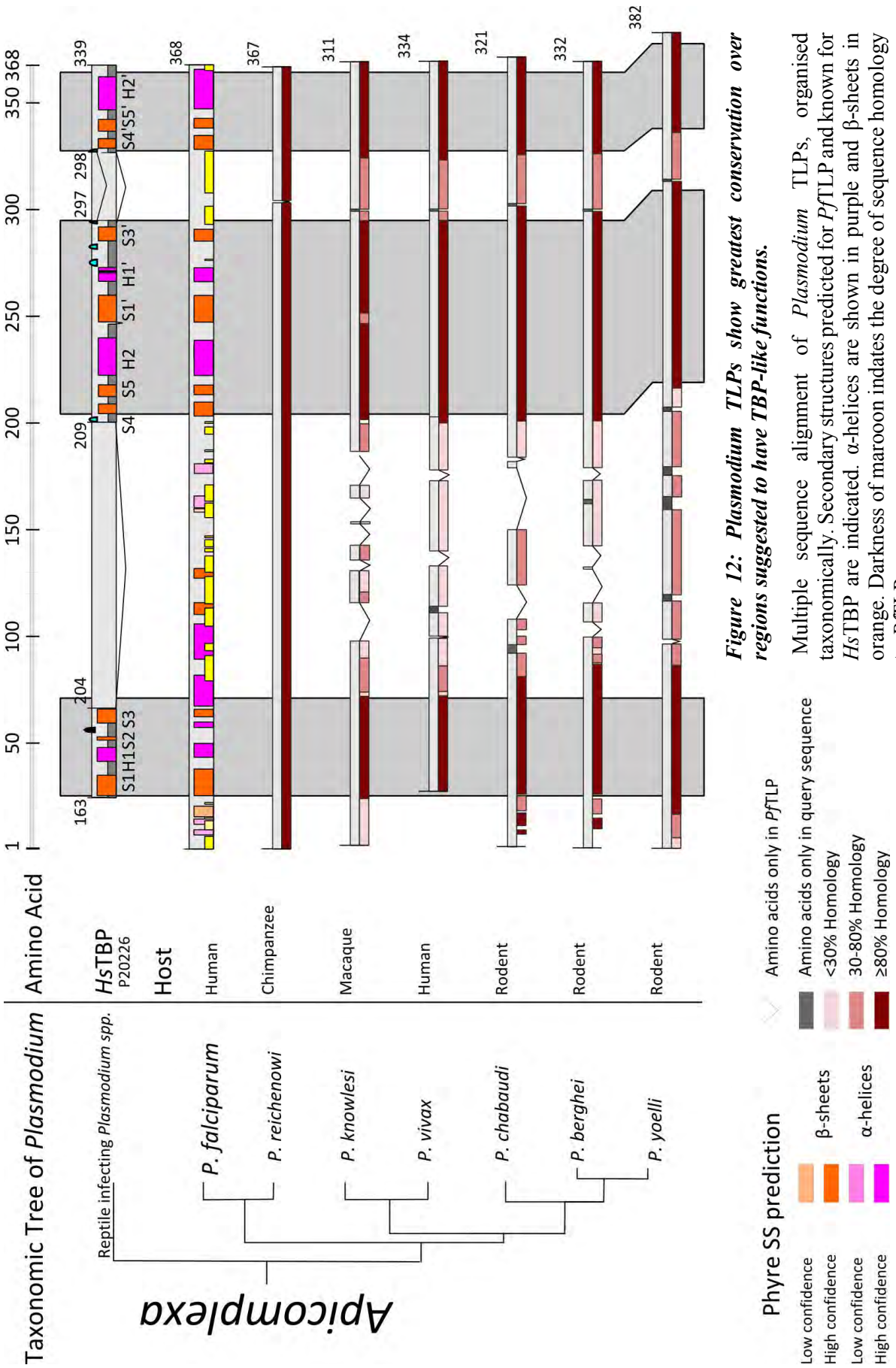



***Figure 12: Plasmodium TLPs show greatest conservation over regions suggested to have TBP-like functions.***

Multiple sequence alignment of *Plasmodium* TLPs, organised taxonomically. Secondary structures predicted for *Pf*TLP and known for *Hs*TBP are indicated. α-helices are shown in purple and β-sheets in orange. Darkness of marooon indates the degree of sequence homology to *Pf*TLP.

positioned loops, at the S3-S4 loop and S3'-S4' loop. It is known that LCRs typically form at turns, and are positioned as a loop out of the main globular structure. Supporting this is that 72% of the residues in the large LCR are polar and the LCR should be water soluble. It should also be noted that the presence of these large low complexity regions may have an effect on the correct folding and expression of the proteins in heterologous systems, such as the prokaryotic *E. coli* (Mehlin et al. 2006).

Tertiary Structure Prediction of *Pf*TLP

The Phyre prediction also generates a tertiary structure from databases of solved protein structures. The predicted structure of *Pf*TLP is shown in Figure 13, along with the known structure of *Hs*TFIIB. Appended to this figure is a 2-dimensional schematic of the proteins, for easy comparison.

The predicted structure of *Plasmodium falciparum* TLP appears to be semi-symmetrical and similar to the canonical TBP structure, albeit with sets of only four β-sheet motifs, as compared to the five in *Hs*TBP. Interestingly, the conserved phenylalanine residues, which prise into the DNA helix are conserved and placed at similar locations. The predicted LCR regions are placed in the turns between the S3-S4 and S3'-S4' sheets. Interestingly, these LCRs are adjacent to where in humans, *Hs*TBP is known to form complexes with TFIIB and TFIIA. The residues responsible for TFIIA interactions have been noted to be less conserved than those seen in TFIIB. This raises interest in the ability for *Pf*TLP to bind to the homologues of these factors identified in *Plasmodium*.

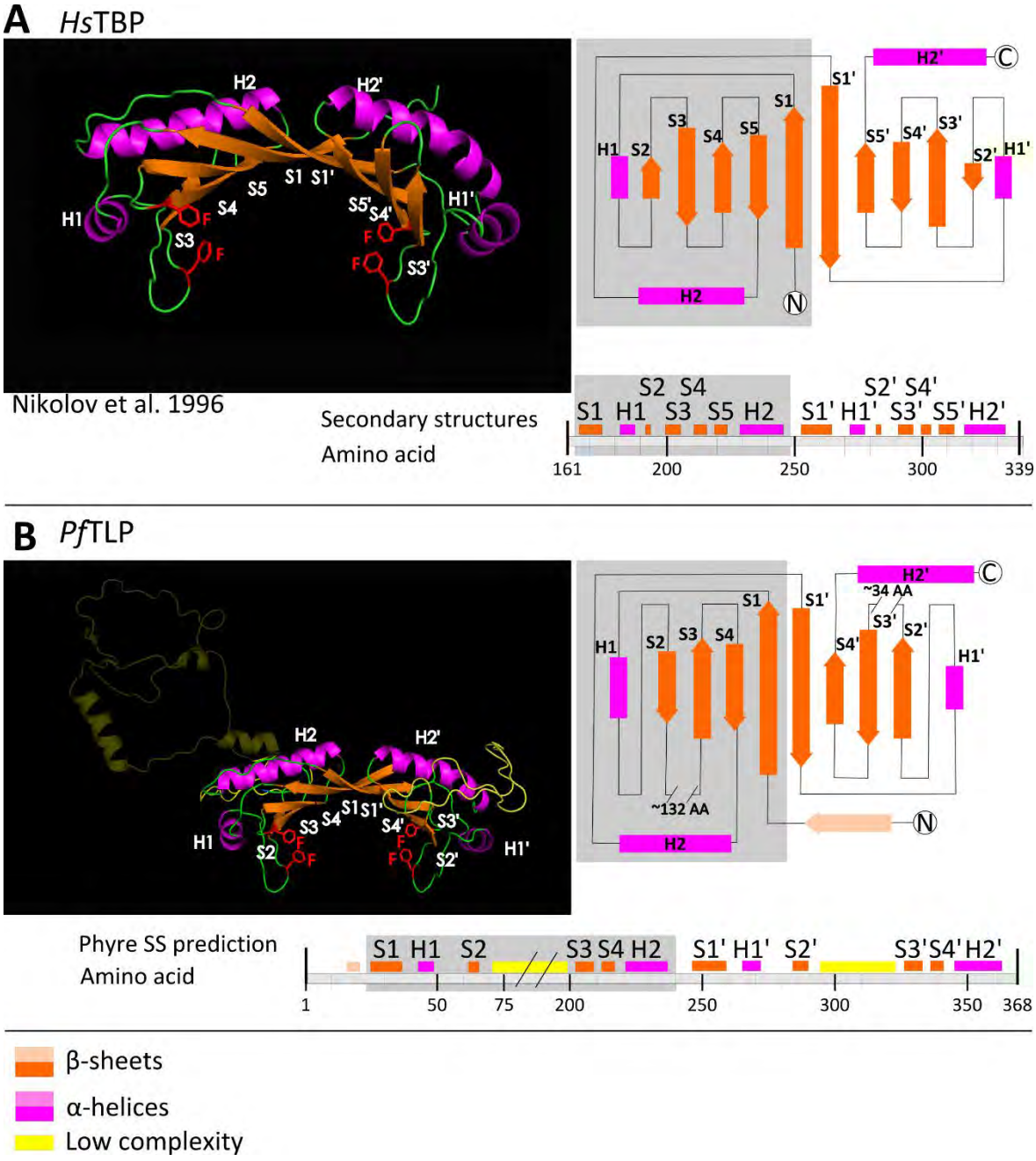



***Figure 13: The predicted PfTLP structure is analogous to HsTBP.***

Schematics shown to the right of 3D structures show the semi-symmetrical nature of the molecules and indicate the placement and direction of alternating β-sheets (orange) which form the β-barrel. α-helices are coloured purple.

### 3.2.3 Summary of *in silico* analysis of *Pf*TLP

The amino acid sequence alignments and the *in silico* analysis of the putative *Pf*TLP structure suggest an overall conservation of the main structural and functional features of canonical TBP/TLP molecules, albeit with some notable *Plasmodium*-specific features. The most obvious of these is the presence of the LCRs in the molecule. Interestingly, while *Pf*TLP displays these LCR structures, putative *Pf*TBP does not. Additionally, *Pf*TBP has the classical long N-terminal region, which is not true for *Pf*TLP, nor other TRF2 molecules (Thomas & Chiang 2006). This suggests that putative *Pf*TLP has correctly been annotated as a TBP-related factor, but as evidenced from a phylogenetic analysis, as well as the presence of the four phenylalanine residues, not necessarily a TRF2 orthologue. The ability for *Pf*TLP to bind DNA seems likely. The protein has an overall basic charge, a pI of 9.71, and appears to have the ability to fold a β-barrel similar to canonical TBP proteins. It should be noted that the first long complexity region has a basic pI of 9.39 (residues 71-202), and may possibly have secondary structures form in this region. This may suggest additional DNA-interaction motifs. Should *Pf*TLP be shown to have clear DNA-binding activity, this region should be investigated for a potential role in DNA-binding.

Interestingly, *in silico* prediction of the general transcription factors, has predicted a single large subunit, and two small subunits (γ-subunit) of putative *Pf*TFIIA (PlasmoDB accession codes *Pf*3d7_1250700 and *PF*3D7_0933700). It is known that in frog oocytes, the TLP orthologue interacts predominantly with an analogue to TFIIA, the TFIIA-like factor (ALF), which has not yet been identified in *P. falciparum*. Similarly, the transcriptome profile of *Pf*TLP and the small TFIIA γ-subunit *PF*3D7_0933700, are both shown to be predominantly expressed in the gametocytes of the parasite (www.PlasmoDB.org). The potential for interactions of *Pf*TLP/*Pf*TBP with alternate TFIIA-gamma subunits should be investigated, and may shed light on the roles of these two proteins.

## 3.3 Expression and Purification of *Plasmodium* TFIIB

An expression vector for *Pf*TFIIB was cloned previously in the lab and provided by Dr Thomas Oelgeschläger. This vector expresses *Pf*TFIIB with an N-terminal hexahistidine tag, and is driven by a T7 promoter (Figure 45). T7 polymerase is expressed BL21-codonPlus® (DE3)-RIL cells through induction of the *lac*UV5 promoter with IPTG.

### 3.3.1 IPTG-induced overexpression of recombinant *Plasmodium* TFIIB leads to a reduction in cell growth in *E. coli*.

IPTG-induced overexpression of 6His-*Pf*TFIIB was performed in order to assess the effects of heterologous gene expression on the *E. coli* host cell, and to set a basis for future expression trials. To this end, two sets of *E. coli* BL21-codonPlus® (DE3)-RIL cell cultures, transformed with the pET11d-6His-*Pf*TFIIB vector, were grown in liquid media at 37°C. The density of the cell cultures were monitored by light spectrophotometry (wavelength of 600nm). At an $OD_{600nm}$ of approximately 0.5, one of the cultures was induced with 1mM IPTG. The cell growth was monitored, and the cell cultures harvested two hours post-induction (Figure 14).

Induction of protein expression by IPTG resulted in a reduction in growth of the induced cell culture, as compared to the non-induced control. This indicates that the IPTG-induced overexpression of 6His-*Pf*TFIIB is detrimental to the host cell-growth. The reduction of cell-growth was reasoned to be due to the metabolic stress arising from the strongly induced overexpression of both T7 polymerase, as well as recombinant 6His-*Pf*TFIIB. The expression of 6His-*Pf*TFIIB protein by the cells was, therefore, investigated.

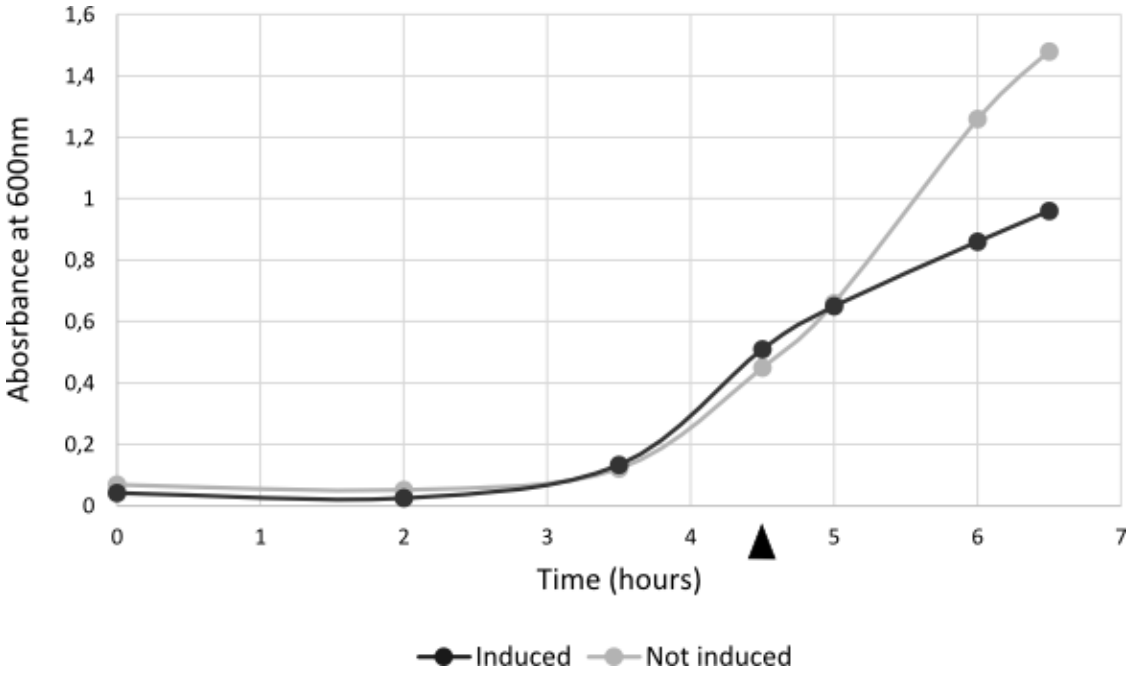



***Figure 14:*** **IPTG-induced *6His-PfTFIIB overexpression in E. coli BL21-CodonPlus-RIL cells leads to growth inhibition.***

Cell densities of cultures, measured at wavelength 600nm, were plotted against the number of hours post-inoculation of expression culture from overnight cultures. Growth of the induced culture (black) is compared to a non-induced culture (grey). The time of induction with 1mM IPTG is indicated by a black arrow.

### 3.3.2 Expression of recombinant *Plasmodium* TFIIB can be achieved from non-induced cell cultures.

In order to determine the expression of recombinant *Plasmodium* TFIIB protein from the harvested cell pellets from the IPTG-induced protein expression experiment, the cells were lysed and the soluble protein fraction (cleared cell lysate) recovered. Note that the cell mass recovered from the IPTG-induced cell culture was less than that of the non-induced cell culture. The cell mass to volume ratio of these cleared cell lysates were prepared identically (5ml of sonication buffer per gram of cell mass). A pilot-scale affinity purification with PureProteome™ Nickel Magnetic Beads (*Merck Millipore*; abbreviation: Ni-beads) was conducted (Figure 15). It was found that there was enrichment of a protein at the expected size for 6His-*Pf*TFIIB protein in both the non-induced (A) and IPTG-induced (B) cultures.

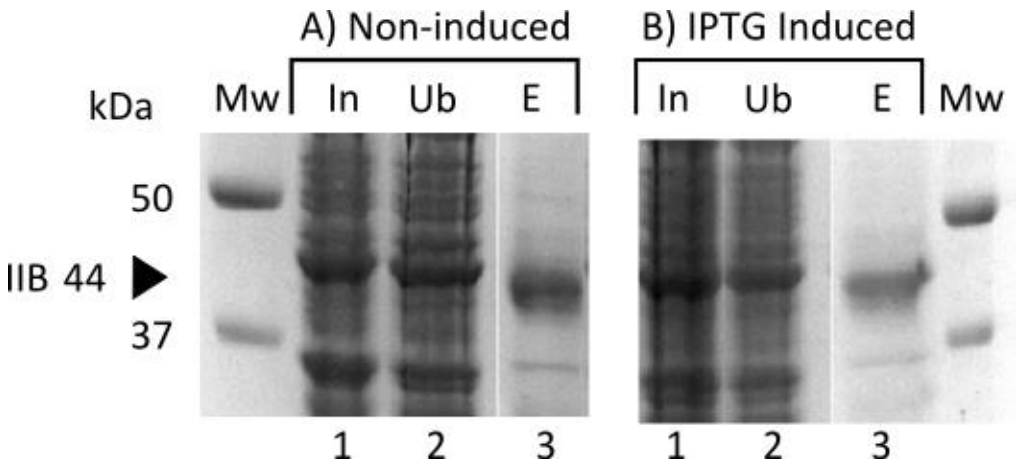



***Figure 15: PfTFIIB protein is expressed in both induced and non-induced E. coli cells.***

*E. coli* BL21-CodonPlus-RIL cells carrying the expression vector pET11d-6His-*Pf*TFIIB were grown and either induced with 1mM IPTG (Panel B) or not (Panel A). Protein was purified by nickel-affinity purification from cleared cell lysates and analysed by SDS-PAGE. Equivalent amounts of cleared lysate input (**In**; 15/1000 vol.) and the protein fraction not bound to the Ni-beads (unbound, **Ub**), and the protein fraction eluted from the Ni-beads with 2×SDS-loading buffer (**E**, ½ vol.) were loaded. A black arrow indicates the expected size of 6His-*Pf*TFIIB protein at 44kDa. Molecular weight marker (**Mw**) – 5µl of Precision Plus Protein All Blue standard (*Bio-Rad*).

The enriched protein in the non-induced sample was similar in molecular mass (kDa) and quantity (relative to the input lanes) to the IPTG-induced sample, which suggests that the enriched protein in both samples was either a contaminant, or 6His-*Pf*TFIIB protein. Additional attempts to express *Pf*TFIIB protein consistently led to similar results as in Figure 15; IPTG-induction of 6His-*Pf*TFIIB protein expression results in both reduced cell growth and cell mass recovered. Additionally, the expression of 6His-*Pf*TFIIB may be achieved from the non-induced cell cultures. Considering that more cell mass can be harvested from the non-induced cell cultures than from induced cell cultures, and that the quantity of 6His-*Pf*TFIIB per gram of cell mass is similar, it was reasoned that expression of 6His-*Pf*TFIIB from non-induced cell cultures would be a more efficient and cost-effective method of recombinant protein production.

The media used for the experiment described in Figures 14 and 15 was prepared from tryptone powder. Tryptone powder is prepared from the enzymatic digestion of casein,

derived from bovine milk and it is known that tryptone powder may be contaminated with traces of lactose. Lactose has been extensively studied in the activation of transcription from the *Lac* operon (Perlman et al. 1969) as part of the innate metabolism regulation pathways in *E. coli*. The *Lac* operon has been incorporated into the *E. coli* BL21 strains in order to regulate the expression of T7 polymerase which in turn drives protein expression from plasmids containing a T7 promoter, in this case pET11d plasmids. The presence of lactose in the media may thus account for the presence of protein expressed in the non-induced sample.

The *Lac* operon is repressed through the action of the catabolite activator protein (CAP) naturally expressed in *E. coli* in the presence of glucose (Perlman et al. 1969). It was reasoned that the addition of glucose to the growth media would repress the leaky expression of T7 polymerase, and therefore inhibit 6His-*Pf*TFIIB expression. Should the expression of T7 and/or 6His-*Pf*TFIIB negatively affect the growth of the cells, it is likely that the transformation of cells on plates without glucose may select against cells which express high levels of the proteins. For this reason, *E. coli* BL21-CodonPlus-RIL cells were transformed and plated onto media that contained 1% (w/v) glucose.

### 3.3.3 Expression of recombinant *Plasmodium* TFIIB can be achieved from the release of catabolite repression in non-induced cell cultures.

Having identified potential recombinant protein expression from non-induced cell cultures, it was investigated whether protein expression through an 'auto-induction' method could be increased.

Diauxic growth of *E.coli* cells grown in glucose-supplemented media

*E. coli* cells grown in media supplemented with glucose exhibit a diauxic growth curve as the bacteria alter their metabolism based on the availability of nutrients (Wong et al. 1997).

Diauxic growth is characterised by the presence of two exponential and stationary growth phases. An initial exponential growth phase occurs as the cells grow on the easily metabolised glucose. As the glucose is depleted the cells enter a short stationary phase as they adjust their metabolic machinery for the utilisation of a secondary nutrient source. A second exponential phase begins as they start to metabolise the alternative energy source.

This experiment sought to investigate whether the *lac*UV promoter driven T7 polymerase expression could be repressed by the supplementation of the growth media with glucose. Additionally, as the cell culture is depleted of glucose, whether the lifting of catabolite-repression would result in the expression of T7 polymerase, and thus recombinant *Pf*TFIIB.

In this experiment, both the overnight cultures used to inoculate the expression cultures, as well as the expression cultures were supplemented with 1% glucose to repress leaky expression from the *lac*UV operon. Two cell cultures were grown, and the growth of the cells measured by spectrophotometry. One culture was harvested late into the first

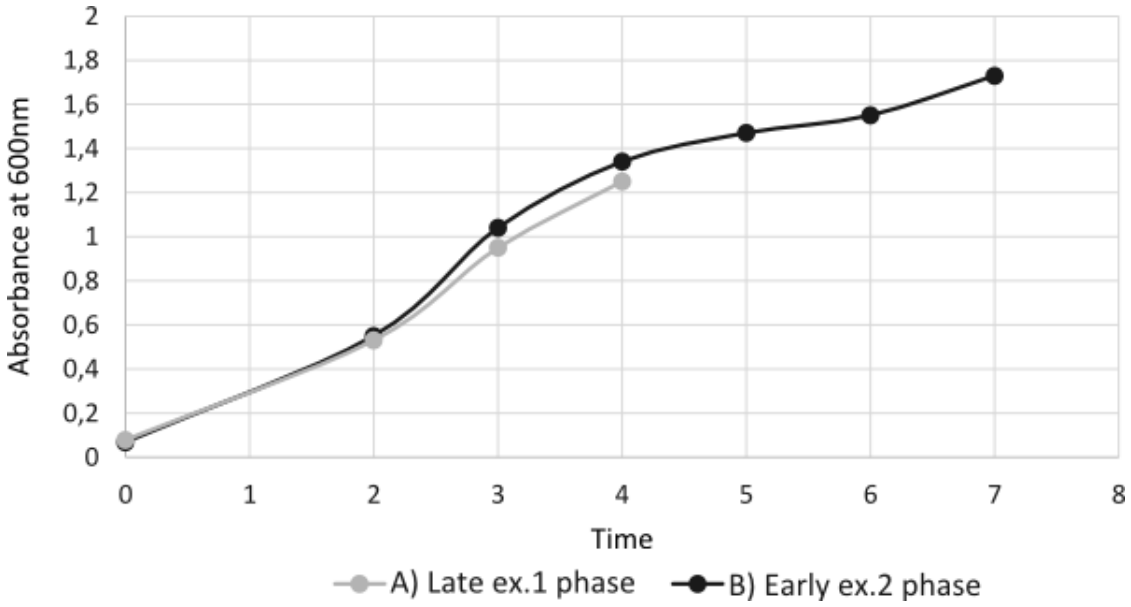



***Figure 16: Diauxic growth of transformed E. coli cells in growth media supplemented with 1%glucose.***

*E. coli* BL21-codonPlus® (DE3)-RIL cells carrying the expression vector pET11d-6His-*Pf*TFIIB were grown and harvested at the late initial exponential phase or early in the second exponential phase. Optical densities of the cultures are plotted against time.

exponential growth phase, when catabolite repression should be in place. A second culture was allowed to grow past the first stationary phase, at which point the glucose-derived catabolite repression would be lifted. As expected, the cell culture allowed to continue to grow past an initial exponential phase displayed a diauxic growth curve (Figure 16).

*Pf*TFIIB protein is expressed through the release of catabolite repression

Expressed 6His-*Pf*TFIIB protein was purified from cleared cell lysates and analysed by SDS-PAGE as described above. It is clear in the comparison of the protein purified from the culture harvested late in the first exponential phase, and that which was grown to the beginning of the second exponential phase (Figure 17), that protein is expressed through 'auto-induction', due to the release of catabolite-repression. Additionally, due to the presence of glucose in the media, the rate of cell-growth was increased, and resulted in higher

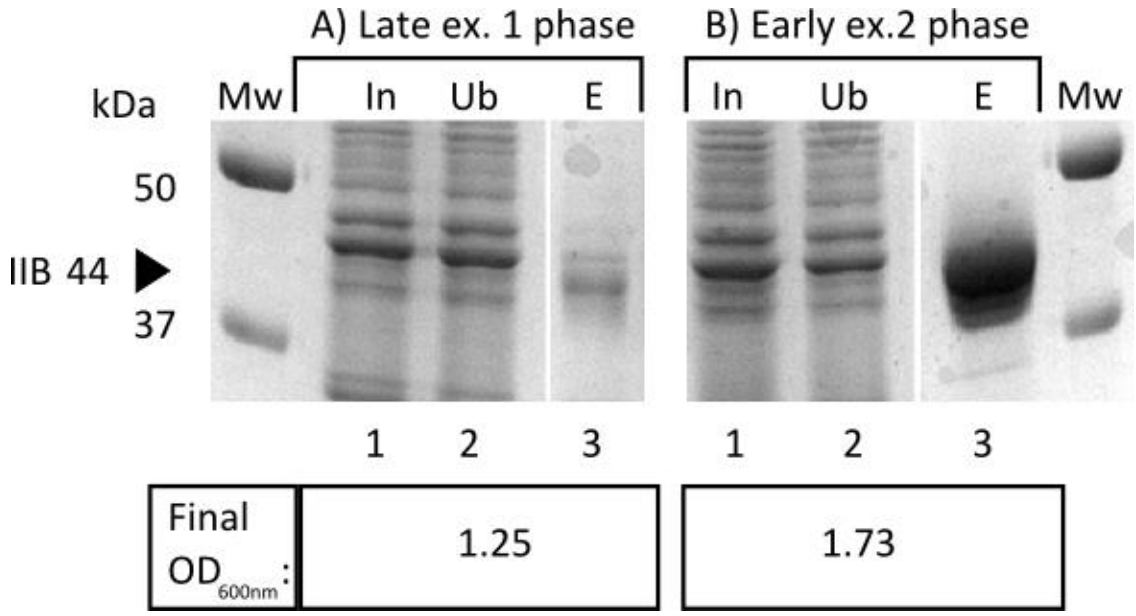



***Figure 17: Increased expression of recombinant PfTFIIB protein in E. coli BL21-codonPlus® (DE3)-RIL cells through the release of catabolite repression.***

*E. coli* cells carrying the expression vector pET11d-6His-*Pf*TFIIB were grown and harvested either at the late initial exponential phase (Panel **A**), or early second exponential phase (Panel **B**). Protein was purified by nickel-affinity purification from cleared cell lysates and analysed by SDS-PAGE. Equivalent amounts of cleared lysate input (**In**, 1/100 vol.), and the protein fraction not bound to the Ni-beads (unbound, **Ub**), and the protein fraction eluted from the Ni-beads with 2×SDS-loading buffer (**E,** ½ vol.) were loaded. A black arrow indicates the expected size of 6His-*Pf*TFIIB protein at 44kDa. Molecular weight marker (**Mw**) – 5µl of Precision Plus Protein All Blue standard (*Bio-Rad*).

cell-density and increased harvested cell mass. These results clearly show that *Pf*TFIIB may be efficiently expressed through auto-induction. The identity of 6His-*Pf*TFIIB was confirmed by mass spectrometry.

