UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Regulation of transcription in *Plasmodium falciparum*, the causative agent of severe malaria: initial characterisation of PfTBP and PfTFIIA

By

Robert A. Milton

Dissertation presented for the degree of

Doctor of Philosophy

Molecular & Cell Biology

MCB 6002W

Supervisor

Dr Thomas Oelgeschläger

Carnegie Foundation

NRF National Research Foundation

Acknowledgements

I am deeply grateful to my supervisor Thomas. Thank you for your invaluable guidance and patience throughout the course of my PhD.

My parents Kevin and Gene Milton. Thank you for your unconditional love and support. Despite my many failings, you have always been there for me.

A big thank you to Daniel, Steven, Sam, and Lizzie for your support along this road.

Gertrud Talvik, my good and loyal friend.

Emmanuel Margolin, a shared sense of humour is our solid foundation. Disappointing movies are not the same without you!

A big thank you to Kesolofetse Kehiloe for all the help with the many plasmid minipreps. Your assistance was invaluable!

Also, thank you to Stephen Schlebusch for your contribution to the SELEX discussion.

Lastly, to the many friends that I have made and lost throughout these Cape Town years. You have all played an important part in my life. I wish you all the very best…

Abstract

Malaria, caused by *Plasmodium* parasites, remains a leading cause for morbidity and mortality worldwide, resulting in more than 430 000 deaths annually. *P. falciparum* is responsible for the vast majority of severe malaria cases, accounting for more than 90% of malaria-related fatalities, predominantly in subsaharan Africa. The parasite has a complex life cycle, which involves transitioning between multiple distinct morphologies. The severity of the disease is brought about by the variable expression of parasite proteins on the surface of infected red blood cells. The substantial morphological changes, together with the variable expression of cell surface proteins, are governed by tightly controlled stage-specific changes in gene expression patterns. Understanding the regulatory mechanisms that govern these changes is crucial to fully understanding the parasites biology and pathology at the molecular level, a key step toward identifying targets for the development of much needed novel antimalarial drugs.

Ultimately, all gene regulatory mechanisms converge to regulate the assembly and function of the RNA polymerase II (RNAP-II) transcription initiation complex composed of RNAP-II and the general transcription factors (GTFs). Bioinformatics analyses show that the RNAP-II GTFs in *P. falciparum* have greatly diverged from those studied in other eukaryotes, suggesting the existence of parasite-specific gene regulatory mechanisms, which have so far not been studied. This research project concerns the structure and function of *P. falciparum* TBP, TLP and TFIIA, key proteins involved in core promoter recognition, the first step in RNAP-II transcription initiation complex assembly. The work provides strong evidence for the existence of two different PfTFIIA complexes containing different PfTFIIA-γ subunits. The data further demonstrate that PfTBP and PfTLP DNA-binding activities differ distinctly from the classical TBP-DNA interactions seen in other eukaryotes and demonstrate interaction with and stimulation of PfTBP and PfTLP DNA-binding activity by one of the two PfTFIIA complexes. The work represents a first step towards understanding the regulation of transcription initiation in *P. falciparum*, gives first insights into *Plasmodium*-specific features, and provides a solid foundation for further investigations into this crucial aspect of malaria biology.

Common Abbreviations

6H- Hexa-histidine tag
A,T,G,C - adenine, thymine, guanine, cytosine
Ad2ML- Adenovirus type 2 major late
Amp - Ampicillin
ATP - Adenosine tri-phosphate
BC-100- BC-buffer; 100mM Salt
BLAST - Basic Local Alignment Search Tool
$BRE^{u/d}$ - upstream/ downstream TFIIB recognition element
BSA - Bovine serum albumin
BTAF1 - TATA-binding protein-associated factor 172
C-, N- terminus - carboxyl-; amino- terminus
Cam- Chloramphenicol
cDNA- Complimentary DNA
COBALT- Constraint-based multiple protein alignment tool
CPE- Core promoter element
cryo-EM - Cryo-electron microscopy
CTD- C-terminal domain
DCE - Downstream core element
DNA- Deoxyribonucleic acid
DPE - Downstream promoter element
DTIE- Downstream transcription initiation element
DTSS- Downstream transcription start site
EMSA- Electrophoretic mobility shift assay
EPD- Eukaryotic promoter database
FISH- Fluorescent *in situ* hybridization
GBP-130 - glycophorin binding protein 130
GCR - Gene coding regions
GTF - General transcription factor
HCA - Hydrophobic cluster analysis
HCS - High-content screening
Inr - Initiator element
IPTG - Isopropyl β-D-1-thiogalactopyranoside
IRS - indoor residual spraying
ITA - Immobilised template assay
KAHRP - Knob associated histidine rich protein
*Lac*UV5 - *E. coli* lac operon promoter
LB - Luria broth
LLIN - Long-lasting insecticidal nets
mRNA - Messenger RNA
MSP1 - Merozoite surface protein-1
MTE - Motif ten element
MW - Molecular weight
NC2 - Negative cofactor 2
Ni-NTA - Nickel- nitrilotriacetic acid resin
OD - Optical density
ORF- Open reading frame
PCR - Polymerase chain reaction
PfEMP1 - *P. falciparum* erythrocyte membrane protein 1
PfTBP - *P. falciparum* TATA-binding protein
PfTFIIA-αβ - *P. falciparum t*ranscription factor IIA αβ subunit
PfTFIIA-γ1s - *P. falciparum t*ranscription factor IIA γ1s subunit
PfTFIIA-γ2s - *P. falciparum t*ranscription factor IIA γ2s subunit
PfTLP - *P. falciparum* TBP-like protein
PIC - Pre-initiation complex
PlasmoDB- *Plasmodium* genome database
poly(dG-dC) - poly(deoxyguanylic-deoxycytidylic) acid
Q-sepharose - quaternary-ammonium-sepharose
SP-sepharose - sulphopropyl-sepharose
RBC - Red blood cell
RNA - Ribonucleic acid
RNAP-II - RNA polymerase II
RPB 1-12 - RNA polymerase II subunits
RTS,S/AS01 - Mosquirix®, GSK malaria vaccine
SELEX - Systematic evolution of ligands by exponential enrichment
SDS-PAGE - Sodium dodecyl sulfate polyacrylamide gel electrophoresis
STF - Specific transcription factors
TAF - TBP-associated factor
TBP - TATA-binding protein
TBST - Tris-buffered saline and Polysorbate 20
TdT - Terminal deoxynucleotidyl transferase
TFIIA - Trancription factor IIA
TFIID - Transcription factor IID
TLP - TBP-like protein
TRF - TBP-related factor
tRNA - Transport RNA
TSS - Transcription start site
UTR - Untranslated region
WHO - World Health Organisation

Contents

Chapter 1

Introduction

1.1 Overview: *Plasmodium spp* Biology and Disease

The malaria disease poses one of the greatest health challenges in the developing world. Each year there are in excess of 200 million reported cases of malaria resulting in at least 500,000 deaths. The disease is caused by a parasitic infection of erythrocyte cells by one of several *Plasmodium* species: *Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale*, *Plasmodium malariae,* and *Plasmodium knowlesi* (Kantele & Jokiranta, 2011). The degree of virulence is largely determined by the parasite species responsible as well as the relative vulnerability of the human host, with *P. falciparum* and *P. vivax* accounting for almost all severe cases of the disease. The 2016 World Health Organisation (WHO) malaria report estimates that 70% of the total malaria deaths in 2015 occurred in children under the age of 5 years, with >90% of these fatal malaria infections being attributed to *P. falciparum,* the malaria parasite prevalent in Africa (WHO, 2016).

1.1.1 *P. falciparum* life cycle

P. falciparum is a unicellular eukaryotic organism that forms part of a diverse group of parasites within the phylum *Apicomplexa*. Typically, *Apicomplexans* are obligate endoparasites characterised by a life cycle consisting of several distinct morphological

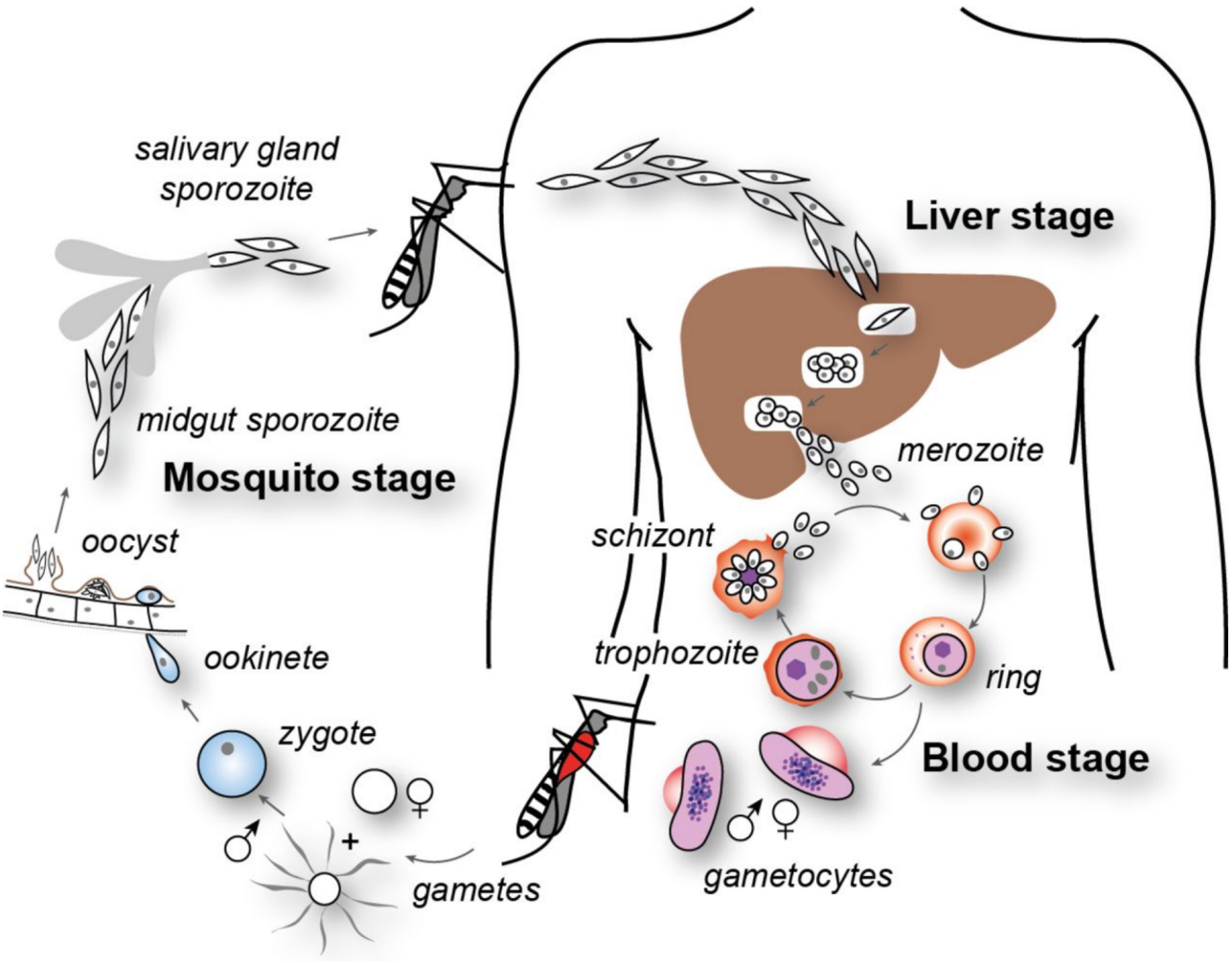


Figure 1: *P. falciparum* life cycle. The life-cycle is marked by the parasite transitioning between several distinct morphologies. The extracellular sporozoite form enters the human host through the saliva of a feeding mosquito before invading the liver hepatocytes and transitioning to an infectious extra-erythrocytic merozoite form. These merozoites are released into the blood system to infect the host erythrocytes. Once inside the erythrocytes, the merozoites grow into trophozoites and finally into mature schizonts. The subsequent lyses of the infected erythrocytes by newly formed merozoites leads to a new wave of infections. A small proportion of parasites in the blood stage undergo differentiation to form male and female gametocytes which can then be ingested by a new mosquito vector. Fertilisation occurs within the mosquito midgut and the resultant zygote encysts on the basal lamina of the midgut epithelium before entering into sporogenesis to complete the life cycle. Figure adapted from Cowman *et al*., 2012.

stages, some of which rely on multiple host organisms for propagation (Fig. 1) (reviewed in Plattner & Soldati-Favre, 2008). *P. falciparum* is not found in any known ecological niches outside its two host systems: the female *Anopheles gambiae* mosquito and humans. Infection of the human host begins when *P. falciparum* sporozoites enter the dermis through the salivary glands of an infected mosquito. The mosquito injects saliva into the dermis through repeated bites while probing for a blood vessel. The injected sporozoites rely on active transport known as gliding motility to migrate from the dermis into the human circulatory system, usually crossing within 1-3 hours post exposure (Keely & Soldati, 2004; Baum *et al*., 2006). Movement from the bloodstream into the hepatocytes occurs as the sporozoite traverses the epithelial and Kupffer cells lining the liver sinusoid to contact and subsequently enter the hepatocytes through the formation of a parasitophorous vacuoles at the cell surface (Plattner & Soldati-Favre, 2008; Tavares *et al*., 2013). Hepatocyte invasion from the bloodstream