Having successfully determined the conditions necessary for 6His-*Pf*TFIIB expression, accumulation of cell mass of cells which had expressed the protein was performed.

Discussion of expression of 6His-*Pf*TFIIB

There is clear growth reduction upon induction with 1mM IPTG in the 6His-*Pf*TFIIB expressing cells as compared to the non-induced cell culture, (Figure 14). This suggests that the induction of the cell cultures with IPTG led directly to either a reduction in cell growth and/or cell death. Upon induction with IPTG the cells are required to utilise many cell-resources in order to synthesise both the T7 polymerase, as well as 6His-*Pf*TFIIB proteins. The metabolic burden of heterologous gene expression and protein synthesis may deter cell growth or prove to be toxic, and this is a known issue with heterologous gene expression the BL21(DE3) *E. coli* cells (Rosano & Ceccarelli 2014; Bentley et al. 1990; Dumon-Seignovert et al. 2004). This protein over-production may result in cell death. It is also known that individual cells may vary widely both in their sensitivity to an inducing agent, as well as in their protein expression. It is suggested that since the protein expression per gram of cell mass in both cultures was the same, the cells which survived induction by IPTG ultimately expressed similar average quantities of protein per cell, however, due to the decrease of harvestable cell mass from these induced cell cultures, it was decided to instead make use of the auto-induction method of protein expression.

### 3.3.4 Bulk purification of recombinant 6His-*Pf*TFIIB protein

Purification of a bulk preparation of recombinant 6His-*Pf*TFIIB protein was performed as described in Chapter 2.5.2. The fractions containing the greatest amount of protein were

pooled, and titrated against a standard curve of BSA protein ( Figure 18 and Figure 49 in the Appendix). The total pooled protein fraction was 1.4ml, at a concentration of 50ng/µl per band of the doublet, or 100ng/µl in total.

Analysis of purified 6His-*Pf*TFIIB

From Figure 18 (full gel is presented in the Appendix 5.6 as Figure 49), it appears as though two bands are present in the region of 44kDa. The two bands are approximately 2-3kDa different in size, with the top band determined by a standard curve against the molecular weight marker being approximately 44kDa, and the lower band approximately 42kDa. This is consistent with the results seen throughout the optimisation of protein expression, during the accumulation of protein expressed cell mass, and in the purification of the sample.

It is unclear if one of the bands of the doublet is the result of a contaminant in the protein preparation. Table 5 contains an abbreviated list of common proteins which may contaminant protein preparations purified through immobilised metal affinity chromatography, as was used here. Several species of proteins may have contaminated the preparation at the molecular weight expected for 6His-*Pf*TFIIB, such as ArgE and ODO2, which are known to frequently contaminate Ni-NTA affinity preparations from *E. coli*

***Table 5: Abbreviated table of metal-affinity purification commonly contaminating proteins***; derived from Bolanos-Garcia & Davies (2006)

| ***Protein*** | ***SwissProt access code*** | ***Molecular Mass (kDa)*** | ***pI*** |
|---|---|---|---|
| CAT | AAA57080 | 25.5 | 5.9 |
| ArgE | P23908 | 42.3 | 5.5 |
| ODO2 | P07016 | 44.0 | 5.5 |
| YfbG | P77398 | 74.2 | 6.3 |

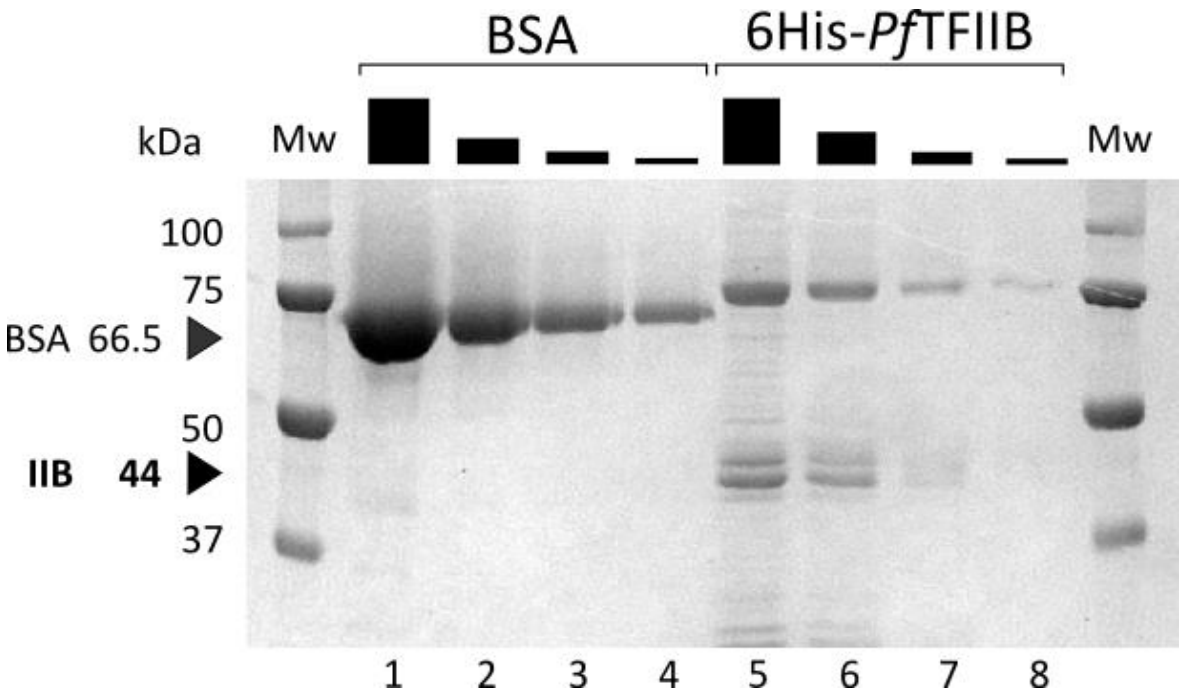



***Figure 18: Determination of concentration and purity of 6His-PfTFIIB.***

BSA standards were loaded in lanes 1-4, 5µg, 2µg, 1µg and 0.5µg respectively. Lanes 5-8, 10µl, 5µl, 2µl, and 1µl of pooled protein fractions respectively (1.4ml total). Expected size for 6His-*Pf*TFIIB is 44kDa, indicated with a black arrow. Molecular weight marker (**Mw**) – 5µl of Precision Plus Protein All Blue standard (*Bio-Rad*).

(Bolanos-Garcia & Davies 2006). However, throughout the protein expression trials the protein bands of the doublet at 44kDa consistently remained at approximately the same quantity, and were enriched well beyond that of other contaminant proteins. Additionally, BL21-codonPlus® (DE3)-RIL *E. coli* cells not carrying an expression plasmid were grown under similar conditions, and the soluble protein analysed through Ni-NTA affinity purification (Figure 48, the Appendix 5.5). There is little indication of enrichment of proteins at the 44kDa molecular weight in these non-transformed bacteria, suggesting that the two bands of the doublet are indeed *Pf*TFIIB.

This suggests a number of possibilities; the protein is expressed as two isoforms of differing lengths perhaps due to a C-terminal degradation or the premature termination of translation. The doublet may also have arisen due to incomplete denaturation of the protein prior to

loading on the SDS-PAGE, perhaps due to highly stable bonds, or that a post-translational modification event has occurred on the protein.

An immunoblot was performed, making use of anti-sera from rabbits immunised against a peptide, located on the C-most terminal end of the predicted *Pf*TFIIB sequence, (residues 353-367) characterised by Gertrud Talvik, an MSc student in the lab. Should the protein have terminated prematurely, or undergone C-terminal degradation, this region should be missing in the lower band of the doublet, and only the upper band of the doublet should be present. However, as seen in Figure 19 there is reactivity against both bands in the 44kDa region. It is noted that a large quantity of protein was loaded onto this SDS-PAGE, and so the detection of a doublet may simply be due to low-specificity of the anti-body. However, additional work in the laboratory has demonstrated that the anti-body is indeed specific to the heterologous *Pf*TFIIB in much reduced concentrations. Additionally, recent work in the laboratory to express a 6His-*Pf*TFIIB-GST protein has shown similar results, with the consistent expression of a doublet at the expected size (Gertrud Talvik, unpublished observations). This suggests that the doublet is the result of structures resistant to denaturation, or due to post-translational modifications. It may also be suggested that the doublet of protein represent a mixed TFIIB population of closed and open form, as has

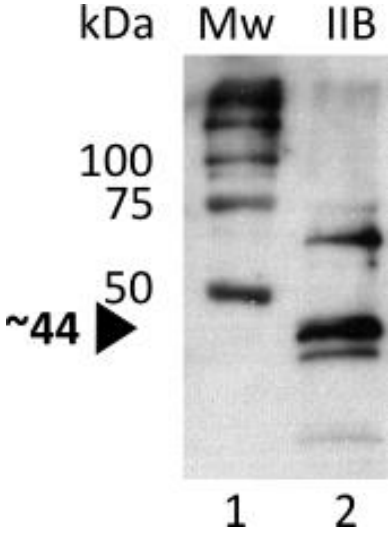



***Figure 19: Immunoblot showing the reactivity of 6His-PfTFIIB to peptide-raised antibody.***

150ng/µl of 6His-*Pf*TFIIB protein preparation (per band of the doublet) was separated on a 5/12% SDS-PAGE and transferred to a PVDF membrane. The membrane was probed with a 1/1000 dilution of anti-*Pf*TFIIB sera, followed by a 1/3000 anti-rabbit goat IgG-coupled peroxidase. Immunoreactivity detected with ECL substrate, and exposure to X-ray film.

Black arrow indicates the doublet at the expected size of 6His-*Pf*TFIIB protein. Molecular weight marker - 4µl of Protein Plus Unstained marker (Bio-Rad).

been described for *Hs*TFIIB (Glossop et al. 2004). This requires additional investigation, and was not pursued further in this project.

### 3.3.5 Further purification of 6His-*Pf*TFIIB

Attempts were made to remove the contaminants present within the 6His-*Pf*TFIIB protein preparation through the use of ion-exchange resins. The predicted pI of 6His-*Pf*TFIIB is 9.81 and the BC-100 buffer the protein eluted in contains a KCl concentration of 100mM, and a pH of 7.8 at 4°C. The majority of the common Ni-NTA affinity preparation contaminating proteins have pIs below 7 (Bolanos-Garcia & Davies 2006). Therefore, the initial purification attempt of the 6His-*Pf*TFIIB protein preparation was performed through batch-purification with the strong cation exchange resin, sulphopropyl- Sepharose® Fast Flow (SP-Sepharose®; *Sigma-Aldrich*®). Proteins were bound at 100mM KCl and eluted with increasing KCl concentrations (Figure 20A). However, the majority of the recombinant *Pf*TFIIB protein as well as the contaminants were not bound to SP-Sepharose® resin. Further attempts to purify 6His-*Pf*TFIIB were conducted with different cation and anion exchange resins. The resins used were: quaternary ammonium-Sepharose® (Q) Fast Flow, a strong anion exchange resin; diethylaminoethanol-Sepharose® (DEAE) Fast Flow, a weak anion exchange; and carboxymethyl-Sepharose® (CM) Fast Flow, a weak cation exchange resin.

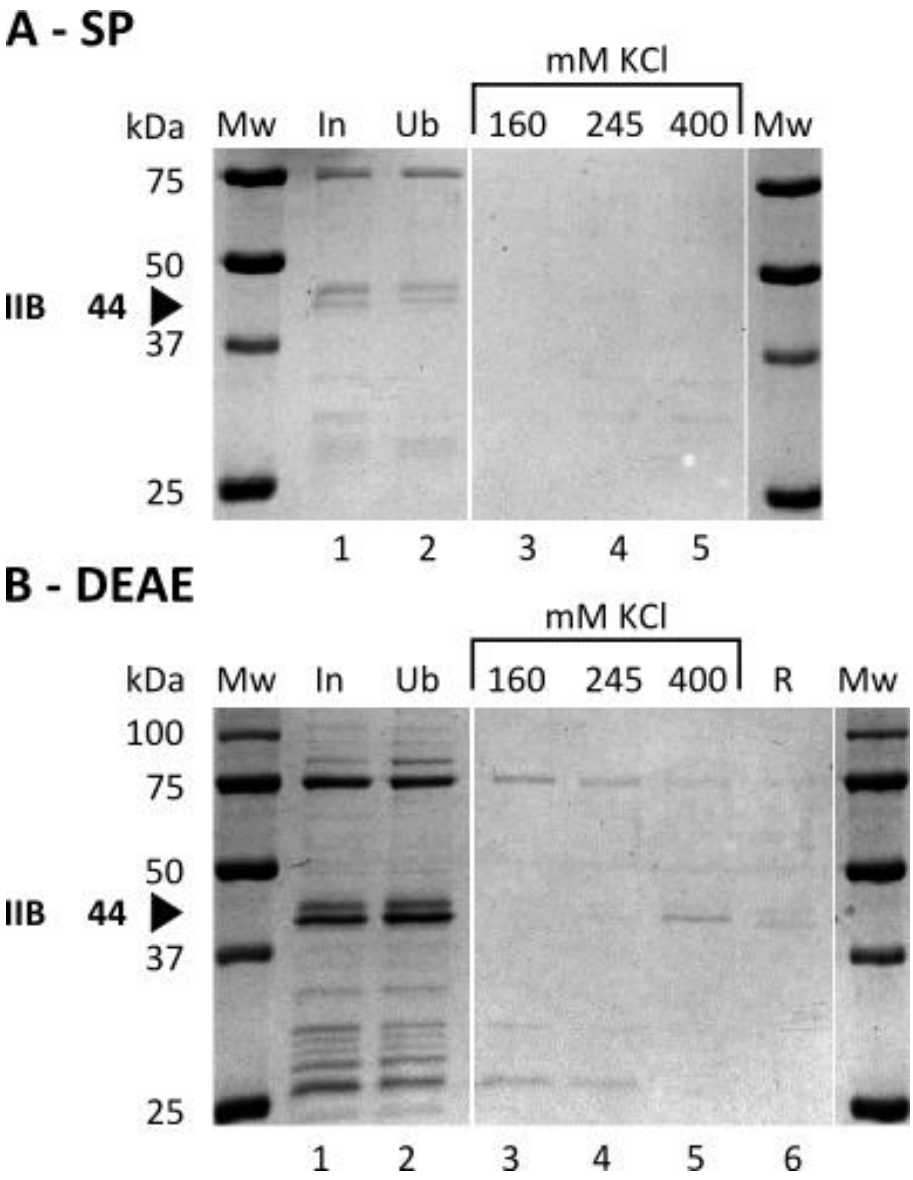



***Figure 20: 6His-PfTFIIB binds partially to weak anion-exchange resin.***

**A.** *Pf*TFIIB and the majority of the contaminants do not bind to SP Sepharose® resin. 100ng of 6His-*Pf*TFIIB was loaded into the input lane (**In**), and an equivalent amount of the protein not bound (unbound, **Ub**). 1.5× of the input is loaded in each of the KCl elution lanes 3-5.
**B.** Partial binding of 6His-*Pf*TFIIB to DEAE resin and partial elution at 400mM KCl. 100ng of 6His-*Pf*TFIIB was loaded into the input lane (**In**), and 1.5× of the equivalent amount of the protein not bound (unbound, **Ub**). 1.5× of the input is loaded in each of the KCl elution lanes 3-5. Lane 6 contains 1/6th of the resin (R), eluted in 2×SDS loading buffer.
Molecular weight marker (**Mw)** - 2µl of Precision Plus Protein All Blue standard (*Bio-Rad*).

A proportion of the 6His-*Pf*TFIIB protein was bound to DEAE- Sepharose® resin and was only partially eluted with high salt concentrations from the resin (see lanes 5 and 6 of Figure 20B). The majority of the protein, however, was still uncaptured by the resin. Only limited binding of contaminant proteins to Q-Sepharose® was seen (not shown). No binding of

*Pf*TFIIB or contaminant proteins to CM-Sepharose® was seen (not shown). The partial binding of *Pf*TFIIB to DEAE resin suggests heterogeneity in the protein preparation. It is possible that some proportion of the protein may have aggregated. The protein contains regions which are predicted to be negatively charged predicted with pIs of less than 4.5 (B-linker and reader), as well as regions predicted to be positively charged, with predicted pIs of over 10. These may interact preferentially with one another over the binding of DNA.

It is known that DNA is strongly bound by the weak anion exchange resin, DEAE-Sepharose®. It was suggested that *Pf*TFIIB may bind weakly to DNA, as it is known to associate in other TFIIB homologues, albeit in the presence of TBP. Additionally, the calculated pI of *Pf*TFIIB (9.73), with several regions of pI above 10 (Table 4), suggests the protein may be highly positively charged, and therefore may bind DNA.

Aliquots of the 6His-*Pf*TFIIB preparation were treated with a nuclease enzyme (Benzonase, *Sigma-Aldrich*®). The nuclease enzyme digests nucleic acids contaminating the protein preparation. Should 6His-*Pf*TFIIB bind to DNA, which in turn is bound to the DEAE resin, then *Pf*TFIIB in the treated protein preparation should no longer bind to the DEAE resin.

In Figure 21B and C, the results of attempting to bind nuclease treated protein preparations to the various Sepharose® resins is shown. It was found that the protein was consistently recovered in the unbound protein fraction of each experiment. Proteins were eluted with 500mM final KCl, similarly to previously described experiments. Notably, this is true also for the experiment with DEAE resin. Note that panel A in the figure is a positive control. The recombinant HsTFIIB protein is known to bind to strong cation exchangers, such as SP-Sepharose®, and was used to confirm that the resin was functional. This strongly suggests that recombinant *Pf*TFIIB was able to interact indirectly to DEAE-Sepharose® due to the presence of DNA.

Due to time constraints, further attempts for the purification of 6His-*Pf*TFIIB were not undertaken.

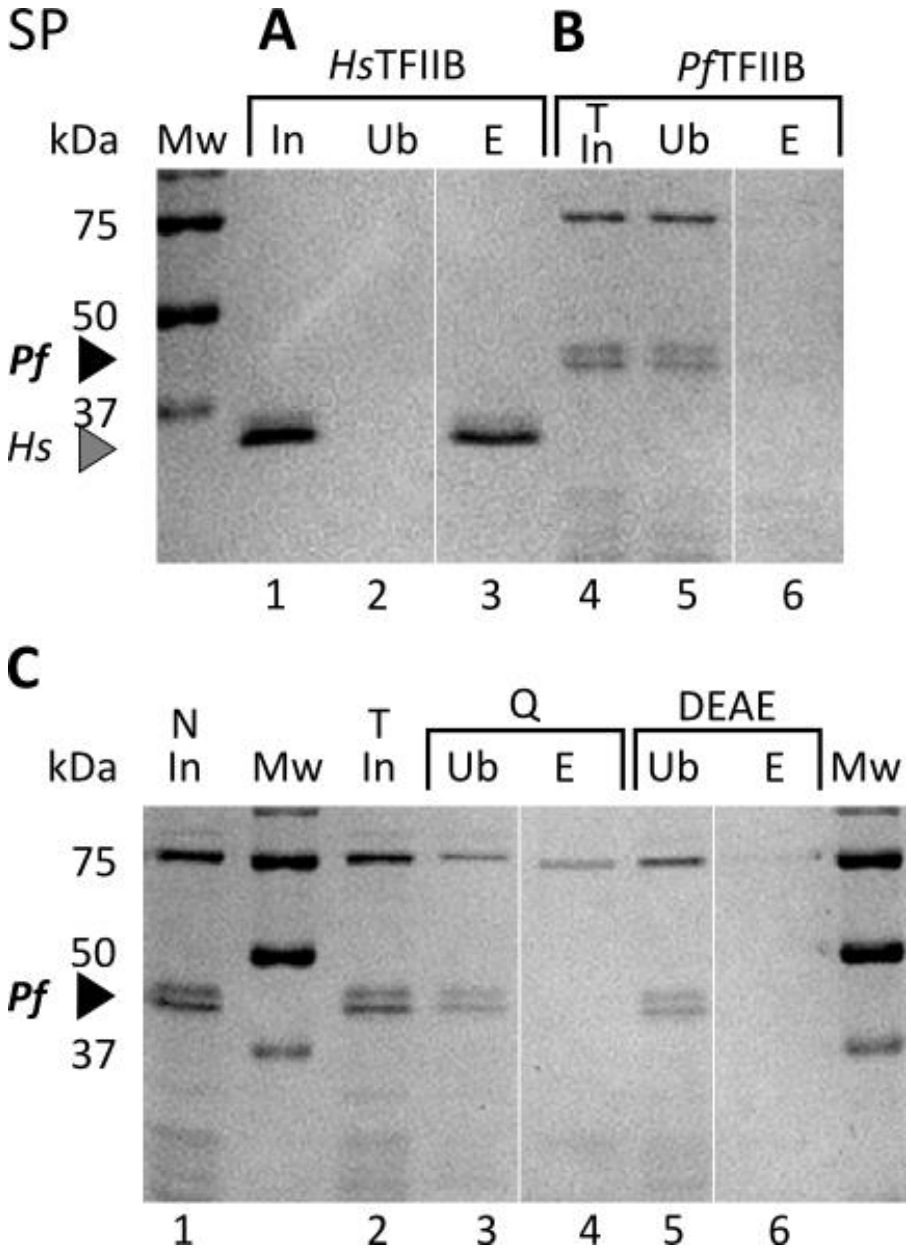



***Figure 21: Nuclease treated 6His-PfTFIIB does not bind to ion-exchange resins.***

**A**. Positive control of 200ng of recombinant 6His-*Hs*TFIIB, previously purified using Mono S Sepharose®. Eluted in 500mM KCl.
**B.** Nuclease treated *Pf*TFIIB protein does not bind to the SP resin.
**C**. Nuclease treated *Pf*TFIIB protein does not bind to either Q or DEAE Sepharose® resins.
Lane 1 is the preparation prior to nuclease treatment, and lane 2 is post-treatment, used with the resins.
Unbound protein (**Ub**) is equivalent of 50% the input (**In**) protein (100ng), hence the reduction in signal. Elution (**E)** with 500mM KCl, is equivalent to 63% of the input loaded. Treated input (**TIn**) is equivalent to **In**. Molecular weight marker (**Mw)** - 2µl of Precision Plus Protein All Blue standard (*Bio-Rad*).

### 3.3.1 Summary of *Pf*TFIIB expression and purification

A number of interesting features of *Pf*TFIIB protein were found. Firstly, *Pf*TFIIB may be efficiently expressed through a method of 'auto-induction', by modification of the cell culture growth media. This is useful not only due to the ease of expression, but also because this method is cost-effective as there is no need for IPTG-induction.

The purification of *Pf*TFIIB shows the presence of a doublet. This may be related to the open and closed conformations that been studied in other TFIIB orthologous, (Elsby & Roberts 2004; Zheng et al. 2004; Glossop et al. 2004; Fairley et al. 2002) and may be implicated in transcription. This is supported by the conservation of the charged cluster domain, described previously (Chapter 3).

Due to the inability to purify 6His-*Pf*TFIIB further, it was decided that the research group would re-clone the putative *Pf*TFIIB protein into an expression vector containing two flanking tags, a 6His-tag, and a GST-tag, to allow for a more stringent purification (Gertrud Talvik, current investigation).

Finally, it is intriguing that evidence from the binding of 6His-*Pf*TFIIB to DEAE resin, and the subsequent abolishment of this upon treatment with nuclease, suggests that *Pf*TFIIB may bind to DNA in a manner not yet described for the protein. Current work in the research group appears to support DNA-binding activity in recombinant *Pf*TFIIB (Gertrud Talvik, unpublished observations).

## 3.4 Expression of recombinant *Pf*TLP protein

### 3.4.1 Cloning of *Pf*TLP

Isolation of *Pf*TLP open reading frame from cDNA libraries

The open reading frame for *Plasmodium* TLP was PCR amplified by using 2µl of each of the cDNA libraries MRA-296 and 297 (described in Chapter 2.3.2), and analysed on a 0.8% TBE/EtBr agarose gel (Figure 22). The MRA 296-299 cDNA phage libraries are derived from mRNA isolated from the blood stage cultures of *Plasmodium falciparum* clone 3D7 extracted from parasites 16, 26, 36 hours post-synchronisation, as well as asynchronous parasites, respectively (Chakrabarti et al. 1994). These correspond to parasites in late ring-stage, late trophozoite stage, schizont stage, as well as asynchronous parasites, respectively. It was found from the PCR analysis that mRNA for *Pf*TLP was most prevalent at 16 hours post-infection (late-ring stage), which corresponds with microarray expression data obtained by the DeRisi lab (Bozdech, Zhu, et al. 2003), hosted on the PlasmoDB database (Aurrecoechea et al. 2009).

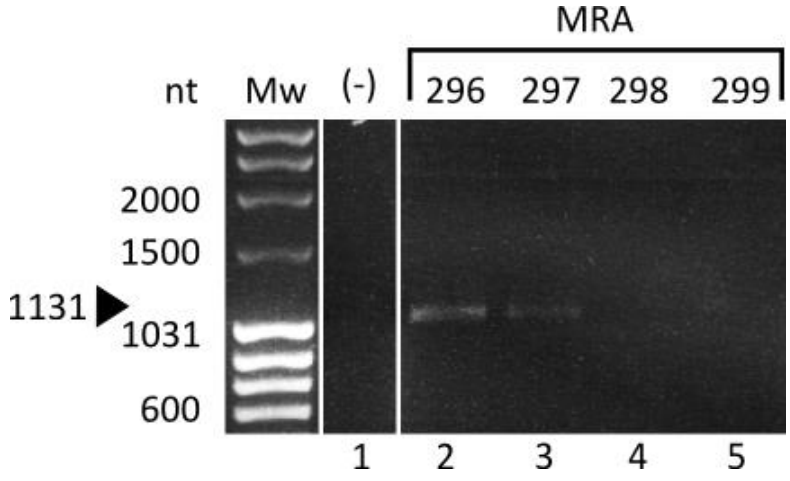



***Figure 22: PCR amplification of the PfTLP open reading frame from MRA cDNA libraries.***

Amplification of the *Pf*TLP gene from the indicated MRA cDNA libraries was analysed on a 0.8% agarose gel. Lane 1 – no cDNA control (-). Lanes 2-5: 1/10 of PCR reactions with MRA libraries 286-299 were loaded. Expected size is 1131bp, indicated by the black arrow. Molecular weight marker (**Mw**) - 10µl of MassRuler™ DNA Ladder Mix (*Thermo Scientific*).