is a rapid process with entry of sporozoites shown to occur in mice within 2 minutes of intravenous injection (Shin *et al*., 1982). Once established in the hepatocyte, the sporozoite matures over 7-10 days to an exo-erythrocytic form called the merozoite. The parasite then multiplies to generate up to 40 000 new merozoites per infected hepatocyte (reviewed in Prodêncio *et al*., 2006; Cowman *et al*., 2016). Merozoites are released as membrane-bound merosomes into the lumen of the liver sinusoid (Sturm *et al*., 2006). Following this, the merosomes rupture, resulting in the re-entry of the parasite into the host circulatory system where they are free to enter the next stage of the life cycle by invading the host erythrocyte cells. Merozoite adherence to the erythrocyte surface is thought to be mediated by merozoite surface protein 1 (MSP1), the most abundant and functionally conserved coat protein. MSP1 has been shown to serve as a platform for other surface peptides, thus forming multiple distinct protein assemblies which may be responsible for the binding of target erythrocyte cells (Lin *et al*., 2016). There are however numerous other surface proteins present (reviewed in Cowman *et al*., 2012) and

the requirement for MSP1 in erythrocyte invasion is found not to be absolute (Das *et al*., 2015). There is also some suggestion that MSP1 may function in the evasion of the host immune responses as opposed to erythrocyte invasion (Cowman *et al*., 2016). Once initial contact with the erythrocyte surface is made, the merozoite is re-orientated through an array of specific ligand-receptor interactions to bring its apical end into direct contact with the erythrocyte membrane and enters the cell, thereby forming a new parasitophorous vacuole (Weiss *et al*., 2015). Once inside the erythrocyte, the parasite development enters the ring stage, the name based on its appearance in Giemsa-stained blood smears. Here, the parasite's nutrient uptake is at first achieved by feeding on small amounts of nutrients from the cell plasma through a nutrient channel located within the parasitophorous vacuole membrane (Desai & Rosenberg, 1997). Continued growth leads to the trophozoite stage where the most pronounced nutrient uptake, growth, and modification to the erythrocyte cell occurs. Faced with the problem of nutrient assimilation, the parasite begins to metabolise the cellular haemoglobin as a source of amino acids, and also begins to alter the permeability of the host cell membrane through the modification of surface anion channels, which subsequently facilitates the transport of purines, sugars and other essential nutrients (Desai *et al*., 2000; Decherf *et al*., 2004). In addition, the parasite forms a tubulovesicular network that extends from its vacuole to the cell surface. This network has also been implicated in the uptake of nutrients such as nucleosides and amino acids (Lauer *et al*., 1997). The haemoglobin breakdown product, ferriprotoporphyrin IX, crystallises to form the darkly pigmented haemozoin crystals within the food vacuole. Secretory compartments known as Maurer's clefts are formed which aid in the transport of parasite-derived proteins to the erythrocyte cell surface (Tilley *et al*., 2008). The presentation of *P. falciparum* proteins - such as those from the *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) family - causes knob-like projections from the erythrocyte surface and facilitates cytoadherence to the endothelial cells of the host blood vessels.

This cytoadherence aids in immune evasion by preventing clearance of infected cells by the spleen (reviewed in Hviid & Jansen, 2015). Lastly, some of the surface antigens are responsible for the binding of healthy erythrocytes by infected cells. This process is known as rosetting and is thought of as a possible strategy to mask infected cells from the host immune system (Plattner & Soldati-Favre, 2008). The parasite next enters the schizont stage. Here it undergoes a series of nuclear divisions and begins to synthesise the molecules needed for erythrocyte invasion. The new parasite nuclei start to move toward the cell periphery as they bud off from the main parasite compartment to form new merozoites. Finally, a signalling cascade involving serine protease (PfSUB1) together with calcium-dependent (PfCDPK5) and cGMP-dependent (PfPKG) protein kinases trigger the release of infective merozoites through the rupturing of the parasitophorous vacuole and erythrocyte membrane (Salmon *et al*., 2001; Dvorin *et al*., 2010; Collins *et al*., 2013). During the transition to the schizont phase, a small proportion of parasites undergo a metabolic switch and are committed to the formation of male and female gametocytes. Gametocyte maturation is a process lasting approximately 6-8 days. The gametocytes are able to avoid clearance by the spleen during maturation by being sequestered in the bone marrow (Joice *et al*., 2014). Mature gametocytes are subsequently released back into the bloodstream, being made available for further transmission through the blood uptake of a feeding mosquito. The exact molecular mechanisms controlling commitment to the sexual cycle are unknown. However, increased gametogenesis has been linked to high levels of parasitemia as well as the presence of antimalarials such as chloroquine, suggesting that it is likely brought about through environmental cues (Cowman *et al*., 2016). Fertilization occurs in the midgut of the mosquito vector. Here, the resultant zygote matures within several hours to form an ookinete, which is able to migrate through the midgut epithelium and then to transition to a mature oocyst on the basal lamina of the epithelial cells. New sporozoites are formed through asexual sporogenic replication and released through the

subsequent rupturing of the oocyst. Migration to the salivary glands is then achieved via the mosquitos haemolymph system (reviewed in Beier, 1998; Cowman *et al*., 2016).

1.1.2 Epidemiology and natural acquired immunity of host

Malaria is endemic across most of Africa, Central and South America, and Southeast Asia. The distribution and the intensity of transmission are determined by a combination of ecological factors involving the mosquito vector, the parasite, and the human host (WHO, 2016). Key environmental factors are moderate temperature and moderate to high humidity, both of which are required for the proliferation of the mosquito vector. The proclivity of the African *Anopheles* species to feed on humans together with its relatively long lifespan is largely responsible for the disproportionately high number of malaria cases in Africa. In areas of extremely high transmission, severe disease and death is largely restricted to young children (WHO, 2016). This is due to a naturally acquired immunity developed in the older population through continuous parasitaemia. The adult acquired immunity is not absolute as the risk of infection continues throughout life. There is however a marked reduction in the disease state symptoms reflecting an acquired ability to tolerate levels of asymptomatic parasitaemia as opposed to immune responses that totally eradicate the parasite from the host system (Cowman *et al*., 2016). The ability of gamma globulins isolated from immune adults to markedly suppress parasitemia in children suggests that the acquired immunity involves antibody-mediated targeting of the blood stage parasite (Cohen *et al*., 1961). However, the extremely slow development of natural immunity together with evidence that a strong immune response can be developed against pre- (Riley & Stewart, 2013) and post- (Wu *et al*., 2015) erythrocytic parasite stages indicate that it is likely the result of a complex immune response developed against a number of targets over many generations of parasite invasion and proliferation.

1.1.3 Pathogenesis of *P. falciparum* malaria

The clinical symptoms of malaria are mainly febrile, involving headaches, nausea, and general muscle fatigue. Early treatment of *P. falciparum*-induced malaria with antimalarial drugs will usually result in the alleviation of symptoms within 5-10 days and the eventual eradication of the parasite from the host system. In the case of an uncontrolled infection or a drug-resistant parasite, the disease can very quickly progress to the state of severe malaria (WHO, 2016). Symptoms of severe malaria differ depending on the age of the infected individual. Young children experience a combination of severe anaemia, respiratory distress, and/or cerebral malaria that can lead to malaria-induced coma and death. In addition, non-immune adults experience hepatic and renal dysfunction together with hypoglycaemia and respiratory distress due to pulmonary edema. The pathogenic effects of *P. falciparum* infection stem almost entirely from erythrocyte dysfunction brought about by parasite-mediated changes during the blood stage of the life cycle (Smith *et al*., 2013). These changes lead to microvascular obstruction through cytoadherence to the epithelial cells of the vasculature together with localised endothelial activation (Wassmer *et al*., 2015). Microvascular obstruction causes direct organ damage through blood flow restriction as well as broader systemic problems such as metabolic acidosis and acute pro-inflammatory responses (Kwiatkowski *et al*., 1998; Dondorp *et al*., 2008). There is also a strong correlation between severe malaria and incidences of bacterial pathogenesis in young children with a marked increase in dual infection mortality (Church & Maitland, 2014; Gómez-Pérez *et al*., 2014). In addition, pregnant woman are very susceptible to the development of severe malaria due to placental tissue providing new surface area for parasite sequestration through the adhesion of infected erythrocytes to placental chondroitin sulphate A (Khunrae *et al*., 2010). This can lead to low-birthweight or still birth

depending on the degree of natural immunity in the expectant mother (McLean *et al*., 2015).