Construction of the pET11d-6His-*Pf*TLP vector

The *PfTLP* gene was then digested and ligated into the pET11d-6His vector frame, as described in Chapter 2.3, to make the pET11d-6His-*Pf*TLP vector.

### 3.4.2 IPTG-induced expression of recombinant *Plasmodium* TLP leads to a reduction in cell growth in *E. coli*.

Similarly to the initial characterisation of recombinant *Pf*TFIIB expression, IPTG-induced expression of 6His-*Pf*TLP was performed in order to assess the effects of heterologous gene expression on the *E. coli* host cell, and to set a basis for future expression trials.

*E. coli* BL21-codonPlus® (DE3)-RIL cell cultures transformed with the pET11d-6His-*Pf*TLP vector, were grown in liquid media at 37°C. The density of the cell

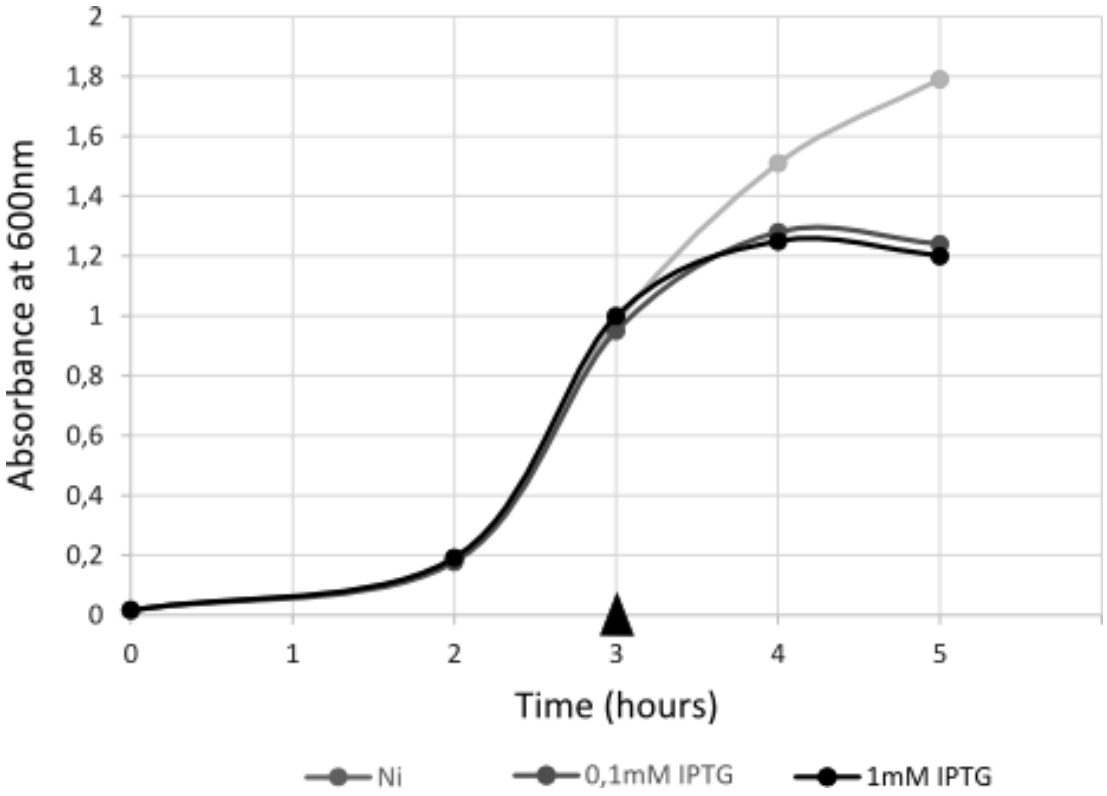



***Figure 23: IPTG-induced expression of 6His-PfTLP on BL21-codonPlus® (DE3)-RIL E. coli leads to growth inhibition.***

Cell densities of cultures, measured at wavelength 600nm, were plotted against the number of hours post-inoculation of expression culture from overnight cultures. Growth of the cultures induced with 1mM and 0.1mM IPTG (black and dark grey, respectively) is compared to a non-induced culture (light grey). The time of induction with IPTG is indicated by a black arrow.

cultures were monitored by light spectrophotometry (wavelength of 600nm). At an $OD_{600nm}$ of approximately 1, the cultures were induced with 0.1mM, or 1mM IPTG or left growing without addition of IPTG. Induction of expression with IPTG was found to have a negative effect on cell growth (Figure 23).

The protein expression from these experiments were analysed by Ni-NTA affinity chromatography of cleared cell lysates. In these experiments, no soluble 6His-*Pf*TLP could be detected (not shown).

Similarly to the expression of *Pf*TFIIB, it was decided to include glucose in the media used to grow *E. coli* cells containing expression vectors for the *Plasmodium* transcription factors. Notably, in an initial transformation with the pET11d-6His-*Pf*TLP vector making use of LB plates which lacked glucose, very few transformants were obtained ($5\times10^{-2}$ cfu/µg of pDNA). Including 1% glucose in the plates increased the transformation efficiency significantly ($2.6\times10^{3}$ cfu/µg pDNA). Such a dramatic increase in transformation efficiency was not seen in the transformation with the recombinant *Pf*TFIIB vector, suggesting that the expression of *Pf*TLP is far more toxic to the *E. coli* host cells than *Pf*TFIIB.

Attempts to express 6His-*Pf*TLP in a similar manner to 6His-*Pf*TFIIB (Section 3.3.3), through the use of glucose supplemented media and auto-induction of expression through the release of catabolite repression did not yield measurable protein expression (not shown). It is known that often recombinant proteins may be expressed and aggregate in the form of inclusion bodies which are insoluble and not detectable from the soluble protein fraction of the cells (Mehlin et al. 2006). This was thought to be especially possible due to the presence of presumptive low-complexity regions in the *Pf*TLP protein. However, making use of purification methods for insoluble proteins, 8M urea denaturing conditions, did not yield detectable protein expression (data not shown).

Cold-shock induction of 6His-*Pf*TLP protein expression

Expressing proteins at low temperatures (below 30°C) in *E. coli* may be beneficial in expressing recombinant proteins (Qing et al. 2004). The rapid reduction in cell temperature (cold-shock), as well as continued growth at low temperatures, may lead to the increased expression and activity of cold-shock chaperone proteins. The expression of the recombinant protein is lowered at low temperatures, but this may lead to a higher ratio of soluble to insoluble protein, as the reduced cellular concentration of protein favours folding, and reduces protein aggregation (Baneyx 1999).

In this experiment, cell cultures, supplemented with glucose, were grown to mid-log phase at 37°C. One of the cultures was cold-shocked to 25°C, induced with 0.5mM IPTG, and maintained at 25°C. The second cell culture was not cold-shocked, induced with 0.5mM IPTG, and maintained at 37°C (Figure 24A). After three hours the cells were harvested. The expression of 6His-*Pf*TLP protein was analysed by affinity purification from the cleared cell lysate with Ni-beads and SDS-PAGE (Figure 24B).

It was found that recombinant *Pf*TLP was preferentially expressed after cold-shock induction, as compared to the cell culture maintained at 37°C. Additionally, although both cultures decreased their rate of cell growth, the cell culture maintained at 25°C has a similar growth reduction to the cell culture maintained at 37°C. This suggests that the colder conditions had a protective effect on the cells after induction. Had the cells experienced a similar toxic effect from *Pf*TLP expression as the cells induced at 37°C, it would be expected that the reduction in cell growth of the 25°C cell culture would be far greater. No protein could be purified under denaturing conditions.

Mass spectrometry analysis of the sample was performed, by extracting bands of protein from an SDS-PAGE gel with the same sample. Peptide analysis of a band doublet seen at 44kDa (Figure 24B) confirmed that the top band is 6His-*Pf*TLP.

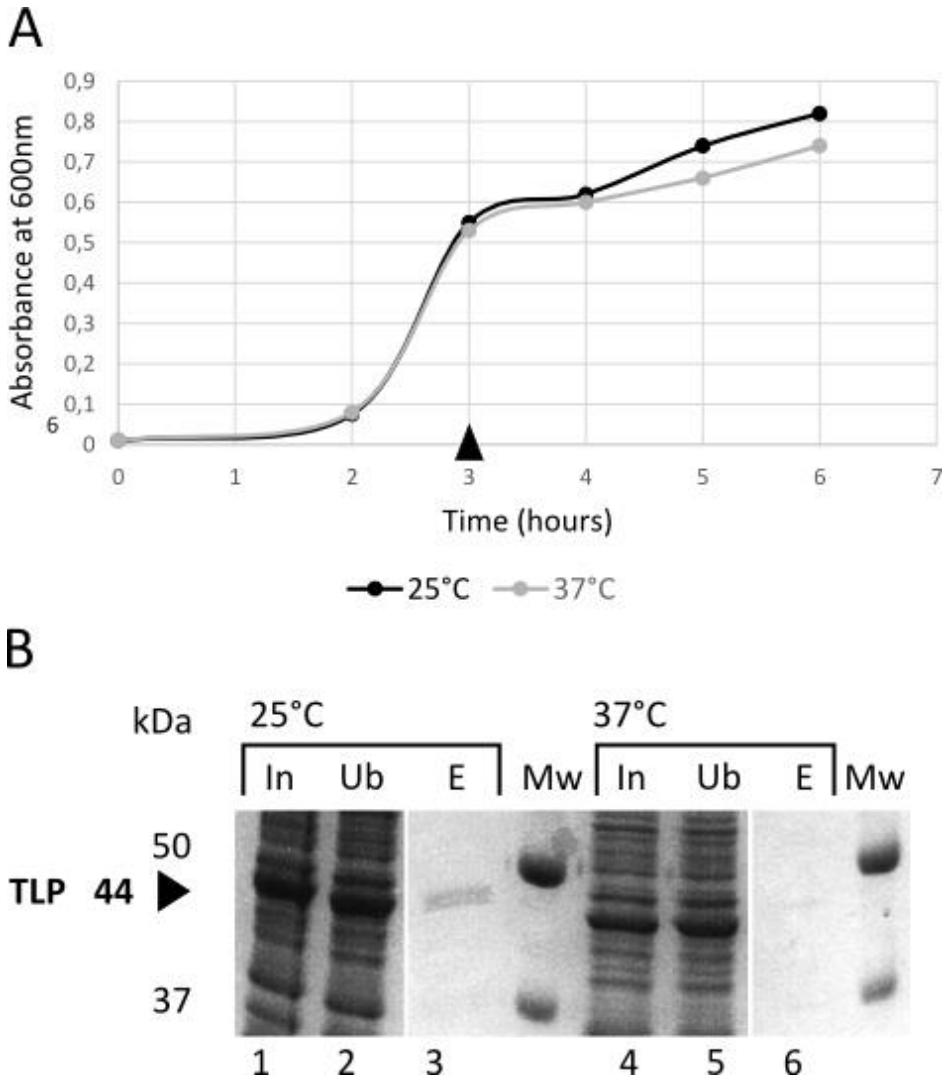



***Figure 24: 6His-PfTLP protein is expressed in BL21-codonPlus® (DE3)-RIL E. coli through cold-shock IPTG-induction.***

**A**. Cell densities of cultures, measured at wavelength 600nm, were plotted against the number of hours post-inoculation of expression culture from overnight cultures. Growth of the culture cold-shocked and grown at 25°C (black) is compared to a culture maintained at 37°C (grey). The time of cold-shock and induction with IPTG is indicated by a black arrow.
**B.** Protein was purified by nickel-affinity purification from cleared cell lysates and analysed by SDS-PAGE. Lanes 1-3 cell culture cold-shocked and grown at 25°C. Lanes 4-5 cell culture maintained at 37°C. Equivalent amounts of cleared lysate input (**In**; 1/100 vol.) and the protein fraction not bound to the Ni-beads (unbound, **Ub**), and the protein fraction eluted from the Ni-beads with 2×SDS-loading buffer (**E**, ½ vol.) were loaded. A black arrow indicates the expected size of 6His-*Pf*TFIIB protein at 44kDa. Molecular weight marker (**Mw**) - 5µl of Precision Plus All Blue standards (*Bio-Rad*).

### 3.4.3 Expression of GST-6His-*Pf*TLP

Recombinant 6His-*Pf*TFIIB protein was purified through metal-affinity purification via a single hexahistidine purification tag. As shown previously, this was insufficient to achieve high purity, discussed in Chapter 3.3.1. Purification of 6His-*Pf*TLP was expected to be similarly challenging. In addition, the toxicity of *Pf*TLP recombinant protein expression resulted in low protein expression, making it logistically difficult to grow enough cell mass to purify sufficient protein for future studies. It is known that the addition of a soluble, easily expressed protein tag to a difficult to express protein often increases the yield of the fusion-protein (Sørensen & Mortensen 2005; Smith & Johnson 1988). To this end, a GST-tag was inserted into the pET11d-6His expression vector to form a pET11d-GST-6His vector (Figure 43 and Figure 47). The putative *Pf*TLP gene was then cloned into this vector, to form the pET11d-GST-6His-*Pf*TLP vector (Figure 46).

Cold-shock induced expression of GST-6His *Pf*TLP

BL21-CodonPlus® (DE3)-RIL *E. coli* cells were transformed with either the pET11d-GST-6His-*Pf*TLP or the empty pET11d-GST-6His vector. Three cell cultures were grown, two expressing GST-6His-*Pf*TLP and another expressing GST-6His. This was done to compare the effect of expressing GST-6His to GST-6His-*Pf*TLP protein, and so determine if the addition of the GST-tag would lower the toxicity of recombinant *Pf*TLP.

The cell cultures were inoculated from overnight cultures, and then grown at 37°C, and the cell culture densities monitored spectrophotometrically. When the cell cultures had an approximate $OD_{600nm}$ of 1, all of the cell-cultures were cold-shocked to 25°C and maintained at this temperature for three hours before harvesting (Figure 25). The cell culture transformed with the pET11d-GST-6His vector was induced with 0.5mM IPTG, as was one

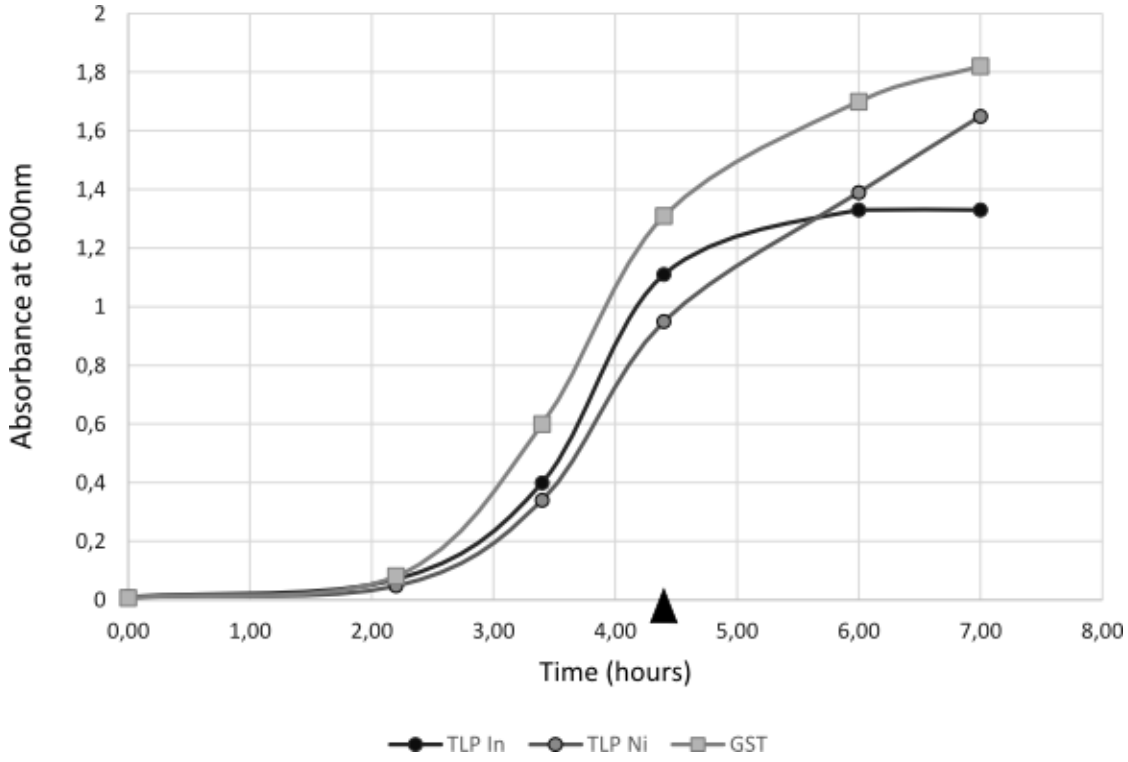



***Figure 25: IPTG-induced expression of GST-6His and GST-6His-PfTLP in E. coli BL21-CodonPlus-RIL cells leads to growth inhibition.***

Cell densities of cultures, measured at wavelength 600nm, were plotted against the number of hours post-inoculation of expression culture from overnight cultures. Growth of the induced (**In**, black) and non-induced (**Ni**, grey) cultures expressing GST-6His-*Pf*TLP are compared to induced GST-6His expressing culture (GST, light grey with squares). Time of cold-shock and induction with 0.5mM IPTG is indicated by a black arrow.

of the GST-6His-*Pf*TLP cell cultures. There was an immediate decline in growth of all of the cell cultures, with the most dramatic decline in growth seen in the induced GST-6His-*Pf*TLP culture.

The decline in cell growth of the cultures is more clearly illustrated in Figure 26**,** where the rate of cell growth over time ($\frac{dOD}{dt}$) is plotted against time. Prior to induction, the GST-6His-*Pf*TLP cell cultures were growing at 0.6-0.7 $OD_{600nm}$ /hour. After the cold-shock, the non-induced *Pf*TLP cell culture growth rate was reduced, but at a much lesser extent than the induced culture, which ceased growth after approximately 1.5 hours. The GST-6His expressing cells also showed a reduction in growth after cold-shock and induction, but not

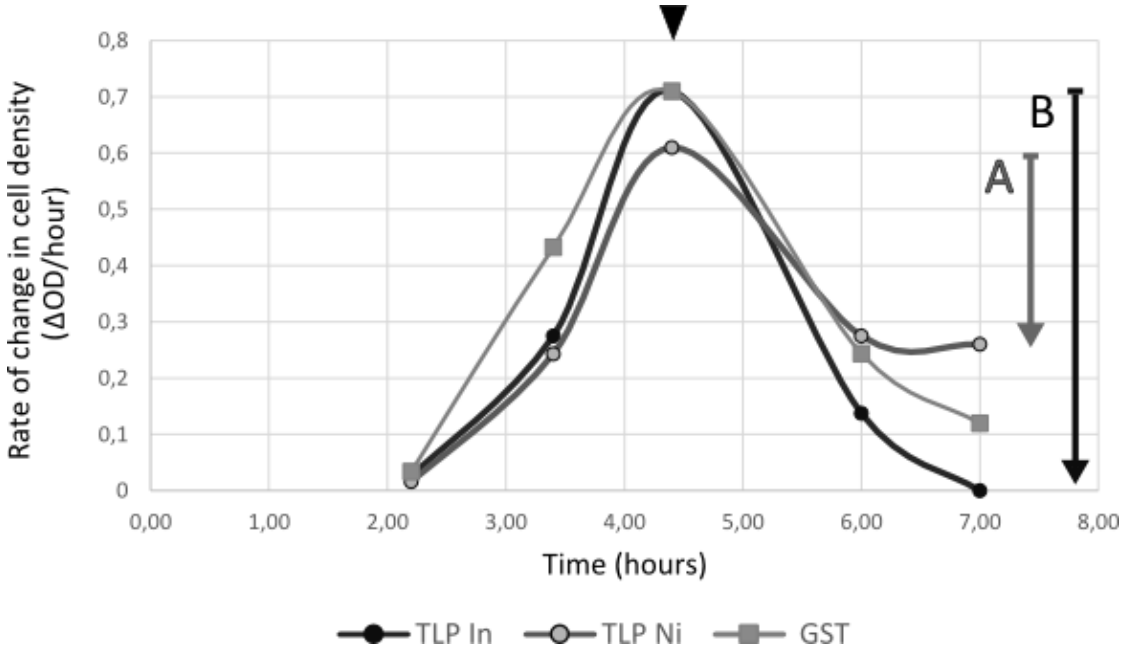



***Figure 26: IPTG-induced expression of GST-6His-PfTLP in E. coli BL21-CodonPlus-RIL cells reduces growth rate.***

Change in the rate of cell growth, calculated as the change in cell density (measured at wavelength 600nm) over time, is plotted the number of hours post-inoculation of expression culture from overnight cultures. The growth rate of the induced and non-induced GST-6His-*Pf*TLP expressing culture (TLP In, black; and TLP Ni, dark grey, respectively) are compared to induced GST-6His expressing culture (GST, light grey with squares). Black and grey arrows on the right indicate the reduction in change of growth of the cultures of the non-induced GST-6His-*Pf*TLP cells (A) and the induced cells (B) respectively. Time of cold-shock and induction with 0.5mM IPTG is indicated by a black arrow.

to the same degree as the expression-induced GST-6His-*Pf*TLP cells. This reduction in growth be attributed to metabolic stress of expressing both T7 polymerase as well as the GST-6His peptide.

The expression of protein was investigated through nickel-magnetic bead affinity purification and SDS-PAGE, as described previously (Figure 27). No expression of GST-6His-*Pf*TLP protein could be detected from either the induced or non-induced cell cultures. In contrast, the expression of GST-6His was successful. It should be noted that the use of the Ni-magnetic beads was not able to fully deplete the cleared cell lysate of the expressed GST-6His tag, determined from the presence of protein in both the input and unbound samples of the GST-6His peptide purification (compare lanes 4, 5 and 6 in Figure 27). This

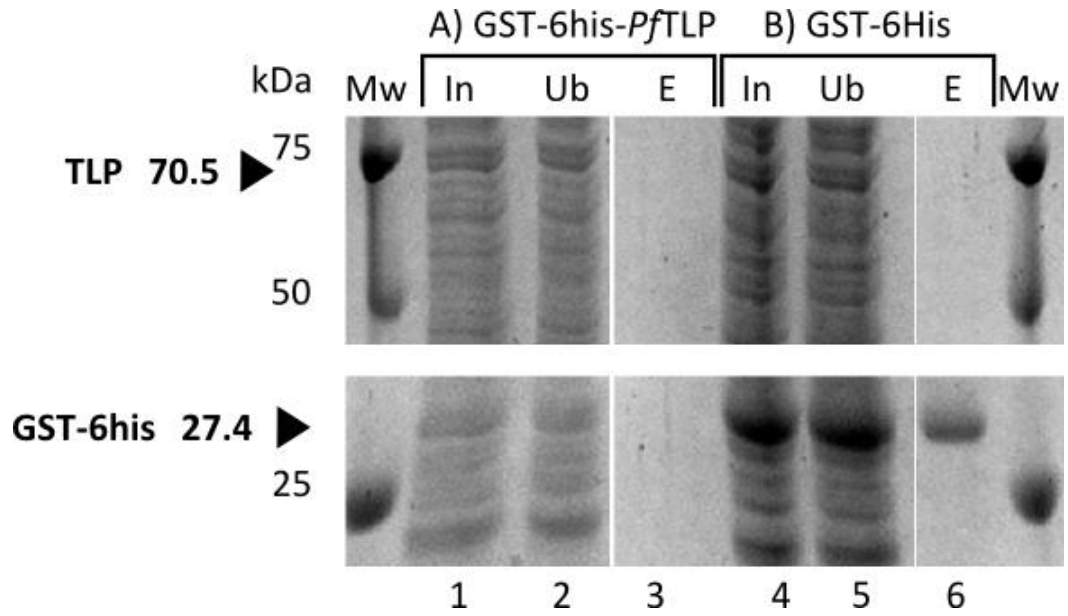


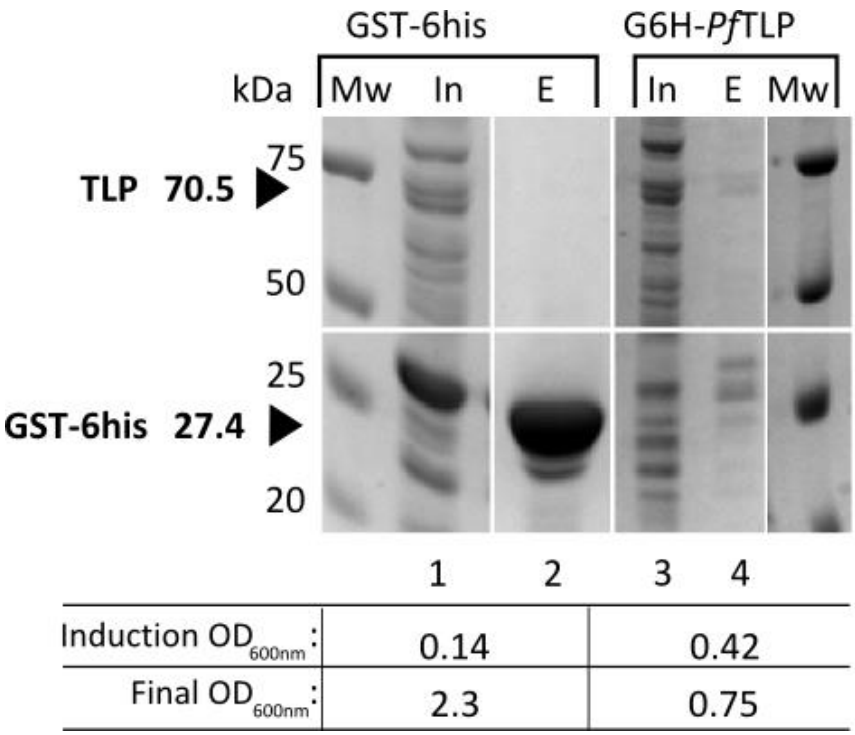

***Figure 27: Cold-shock IPTG-induction does not lead to detectable expression of GST-6His-PfTLP in BL21-codonPlus® (DE3)-RIL E. coli cells.***

Protein was purified by nickel-affinity purification from cleared cell lysates and analysed by SDS-PAGE. **A**. GST-6His-*Pf*TLP expression induced cell culture, calculated molecular weight 70.5kda. **B**. GST-6His expression induced cell culture, calculated molecular weight 27.4kDa. Equivalent amounts of cleared lysate input (**In**, 7.5/1000 volume), and the protein fraction not bound to the Ni-beads (unbound, **Ub**), and the protein fraction eluted from the Ni-beads with 2×SDS-loading buffer **(E**, 1/3 vol.) were loaded. Molecular weight marker (Mw) - 5µl of Precision Plus Protein™ All Blue Standard (*Bio-Rad*).

prompted the usage of glutathione-agarose resin, to which the GST-tag has strong affinity, for further affinity-purification experiments.