1.1.4 Transmission control and current treatments

Reductions in the rates of transmission within endemic areas are achieved mainly by way of vector control strategies such as long-lasting insecticidal nets (LLINs) and indoor residual spraying (IRS) within homes (WHO, 2016). The LLINs function as a physical barrier placed over the bed to prevent vector contact at night while insecticide residue on the walls and ceilings of homes help to reduce the vector population within a given area if large scale coverage (> 80%) is achieved. In addition, travellers to high risk areas may take preemptive antimalarial medications such as sulfadoxine-pyrimethamine to reduce the risk of infection. Control strategies have proved effective. However, resistance to the insecticides has started to emerge in vector populations, prompting the WHO to release the “Global plan for insecticide resistance management in malaria vectors” in May 2012, which is aimed at facilitating collaboration between the relevant stakeholders in the development of new tools and strategies for vector control (Mnzava *et al*., 2015). In the case of active malaria infections, the most effective treatments currently available are artemisinin-based combination therapies. Parasite resistance to many of the first generation malaria drugs such as chloroquine has necessitated the use of these multi-drug strategies whereby artemisinin or a semi-synthetic derivative is administered in conjunction with a second relatively long-lasting antimalarial to counteract the short half-life of artemisinin (Vugt *et al*., 1999). The mechanism of action is thought to involve a haem-mediated activation of an endo-peroxide ring within the artemisinin structure resulting in increased oxidative stress and parasite death (Winzeler & Manary, 2014). Unfortunately, signs of artemisinin resistance have already begun to emerge in Southeast Asia (Dondorp

et al., 2008). Whole-genome studies of resistant parasites have identified at least 20 polymorphisms in the K13-propellar gene as a set of useful genetic markers for resistance (Ariey *et al*., 2014; Miotto *et al*., 2015) with a concurrent study demonstrating that insertion of some of these mutations was indeed able to confer a measure of resistance to non-resistant reference strains (Straimer *et a*l., 2015). In addition, transcriptional analysis has indicated that resistant parasites display an up-regulation of their unfolded-protein response pathways together with a decrease in early ring-stage parasite growth rate, suggesting an increased tolerance of blood-stage parasites to the negative effects of the artemisinin induced oxidative stress (Mok *et al*., 2015).

1.1.5 Development of novel antimalarial compounds

Widespread resistance to current treatments have been met with a concerted effort by both public and private entities to find and develop novel antimalarial compounds (Gamo *et al*., 2010; Guiguemde *et al*., 2010; Meister *et al*., 2011; Spangenberg *et al*., 2013). To this end, both high-content screening (HCS) together with synthetic approaches are currently being used. HCS, also known as whole-cell screening, is a high-throughput method of drug screening that combines multi-colour fluorescence microscopy together with cellular imaging and data analysis software to examine phenotype responses to a large number of candidate compounds in order to identify possible lead molecules for further development (Zanella *et al*., 2010). Synthetic approaches form part of the rational drug design method which make use of computer modelling together with known structural aspects of either the drug target or available drugs to produce a blueprint for the chemical synthesis of novel compounds (Mandal *et al*., 2009). There is increasing overlap in the employment of these approaches as HCS technologies are often used to further test for efficacy and toxicity of newly synthesised compounds (Zanella *et al*., 2010). A number of

new drugs which have been found to disrupt the parasites Na^+ homeostasis by binding to a Na^+-ATPase (PfATP4) are currently in various stages of clinical trials (Spillman *et al*., 2013; Vaidya et al., 2014; Jiménez-Díaz *et al*., 2014). Further development is also underway for a more recent compound that has been identified as a potent inhibitor of protein synthesis through its inhibition of translation elongation factor 2 (eEF2) (Baragaña *et al*., 2015). In addition, a new class of compounds named imidazopyrazines have been identified, and act by inhibiting the parasite lipid kinase PI4K (McNamara *et al*., 2013). Significantly, the imidazopyrazines also show activity against the dormant liver stage of the simian parasite *P. cynomolgi.* Although not present in the *P. falciparum* life cycle*,* these dormant parasite forms known as hypnozoites are features of both *P. vivax* and *P. ovale.* Hypnozoites can lead to a series of relapse infections spanning a number of weeks or even months and therefore pose a serious challenge to the eradication of malaria. There is currently only one commercially available treatment active against these dormant cells. However, newly identified compounds are currently being tested (Ashley *et al*., 2014; Llanos-Cuentas *et al*., 2014; Zeeman *et al*., 2014). Rational design approaches have also led to the development of a class of antimalarial molecules known as ozonides. These molecules are based on the conservation of the endoperoxide bridge found in artemisinin and thus likely cause oxidative damage to the cellular components of the blood stage parasite upon activation by the ferrous iron of the metabolised haemoglobin (O'neill, 2004; Vennerstrom *et al*., 2004; Charman *et al*., 2011). Finally, the *P. falciparum* dihydrooratate dehydrogenase enzyme from the parasites pyrimidine biosynthesis pathway has been used as a target for the identification of potent triazolopyrimidine-based inhibitors, one of which has recently been advanced to human trials (Coteron *et al*., 2011; Phillips *et al*., 2015).

1.1.6 Vaccine development

There is currently only one commercially available anti-malaria vaccine. RTS,S/AS01 (Mosquirix) is a sporozoite-based vaccine that utilises a chimeric protein formed from the C-terminal half of the *P. falciparum* circumsporozoite protein (CSP) fused to the hepatitis B surface antigen (Cohen *et al*., 2010). Unfortunately, despite very promising phase II clinical trails, the phase III trials demonstrated only a very modest efficacy with a three dose regimen resulting in only a 28% reduction in malaria cases for immunised children aged 5-17 months and with a further drop in the efficacy for infants (Greenwood & Doumbo, 2016). Overall, the vaccine has however been judged by several disease modelling studies to be a cost effective tool when compared to the costs incurred by alternative control measures, thus making it an important milestone by which all future vaccine candidates will be judged (Penny *et al*., 2016).

The difficulty in producing an effective anti-malaria vaccine is largely due to the complexity of the parasite life cycle and the high degree of antigenic variation brought about by the transient expression of multiple PfEMP1 gene variants during erythrocyte infection (reviewed in Scherf *et al*., 2008; Hviid & Jensen 2015; further discussed in Section 1.1.8). The current vaccine development strategies are mainly focused on the pre-erythrocytic and sexual stages of parasite life cycle as both of these stages represent a bottleneck in parasite numbers and a comparative lack in immune evasion capabilities relative to the blood stage merozoite form. There are currently a number of live-cell sporozoite vaccines under development using radiation attenuated (Seder *et al*., 2013) and genetically modified (Spring *et al*., 2013) sporozoites. These however all suffer the major drawback of only producing strain-specific immune protection. In addition, the most notable problem with strategies of this nature is that the immunogenic reactions elicited from different stages of the parasite life cycle are very specific to that stage (Felgner *et al*., 2013). This means that if a single parasite is able to evade the immune response long enough to reach

the red blood cells then the resistant parasite would be propagated and could result in a potential resurgence of infections. This eventuality seems even more likely when considering the declining efficacy of current antimalarial drugs. A possible strategy to prevent this would be to produce a vaccine using multiple antigenic targets from various stages in the parasite life cycle. One such candidate makes use of the blood stage antigen - glutamate-rich protein (GLURP), fused to a fragment of the sexual stage antigen Pfs48/45 to form a chimera that produced a strong antibody response in rodents (Theisen *et al*., 2014). This strategy will likely be central to vaccine development in the coming years as candidates for a broad stage anti-malaria vaccine have already been identified in both the sporozoite and erythrocytic parasite forms (Cohen *et al*., 2010; Douglas *et al*., 2015).

1.1.7 Gene regulation in *P. falciparum*

By its very nature, the complexity of the *P. falciparum* life cycle necessitates a highly controlled and equally complex gene expression program that sees the parasite transitioning between multiple morphological states in both intracellular and extracellular environments within two very different host organisms. This is indeed evidenced by the patterns of stage-specific gene expression seen throughout the parasites life cycle (Florens *et al*., 2002; Bozdech *et al*., 2003; Le Roch *et al*., 2003; Adjalley *et al*., 2016). Following on from the original publication of the *P. falciparum* genome sequence (Gardner *et al*., 2002), studies which investigated the changes in gene expression in response to external stimuli and physiological stresses such as temperature changes and glucose starvation found evidence of flexibility in the parasites expression profiles (Fang & McCutchan, 2002; Fang *et al*., 2004; Oakley *et al*., 2007). These observations, together with the periodic nature of mRNA levels during the sporozoite, intra-erythrocytic, and sexual stages, which correlated with the expression of discrete gene clusters, related by

function and/or cellular process to the immediate physiological demands of the parasite, led to the hypothesis that transcription immediately proceeded to the translation of mature protein and thus implicated transcription initiation as the primary point of gene regulation in *P. falciparum* (Bozdech *et al*., 2003; Le Roch *et al*., 2003). Subsequent studies questioned this idea. Quantitative microarray data together with a semi-quantitative protein analysis techniques employing multidimensional liquid chromatography coupled to mass spectrometry did not confirm a tight correlation between mRNA and protein abundance but instead found a distinct lag between mRNA production and translation. This finding led to the proposal that post-transcriptional mechanisms may play a predominant role in gene regulation (Le Roch *et al*., 2004). This hypothesis is supported by bioinformatics analysis of the *P. falciparum* genome sequence, showing a striking paucity of specific transcription factors (STF) and a comparative over-representation of CCCH zinc finger domain proteins involved in the stability, localisation, and translation of mRNA (Coulson *et al*., 2004; reviewed in Deitsch *et al*., 2007; Horrocks *et al*., 2009). In addition, studies into the mechanisms underlying the parasite antigenic variation through the differential expression of the *var* gene family (reviewed in Scherf *et al*., 2008), together with the determination of the *P. falciparum* genomic histone modification profile has placed much emphasis on the epigenetic aspects of gene regulation (Scherf *et al*., 1998; Horrocks & Lanzer., 1999; Deitsch *et al*., 2001; Miao *et al*., 2006; reviewed in Merrick & Duraisingh, 2010). The importance of gene regulation at the level of transcriptional initiation was, however, again highlighted by the discovery of a range of *Apicomplexa* STFs containing a novel Apetala 2 (AP2) DNA-binding domain (Balaji *et al*., 2005; De Silva *et al*., 2008). This was followed shortly by the identification of a number of key RNA polymerase II (RNAP-II) general transcription factors that had up to that point been missed by traditional bioinformatics searches due to the extremely high (> 80%) A/T nucleotide base content of the parasite genome (Callebaut *et al*., 2005). Largely owing to the atypical nucleotide content of the

parasite genome, there has up until very recently been a distinct lack of data with respect to the regulation of basal RNAP-II transcription initiation beyond that of the *in silico* prediction of putative protein candidates for some conserved eukaryotic transcriptional machinery. So far, the *P. falciparum* TATA-binding (PfTBP) protein is the only general transcription factor that has been functionally characterised, albeit in a very preliminary manner. In these studies, results of electrophoretic mobility shift assays and *DNase* I footprinting assays are presented to suggest that PfTBP binds a consensus TATA-sequence (TATAA) upstream of the *kahrp* gene and an atypical TATA-sequence (TGTAA) upstream of the *gbp-130* gene (Ruvalcaba-Salazar *et al*., 2005). The notion that PfTBP binds to distinct sequences in parasites core promoters seem however not compatible with results of this work and with recently published work which produced a comprehensive map of transcription start site positions in the *P. falciparum* genome through RNA-seq data. These studies found little evidence of traditionally defined promoter elements but rather that transcription initiation events where correlated to the local A/T basepair content as opposed to a distinct sequence motif (Adjalley *et al*., 2016).