Extended expression of G6H*Pf*TLP at low temperature conditions

It was reasoned that at lower temperatures, the rate of protein production is substantially slowed. As GST-6His-PfTLP is considerably longer than 6His-PfTLP, additional time may be required for the expression of the fusion-protein. Therefore, a similar experiment was attempted, making use of a lower expression temperature, but with a much longer time to express. In this experiment, cells transformed with either pET11d-GST-6His or pET11d-GST-6His-PfTLP vectors were grown to a low optical density, cold-shocked to 20°C, induced with 0.5mM IPTG, and then incubated for 14.5 hours at 20°C. It was found that the growth of cells expressing G6H-PfTLP was severely restricted, as compared to the cells




| | | |
|---|---|---|
| Induction $OD_{600nm}$: | 0.14 | 0.42 |
| Final $OD_{600nm}$: | 2.3 | 0.75 |

***Figure 28: Cold-shock IPTG-induction leads to low expression of soluble GST-6His-PfTLP in BL21-codonPlus® (DE3)-RIL E. coli cells.***

Protein was purified by glutathione-affinity purification from cleared cell lysates and analysed by SDS-PAGE. Lanes 1-2 IPTG-induced GST-6His expression, expected molecular weight 27.4kDa. Lanes 3-4 IPTG-induced GST-6His-*Pf*TLP expression, expected molecular weight 70.5kDa. Equivalent amounts of cleared lysate input (**In**; 7.5/1000 vol.) and the protein fraction eluted from the glutathione-agarose resin with 2×SDS-loading buffer (**E**, 1/3 vol.) were loaded. Molecular weight marker (**Mw**) - 5µl of Precision Plus Protein™ All Blue Standard (*Bio-Rad*).

expressing the GST-6His peptide, suggesting the continued toxicity of the expressed protein. The protein expression was then investigated using affinity purification of the cleared cell lysate with glutathione-agarose resin (Sigma-Aldrich®), and SDS-PAGE analysis (Figure 28). Not induced control cultures did not express protein (data not shown). An extremely large quantity of the GST-6His peptide was expressed, lane 2, however, very little GST-6His-*Pf*TLP was expressed, lane 5. Additionally, the presence of several contaminants at approximately the size of the GST-6His tag were enriched by purification of the protein from the GST-6His-*Pf*TLP expressing cells.

Conclusion of the expression of GST-6His-*Pf*TLP

It was found that the use of cold-shock induction had a positive effect on protein expression, as seen with 6His-*Pf*TLP expression. However, the addition of a GST-tag did not reduce the toxicity of *Pf*TLP expression to the host cell.

As with many *Plasmodium* genes, *PfTLP* contains codons which are not commonly used for expression by the *E. coli* host (Birkholtz et al. 2008a). It is thought that the usage of these codons in quick succession results in metabolic strain on the host cell to generate sufficient tRNAs, which in turn leads to ribosomal stalling.

In the *PfTLP* gene, 26 codons are reportedly rare in *E. coli* (Zhang et al. 1991, the Appendix 5.7). Of these rare codons, two are the CCC codon coding for proline. The rest of the codons are predominantly AUA, coding for isoleucine (18/26), and the rest code for arginine and leucine. The recombinant *PfTLP* gene has been expressed in BL21-CodonPlus® (DE3)-RIL *E. coli* cells, which posses additional tRNAs gene for these rare codons (*argU* [AGA, AGG], *ileY* [AUA], *leuW* [CUA]; Stratagene 2005), which are meant to alleviate the effects of rare codons in heterologous gene expression. However, it has been reported that expression of proteins in BL21-CodonPlus® (DE3)-RIL *E. coli* cells, where more than 5% of the amino acids in the heterologous protein are supplied by the RIL strain, results in growth retardation of the cells, low cell yield, and insoluble protein (Rosano & Ceccarelli 2009). The rare codons in the *Pf*TLP make up a little over 6.5% of the gene sequence, and thus are likely to be responsible for the low protein yield.

As has been shown in Chapter 3.2, the *Pf*TLP protein also has stretches of low amino acid sequence complexity. *Plasmodium p*roteins with LCRs are reportedly difficult to express in *E. coli*, most likely due to the over-utilisation of a small subset of tRNAs (Birkholtz et al. 2008a).

### 3.4.4 Expression of codon optimised GST-6His-*Pf*TLP

In order to investigate whether the sequence of the gene is the predominant cause of toxicity in heterologous expression, the *Pf*TLP gene codon was optimised for expression in *E. coli*.

Codon optimisation of *Pf*TLP

The *PfTLP* gene was codon optimised by GenScript, according to their OptimumGene™ Codon Optimization Analysis (http://www.genscript.com/). Codon optimisation consisted of adjusting the frequency of codon usage across the gene sequence, including the replacement of all rare codons with those more frequently used by *E. coli*. Additionally, runs of identical codons were replaced, to prevent over-utilisation of the related tRNAs. The G+C content of the gene was raised from 22% to 38%, to favour expression in *E. coli*. Additionally, runs of polyT and polyA (TTTTTT and AAAAAA) were removed, as these may form cryptic translational termination sites. Finally, an internal *BamHI* site was removed, for ease of cloning. A comparison of the codon optimised ORF and the original ORF of the *PfTLP* the gene are shown in the Appendix 5.7.

Construction of the pET11d-GST-6His-*Pf*TLPco vector

The codon optimised *PfTLP* gene (*PfTLPco*) was PCR amplified from the pUC57 vector supplied by *GenScript*, and cloned into the pET11d-GST-6His vector (Figure 46). In the first attempt to clone the pET11d-GST-6His-*Pf*TLPco vector, an insertion mutation occurred, and two additional codons were inserted in front of the initial methionine residue (pET11d-GST-6His-*Pf*TLPco(aa), see Table 10 in the Appendix). Expression of this mutated gene was not expected to differ substantially from the expression of the non-mutated gene, and so this vector was used to determine expression conditions while a vector containing a non-mutated *Pf*TLP insert was constructed.

Expression of GST-6His-*Pf*TLPco protein

BL21-CodonPlus® (DE3)-RIL competent cells were transformed with the pET11d-GST-6His-*Pf*TLPco(aa) vector. The cells were grown at 37°C to a cell density of $OD_{600nm}$ ~1.4, before cold-shocking to 25°C and inducing protein expression with 0.5mM IPTG. The cell cultures were then incubated at 25°C for 20 hours. It should be noted that the cells grew much more rapidly at 37°C than cells transformed with the original pET11d-GST-6His-*Pf*TLP vector (compare Figure 25 and Figure 29A). Additionally, the cells induced with IPTG did not have a reduction in cell growth. This stands in contrast to both the non-induced culture, as well as to cells expressing the non-codon optimised gene, suggesting that the toxicity of *Pf*TLP lies in the difficulty of translating the mRNA transcript.

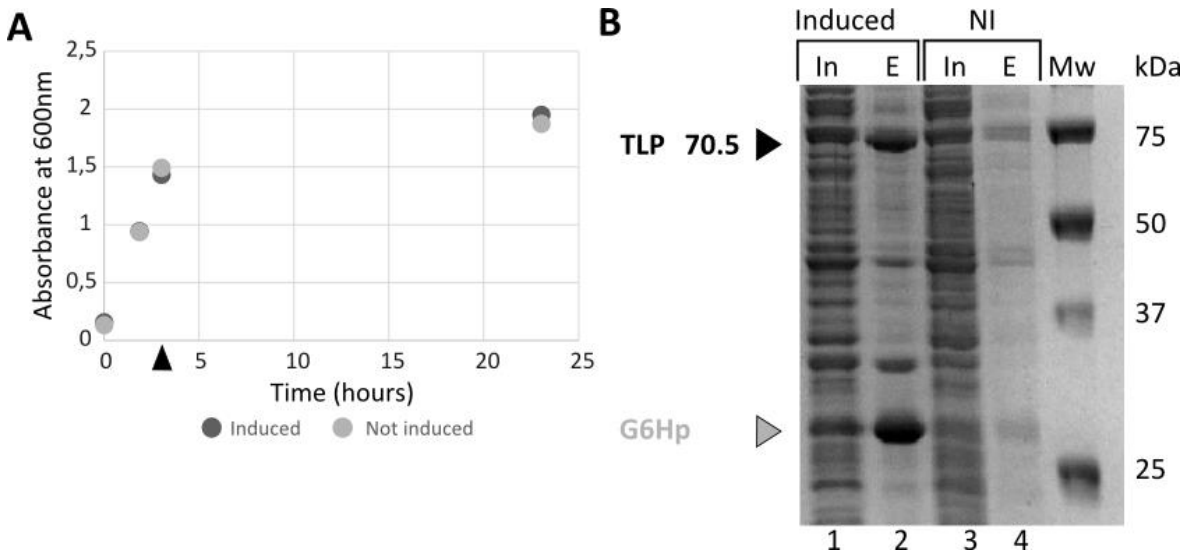



***Figure 29: Expression of codon-optimised GST-6His-PfTLP through cold-shock induction in BL21-CodonPlus® (DE3)-RIL E. coli cells.***

**A.** Cell densities of cell cultures carrying the expression vector pET11d-GST-6His-*Pf*TLPco(aa), measured at wavelength 600nm, were plotted against the number of hours post-inoculation of expression culture from overnight cultures. Cultures were grown at 37°C and then cold shocked to 25°C. Cells induced with 0.5mM IPTG are in black, non-induced cells in grey. Black arrow indicates the time of cold-shock and induction. **B.** Protein was purified by glutathione-affinity purification from cleared cell lysates and analysed by SDS-PAGE. Lanes 1-2 induced GST-6His-*Pf*TLPco protein expression. Lanes 3-4 Non-induced (**Ni**) protein expression. Equivalent amounts of cleared lysate input (**In**; 7.5/1000 vol.) and the protein fraction not bound to the Ni-beads (unbound, **Ub**), and the protein fraction eluted from the glutathione-agarose resin with 2×SDS-loading buffer (**E**, 1/3 vol. loaded) were loaded. Molecular weight marker (**Mw**) - 2µl of Precision Plus Protein™ All Blue Standard (*Bio-Rad*).

The expression of G6H*Pf*TLPco protein was analysed using affinity purification of the cleared cell lysate by glutathione-agarose resin, and SDS-PAGE (Figure 29B). GST-6His-*Pf*TLP protein was successfully expressed in the induced cell culture, and not in the non-induced cell culture. Similarly to the expression from the pET11d-GST-6His vector, there appears to be enrichment of protein at approximately the size of the GST-tag. It is thought that the protein in this region is truncated peptide, due to premature translational termination. Confirmation that this protein contains a 6His-tag, and so is a truncated GST-6His-*Pf*TLP peptide (G6Hp), is discussed below.

Expression of the GST-6His-*Pf*TLP protein from the pET11d-GST-6His-*Pf*TLPco vector was found to be the same as from the pET11d-GST-6His-*Pf*TLPco(aa) vector (not shown). In the expression from this vector, G6Hp was also present. The size of the peptide was

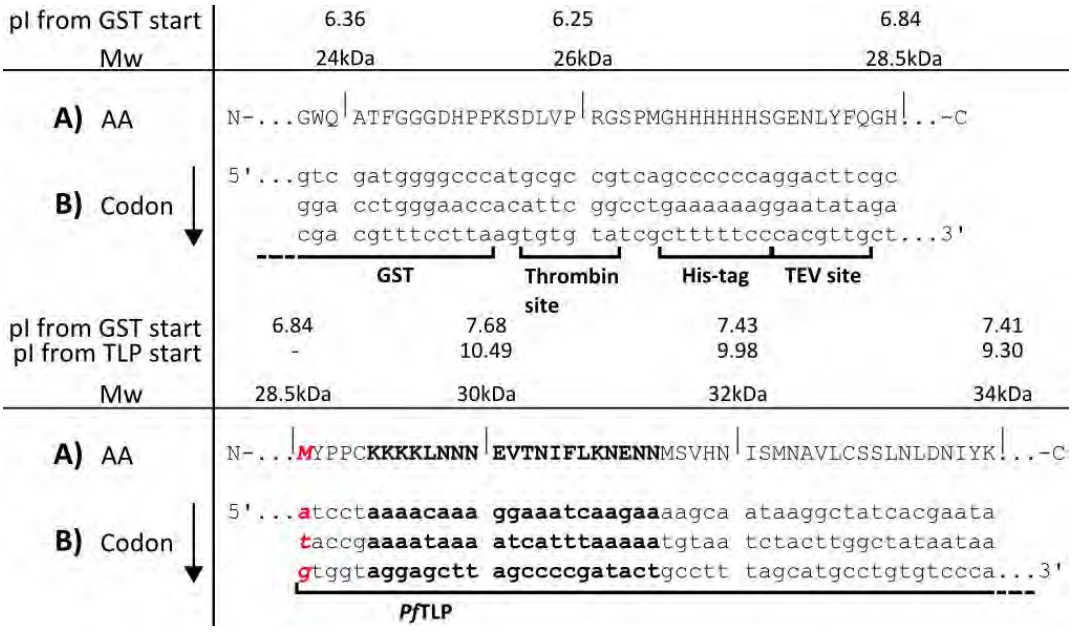



***Figure 30: Examination of the proposed region for premature translational termination.***

Sequences A and B are the amino acid and the codon coding for the amino acid respectively. Note that codons are written vertically, direction indicated by the arrow. Features of the expressed gene are underlined and indicated. The increasing molecular weight of the protein is indicated above the amino acid sequence in kDa. The calculated pIs of the growing polypeptide chain, measured from the beginning of the GST-Tag, and from the start of the *Pf*TLP peptide are indicated above the sequence. The start codon of *Pf*TLP is indicated in red. The region thought to be implicated in premature termination is indicated in **bold.**

estimated by using the molecular weight marker to generate a standard curve, and then inferring the size of the peptide to be 29 - 32kDa. The region in which the peptide may be truncated is mapped onto the fusion-protein sequence in Figure 30.

G6Hp was determined to be greater than 28kDa, and includes both the thrombin and TEV sites, as well as the GST and 6His-tags.

A large scale expression study investigating recombinant *Plasmodium* proteins expression has found that proteins with high pI (>6) are difficult to express in *E. coli* (Mehlin et al. 2006). The calculated pI of *Pf*TLP is 9.71. The addition of the GST-tag to 6His-*Pf*TLP lowers this to 8.97, as the pI of the tag alone is 6.51. It is possible that as the protein translates the fusion-protein, the rapid increase of the local pI to approximately 10, combined with the utilisation codons for the same amino acids in the region, leads to premature translational truncation. G6Hp is smaller than GST-6His-*Pf*TLP (~28kDa versus ~71kDa), and is likely to be more easily translated, hence there would be a higher rate of translation of the truncated peptide than the full-length fusion protein, which would result in a high ratio of G6Hp to GST-6His-*Pf*TLP produced.

Overall, the codon optimisation of the *Pf*TLP gene for expression in *E. coli* was successful in allowing for soluble expression of the protein, with lowered toxicity to the host cell. Despite the presence of contaminating G6Hp, the expression of *Pf*TLP protein from the codon-optimised vector was substantially greater than seen previously. Additionally, the presence of two epitope tags allows for much more stringent purification than a single-6His tagged protein (Figure 29).

### 3.4.5 Purification of *Pf*TLP protein

Purification from accumulated cell mass

*E. coli* cells containing the expression vector pET11d-GST-6His-*Pf*TLPco were grown, protein induced through the cold-shock induction method, and cell mass accumulated. The cell mass was then used to purify GST-6His-*Pf*TLP protein. Protein was purified from the cleared cell lysate through affinity purification on glutathione-agarose resin as described in Chapter 2.5.2.

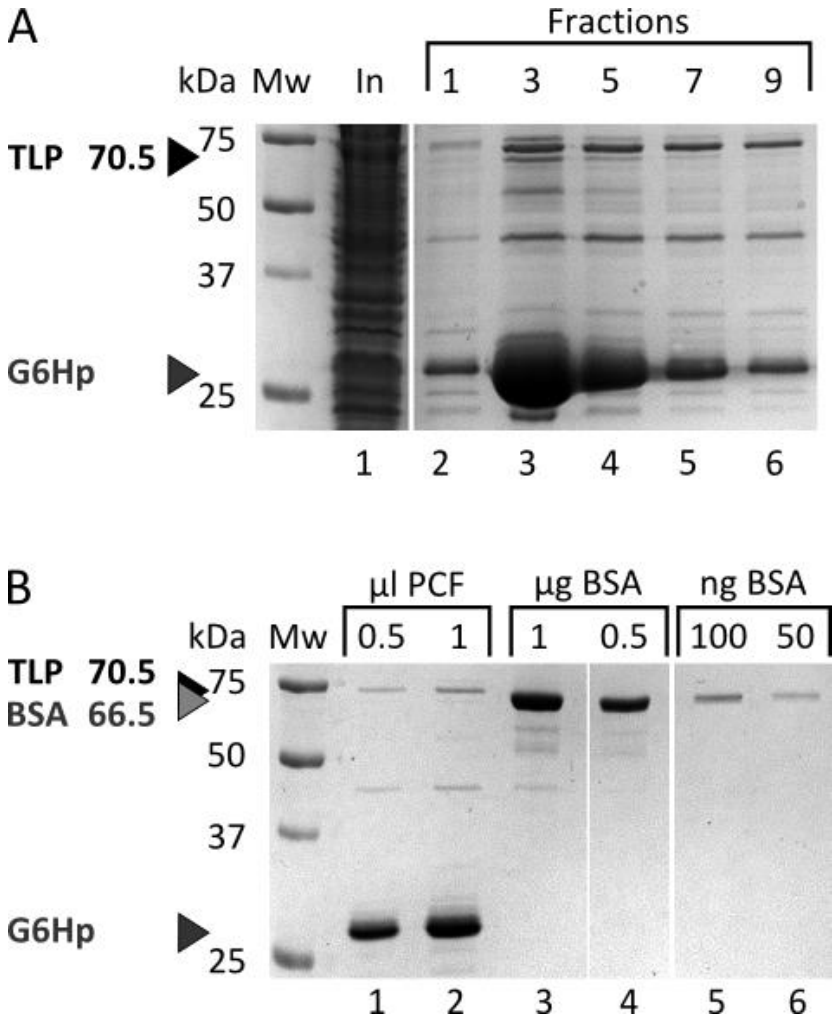



***Figure 31: Purification of GST-6His-PfTLP by glutathione-affinity agarose.***

**BSA** standards were loaded in lanes 3-6, 1µg, 0.5µg, 100ng and 50ng respectively. Lanes 1-2, contain 0.5µl and 1µl of pooled protein containing fractions (**PCF**, 4.2ml total) respectively. Expected size for GST-6His-*Pf*TLP is 70.5kDa, indicated with a black arrow. G6Hp is present and indicated by a dark grey arrow. Molecular weight marker (**Mw**) – 5µl of Precision Plus Protein™ All Blue Standard (*Bio-Rad*).

The glutathione-eluted protein fractions were collected as shown in Figure 31A. The presence of an enriched band in the region of the expected size of GST-6His-*Pf*TLP was seen. Evidence from the purification of other GST-tagged *Plasmodium* proteins in the lab have shown that some *Plasmodium* proteins are not eluted rapidly (in the first few fractions) from the GST-agarose resin by glutathione, and instead tend to elute slowly over many fractions. This is most likely due to non-specific interactions between GST-6His-*Pf*TLP and the glutathione-agarose resin, which are not disrupted by glutathione in the elution buffer. In contrast, the elution of G6Hp is very rapid, with the bulk of the protein eluted in fraction 3 (lane 3, Figure 31A. This appears to be case with GST-6His-*Pf*TLP, and so the majority of the fractions were pooled, in order to increase the yield of partly purified protein. This had the effect of diluting the protein preparation. The pooled preparation was then titrated against BSA protein to estimate the quantity of protein purified (Figure 31).

A contaminant band was present at the 44kDa marker. It is suggested that the band present at the 44kDa marker is the native *E. coli* ODO2 protein (Table 5). It has been reported that ODO2 may contaminate protein preparations purified by both glutathione agarose and metal-affinity resins (Bolanos-Garcia & Davies 2006).

It is clear from the purification that although GST-6His-*Pf*TLP is recovered, at about 100ng/µl, there is also a great deal of contaminating G6Hp present, at approximately 1µg/µl, or 10-fold the concentration of the full-length G6H*Pf*TLP.

Thrombin cleavage of GST-6His-*Pf*TLP protein

The presence of the bulky GST-tag may interfere with *Pf*TLP protein function. For this reason, the GST-tag was removed from the protein through thrombin cleavage.

Thrombin cleavage was performed as described in Chapter 2.6.4. The protein preparation was incubated with thrombin enzyme at 4°C, and the cleavage monitored through analysis

of aliquots by SDS-PAGE. The protein was incubated for ~40 hours at 4°C, due to the large amount of GST-6His-peptide present within the protein preparation, which competed with GST-6His-*Pf*TLP protein for cleavage by the thrombin enzyme. It is clear from Figure 32 that over time 6His-*Pf*TLP appeared at the 44kDa region, along with cleaved GST-tag at ~26kDa.

The presence the cleaved G6Hp peptide presented further issues for the purification of 6His-*Pf*TLP protein from the protein preparation. Attempts to capture 6His-*Pf*TLP from the protein preparation through Ni- NTA affinity column purification were largely unsuccessful (Figure 33). The protein preparation (In1, lane 1) was passed through a Ni-NTA affinity-purification column. Protein was eluted with 250mM imidazole, and fractions collected, analysed by SDS-PAGE, and the protein containing fractions pooled. As can be seen in Figure 33, lane 2, the unbound fraction still contained a large amount of 6His-*Pf*TLP protein.

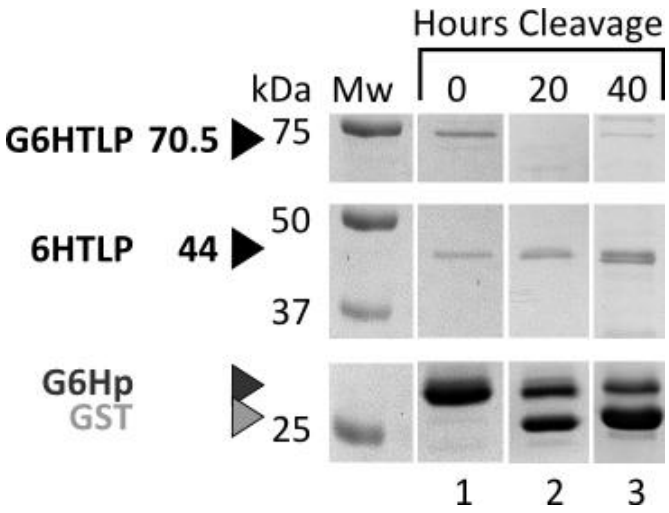



***Figure 32: Time course of thrombin cleavage of the GST-6His-PfTLP.***

Lane 1, GST-6His-*Pf*TLP preparation prior to thrombin cleavage (**0**, 1µl of 4ml total), Lane 2, **20** hours of thrombin cleavage (2µl of 4.2ml total), Lane 3, **40** hours of thrombin cleavage (5µl of total 4.2ml, an increased volume was loaded in order to better visualise cleavage).

Expected size of: GST-6His-*Pf*TLP 70.5kD; 6His-*Pf*TLP 44.4kDa, indicated by black arrows. GST-tag and GST-6His-peptide indicated by grey arrows. Molecular weight marker (**Mw**) - 2µl Precision Plus Protein™ All Blue Standard(*Bio-Rad*).

This unbound fraction was Ni-NTA affinity-purified a second time, resulting in the second unbound fraction (lane 3), and the imidazole-eluted protein fractions pooled.

It is clear that the GST-6His-peptide competed against the 6His-*Pf*TLP protein for binding to the Ni-NTA resin. Additionally, the G6Hp which was successfully cleaved would have also generated free 6His-peptides which may also have competed for binding to the Ni-NTA resin. It is thought that the GST-tag and G6Hp could form dimers in solution, which explains the presence GST-tags cleaved by thrombin digestion present in the eluted fractions (lane 4, PCF). The presence of two bands at the 44kDa region are most likely 6His-*Pf*TLP, and the *E. coli* contaminant ODO2.

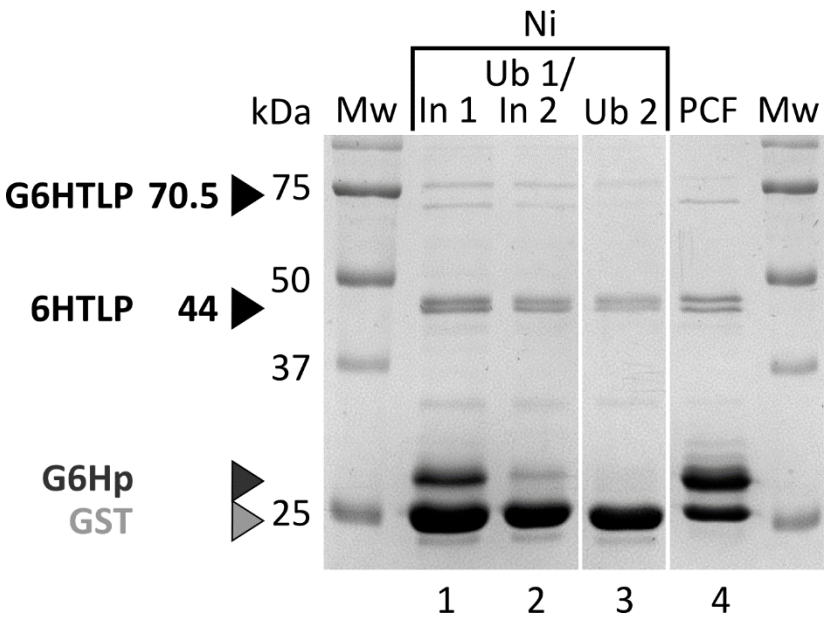



***Figure 33: Ni-affinity purification of 6His-PfTLP after thrombin cleavage.***

Lane 1 contains 5µl of input (**In1,** total 3.7ml), an equivalent amount of the protein fraction not bound to the resin is in lane 2, which was passed through another round of purification (unbound, **Ub1/In2**), the protein fraction not bound is loaded in lane 3 (**Ub2**). The protein eluted from each column with 250mM imadazole, was pooled and 10µl loaded in lane 4 (pooled column fractions, **PCF**, total 2.2ml). Molecular weight marker (**Mw**) - 2µl Precision Plus Protein™ All Blue Standard(*Bio-Rad*).

Attempts at further purification with SP-Sepharose® resin

The constituents of the *Pf*TLP protein have differing calculated pI values; GST-tag and the presumed GST-6His-peptide have predicted pIs of 6.51 and 7.4 respectively, 6His-*Pf*TLP has a predicted pI of 9.63. The protein preparation was stored in a buffer with a pH of 7.9 at 4°C, suggesting that 6His-*Pf*TLP should be positively charged while the GST-tag and G6Hp negatively charged. Therefore, an attempt was made to purify 6His-*Pf*TLP from the protein preparation using the cation-exchange resin SP-Sepharose® (Figure 34). For this experiment, the pooled column fraction eluted from the second purification by nickel-affinity was used. Proteins were bound at 100mM KCl and eluted with increasing KCl concentrations. It was found that 6His-*Pf*TLP binds partially to the SP-Sepharose® resin, with some proportion of the protein remaining bound under high KCl concentrations. This suggests that there is heterogeneity within the 6His-*Pf*TLP protein preparation.