1.1.8 *P. falciparum* antigenic variation

A defining characteristic of the *Plasmodium* species is a highly evolved immune evasion strategy mediated by antigenic variation at the surface of infected erythrocyte cells. In the *P. falciparum* (3D7) genome there are 58 *var* genes (and 1 *var* pseudogene) encoding a clonally variant and immunodominant family of adhesion molecules known as *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) (reviewed in Kraemer & Smith, 2006; Scherf *et al*., 2008). These are presented at the cell surface of infected erythrocyte molecules, causing the infected erythrocytes to adhere to various cell types within the host vasculature as a way of avoiding clearance by the spleen. In addition, there is evidence

that these molecules are able to interact directly with the host immune system, a characteristic which likely leads to immune suppression (Gilbert *et al*., 1998; Urban *et al*., 1999; Donati *et al*., 2004).

All *var* genes consist of two exons separated by a conserved intron. Exon 1 codes for a polymorphic extracellular region that contains multiple Duffy-binding-like (DBL) adhesive domains as well as cysteine-rich inter-domain regions (CIDR), while the second exon codes for a semi-conserved cytoplasmic domain (Kramer & Smith, 2006). These genes are distributed across four of the fourteen parasite chromosomes (chromosome 4, 7, 8, and 12). Most *var* genes are positioned at the highly variable chromosomal ends, while approximately 40% are located near the chromosome centres (reviewed in Kirkman & Deitsch, 2012). The subtelomeric *var* genes are flanked by six distinct regions of non-coding DNA known as telomere-associated repeat elements (TARE 1-6). Subtelomeric *var* genes are typically arranged in a head to head orientation with the transcription of adjacent genes proceeding in opposite directions while the centrally located *var* genes are found clustered in tandem-arrays. The *var* genes of the *P. falciparum* 3D7 strain are further characterised by five possible promoter sequences: upsA, -B,- C, -D, and -E, as well as a sixth combinatorial promoter sequence designated upsBC. There is a strong correlation between these various promoter sequences with both the chromosomal positioning and the orientation of transcription associated with the cognate gene (Lavstsen *et al*., 2003; Kirkman & Deitsch, 2012). Subtelomeric *var* genes that are transcribed toward the centre of the chromosome are typically flanked by upsB, those in the same region but transcribed away from the centre are typically flanked by upsA but can also be found associated with upsD and/or upsE sequences. The internal *var* genes however are typically associated with the upsC promoter. Finally, upsBC *var* genes can be found at both chromosomal ends as well as within the central regions. Interestingly, despite these differences in promoter

sequence and gene orientations, all of the *var* gene variants are subject to mutually exclusive expression. The molecular mechanisms underlying this system of expression has been a central research focus for many years in the malaria field (Scherf *et al*., 2008). Fluorescent *in situ* hybridization (FISH) data showing the localisation and clustering of both subtelomeric and centralised *var* genes to the nuclear periphery provide clues to the mechanisms behind the high rates of genetic recombination within the *var* gene family as well as the promoter-dependent regulation of transcription from these loci (Freitas-Junior *et al*., 2000; Ralph *et al*., 2005). Indeed, PCR analysis of the *var* genes from two *P. falciparum* genotypes (3D7 and HB3) have demonstrated phenotypic variations at rates exceeding that which could be explained by genetic drift. Subsequent analysis of *var* gene sequences in the progeny of recombinant laboratory strains provided further evidence of intergenic recombination occurring at the sites of *var* gene clustering (Taylor *et al*., 2000). Although epigenetic control of *var* gene expression is well established (reviewed in Ralph & Scherf, 2005; Cui & Miao, 2010; Gupta *et al*., 2013), the principle of context-dependent activation of transcription from promoters influenced by proximity to other core promoters (reviewed in Andersson *et al*., 2015) is strongly indicated in the repression of the *var* gene family. A series of experiments by several groups using reporter gene constructs under the control of *var* gene promoters demonstrated that promoter pairing is likely to play a significant role in transcriptional silencing. These studies began with the observation that the conserved *var* intron was able to repress transcription when placed at the 3' end of a *var* gene, and that this intron in fact possessed transcriptional activity (Deitsch *et al*., 2001; Calderwood *et al*., 2003). It was then demonstrated that activated *var* promoters (upsB and upsC) in a stably transfected episome were localised to the nuclear periphery and were sufficient to elicit total endogenous *var* gene silencing in the absence of any virulence factor expression. This demonstrated that promoter activity, and not the gene product, is central to the mutually exclusive expression of the *var* gene variants (Voss *et al*., 2006;

Voss et al., 2007). These important findings were confirmed independently by a second laboratory using similarly modified parasite lines (Dzikowski *et al*., 2006). Leading on from these results, the same group demonstrated that the apparent intron-mediated silencing could be achieved by pairing a *var* promoter with a number of *P. falciparum* promoters at the 3' of the reporter gene and that uncoupling the *var* gene from a downstream promoter would remove it from the mutually exclusive *var* gene expression patterns (Dzikowski *et al*., 2007). In summary, the data as a whole supports the notion that in addition to epigenetic regulation, transcriptional silencing of inactive *var* genes is mediated by a novel coupling mechanism that sees a second downstream promoter sequence functioning as a type of repressor in the context of antigenic switching of the PfEMP1 gene family in *P. falciparum*.