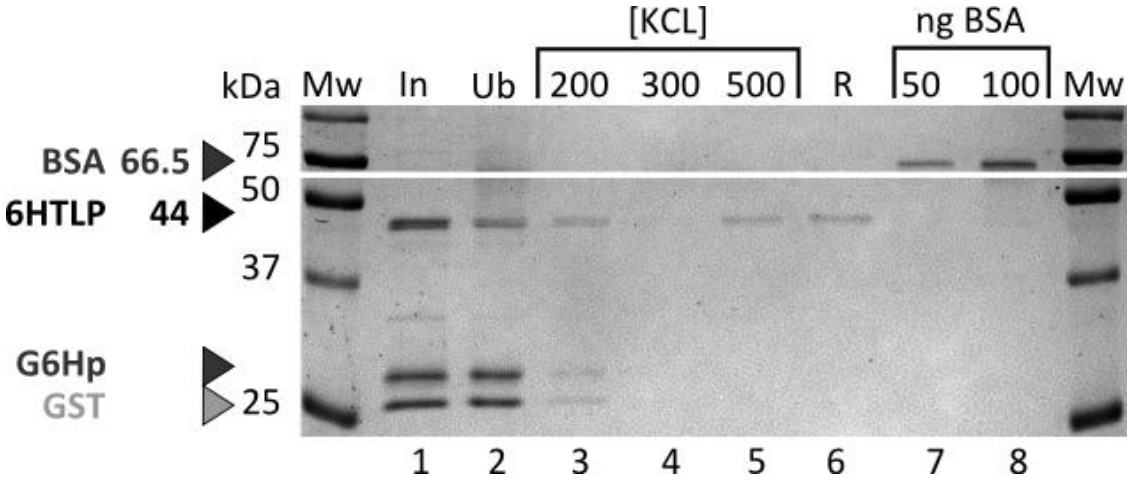



***Figure 34: 6His-PfTLP partially binds to SP-Sepharose strong-cation exchange resin.***

10µl of the 6His-*Pf*TLP protein preparation was loaded into the input lane (**In**, total volume 100µl), and an equivalent amount of the protein not bound (unbound, **Ub**). KCl elutions, equivalent to the input, are loaded lanes 3-5. Lane 6 contains ½ of the resin resuspended in 2×SDS loading buffer (**R**, total volume of 10µl settled resin volume). Lanes 7 and 8 contain 50 and 100ng of BSA protein respectively. Molecular weight marker (**Mw**) - 2µl of Precision Plus All blue Standard (*Bio-Rad*).

It is unclear whether this binding is largely non-specific, or due to interactions of basic regions (Table 8) in the *Pf*TLP structure to the resin. However, it is also difficult to distinguish between the binding of the contaminant protein (presumably ODO2 protein) and 6His-*Pf*TLP, as they have similar sizes, and ODO2 has been shown to bind to Sepharose® resins at salt concentrations above 600mM KCl (Bolanos-Garcia & Davies 2006).

Immunoblot detection of purified 6His-*Pf*TLP

In order to verify the identity of purified 6His-*Pf*TLP, immunoblot analysis was performed making use of an anti-hexahistidine antibody to detect the 6His-*Pf*TLP protein (Figure 35). In this experiment, an SDS-PAGE gel was run containing multiple lanes of both the *Pf*TLP protein preparation as well as a positive control, 6His-*Hs*TFIIB.

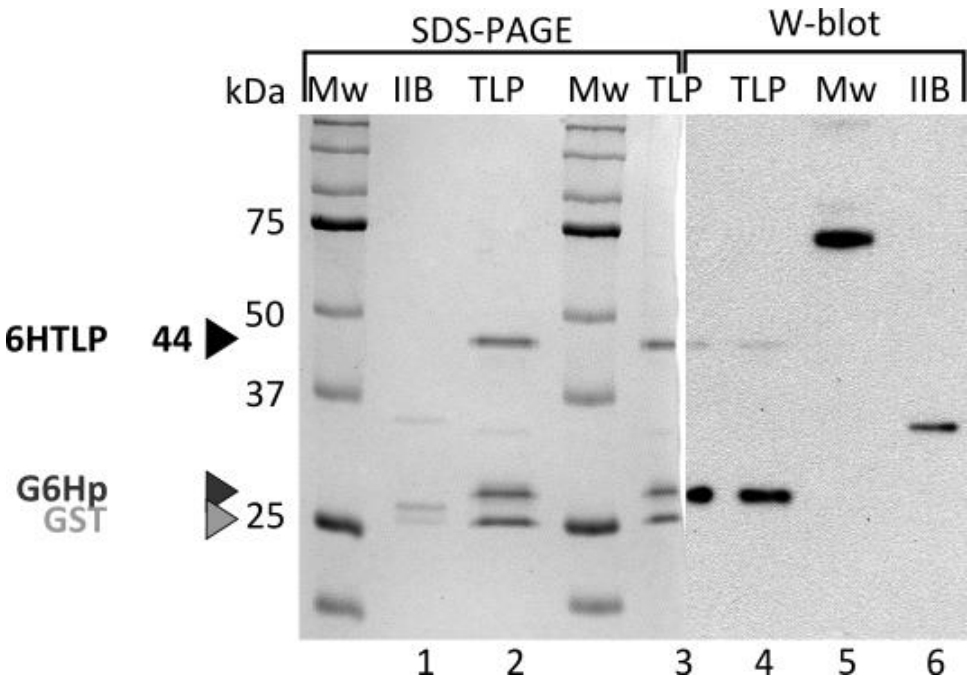



***Figure 35: Immunoblot analysis of 6His-PfTLP protein with anti-6His antibody.***

Lanes 1-3: SDS-PAGE and coomassie staining analyses of *Pf*TLP protein preparation. Lane 1 contains 100ng of 6His-HsTFIIB protein. Lane 2 contains 100ng of 6His-*Pf*TLP protein. Molecular weight marker (**Mw**) is 2µl of of Precision Plus Protein™ All Blue Standard (*Bio-Rad*). Lanes 3-6: Immunoblot analysis of *Pf*TLP protein preparation. In lanes 3 and 4, 100ng of 6His-*Pf*TLP protein protein preparation was loaded. Lane 6 conatins 50ng of 6His-*Hs*TFIIB protein. 4µl of Mw was loaded. The SDS-PAGE gel was transferred to PVDF membrane, probed with anti-6hHis mouse monoclonal antibody.

Expected size of: 6His-*Pf*TLP ~44.4kDa, recombinant 6His-*Hs*TFIIB as ~37 kDa. Truncated GST-6His-*Pf*TLP at ~28kDa, GST from thrombin cleavage at 26.1 kDa.

The gel was sliced in half, and one half was transferred to PVDF membrane and probed. This allowed side-by-side comparisons of the proteins detected in the gel and immunoblot. Unfortunately, the gel did not separate the bands at the 44kDa region successfully. However, ODO2 does not have clusters of histidines, and is thought to be non- reactive to anti-hexahistidine antibody. It was found that in the *Pf*TLP preparation, two bands were immune-reactive against the antibody, at the sizes expected for the truncated GST-6His-peptide and that of 6His-*Pf*TLP. The cleaved GST-tag was present at a similar quantity to the GST-6His-peptide, but was not immune-reactive, suggesting that the antibody was specific to the 6His-tag.

Depletion of GST-tag by glutathione-agarose

The purification of the 6His-*Pf*TLP protein preparation through Ni-NTA affinity resin allowed for a buffer exchange and removal of glutathione from the protein preparation. Therefore, an attempt to deplete the contaminating GST-tag/GST6Hp peptides from the

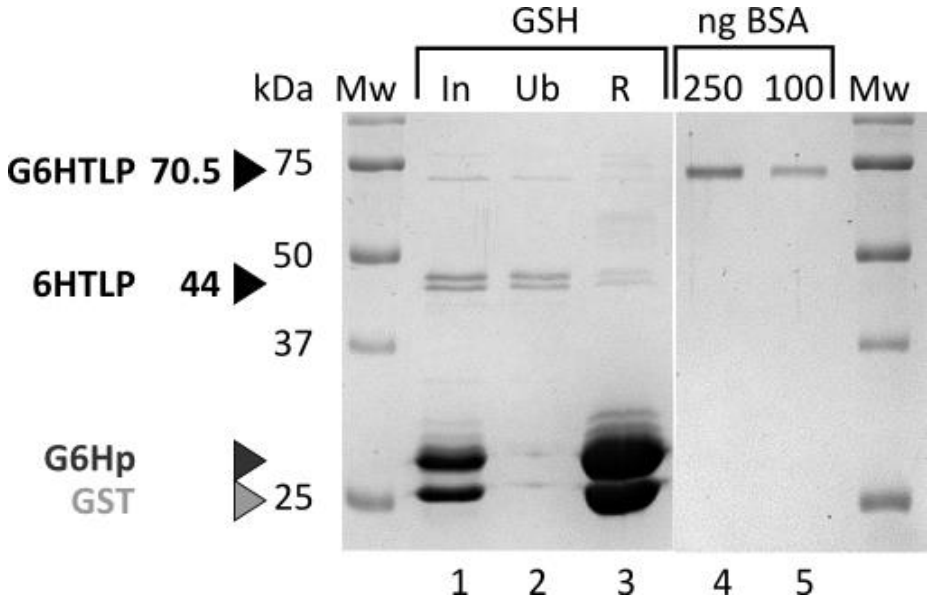



***Figure 36: Glutathione-affinity based depletion of GST/G6Hp from the GSH Sepharose PfTLP protein preparation.***

Input (**In**, 10µl of 2.2ml protein preparation, lane 1), the unbound flow-through (**Ub**, 10µl of 2.7ml), a sample of the resin (**R**, 10µl of resin, denatured with 10µl of 2×SDS loading dye, total resin volume of 0.5ml), and 250ng and 100ng of **BSA** standard were loaded. Molecular weight marker (**Mw**) - 2µl of Precision Plus Protein All Blue standard (*Bio-Rad*).

preparation was attempted by passing the preparation through a glutathione-agarose (GSH) resin column. This successfully depleted the majority of the contaminating GST-peptide species from the preparation (Figure 36). However, as may be seen from Figure 36, lane 3, where a small percentage of the resin was heat-denatured, a great deal of the 6His-*Pf*TLP protein was found to bind irreversibly to the column, presumably through non-specific interactions with the resin. This severely reduced the total amount of available 6His-*Pf*TLP protein purified. The concentration 6His-*Pf*TLP in the preparation shown in Figure 36 was estimated to be approximately 10 ng/µl (2.6ml total) based on a colorimetric protein assay (*Bio-Rad* protein assay). The protein preparation was supplemented with 100 ng/µl BSA, snap-frozen in liquid $N_2$ and stored at -80˚C.

### 3.4.6 Summary of *Pf*TLP expression and purification

The expression of recombinant *Plasmodium* proteins is known to be challenging. In a large study of recombinant expression of 1000 genes in *E. coli* cells, only 337 proteins were successfully expressed, and the majority of these were insoluble (Birkholtz et al. 2008b). As with the majority of the proteins which have been attempted to be expressed in *E. coli*, the initial attempts at expressing *Pf*TLP protein showed that over-expression was toxic to the *E. coli* host cell. In this case, the toxicity of the protein expression was alleviated when the protein was codon optimised for *E. coli.* The expression of a GST-tagged *Pf*TLP protein demonstrated premature translational termination of the protein. This is likely due to combination of factors which include the high pI of the *Pf*TLP protein, the A+T richness of the ORF coding for the protein, as well as the repetitive usage of a small subset of codons. In the purification of *Pf*TLP, the protein preparation was shown to be heterogeneous, and would partially bind to agarose-based resins (SP-Sepharose®, glutathione-agarose®). The binding to these resins is most likely to be non-specific, as is shown by the elution profile of the protein from glutathione-agarose resin, and may be due to the presence of the disordered

LCR regions. It has been found in other recombinantly expressed *Plasmodium* proteins which contain LCRs, that during over-expression in *E. coli* cells there is a propensity for the host-cell to sporadically substitute amino acids into the LCR, which cannot be detected by SDS-PAGE analysis (Mehlin et al. 2006). The presence of these amino acid substitutions may influence the binding of the expressed protein to resins during purification. Additionally, the LCRs are also known to lead to recombinant *Plasmodium* protein aggregation, which would reduce the binding of the protein to affinity resins, and be a cause for the heterogeneity in the protein preparation seen (Mehlin et al. 2006). Finally, purification of recombinant *Pf*TLP was achieved, and is sufficiently pure to continue with DNA-binding assays.

## 3.5 Examination of the DNA-binding potential of *Pf*TLP and *Pf*TFIIB

Having achieved expression and purification of *Pf*TLP protein, it was decided to begin investigating the potential for DNA-binding activity in *Pf*TLP.

The DNA-binding activity of TLP/TRF2/TRP molecules has not been extensively studied. It is known that human TLP is able to bind DNA, but lacks the specificity for the TATA-box characteristic of TBP (Ohbayashi et al. 2003). Given the divergence of *Pf*TLP to human TBP/TLPs and even the putative *Plasmodium* TBP, the DNA-binding activity of TLP, and the specificity thereof, could not be fully investigated in the scope of this research project. Should TLP have DNA-sequence specific binding, and whether DNA-binding may rely on the association of other factors, will all form part of extensive future investigations in the lab. Additionally, the conditions that are suitable for human transcription *in vitro* binding may not be suitable or optimal for *Plasmodium* transcription factor DNA-binding, given the potential differences between the organismal cell environments of *Plasmodium* and human. However, as the research group has produced probes for the continued investigations into

*Pf*TBP DNA-binding assays, it was a natural extension to begin the investigation into potential *Pf*TLP DNA-binding. Given the lack of adequately purified *Pf*TFIIB and/or *Pf*TFIIA protein, the majority of the assays were to be performed with *Pf*TLP alone.

The focus of future work in the laboratory will be to determine sequence specificity of *Pf*TBP/TLP to sequences in putative *Plasmodium* core promoter elements. Only a single study has been performed with *Pf*TBP, finding sequence specificity to TATA-like sequences present upstream of the TSSs of the *kahrp* and *gbp-130* genes (Ruvalcaba-Salazar et al. 2005). These putative promoter regions were studied in this research project, as they continue to be investigated by others in the lab. These genes have been targeted due to their role in the pathogenicity of the parasite. G*bp-130* encodes a glycophorin binding protein, which binds to the erythrocyte glycophorin receptor, and is instrumental in the invasion of the RBC by the parasite (Perkins 1984). The knob-associated histidine rich protein (KAHRP) interacts with iRBC cellular structures to form a platform for the presentation of the *Pf*EMP1 protein (Lanzer et al. 1992a). The transcription start sites of these genes have been mapped (Lanzer et al. 1992b; Lanzer et al. 1992a) and so may be useful in the identification of *Pf*TBP/TLP recognition sites in the core promoter.

### 3.5.1 Electrophoretic mobility shift assays (EMSAs)

Electrophoretic mobility shift assays (EMSAs, or 'gel shift') are a commonly used method of detecting protein complexes with nucleic acids (Garner & Revzin 1981; Hellman & Fried 2007). In this assay, nucleic acid probes containing sequences of interest are incubated with a DNA-binding protein. The protein-DNA complexes are then separated from the unbound DNA by electrophoresis on a non-denaturing gel, typically polyacrylamide. DNA which is bound to protein has retarded mobility compared to the free probe, and so migrates more slowly through the gel matrix (the 'gel shift'), which can be seen through visualising the probes, typically through radiographic means (Hellman & Fried 2007). However, for the

experiments discussed below an alternative non-radioactive detection method was performed that makes use of biotin-labelled probes. In this protocol, the electrophoresed probes are transferred to positively charged nylon membrane, fixed by UV-crosslinking, and then visualised by chemiluminescence from streptavidin-conjugated horseradish peroxidase with chemiluminescent substrate, and resultant luminescence detected by X-ray film.

The binding of *Pf*TLP to the putative promoter regions of *gbp-130* and *kahrp* was investigated. Single-biotin labelled probes were designed to cover approximately 60bp downstream and 170bp upstream of the reported transcription starts sites of the two genes and were similar to those used by Ruvalcaba-Salazar et al. (2005) in their investigation of *Pf*TBP DNA binding. In total, the probes are approximately 280bp in length (Chapter 2.9.1). Additionally, a ~200bp probe was synthesised for the adenovirus 2 major late promoter (Ad2ML) promoter. The Ad2ML promoter has been widely used to investigate the formation of the human PIC (Thomas & Chiang 2006), and here is used in conjunction with *Hs*TBP/TFIIA to confirm the assay conditions may allow for visualisation of DNA-binding.

Human transcription factors bind to the Ad2ML probe

The initial DNA-binding conditions were determined for the recombinant human transcription factors TBP and TFIIA with the Ad2ML probes. The DNA-complexes were resolved on a native polyacrylamide gel, and visualised. The binding of TBP to the Ad2ML containing probe is significantly enhanced in the presence of TFIIA, and so this recombinant factor was included in the protein-DNA reaction mix.

It was found that interactions between *Hs*TBP and the Ad2ML probe could be resolved on a native polyacrylamide gel, and probe shifts identified (Figure 37A). Increasing amounts of protein prevented the migration of the protein into the polyacrylamide gel, as seen by the increased retention of probe in the well.

Polyacrylamide gels cannot resolve *Pf*TLP-DNA complexes

A *Pf*TLP concentration-dependent decrease of free probe was seen in the EMSAs with Ad2ML and *gbp-130* probes, with a corresponding increase of protein-probe complex retention in the wells of the polyacrylamide gel (Figure 37B). This suggests that the *Pf*TLP protein preparation is capable of binding to DNA. Several alterations to the gel conditions were investigated in order to allow potential *Pf*TLP-DNA complexes to enter the gel matrix.

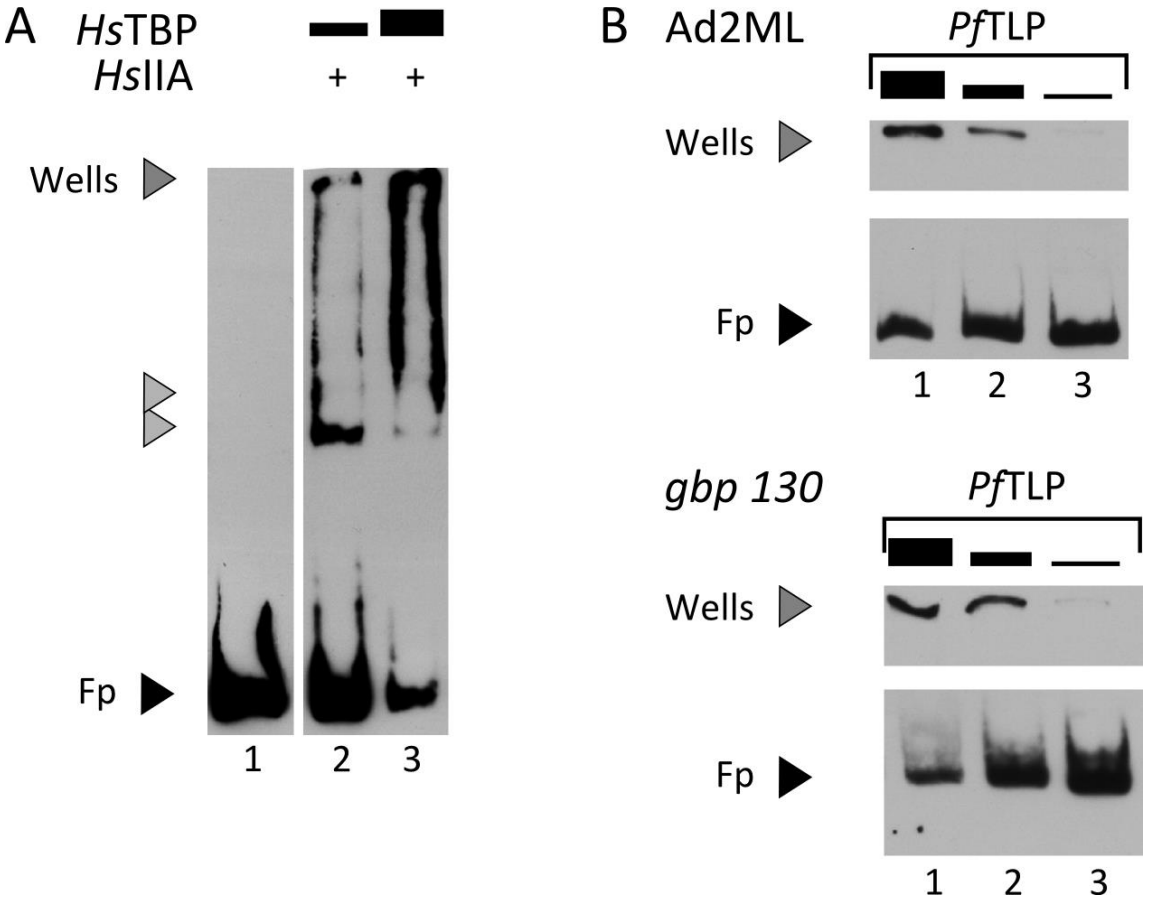



***Figure 37: PfTLP DNA-binding cannot be resolved by polyacrylamide gel electrophoresis.***

**A.** *Human* TBP/TFIIA binding to Ad2ML probe. Complexes were separated on a 3.5% 49:1 native polyacrylamide gel. 20µl binding reactions contained 2fmol Ad2ML probe. Binding reactions loaded in lanes 2-3 contain 2.5 and 5fmol/µl of *Hs*TBP protein and 17fmol/µl *Hs*TFIIA. The position of the *Hs*TBP/TFIIA/promoter complex is indicated by a grey arrow. **B.** Binding of *Pf*TLP protein preparation (GSH-Sepharose® purified) to adenovirus 2 major late (Ad2ML) and *Plasmodium falciparum gbp-130* promoter probe, indicated by probe retention in the wells in a *Pf*TLP concentration-dependent manner. 20 µl binding reactions contained 5fmol Ad2ML or *gbp-130* probe as indicated. Binding reactions loaded in lanes 1-3 contain 70, 35 and 17.5fmol/µl of 6His-*Pf*TLP respectively.

These included: gels prepared with lower acrylamide percentages, altering the *bis*-acrylamide/acrylamide ratios of the gels, inclusion of chaotropic or non-ionic detergents in the protein-DNA reaction mix, as well as variations to the electrophoresis conditions. Unfortunately none of these alterations were successful in permitting the entry of the protein-DNA complex to enter the gel matrix (not shown).

There may be multiple reasons for impaired mobility/entry into the gel matrix of the *Pf*TLP-DNA complex. It is has been shown in Chapter 3.2 that *Pf*TLP contains a large LCR region which may result in undesirable protein-protein interactions, such as aggregation, and so impede movement of the complexes into the gel matrix. Additionally, should the DNA-binding of *Pf*TLP be non-specific to the DNA-sequence, this may result in multiple *Pf*TLP proteins binding to the same probe, and so diminish the ability for the probes to enter the gel-matrix.

### 3.5.2 Immobilised template assays

In order to investigate whether the interaction between the *Pf*TLP protein preparation and the Ad2ML/*gbp-130* probes were indeed dependent on the recombinant *Pf*TLP protein and not another contaminant in the protein preparation, an immobilised template assay was used.

In the immobilised template assay, biotinylated DNA probes are bound to streptavidin-magnetic beads. The protein and DNA are allowed to interact under similar conditions to those used in the EMSA assay. The unbound protein is then removed from the beads, and bound protein is eluted in SDS-PAGE loading buffer and analysed by immunoblotting. This assay serves as a direct measure of the binding of protein to the DNA-probe. At this point the lab had obtained polyclonal antibodies against several of the putative *Plasmodium* general transcription factors, and these needed to be characterised prior to use in the immobilised template assays.

Characterisation of antibodies for *Plasmodium* TLP

The production of polyclonal antibodies was performed by *BioGenes* (Berlin, Germany). Peptide-antibodies were raised in rabbits against the peptide C-NDENKSNDNKEQND, mapping to a small LCR in *Pf*TLP (amino acid positions 294-307).This region was chosen as LCR loops should extend outside of the main globular structure, to provide an accessible epitope.

As described in the Chapter 2.8, 5.4µg of total protein from soluble cell lysate from *E. coli* containing full length GST-6His-*Pf*TLP was mixed with 30ng of purified 6His-*Pf*TLP (making up 0.6% of the total protein mass). Cleared lysate containing both GST-6His-*Pf*TLP and added 6His-*Pf*TLP were loaded onto several lanes of an SDS-PAGE gel, electrophoresed, and transferred to a PVDF membrane. The membrane was then cut into strips and probed with increasing dilutions of the rabbit anti-*Pf*TLP anti-sera (Figure 38).

It is clear from this experiment that the polyclonal antibodies have high specificity for *Pf*TLP, being able to detect in cleared lysate both added purified 6His-*Pf*TLP as well as expressed GST-6His-*Pf*TLP. The cross-reactivity with the *E. coli* proteins in the enriched sample is minimal. Note that there appears to be some cross reactivity with contaminant

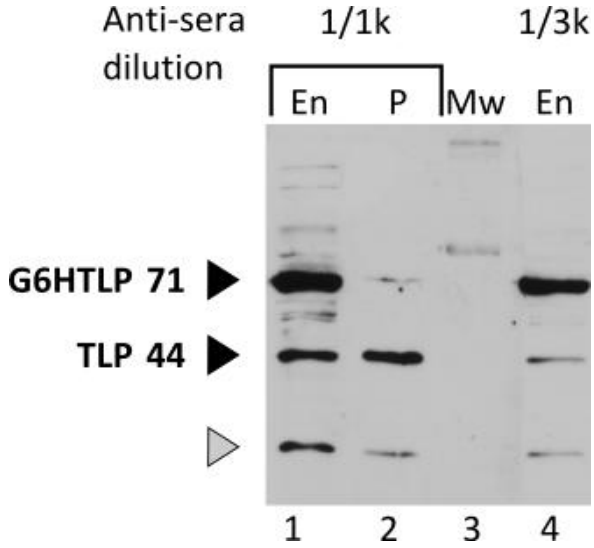



***Figure 38: Anti-PfTLP antibody is specific to PfTLP.***

Cleared cell lysate containing GST-6His-*Pf*TLP and 30 ng added 6His-*Pf*TLP (**En**, lanes 1 and 4) and 30 ng purified *Pf*TLP (lane 2) were electrophoresed by SDS-PAGE, and transferred to PVDF membrane. The membrane was probed with anti-*Pf*TLP anti-sera dilutions as indicated. Molecular weight marker (**Mw**) - 4µl Precision Plus Protein™ Unstained Standards (*BioRad*).

protein at approximately 25kDa. This would correspond to the G6Hp (see Chapter 3.4.5) known to be in high abundance in the enriched sample, and which may contaminate the purified protein preparation. Note also that a similar experiment was performed with equivalent amount of the pre-immune serum from the rabbits, where no *Pf*TLP-specific reactivity could be detected (not shown).

Human transcription factors bind to Ad2ML in a *Hs*TBP dependent manner

For the immobilised template assay, the visualisation of the experiment is performed through immunoblotting. In order to visualise protein-DNA interactions, far more probe and protein is required than in EMSAs. Immobilised template reactions were therefore set to contain 1000fmol of DNA-probe and the protein(s) tested in 40µl total reaction volume. Immobilised template assays were first carried out using recombinant human GTFs with the Ad2ML probe (Figure 39). Human TBP is shown to bind to the probe (Figure 39A). As expected, human TFIIB can only bind to the probe in the presence of TBP, but not on its own (compare Figure C and B). Like TFIIB, TFIIA binding to the Ad2ML probe was dependent on the presence of TBP (data not shown).

*Pf*TLP binds to putative *Plasmodium* promoters

As seen in Figure 39 D-F, *Pf*TLP protein was found to bind directly to both the *gbp-130* and *kahrp* putative promoters (Figure 39 D and E), but not to the Ad2ML probe (F).

The results of the immobilised template assays suggest that purified bacterially expressed *Pf*TLP indeed possesses DNA-binding activity, consistent with the results of the EMSA experiments (Figure 37). This result might indicate some sequence-specificity of *Pf*TLP. Further studies are needed to investigate this interesting possibility. It is unclear why binding to the Ad2ML probe under these conditions could not be detected. It is possible that the wash step in the immobilized template assay may be too stringent, causing weak protein-DNA

associations to dissociate. In the EMSA assays the protein-DNA mixtures are loaded directly after being allowed to associate. The stabilising effect of the EMSA gel matrix could, therefore, ensure that the protein-DNA interactions are able to be detected.

Potential interactions between *Pf*TLP and *Pf*TFIIB could not be investigated with this assay and the conditions used, due to strong non-specific binding of *Pf*TFIIB to the streptavidin-magnetic beads in the absence of DNA (not shown). This problem could not be overcome in the course of this thesis work.

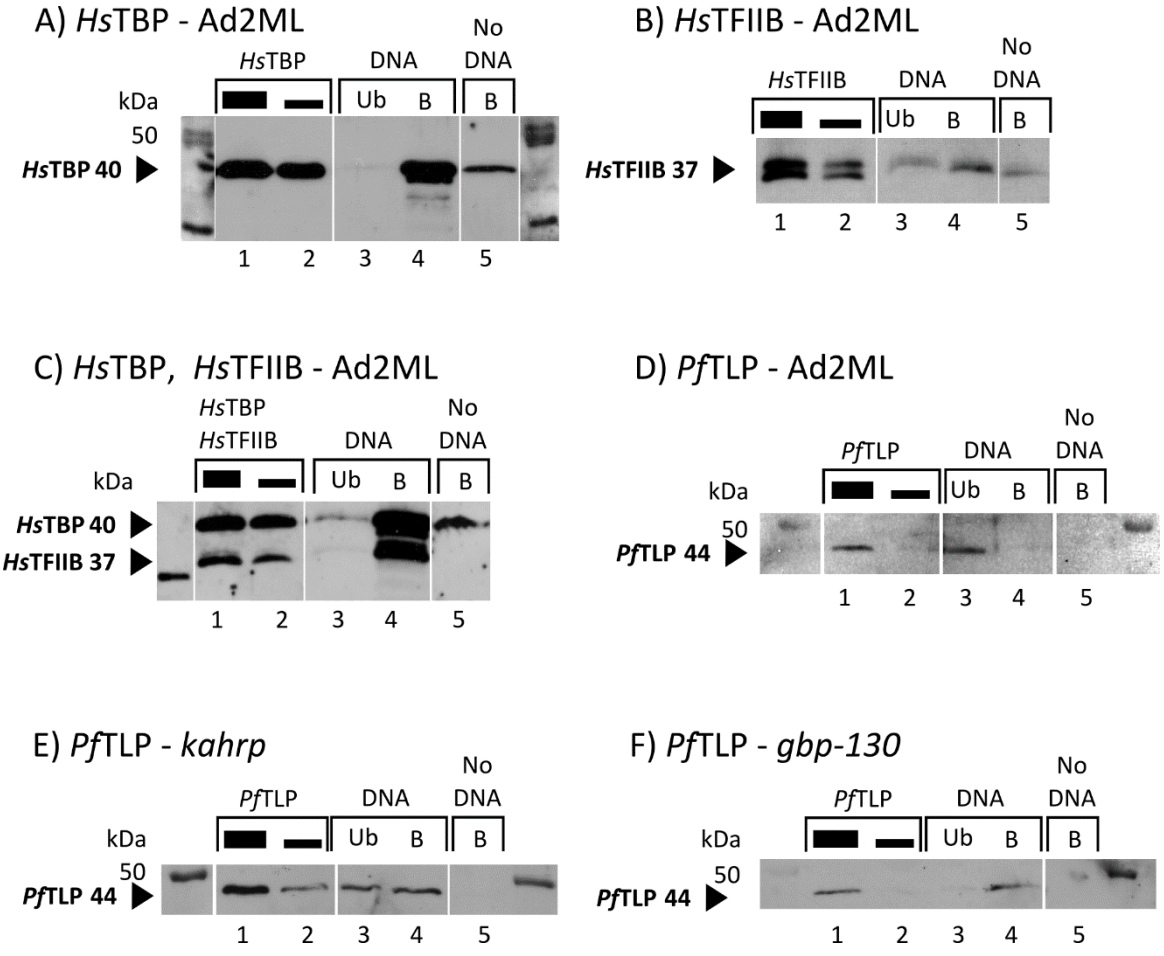



***Figure 39: DNA binding of human and Plasmodium general transcription factors in immobilised template assays.***

Immunoblot analysis of bound (**B**) and unbound (**Ub**) proteins. **A.** Binding of *Hs*TBP to Ad2ML probe. **B** and **C.** Binding of *Hs*TFIIB to the Ad2ML probe is dependent on the presence of HsTBP **D.** *Pf*TLP does not bind to Ad2ML probe. **E** and **F.** *Pf*TLP binding to *kahrp* and *gbp-130* probes, respectively. Protein input equivalent to ½ and ¼ of protein present in binding reactions was loaded in lanes 1 and 2. ¼ of total unbound protein (Ub) and total of DNA-bound protein (B) were loaded in lanes 4 and 5.

### 3.5.3 Agarose EMSAs

Due to the difficulty in resolving the *Plasmodium* protein-DNA complexes in polyacrylamide gel electrophoresis, an alternative agarose gel system was developed. Magnesium-containing agarose gel electrophoresis has been used previously for DNA-binding assays with large multi-protein complexes (Zerby & Lieberman 1997). The larger pore size of the agarose matrix facilitate the migration of larger complexes into the gel.

*Pf*TLP forms complexes with Ad2ML promoter

In order to test the conditions for agarose EMSA, binding reactions containing recombinant human transcription factors *Hs*TBP and *Hs*TFIIA and Ad2ML probe were carried out. Complexes were separated in 1.4% agarose gels in a running buffer composed of 0.5x TBE containing 5 mM $MgCl_2$, transferred to positively charged nylon membrane, and visualised by chemiluminescence as described above (see Chapter 43). As expected, *Hs*TBP formed a clearly detectable complex with the Ad2ML probe in a TFIIA-dependent manner.

Incubation of bacterially expressed purified 6His-*Pf*TLP protein with the Ad2ML probe resulted in the appearance of several gel shifts. Binding reactions contained 5fmol/µl *Hs*TBP and 13.3fmol/µl *Hs*TFIIA (Figure 40, lane 3) whereas *Pf*TLP was present at a concentration of approximately 70 fmol/µl (Figure 40, lanes 4). While *Hs*TBP bound the available probe (10fmol) quantitatively under the reaction conditions, *Pf*TLP at a much higher concentration did not. This result suggests that *Pf*TLP associates with the Ad2ML probe with lower affinity than *Hs*TBP/*Hs*TFIIA, consistent with the results of the immobilised template assays (see Figure 39).

Interestingly, the profile of *Pf*TLP-Ad2ML binding to the Ad2ML probe suggests at least three gel shifts along the length of the gel, along with a smeared region. The presence of the smear and multiple gel-shifts may indicate that *Pf*TLP binds to the probe with low affinity at multiple locations. This would suggest that *Pf*TLP has low or no sequence specificity for the DNA sequence of the Ad2ML promoter, and instead binds non-specifically.

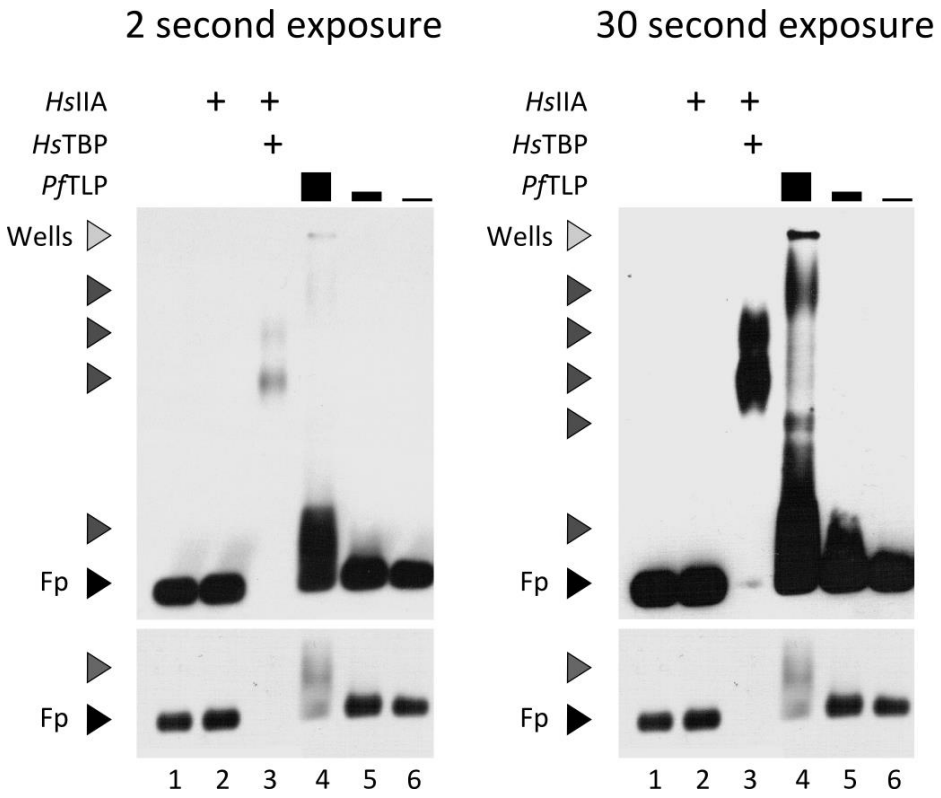



***Figure 40: PfTLP binding to the Ad2ML promoter in agarose EMSA.***

Two different exposures of the same experiment are shown, with a one second exposure of the lower part of the gel containing free probe as a separate panel underneath. 10µl binding reactions contained 10fmol Ad2ML probe, 5fmol/µl *Hs*TBP, 13.3fmol/µl, *Hs*TFIIA and ~70, 23, 7fmol/µl *Pf*TLP as indicated. Free probe is indicated by black arrows and protein/DNA complexes are indicated by grey arrows.

*Pf*TLP forms complexes with putative *Plasmodium gene promoters*

Next, the binding of *Pf*TLP to the putative *Plasmodium* promoters for the *kahrp* and *gbp-130* genes was investigated using agarose EMSA (Figure 41). As seen in the immobilised template assays Figure 39B) *Pf*TLP binds both the *gbp-130* and *kahrp* probes. It remains to be determined if the presence of multiple complexes seen in these agarose EMSAs is due to the presence of several *Pf*TLP molecules bound to the same probe, or the binding of *Pf*TLP molecules at different locations on the DNA probe.

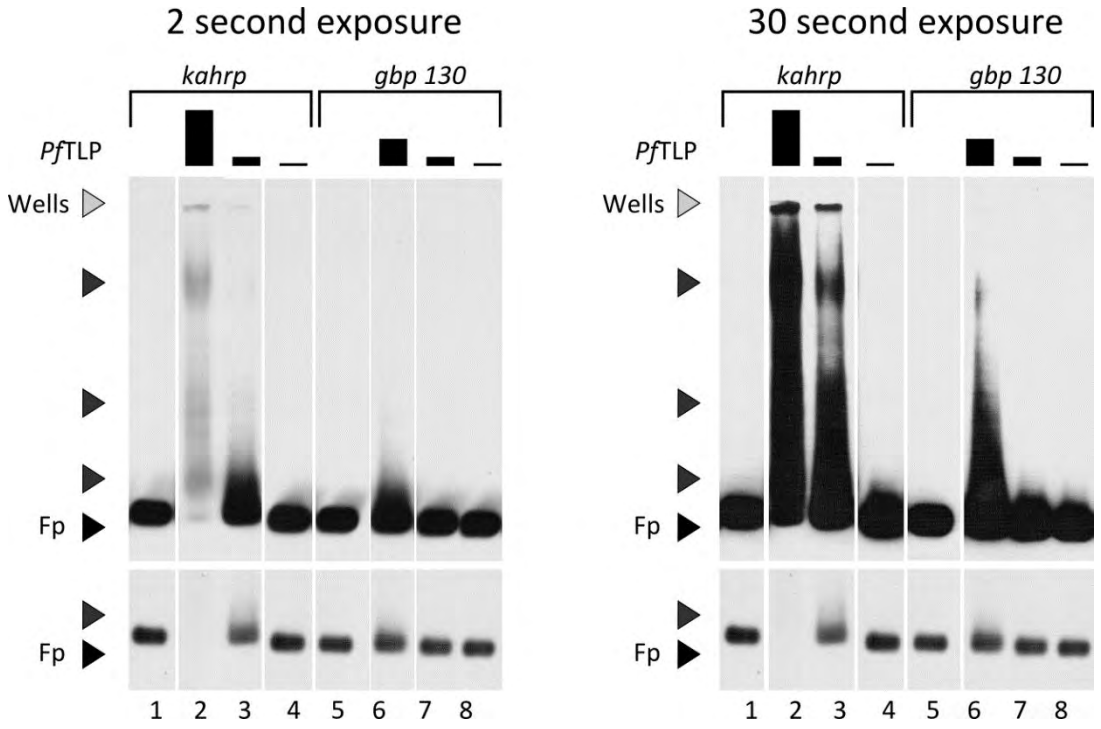



***Figure 41: PfTLP forms complexes with putative Plasmodium promoter sequences.***

Two different exposures of the same experiment are shown, with a one second exposure of the lower part of the gel containing free probe as a separate panel underneath. 10µl binding reactions contained 10fmol *kahrp* or *gbp-130* probe as indicated. *Kahrp* probe was incubated with 140, 70, 7fmol *Pf*TLP as indicated. *Gbp-130* probe was incubated with 70, 23, 7fmol of *Pf*TLP as indicated. Free probe is indicated by black arrows and protein/DNA complexes are indicated by grey arrows.

Investigating the DNA-binding potential of *Pf*TFIIB

In the purification of *Pf*TFIIB, the addition of nucleases appeared to abolish the affinity of *Pf*TFIIB for DEAE-Sepharose resin, suggesting a *Pf*TFIIB-DNA-DEAE interaction. An agarose EMSA was performed with partially purified *Pf*TFIIB protein preparation and the Ad2ML probe (Figure 42). As shown in Figure 42, a dose-dependent decrease of free-probe with increasing amounts of 6His-*Pf*TFIIB preparation was observed. Unfortunately, no clearly defined *Pf*TFIIB gel-shifts could be identified. This result further corroborates the notion that *Pf*TFIIB might possess DNA-binding activity, presumably mediated by its highly basic N-terminus.

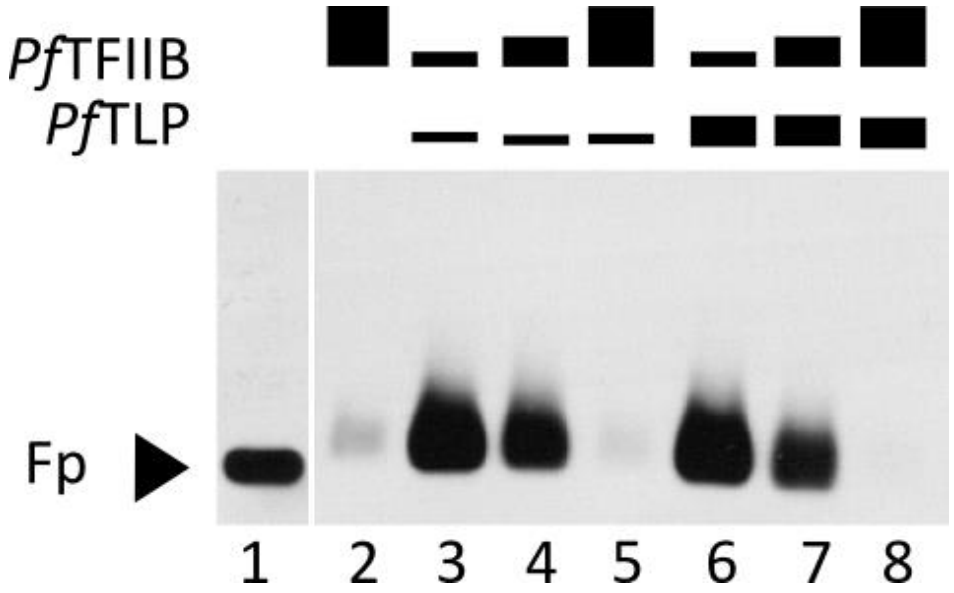



***Figure 42: Binding of partially purified PfTFIIB protein to the Ad2ML promoter in agarose EMSA.***

10µl binding reactions contained 10fmol Ad2ML probe, ~23 or 70fmol/µl *Pf*TLP, 50, 100, or 200fmol/µl *Pf*TFIIB as indicated.

Free probe is indicated by black arrows and protein/DNA complexes are indicated by grey arrows.

### 3.5.4 Summary of the DNA binding potential of *Pf*TLP and *Pf*TFIIB

The current evidence from both immobilised template assays and EMSAs indicates that *Pf*TLP is able to bind to DNA (Figures 39, 40, and 41). The specificity of this binding is not known, however, *Pf*TLP appears to possess greater binding affinity for the putative *Plasmodium* promoters than the Ad2ML promoter. *Pf*TLP appears to bind in multiple locations on the DNA-probes. This is supported both by the inability of the protein-DNA complex to migrate into the native polyacrylamide gels, as well as the appearance of multiple gel-shifts in agarose EMSAs. Agarose EMSA supports previous experimental evidence which suggests *Pf*TFIIB has DNA-binding activity, most likely mediated by its basic N-terminal region. The specificity of this binding requires further investigation.

# Chapter 4

# Conclusions

## 4.1 In silico analysis of *Pf*TFIIB and *Pf*TLP

Putative *Pf*TFIIB has low amino acid sequence homology to that of other eukaryotic TFIIBs studied, however, secondary and tertiary predictions show that the molecule has conservation of the main structural and functional residues important for TFIIB function. These include the presence of many functionally important structures and residues of the N-terminus, including residues responsible for intermolecular interactions with RNAPII, as well as intramolecular interactions. Less conserved are residues associated with DNA and/or TBP interactions. The canonical BREd-interaction region on the N-terminal cyclin domain is relatively well conserved in *Pf*TFIIB, suggesting similar conservation of the associated core promoter elements in the *Plasmodium* core promoter, but there is very little conservation of the BREu-interacting region in the C-terminal cyclin domain. Similarly, TBP-interaction sites are more conserved on the N-terminal cyclin domain than in the C-terminal domain. Whether the corresponding residues in *Pf*TBP have undergone complementary changes to retain *Pf*TFIIB binding requires further investigation. An interesting feature of the molecule is the presence of highly basic N- and C- terminal regions, which are not conserved between the classically studied TFIIBs. These basic regions may have DNA-binding potential, not seen in studied eukaryotic TFIIBs.

Although putative *Pf*TLP has the low amino acid sequence homology to other TBP/TLP molecules, it has been shown to contain the main structural regions associated with TBP/TLPs molecules. The structure of *Pf*TLP contains large low complexity regions which are thought to extend out of the main globular structure of the molecule. Interestingly, while

many residues implicated in DNA and TFIIB binding are conserved, there is far less conservation of the TFIIA interaction sites. These interaction sites lie adjacent to the long (~130bp) LCR, suggesting that TLP may have reduced affinity for TFIIA. This stands in contrast to *Pf*TBP which has retained conservation of these TFIIA interaction sites. However, the identification of two distinct TFIIA γ-subunits expressed in *P. falciparum* (Robert Milton and Thomas Oelgeschläger, unpublished) may indicate that *Pf*TBP and *Pf*TLP interact with alternate forms of *Pf*TFIIA.

## 4.2 Expression and purification of recombinant *Pf*TFIIB

The over-expression of recombinant 6His-*Pf*TFIIB was found to be toxic to the *E. coli* host cells. A method of expressing the protein was developed, whereby protein was expressed through auto-induction by release of catabolite repression. This method allowed for a large quantity of *Pf*TFIIB protein to be expressed for purification. Purification attempts were hampered by the apparent DNA-binding activity of the recombinant protein, where interactions with DNA reduced the availability of purified protein. Additionally, affinity purification with a 6His-tag was not effective in generating sufficiently pure *Pf*TFIIB for future assays. Finally, purification of *Pf*TFIIB shows the consistent presence of a doublet of protein, and while this may be indicative of the open and closed conformations studied in other TFIIB orthologues, the clear heterogeneity of the protein preparation was of concern. The inability to purify 6His-*Pf*TFIIB sufficiently has led to current work in the lab to express the protein with flanking GST- and 6His-tags.

## 4.3 Expression and purification of recombinant *Pf*TLP

Expression of recombinant *Pf*TLP was found to be extremely toxic to the host *E. coli* cells, leading to cessation of growth in cells induced to express the protein. By codon optimising the gene for expression in *E. coli*, and expressing the gene at low-temperatures, this toxicity was alleviated. The cause of the toxicity of the gene is most likely due to the over-utilisation

of codons rare in *E. coli*, as well as the A+T richness of the gene. The high pI of the protein also led to premature translational termination of the protein, and so the accumulation of GST-6His peptides as a significant contaminant in the protein preparation. Expression of the double-tagged GST-6His-*Pf*TLP allowed for sequential affinity-purification of the protein. However, heterogeneity of the protein led to a significant reduction of recovered purified protein, as large quantities of the protein were found to bind irreversibly to agarose-based resins. Finally, a preparation of sufficiently purified 6His-*Pf*TLP protein was obtained for DNA-binding assays.

## 4.4 DNA-binding potential of *Pf*TLP and *Pf*TFIIB

Immobilised template assays and EMSAs indicate that *Pf*TLP is able to bind to DNA, with greater affinity for the putative *Plasmodium* promoters of the *kahrp* and *gbp-130* genes, as compared to the Ad2ML probe. Evidence from polyacrylamide and agarose EMSAs suggest that *Pf*TLP binds to multiple locations on the DNA-probe. This is in contrast to findings that *Pf*TBP binds to distinct TATA-box like sequences in these same probes (Ruvalcaba-Salazar et al. 2006). Other eukaryotic TLPs characterised do not have affinity to the TATA-box (Akhtar & Veenstra 2011), and this may be an indication that *Pf*TLP has sequence specificity alternate to that of *Pf*TBP.

Finally, the suggestion that *Pf*TFIIB has binding activity is supported by agarose EMSA, although the underlying mechanism, the specificity of this binding, and the functional relevance requires further investigation.

## 4.5 Overall conclusions and future work

This research project was undertaken to investigate the similarity and/or divergence of *Pf*TFIIB and *Pf*TLP to their well-characterised orthologous in human. To this end, *in silico* analysis of the proteins was performed and has found that the putative *Plasmodium*

orthologues share many of the main structural and, potentially, functional features of the classically studied molecules.

Expression of the epitope-tagged molecules was also performed, and the conditions for expression of soluble forms of the molecules in *E.coli* BL21-CodonPlus RIL cells established. However, future work into optimising the expression conditions of recombinant *Pf*TLP is required, in order to reduce the expression of truncated species, and so increase the yield of purified protein.

The methodology for the affinity purification of the molecules was investigated. It was found that the addition of only a single His-tag is insufficient for the purification of recombinantly expressed protein, due to very low expression levels and multiple contaminants present after metal-ion affinity purification. The addition of a second affinity-purification tag, in this case GST, allows for far superior purification, and is recommended in the expression of other recombinant putative *Plasmodium* GTFs.

The initial investigation into the DNA-binding potential of *Pf*TLP and *Pf*TFIIB has revealed that both molecules appear to have affinity for DNA. The specificity of this binding, as well as the development of optimal binding conditions will be the focus of future work. Additionally, the interplay of various recombinant *P. falciparum* GTFs will also be investigated, in the modulation of both DNA-binding affinity and specificity.

Future work in the research group will focus on determining the sites of PIC assembly, through the mapping of sites bound by *Pf*TBP with *Pf*TFIIA and/or *Pf*TFIIB, as well as *Pf*TLP *Pf*TFIIA and/or *Pf*TFIIB, in characterised putative *Plasmodium* promoters, as well as by a SELEX strategy. The work presented in this research project has provided a basis for the expression and purification of these factors, as well as provided valuable insights into the possible structures and functions of *Pf*TLP and *Pf*TFIIB.

# Appendix

## 5.1 Accession numbers of proteins used in bioinformatics analysis

***Table 6: Protein accession numbers used for multiple sequence alignments***

| Protein | Organism | Accession Number | | Additional Notes |
|---|---|---|---|---|
| | | **NCBI** | **PlasmoDB** | |
| TFIIB | *Plasmodium berghei* ANKA | XP_678854.1 | PBANKA _020290 | Hypothetical |
| | *Plasmodium chabaudi chabaudi* | XP_740955.1 | PCHAS_020130 | Hypothetical |
| | *Plasmodium cynomolgi* strain B | XP_004220534.1 | PCYB_021360 | Hypothetical |
| | *Plasmodium falciparum* strain 3D7 | XP_001351037.1 | *PF*3D7_0110800 | Hypothetical |
| | *Plasmodium knowlesi* strain H | XP_002257702.1 | PKH_020280 | Hypothetical |
| | *Plasmodium vivax* Sal-1 | XP_001613527.1 | PVX_081260 | Hypothetical |
| | *Plasmodium yoelii yoelii* YM | ETB60715.1 | PYYM_0205800 | Hypothetical |
| | | | | |
| | *Schizosaccharomyces pombe* 972h- | NP_594229.1 | | |
| | *Homo sapiens* | NP_001505.1 | | |
| | *Drosophila melanogaster* | NP_001260349.1 | | |
| | *Encephalitozoon cuniculi* | NP_585866.1 | | Putative |
| | *Saccharomyces cerevisiae* S288c | NP_015411.1 | | Sua7p |
| | *Arabidopsis thaliana* | NP_181694.1 | | TFIIB-1 |
| | *Toxoplasma gondii ME49* | XP_002368543.1 | | Putative |
| | *Pyrococcus furiosus* DSM 3638 | NP_579106.1 | | TFB |
| | | | | |
| TLP/TBP | *Plasmodium berghei* ANKA | XP_680184.1 | PBANKA_101590 | Putative TLP |
| | *Plasmodium chabaudi chabaudi* | CDU21643.1 | PCHAS_101670 | Putative TLP |
| | *Plasmodium cynomolgi* strain B | XP_004224372.1 | PCYB_132990 | Putative TLP |
| | *Plasmodium falciparum* strain 3D7 | XP_001348441.1 | *PF*3D7_1428800.1 | Putative TLP |
| | *Plasmodium knowlesi* strain H | XP_002260735.1 | PKH_132150 | Putative TLP |
| | *Plasmodium vivax* Sal-1 | XP_001616687.1 | PVX_085105 | Putative TLP |
| | *Plasmodium yoelii yoelii* 17XNL/YM | XP_726291.1 | PYYM_1017400 | Putative TLP |
| | *Plasmodium reichenowi* CDC | CDO66755.1 | PRCDC_1428100.1 | Putative TLP |
| | *Toxoplasma gondii* ME49 | XP_002365082.1 | | Putative TLP* |
| | *Homo sapiens* | NP_004856.1 | | TLP |
| | *Xenopus laevis* | AAH90226.1 | | TLF |
| | *Danio rerio* | NP_991285.1 | | TLF |
| | *Caenorhabditis elegans* | CAB03082.2 | | TLF-1, isoform a |
| | | | | |
| | *Homo sapiens* | NP_950248.1 | | TBP2 |

| | | | |
|---|---|---|---|
| *Arabidopsis thaliana* | P28148.1 | | TBP2 |
| | | | |
| *Homo sapiens* | NP_003185.1 | | TBP |
| *Drosophila melanogaster* | NP_523805.1 | | TBP |
| *Saccharomyces cerevisiae S288c* | NP_011075.3 | | TBP |
| *Arabidopsis thaliana* | NP_187953.1 | | TBP |
| *Plasmodium falciparum strain 3D7* | P32086.1 | *PF*3D7_0506200 | TBP |
| *Toxoplasma gondii ME49* | XP_002368492.1 | | Putative TBP* |
| *Xenopus laevis* | P27633.1 | | TBP |
| *Danio rerio* | NP_956390.1 | | TBP |
| *Caenorhabditis elegans* | P32085.1 | | TBP |
| *Pyrococcus furiosus DSM 3638* | WP_011012439.1 | | TBP (archaebacterial) |
| *Methanococcus maripaludis C6* | A9A840.1 | | TBP (archaebacterial) |
| | | | |
| *Drosophila melanogaster* | AAF52600.1 | | TRF1 |

* - Conflicting descriptions for predicted *Tg*TLP and *Tg*TBP.

## 5.2 *Plasmodium* sequences used for alignment purposes

**_PfTFIIB_ amino acid sequence:**

*Plasmodium falciparum* - *PF*3D7_0110800; XP_001351037.1

MSSISPNKKFALTSASDKPITLIRTNVSNENNNVRRKMLRNFRNSRNTTCPDCKEKGIIICDTSEGTQICNGC
GMVVESKILSEEQEWRNFQNDGQSKNTDRNRVGEVSDIWLENNLTSTTFIKSSKKLQHLNMMTQINKNDQTLI
SAFNILKLICDTFFLRSNVIERAKEITKELQDMEQLKNRINNLNMLAVVYLACREAGHIKSIKELITFDRSYK
EKDLGKTINKLKKVLPSRAFVYNENISHLIYSLSNRLQLSIDLIEAIEYVVKKASTLITTSHRLNSLCGGSIH
LIVELNTNEEKNMKLPNLSQIATVCGVTTNTLKTTFKELLNAAEYIIPKYYLSEDNPKLSMLKQKYLSEDKRK
KN

Sequence Length: 367 aa

**_Pf_TLP amino acid sequence:**

*PF*3D7_1428800.2; Product: transcription initiation TFIID-like, putative

MYPPCKKKKLNNNEVTNIFLKNENNNMSVHNISMNAVLCSSLNLDNIYKYFSNCIYNPREFKCMRIDVPVTLST
VQKYIKYLNNKKEKQNDDICKMKNEQVNQTKHSNNNIKEEIKTNGSNKLSDNKLLDSSNNSSSNKILTNNKSH
ENLINISKNNSYLDDNVNKNNFFINDNNSDNKKIDTSDIKNMD*PF*IESEHIINKKLIINVSIFSNGKIICTGN
NSIEACKIAMKKIEKKLKQLNFKNIKLKKITITNILAVYNVGFSIVLPLFAQYYKSVDYDPNVFPACKVKIAL
MNDENKSNDNKEQNDNNFAWCNAKNTIDKDKSKVDIVSASIFSTGNITLTGGKSYENLQKCINILLPYLIKSK
SQH

Sequence Length: 368 aa

Underlined regions are thought to be low-complexity.

## 5.3 Predicted secondary structures for *Pf*TFIIB and *Pf*TLP

***Table 7: Secondary Structure locations predicted for PfTFIIB by Phyre***

| Region | Residue Number | Secondary Structure |
|---|---|---|
| | | |
| N-terminus | 20 - 22 | S |
| | 23 - 27 | H |
| | 36 - 43 | H |
| | | |
| Ribbon; reader; linker | 59 - 62 | S |
| | 68 - 70 | S |
| | 75 - 78 | S |
| | 121 - 122 | S |
| | 127 - 135 | H |
| | | |
| (Core) BH1 | 140 - 159 | H |
| BH2 | 163 - 178 | H |
| BH3 | 187 - 201 | H |
| BH4 | 208 - 215 | H |
| BH5 | 220 - 233 | H |
| BH1’ | 245 - 255 | H |
| BH2’ | 260 - 275 | H |
| BH3’ | 284 - 297 | H |
| BH4’ | 310 - 316 | H |
| BH5’ | 321 - 338 | H |
| | | |
| C-terminus | 341 - 343 | H |
| | 351 - 365 | H |

***Table 8: Boundaries and calculated pIs of predicted secondary structures of PfTLP***

| Region | Residues | Secondary Structure | pI |
|---|---|---|---|
| N-terminus | 6 to 8 | H | 11.11 |
| | 12 to 14 | H | 3.84 |
| | 16 to 21 | S | 9.71 |
| | 1 - 24 | | 9.98 |
| | | | |
| Structured 1 | 25 to 37 | S | 7.56 |
| | 43 - 49 | H | 6.32 |
| | 57-59 | H | 6.41 |
| | 62 - 64 | S | 8.55 |
| | 25 - 66 | | 8.05 |
| | | | |
| LCR 1 | 67 - 81 | H | 9.94 |
| | 89 - 105 | H | 7.36 |
| | 110 - 115 | S | 6.53 |
| | 127 - 131 | S | 3.74 |
| | 158 - 159 | S | 6.09 |
| | 160 -165 | H | 4.10 |
| | 176 -180 | H | 9.54 |
| | 67 - 202 | | 9.39 |
| | | | |
| Structured 2 | 203 - 209 | S | 6.1 |
| | 212 - 217 | S | 8.55 |
| | 222 - 238 | H | 10.51 |
| | 247 -259 | S | 10.25 |
| | 266 - 272 | H | 6.09 |
| | 285 - 290 | S | 9.68 |
| | 203 - 290 | | 10.34 |
| | | | |
| LCR 2 | 291-327 | | 4.46 |
| | | | |
| Structured 3 | 328 - 334 | S | 6.10 |
| | 338 - 342 | S | 6.10 |
| | 347 - 365 | H | 9.26 |
| | 328 - 368 | | 9.84 |
| | | | |
| | 1 - 368 | | 9.71 |

## 5.4 Vector information for this study

***Table 9: Vectors used for gene cloning and protein expression***

Vectors which were used for protein expression purposes are in **bold**. Vectors used for further vector construction are in *italics*. *Pf* stands for *P. falciparum*. *Hs* stands for *H. sapiens.*

| # | Vector | Purification Tag(s) | Protease Site(s) | Size (bp) |
|---|---|---|---|---|
| 1 | *pET11d-6His-HsTBP* | His | Thrombin | 7331 |
| 2 | **pET11d-6His-*Pf*TLP** | His | Thrombin | 6810 |
| 3 | **pET11d-6His-*Pf*TFIIB** | His | Thrombin | 6807 |
| 4 | *pET11d-GST-6His* | GST; His | Thrombin; TEV | 6402 |
| 5 | *pET11d-GST-6His-PfTBP'FL'* | GST; His | Thrombin; TEV | 7362 |
| 6 | **pET11d-GST-6His-*Pf*TLP** | GST; His | Thrombin; TEV | 7485 |
| 7 | **pET11d-GST-6His-*Pf*TLPco*** | GST; His | Thrombin; TEV | 7485 |
| 8 | **pET11d-GST-6His-*Pf*TLPco(aa)*** | GST; His | Thrombin; TEV | 7485 |

* - vector 8 and 7 are similar, with an additional two amino acids (T and A) present in 8, preceding the starting amino acid (M) of the protein sequence in 7, see Table 10 in the Appendix.

The pET11d-6His-*Pf*TFIIB and pET11d-6His-*Pf*TLP vectors express the target genes fused to an N-terminal hexa-histidine (6His) epitope tag for metal-ion affinity purification. A thrombin cleavage site is located between the hexa-histidine tag and the target protein which allows for the removal of the purification tag through thrombin digestion (Figures 43-45).

The pET11d-GST-6His-*Pf*TLP and pET11d-GST-6His-*Pf*TLPco vectors express the target gene fused to an N-terminal glutathione S-transferase (GST) tag followed by a hexa-histidine tag. A thrombin cleavage site is located between the GST and hexa-histidine tags. A tobacco etch virus (TEV) protease cleavage site is positioned between the hexa-histidine site and the target protein and allows for removal of the purification tag(s) by recombinant TEV protease digestion (Figure 47).

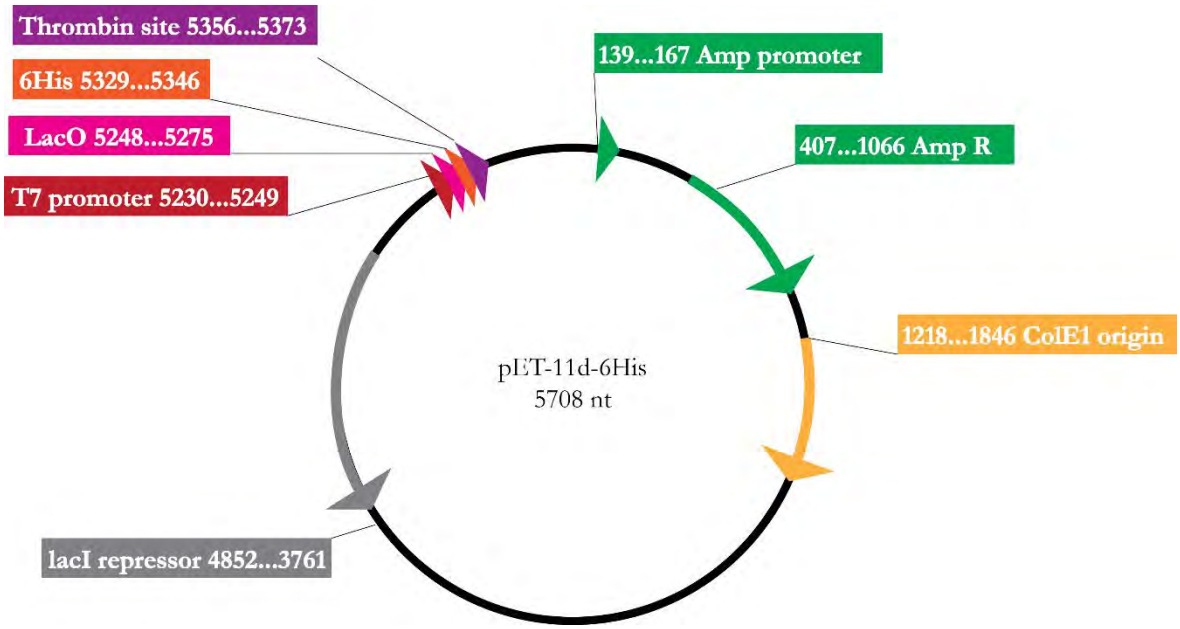



***Figure 43: Plasmid maps of pET11d-6His.***

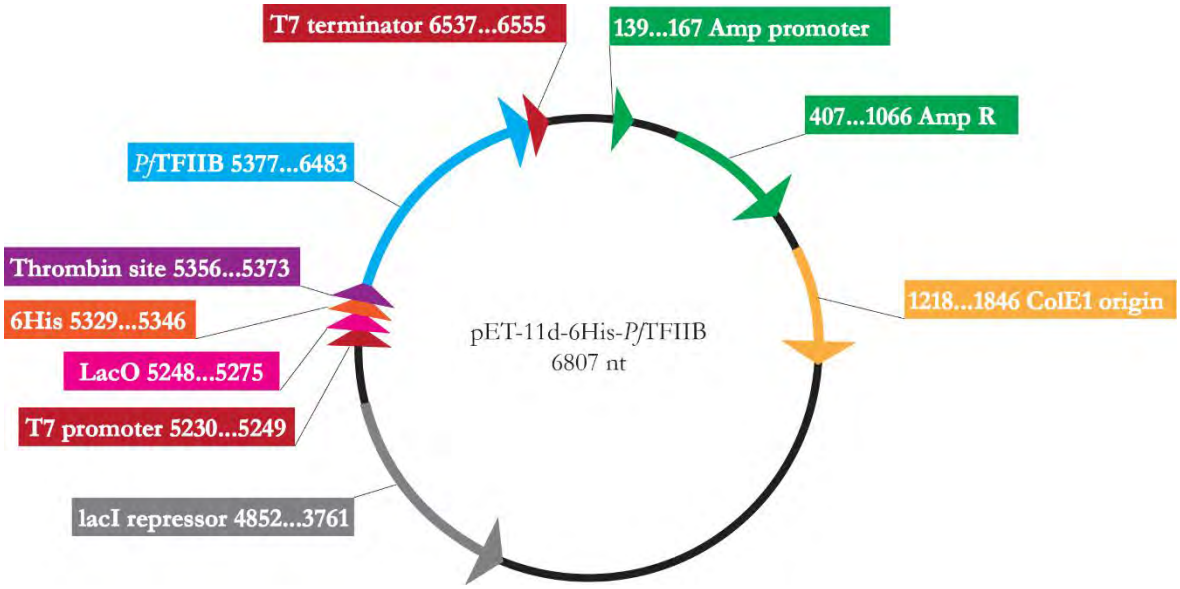



***Figure 44: Plasmid map of pET11d-6His-PfTFIIB.***

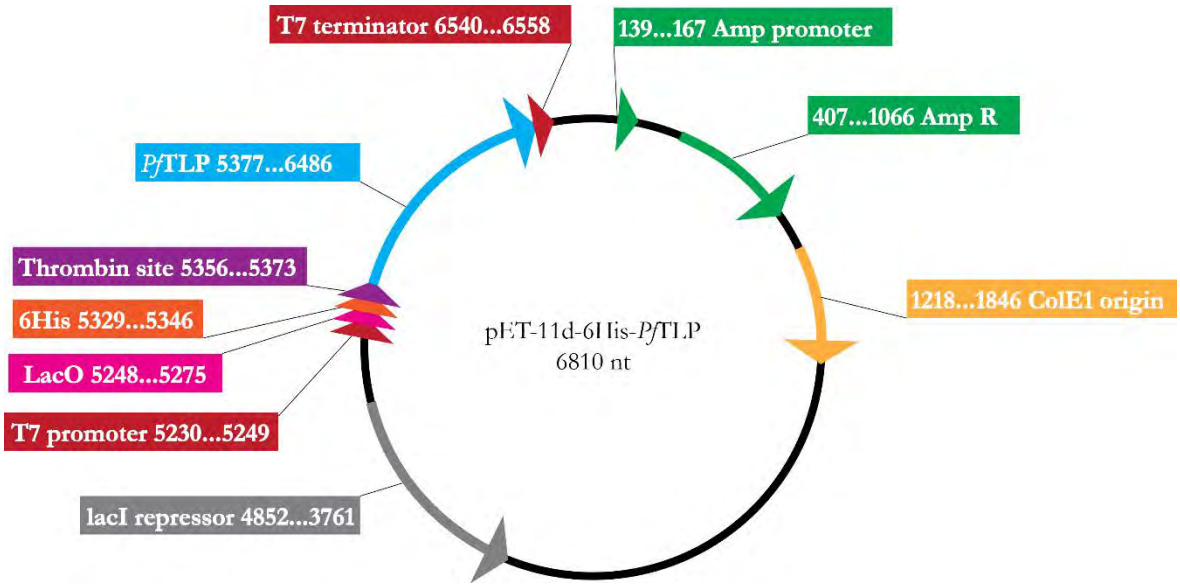



*Figure 45: pET11d-6His-PfTLP.*

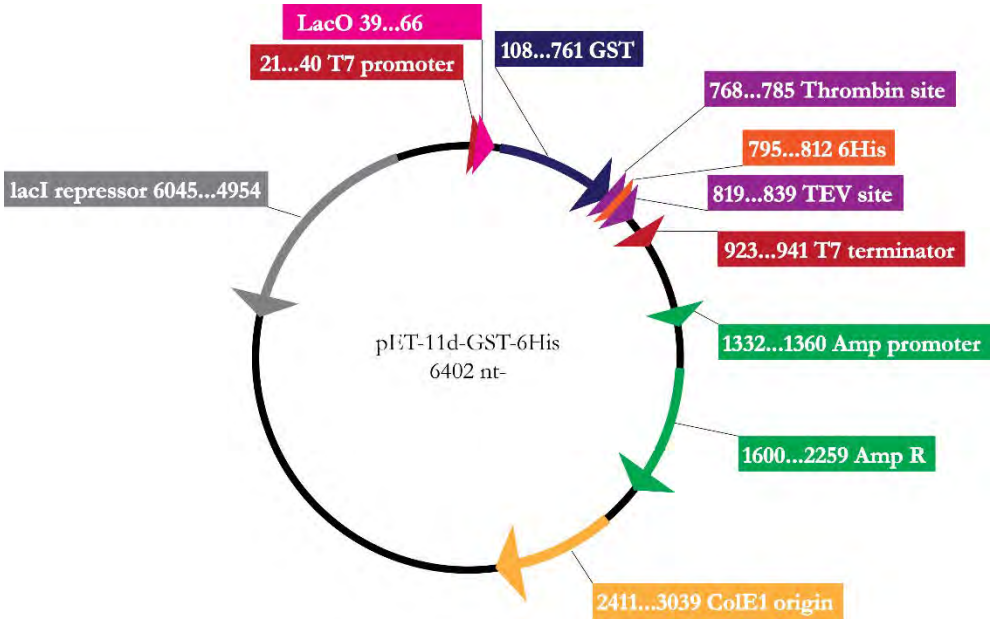



*Figure 46: Plasmid map of pET11d-GST-6His.*

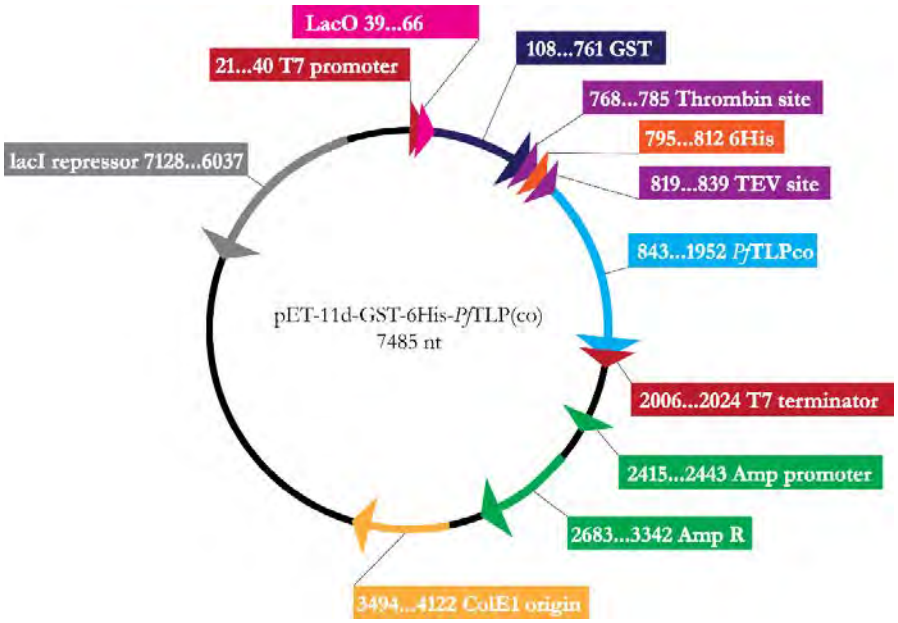



***Figure 47: pET11d-GST-6His-PfTLP[(co)].***

## 5.5 Nickel affinity purification of soluble proteins expressed in *E.coli* BL21-codonPlus® (DE3)-RIL *E. coli* cells not carrying an expression vector

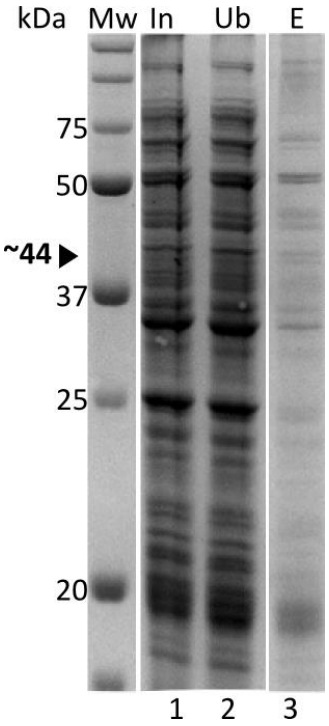



***Figure 48: Ni-affinity purification of E. coli BL21-codonPlus® (DE3)-RIL E. coli cells not carrying an expression vector.*** Cells were grown and induced for protein expression with 1mM IPTG). Soluble protein was purified with nickel-affinity purification and analysed by SDS-PAGE. Lanes 1-3 represent the cleared lysate input (**In**, 25/1000 vol.), the protein fraction not bound to the Ni-beads (**Ub**, equivalent to In), and the protein fraction eluted from the Ni-beads with 2×SDS-loading buffer (**E**, ½ vol. loaded) respectively. A black arrow indicates 44kDa.

## 5.6 SDS-PAGE showing the contaminants present in 6His-*Pf*TFIIB protein preparation

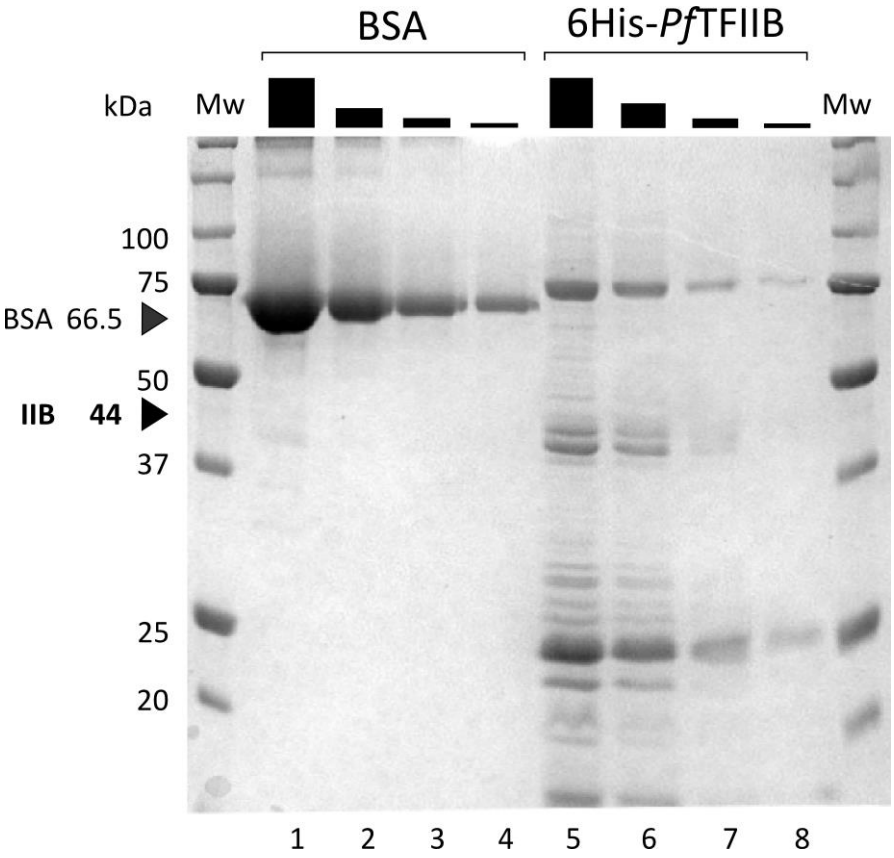



***Figure 49: Analysis of quantity and purity of 6His-PfTFIIB protein purified.***

BSA standards were loaded in lanes 1-4, 5µg, 2µg, 1µg and 0.5µg respectively. Lanes 5-8, 10µl, 5µl, 2µl, and 1µl of pooled protein fractions respectively (1.4ml total). Expected size for 6His-*Pf*TFIIB is 44kDa, indicated with a black arrow. Molecular weight (**Mw**) – 5µl of Precision Plus Protein All Blue standard (*Bio-Rad*).

## 5.7 Rare codon analysis of *Pf*TLP ORF

atg tat CCC cct tgt aaa aag aaa aaa ctt aat aat aac gag gta aca
aat att ttt tta aaa aat gaa aat aac atg agt gtt cat aat att tcc
atg aat gct gtt tta tgt tcc tca tta aac CTA gat aat att tat aaa
tat ttt tca aat tgt ATA tat aac cct aga gaa ttt aaa tgt
atg aga att gat gtc CCC gtc act tta agt aca gta caa aaa tat ATA
aaa tat ttg aat aat aaa aag gaa aaa caa aat gat gat ATA tgt aaa
atg aaa aat gaa caa gtc aat caa aca aaa cat tct aat aat aat att
aaa gaa gag att aaa acc aat gga agt aat aaa tta agt gat aat
aaa CTA tta gat agt agt aac aat tca tcg agt aat aaa ATA tta aca
aat aat aaa tca cat gaa aat CTA ATA aat ATA tct aaa aat aat tca
tat tta gat gat aat gta aat aag aat aat ttt ttt ATA aat gat aat
aat agt gat aat aaa aag ATA gat act tct gat att aaa aac atg gat
cct ttt att gaa tct gaa cat ATA ATA aat aaa aaa tta att att aat
gtt tcc att ttt tca aat gga aaa att ATA tgt aca ggt aat aac
tca ATA gaa gct tgt aaa att gct atg aaa aaa att gaa aag aaa tta
aaa caa tta aat ttt aaa aat att aaa tta aaa aaa ATA act att aca
aat att ttg gct gta tat aat gtg gga ttt tct ATA gtc ttg
ccg CTA ttt gct caa tat tat aaa agt gtg gat tat gat cca aat gta
ttt cca gcg tgc aaa gta aaa att gct ctt atg aat gat gag aat aag
tct aat gac aat aag gag cag aat gat aat aat ttt gct tgg tgt aat
gct aaa aac aca ATA gac aag gat aaa agc aaa gtg gac att gtt tct
gct agt att ttt tca aca ggt aat ATA aca tta act ggg ggg aaa agt
tac gaa aac ctt caa aaa tgt ATA aat ATA tta tta cca tat tta att
aaa agt aaa tcg caa cat taa tag

Red = rare Arg codons AGG, AGA, CGA - 2
Green = rare Leu codon CTA - 4
Blue = rare Ile codon ATA - 18
Orange = rare Pro codon CCC - 2

Total number of rare codons: - 26

Rare codons derived from (Zhang et al. 1991)

## 5.8 DNA alignment of the optimized regions (red) of the *Pf*TLP gene

Optimized 1
ATGTATCCGCCGTGTAAAAAGAAGAAACTGAACAATAATGAAGTGACCAACATCTTCCTG
Original 1
ATGTATCCCCCTTGTAAAAAGAAAAAACTTAATAATAACGAGGTAACAAATATTTTTTTA
Optimized 61
AAAAATGAAAACAATATGAGCGTCCATAATATTTCAATGAACGCAGTTCTGTGCAGCTCT
Original 61
AAAAATGAAAATAACATGAGTGTTCATAATATTTCCATGAATGCTGTTTTATGTTCCTCA
Optimized 121
CTGAATCTGGATAACATCTACAAATACTTCTCGAACTGCATCTACAACCCGCGTGAATTC
Original 121
TTAAACCTAGATAATATTTATAAATATTTTTCAAATTGTATATATAACCCTAGAGAATTT
Optimized 181
AAATGTATGCGCATTGATGTGCCGGTTACCCTGAGTACGGTGCAGAAATACATCAAATAC
Original 181
AAATGTATGAGAATTGATGTCCCCGTCACTTTAAGTACAGTACAAAAATATATAAAATAT
Optimized 241
CTGAACAACAAAAAAGAAAAACAAAACGATGACATTTGTAAAATGAAAAATGAACAGGTT
Original 241
TTGAATAATAAAAAGGAAAAACAAAATGATGATATATGTAAAATGAAAAATGAACAAGTC

```
Optimized                                                              301
AACCAAACCAAACATAGCAACAACAACATCAAAGAAGAAATCAAAACGAACGGCTCGAAC
Original                                                               301
AATCAAACAAAACATTCTAATAATAATATTAAAGAAGAGATTAAAACCAATGGAAGTAAT
Optimized                                                              361
AAACTGAGCGATAACAAACTGCTGGACAGTTCCAATAACTCATCGAGCAATAAAATTCTG
Original                                                               361
AAATTAAGTGATAATAAACTATTAGATAGTAGTAACAATTCATCGAGTAATAAAATATTA
Optimized                                                              421
ACCAATAACAAATCCCACGAAAACCTGATCAACATCTCTAAAAACAACAGTTACCTGGAT
Original                                                               421
ACAAATAATAAATCACATGAAAATCTAATAAATATATCTAAAAATAATTCATATTTAGAT
Optimized                                                              481
GACAATGTCAACAAAAACAACTTTTTCATCAACGATAACAACTCCGACAACAAGAAAATT
Original                                                               481
GATAATGTAAATAAGAATAATTTTTTTATAAATGATAATAATAGTGATAATAAAAAGATA
Optimized                                                              541
GATACCTCAGACATCAAAAACATGGACCCGTTTATTGAAAGCGAACATATCATCAACAAA
Original                                                               541
GATACTTCTGATATTAAAAACATGGATCCTTTTATTGAATCTGAACATATAATAAATAAA
Optimized                                                              601
AAACTGATCATCAACGTGTCCATCTTCTCAAATGGCAAAATTATCTGCACCGGTAATAAC
Original                                                               601
AAATTAATTATTAATGTTTCCATTTTTTCAAATGGAAAAATTATATGTACAGGTAATAAC
Optimized                                                              661
TCTATTGAAGCGTGTAAAATCGCCATGAAGAAAATTGAGAAAAAACTGAAACAGCTGAAC
Original                                                               661
TCAATAGAAGCTTGTAAAATTGCTATGAAAAAAATTGAAAAGAAATTAAAACAATTAAAT
Optimized                                                              721
TTCAAAAACATCAAACTGAAGAAAATTACCATCACGAACATTCTGGCAGTCTATAATGTG
Original                                                               721
TTTAAAAATATTAAATTAAAAAAAATAACTATTACAAATATTTTGGCTGTATATAATGTG
Optimized                                                              781
GGCTTTAGTATCGTGCTGCCGCTGTTCGCTCAATATTACAAATCCGTGGATTACGACCCG
Original                                                               781
GGATTTTCTATAGTCTTGCCGCTATTTGCTCAATATTATAAAAGTGTGGATTATGATCCA

Optimized                                                              841
AACGTTTTCCCGGCGTGCAAAGTCAAAATTGCCCTGATGAACGATGAAAACAAATCTAAC
Original                                                               841
AATGTATTTCCAGCGTGCAAAGTAAAAATTGCTCTTATGAATGATGAGAATAAGTCTAAT
Optimized                                                              901
GACAACAAAGAACAGAACGATAACAACTTCGCATGGTGTAACGCTAAAAACACCATCGAT
Original                                                               901
GACAATAAGGAGCAGAATGATAATAATTTTGCTTGGTGTAATGCTAAAAACACAATAGAC
Optimized                                                              961
AAAGACAAATCGAAAGTTGATATCGTCTCTGCCAGTATTTTCAGCACGGGTAACATCACC
Original                                                               961
AAGGATAAAAGCAAAGTGGACATTGTTTCTGCTAGTATTTTTTCAACAGGTAATATAACA
Optimized                                                             1021
CTGACGGGCGGTAAATCTTATGAAAATCTGCAAAAATGTATTAACATCCTGCTGCCGTAC
Original                                                              1021
TTAACTGGGGGGAAAAGTTACGAAAACCTTCAAAAATGTATAAATATATTATTACCATAT
Optimized 1081  CTGATTAAAAGTAAAAGCCAACACTGATAA
Original  1081  TTAATTAAAAGTAAATCGCAACATTAATAG
```

## 5.9 Analysis of additional codons in pET11d-GST-6His-*Pf*TLPco(aa)

***Table 10: Mutated sequence in pET11d-GST-6His-PfTLPco*(aa)**

Underlined – *Pf*TLP ORF

Bold – inserted codon/amino acids

| | **Molecule** | **Sequence** |
|---|---|---|
| Wild type | cDNA | 5'-…cat atg tat ccg ccg…-3' |
| Mutated | cDNA | 5'-…cat a**CA GCC A**tg tat ccg ccg…-3' |
| Wild type | AA | N-… H M Y P P …-C |
| Mutated | AA | N-… H **T A** M Y …-C |

# References


Akhtar, W. & Veenstra, G.J.C., 2009. TBP2 is a substitute for TBP in Xenopus oocyte transcription. *BMC biology*, 7, p.45.

Akhtar, W. & Veenstra, G.J.C., 2011. TBP-related factors: a paradigm of diversity in transcription initiation. *Cell & bioscience*, 1(1), p.23.

Alano, P., 2007. Plasmodium falciparum gametocytes: Still many secrets of a hidden life. *Molecular Microbiology*, 66(September), pp.291–302.

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

Aurrecoechea, C. et al., 2009. PlasmoDB: a functional genomic database for malaria parasites. *Nucleic acids research*, 37(Database issue), pp.D539–43.

Bachmann, A. et al., 2012. Temporal expression and localization patterns of variant surface antigens in clinical Plasmodium falciparum isolates during erythrocyte schizogony. *PloS one*, 7(11), p.e49540.

Baneyx, F., 1999. Recombinant protein expression in Escherichia coli. *Current Opinion in Biotechnology*, 10(5), pp.411–421.

Bartfai, R. et al., 2004. TBP2, a vertebrate-specific member of the TBP family, is required in embryonic development of zebrafish. *Current biology : CB*, 14(7), pp.593–598.

Bentley, W.E. et al., 1990. Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnology and bioengineering*, 35(7), pp.668–681.

Birkholtz, L.-M. et al., 2008a. Heterologous expression of plasmodial proteins for structural studies and functional annotation. *Malaria journal*, 7, p.197.

Birkholtz, L.-M. et al., 2008b. Heterologous expression of plasmodial proteins for structural studies and functional annotation. *Malaria journal*, 7, p.197.

Bischoff, E. & Vaquero, C., 2010. In silico and biological survey of transcription-associated proteins implicated in the transcriptional machinery during the erythrocytic development of Plasmodium falciparum. *BMC genomics*, 11, p.34.

Bolanos-Garcia, V.M. & Davies, O.R., 2006. Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. *Biochimica et biophysica acta*, 1760(9), pp.1304–13.

Bozdech, Z., Zhu, J., et al., 2003. Expression profiling of the schizont and trophozoite stages of Plasmodium falciparum with a long-oligonucleotide microarray. *Genome biology*, 4, p.R9.

Bozdech, Z., Llinás, M., et al., 2003. The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. *PLoS Biology*, 1(1), pp.85–100.

Brick, K., Watanabe, J. & Pizzi, E., 2008. Core promoters are predicted by their distinct physicochemical properties in the genome of Plasmodium falciparum. *Genome biology*, 9(12), p.R178.

Buendía-Orozco, J., Guerrero, A. & Pastor, N., 2005. Model of the TBP-TFIIB complex from Plasmodium falciparum: Interface analysis and perspectives as a new target for antimalarial design. *Archives of Medical Research*, 36, pp.317–330.

Buffet, P.A. et al., 2010. The pathogenesis of Plasmodium falciparum malaria in humans: insights from splenic physiology. *Blood*, 117(2), pp.381–392.

Bushnell, D. a et al., 2004. Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science (New York, N.Y.)*, 303(August 1999), pp.983–988.

Butler, J.E.F. & Kadonaga, J.T., 2002. The RNA polymerase II core promoter : a key component in the regulation of gene expression. , pp.2583–2592.

Chakrabarti, D. et al., 1994. Analysis of expressed sequence tags from Plasmodium falciparum. *Mol Biochem Parasitol*, 66(1), pp.97–104.

Chen, B. & Hampsey, M., 2004. Functional Interaction between TFIIB and the Rpb2 Subunit of RNA Polymerase II : Implications for the Mechanism of Transcription Initiation. *Society*, 24(9), pp.3983–3991.

Chen, H.-T. & Hahn, S., 2003. Binding of TFIIB to RNA polymerase II: Mapping the binding site for the TFIIB zinc ribbon domain within the preinitiation complex. *Molecular cell*, 12(2), pp.437–47.

Chen, H.-T. & Hahn, S., 2004. Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: a model for the structure of the PIC. *Cell*, 119(2), pp.169–80.

Cianfrocco, M.A. & Nogales, E., 2013. Regulatory interplay between TFIID's conformational transitions and its modular interaction with core promoter DNA. *Transcription*, 4(June), pp.1–7.

Coulson, R., Hall, N. & Ouzounis, C., 2004. Comparative genomics of transcriptional control in the human malaria parasite Plasmodium falciparum. *Genome research*, pp.1–7.

Cowman, A.F. & Crabb, B.S., 2006. Invasion of red blood cells by malaria parasites. *Cell*, 124, pp.755–766.

Cui, L. & Miao, J., 2010. Chromatin-Mediated epigenetic regulation in the malaria parasite Plasmodium falciparum. *Eukaryotic Cell*, 9(8), pp.1138–1149.

Dantonel, J. et al., 2000. TBP-like factor is required for embryonic RNA polymerase II transcription in C. elegans. *Molecular cell*, 6, pp.715–722.

Decker, K.B. & Hinton, D.M., 2013. Transcription regulation at the core: similarities among bacterial, archaeal, and eukaryotic RNA polymerases. *Annual review of microbiology*, 67, pp.113–39.

Deng, W. et al., 2009. TFIIB recognition elements control the TFIIA-NC2 axis in transcriptional regulation. *Molecular and cellular biology*, 29(6), pp.1389–400.

Deng, W. & Roberts, S., 2005. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes & development*, pp.2418–2423.

Deng, W. & Roberts, S.G.E., 2007. TFIIB and the regulation of transcription by RNA polymerase II. *Chromosoma*, 116(5), pp.417–29.

Dumon-Seignovert, L., Cariot, G. & Vuillard, L., 2004. The toxicity of recombinant proteins in Escherichia coli: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein expression and purification*, 37(1), pp.203–206.

Elsby, L.M. & Roberts, S.G.E., 2004. The role of TFIIB conformation in transcriptional regulation. *Biochemical Society transactions*, 32(Pt 6), pp.1098–9.

Fairley, J. a et al., 2002. Core promoter-dependent TFIIB conformation and a role for TFIIB conformation in transcription start site selection. *Molecular and cellular biology*, 22(19), pp.6697–6705.

Florens, L. et al., 2002. A proteomic view of the Plasmodium falciparum life cycle. *Nature*, 419(6906), pp.520–6.

Francis, S.E., Sullivan, D.J. & Goldberg, D.E., 1997. Hemoglobin metabolism in the malaria parasite Plasmodium falciparum. *Annual review of microbiology*, 51, pp.97–123.

Frugier, M. et al., 2010. Low Complexity Regions behave as tRNA sponges to help co-translational folding of plasmodial proteins. *FEBS letters*, 584(2), pp.448–54.

Gardner, M.J. et al., 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*, 419.

Garner, M.M. & Revzin, A., 1981. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic acids research*, 9(13), pp.3047–3060.

Gazdag, E. et al., 2007. Analysis of TATA-binding protein 2 (TBP2) and TBP expression suggests different roles for the two proteins in regulation of gene expression during oogenesis and early mouse development. *Reproduction (Cambridge, England)*, 134(1), pp.51–62.

Geiger, J.H. et al., 1996. Crystal Structure of the Yeast TFIIA/TBP/DNA Complex. *Science* , 272 (5263 ), pp.830–836.

Gietl, A. et al., 2014. Eukaryotic and archaeal TBP and TFB/TF(II)B follow different promoter DNA bending pathways. *Nucleic Acids Research*, 42(10), pp.6219–6231.

Gissot, M. et al., 2005. PfMyb1, a Plasmodium falciparum transcription factor, is required for intra-erythrocytic growth and controls key genes for cell cycle regulation. *Journal of Molecular Biology*, 346, pp.29–42.

Glossop, J. a., Dafforn, T.R. & Roberts, S.G.E., 2004. A conformational change in TFIIB is required for activator-mediated assembly of the preinitiation complex. *Nucleic Acids Research*, 32(5), pp.1829–1835.

Greenberg, A.E. & Lobel, H.O., 1990. Mortality from Plasmodium falciparum Malaria in Travelers from the United States, 1959 to 1987. *Annals of Internal Medicine*, 113(4), pp.326–327.

Grünberg, S. & Hahn, S., 2013. Structural insights into transcription initiation by RNA polymerase II. *Trends in biochemical sciences*, 38(12), pp.603–11.

Grünberg, S., Warfield, L. & Hahn, S., 2012. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nature structural & molecular biology*, 19(8), pp.788–96.

Guizetti, J. & Scherf, A., 2013. Silence, activate, poise and switch! Mechanisms of antigenic variation in Plasmodium falciparum. *Cellular Microbiology*, 15(February), pp.718–726.

Hellman, L.M. & Fried, M.G., 2007. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat. Protocols*, 2(8), pp.1849–1861.

Hobbs, N.K. et al., 2002. Removing the Vertebrate-Specific TBP N Terminus Disrupts Placental β2m-Dependent Interactions with the Maternal Immune System. *Cell*, 110(1), pp.43–54.

Hochheimer, A. & Tjian, R., 2003. Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. *Genes and Development*, 17, pp.1309–1320.

Hoeijmakers, W. a M., Stunnenberg, H.G. & Bártfai, R., 2012. Placing the Plasmodium falciparum epigenome on the map. *Trends in Parasitology*, 28(11), pp.486–495.

Horrocks, P. et al., 2009. Control of gene expression in Plasmodium falciparum - Ten years on. *Molecular and Biochemical Parasitology*, 164, pp.9–25.

Horrocks, P., Dechering, K. & Lanzer, M., 1998. Control of gene expression in Plasmodium falciparum. *Molecular and biochemical parasitology*, 95, pp.171–181.

Imbalzano, A.N., Zaret, K.S. & Kingston, R.E., 1994. Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. *Journal of Biological Chemistry*, 269(11), pp.8280–8286.

Jallow, Z. et al., 2004. Specialized and redundant roles of TBP and a vertebrate-specific TBP paralog in embryonic gene regulation in Xenopus. *Proceedings of the National Academy of Sciences of the United States of America*, 101(37), pp.13525–13530.

Jurgelenaite, R. et al., 2009. Gene regulation in the intraerythrocytic cycle of Plasmodium falciparum. *Bioinformatics*, 25(12), pp.1484–1491.

Juven-Gershon, T. et al., 2008. The RNA polymerase II core promoter — the gateway to transcription. *Current Opinion in Cell Biology*, 20(3), pp.253–259.

Karunaweera, N.D. et al., 1992. Dynamics of fever and serum levels of tumor necrosis factor are closely associated during clinical paroxysms in Plasmodium vivax malaria. *Proceedings of the National Academy of Sciences* , 89 (8 ), pp.3200–3203.

Kelley, L. a & Sternberg, M.J.E., 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nature protocols*, 4(3), pp.363–71.

Khoury, D.S. et al., 2014. Effect of mature blood-stage Plasmodium parasite sequestration on pathogen biomass in mathematical and in vivo models of malaria. *Infection and Immunity*, 82(1), pp.212–220.

Kim, J.L. & Burley, S.K., 1994. 1.9 A resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nat Struct Mol Biol*, 1(9), pp.638–653.

Kokubo, T. et al., 1998. The yeast TAF145 inhibitory domain and TFIIA competitively bind to TATA-binding protein. *Molecular and cellular biology*, 18(2), pp.1003–1012.

Komaki-Yasuda, K. et al., 2013. Identification of a Novel and Unique Transcription Factor in the Intraerythrocytic Stage of Plasmodium falciparum. *PLoS ONE*, 8(9).

Kostrewa, D. et al., 2009. ARTICLES RNA polymerase II – TFIIB structure and mechanism of transcription initiation. *Nature*, 462(7271), pp.323–330.

Lagrange, T. et al., 1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes & development*, 12(1), pp.34–44.

Lanzer, M. et al., 1993. Plasmodium: Control of Gene Expression in Malaria Parasites. *Experimental Parasitology*, 77(1), pp.121–128.

Lanzer, M., de Bruin, D. & Ravetch, J. V, 1992a. A sequence element associated with the Plasmodium falciparum KAHRP gene is the site of developmentally regulated protein-DNA interactions. *Nucleic acids research*, 20(12), pp.3051–3056.

Lanzer, M., de Bruin, D. & Ravetch, J. V, 1992b. Transcription mapping of a 100 kb locus of Plasmodium falciparum identifies an intergenic region in which transcription terminates and reinitiates. *The EMBO journal*, 11(5), pp.1949–1955.

Larkin, M. a et al., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, 23(21), pp.2947–8.

Lescure, A. et al., 1994. The N-terminal domain of the human TATA-binding protein plays a role in transcription from TATA-containing RNA polymerase II and III promoters. *The EMBO Journal*, 13(5), pp.1166–1175.

Marchler-Bauer, A. et al., 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research*, 39(suppl 1), pp.D225–D229.

Marchler-Bauer, A. et al., 2009. CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Research*, 37(suppl 1), pp.D205–D210.

Marchler-Bauer, A. & Bryant, S.H., 2004. CD-Search: protein domain annotations on the fly. *Nucleic Acids Research*, 32(suppl 2), pp.W327–W331.

Martianov, I. et al., 2001. Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene. *Molecular Cell*, 7, pp.509–515.

McAndrew, M.B. et al., 1993. Characterisation of the gene encoding an unusually divergent TATA-Binding Protein (TBP) from the extremely A+T-rich human malaria parasite Plasmodium falciparum. *Gene*, 124(2), pp.165–171.

McWilliam, H. et al., 2013. Analysis Tool Web Services from the EMBL-EBI. *Nucleic acids research*, 41(Web Server issue), pp.W597–600.

Mehlin, C. et al., 2006. Heterologous expression of proteins from Plasmodium falciparum: results from 1000 genes. *Molecular and biochemical parasitology*, 148(2), pp.144–60.

Miller, L.H. et al., 2013. Malaria biology and disease pathogenesis: insights for new treatments. *Nature medicine*, 19(2), pp.156–67.

Monera, O.D. et al., 1995. Relationship of sidechain hydrophobicity and α-helical propensity on the stability of the single-stranded amphipathic α-helix. *Journal of Peptide Science*, 1(5), pp.319–329.

Moore, P.A. et al., 1999. A human TATA binding protein-related protein with altered DNA binding specificity inhibits transcription from multiple promoters and activators. *Molecular and cellular biology*, 19(11), pp.7610–7620.

Neva, F.A. & Brown, H.W., 1994. *Basic Clinical Parasitology*, Appleton & Lange.

Nikolov, D.B. et al., 1996. Crystal structure of a human TATA box-binding protein/TATA element complex. *Proceedings of the National Academy of Sciences of the United States of America*, 93(10), pp.4862–4867.

Nikolov, D.B. et al., 1995. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature*, 377(6545), pp.119–128.

Ogawa, N. & Biggin, M.D., 2012. High-throughput SELEX determination of DNA sequences bound by transcription factors in vitro. In *Gene Regulatory Networks*. Springer, pp. 51–63.

Ohbayashi, T. et al., 2003. Vertebrate TBP-like protein (TLP/TRF2/TLF) stimulates TATA-less terminal deoxynucleotidyl transferase promoters in a transient reporter assay, and TFIIA-binding capacity of TLP is required for this function. *Nucleic Acids Research*, 31(8), pp.2127–2133.

Ohbayashi, T. et al., 2003. Vertebrate TBP-like protein (TLP/TRF2/TLF) stimulates TATA-less terminal deoxynucleotidyl transferase promoters in a transient reporter assay, and TFIIA-binding capacity of TLP is required for this function. *Nucleic acids research*, 31(8), pp.2127–2133.

Painter, H.J., Campbell, T.L. & Llinás, M., 2011. The Apicomplexan AP2 family: Integral factors regulating Plasmodium development. *Molecular and Biochemical Parasitology*, 176(1), pp.1–7.

Papadopoulos, J.S. & Agarwala, R., 2007. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* , 23 (9 ), pp.1073–1079.

Perkins, M.E., 1984. Surface proteins of Plasmodium falciparum merozoites binding to the erythrocyte receptor, glycophorin. *The Journal of Experimental Medicine* , 160 (3 ), pp.788–798.

Perlman, R.L., de Crombrugghe, B. & Pastan, I., 1969. Cyclic AMP regulates catabolite and transient repression in E. coli. *Nature*, 223, pp.810–812.

Pettersen, E.F. et al., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13), pp.1605–1612.

Qing, G. et al., 2004. Cold-shock induced high-yield protein production in Escherichia coli. *Nature biotechnology*, 22(7), pp.877–882.

Qureshi, S. a & Jackson, S.P., 1998. Sequence-specific DNA binding by the S. shibatae TFIIB homolog, TFB, and its effect on promoter strength. *Molecular cell*, 1, pp.389–400.

Van Roon-Mom, W.M.C. et al., 2005. TATA-binding protein in neurodegenerative disease. *Neuroscience*, 133(4), pp.863–872.

Rosano, G.L. & Ceccarelli, E. a, 2009. Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted Escherichia coli strain. *Microbial cell factories*, 8, p.41.

Rosano, G.L. & Ceccarelli, E.A., 2014. Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology*, 5, p.172.

Ruvalcaba-Salazar, O.K. et al., 2005. Recombinant and native Plasmodium falciparum TATA-binding-protein binds to a specific TATA box element in promoter regions. *Molecular and biochemical parasitology*, 140(2), pp.183–96.

Ruvalcaba-Salazar, O.-K. et al., 2006. Preparation and characterization of a monoclonal antibody specific to Plasmodium falciparum TATA binding protein. *Hybridoma (2005)*, 25(6), pp.367–71.

Sambrook, J., Fritsch, E.F. & Maniatis, T., 1989. *Molecular cloning*, Cold spring harbor laboratory press New York.

Schmidt, E.E. et al., 2003. Fundamental cellular processes do not require vertebrate-specific sequences within the TATA-binding protein. *The Journal of biological chemistry*, 278(8), pp.6168–6174.

Schrödinger, LLC, 2010. *The {PyMOL} Molecular Graphics System, Version~1.3r1*,

Shaw, S.P. et al., 1996. Identifying a species-specific region of yeast TF11B in vivo. *Molecular and cellular biology*, 16(7), pp.3651–3657.

Shaw, S.P. et al., 1997. Mutational studies of yeast transcription factor IIB in vivo reveal a functional surface important for gene activation. *Proceedings of the National Academy of Sciences of the United States of America*, 94(6), pp.2427–32.

Shock, J.L., Fischer, K.F. & DeRisi, J.L., 2007. Whole-genome analysis of mRNA decay in Plasmodium falciparum reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle. *Genome biology*, 8(7), p.R134.

Sievers, F. et al., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(539), p.539.

De Silva, E.K. et al., 2008. Specific DNA-binding by apicomplexan AP2 transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, 105(16), pp.8393–8398.

Sinha, A. et al., 2014. A cascade of DNA-binding proteins for sexual commitment and development in Plasmodium. *Nature*, 507(7491), pp.253–7.

Smith, D.B. & Johnson, K.S., 1988. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene*, 67(1), pp.31–40.

Sørensen, H.P. & Mortensen, K.K., 2005. Advanced genetic strategies for recombinant protein expression in Escherichia coli. *Journal of Biotechnology*, 115, pp.113–128.

States, D.J. & Gish, W., 1994. Combined use of sequence similarity and codon bias for coding region identification. *Journal of computational biology : a journal of computational molecular cell biology*, 1(1), pp.39–50.

Stratagene, 2005. BL21-CodonPlus Competent Cells Instruction Manual.

Su, X.Z. et al., 1996. Reduced extension temperatures required for PCR amplification of extremely A+T-rich DNA. *Nucleic acids research*, 24(8), pp.1574–5.

Tamura, K. et al., 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30(12), pp.2725–2729.

Thomas, M.C. & Chiang, C.-M., 2006. The general transcription machinery and general cofactors. *Critical reviews in biochemistry and molecular biology*, 41(3), pp.105–78.

Tsai, F.T. & Sigler, P.B., 2000. Structural basis of preinitiation complex assembly on human pol II promoters. *The EMBO journal*, 19(1), pp.25–36.

Tubon, T.C., Tansey, W.P. & Herr, W., 2004. A Nonconserved Surface of the TFIIB Zinc Ribbon Domain Plays a Direct Role in RNA Polymerase II Recruitment. , 24(7), pp.2863–2874.

Veenstra, G.J., Weeks, D.L. & Wolffe, a P., 2000. Distinct roles for TBP and TBP-like factor in early embryonic gene transcription in Xenopus. *Science (New York, N.Y.)*, 290(December), pp.2312–2315.

Wakaguri, H. et al., 2009. Inconsistencies of genome annotations in apicomplexan parasites revealed by 5'-end-one-pass and full-length sequences of oligo-capped cDNAs. *BMC genomics*, 10, p.312.

White, N.J., 2008. Qinghaosu (artemisinin): the price of success. *Science (New York, N.Y.)*, 320(April), pp.330–334.

Winzeler, E.A., 2009. Malaria research in the post-genomic era. *Cell*, 455(7214), pp.751–756.

Wong, P., Gladney, S. & Keasling, J.D., 1997. Mathematical Model of the lac Operon: Inducer Exclusion, Catabolite Repression, and Diauxic Growth on Glucose and Lactose. *Biotechnology Progress*, 13(2), pp.132–143.

World Health Organisation, 2014. World Malaria Report 2014. *WHO Press*.

Xin Liu, David A. Bushnell, Dong Wang, Guillermo Calero, R.D.K., 2010. Structure of an RNA Polymerase II–TFIIB Complex and the Transcription Initiation Mechanism. *Science*, 327(January), pp.206–209.

Yamagishi, J. et al., 2014. The Babesia bovis gene and promoter model: an update from full-length EST analysis. *BMC genomics*, 15(1), p.678.

Young, J.A. et al., 2008. In silico discovery of transcription regulatory elements in Plasmodium falciparum. *BMC genomics*, 9, p.70.

Yuda, M. et al., 2010. Transcription factor AP2-Sp and its target genes in malarial sporozoites. *Molecular Microbiology*, 75(4), pp.854–863.

Zerby, D. & Lieberman, P.M., 1997. Functional Analysis of TFIID–Activator Interaction by Magnesium-Agarose Gel Electrophoresis. *Methods*, 12(3), pp.217–223.

Zhang, D. et al., 2001. Spermiogenesis deficiency in mice lacking the Trf2 gene. *Science (New York, N.Y.)*, 292(5519), pp.1153–1155.

Zhang, D.Y. et al., 2000. Intramolecular interaction of yeast TFIIB in transcription control. *Nucleic acids research*, 28(9), pp.1913–20.

Zhang, S., Zubay, G. & Goldman, E., 1991. Low-usage codons in Escherichia coli, yeast, fruit fly and primates. *Gene*, 105(1), pp.61–72.

Zhao, X. & Herr, W., 2002. A regulated two-step mechanism of TBP binding to DNA: A solvent-exposed surface of TBP inhibits TATA box recognition. *Cell*, 108, pp.615–627.

Zheng, L. et al., 2004. FRET evidence for a conformational change in TFIIB upon TBP-DNA binding. *European Journal of Biochemistry*, 271(4), pp.792–800.

Zilversmit, M.M., Volkman, S.K., Depristo, M.A., et al., 2010. Low-Complexity Regions in Plasmodium falciparum : Missing Links in the Evolution of an Extreme Genome Research article. , 27(9), pp.2198–2209.

Zilversmit, M.M., Volkman, S.K., DePristo, M. a, et al., 2010. Low-complexity regions in Plasmodium falciparum: missing links in the evolution of an extreme genome. *Molecular biology and evolution*, 27(9), pp.2198–209.