1.2 Overview of eukaryotic transcription

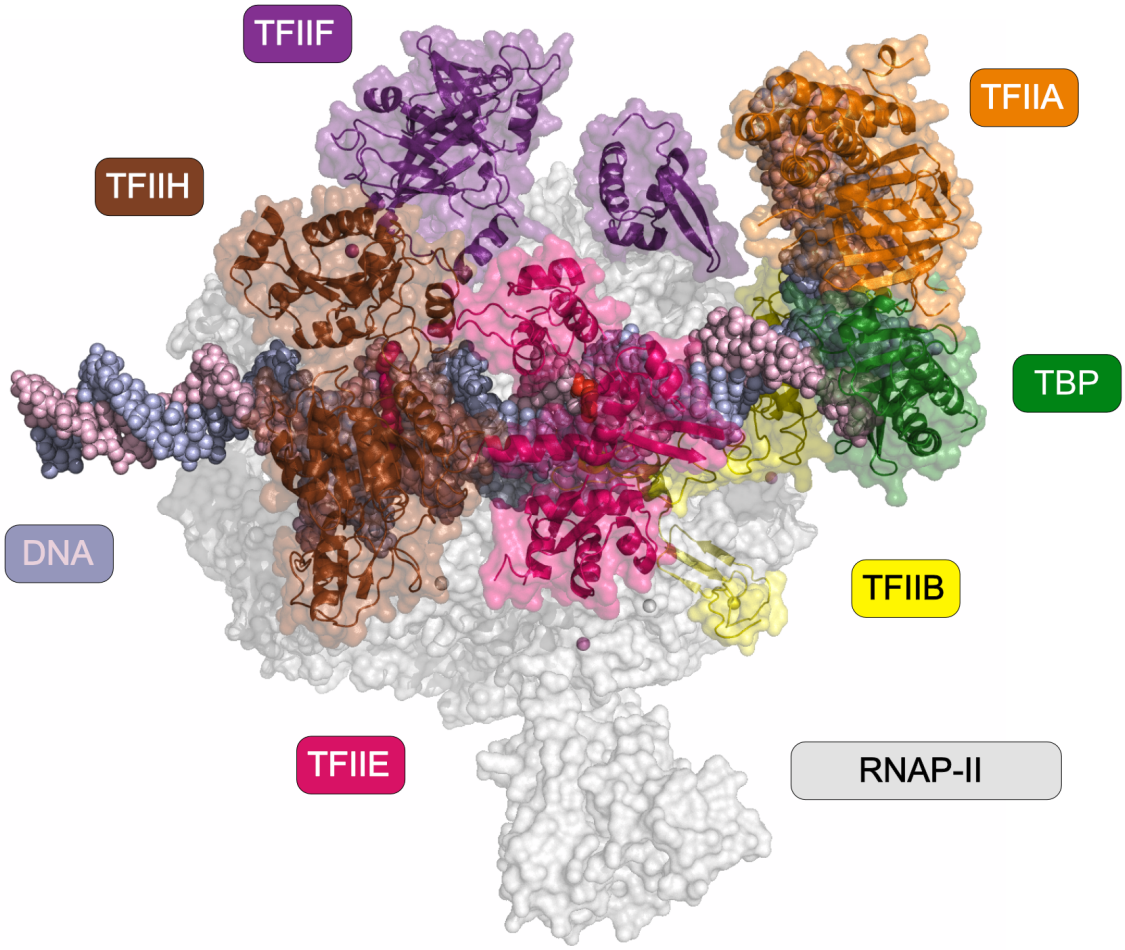
The process of transcription is defined as the DNA-dependent synthesis of RNA molecules by cellular RNA polymerases. With respect to nuclear gene expression in humans, there are four distinct RNA polymerases that take part in this highly co-ordinated process, each of which is responsible for the synthesis of a distinct sub-class of RNA molecules (Roeder & Rutter, 1969; Roeder & Rutter, 1970; Weinmann & Roeder, 1974; Kravchenko *et al*., 2005). RNA polymerase I (RNAP-I) is responsible for the transcription of 18S, 28S, and 5.8S ribosomal RNA. RNA polymerase II (RNAP-II) transcribes the bulk of the cellular messenger RNA (mRNA) molecules as well as microRNAs and long non-coding RNAs from intergenic and enhancer regions. RNA polymerase III (RNAP-III) synthesises the 5S ribosomal RNA as well as the various transport RNAs (tRNAs). More recently, a single-polypeptide nuclear RNA polymerase (spRNAP-IV) was discovered in human HeLa cells (Kravchenko *et al*., 2005). Structurally, this protein is an N-terminal truncation of the

mitochondrial RNA polymerase (mtRNAP) peptide that is coded for by the nuclear *POLRMT* gene. It has been found that spRNAP-IV is responsible for the transcription of a subset of nuclear protein-coding genes that are not recognised by the RNAP-II transcription machinery. In addition to the essential RNA polymerases found in all eukaryotes (RNAP-I,-II, and -III), plants contain two more RNAP variants, RNAP-IV and RNAP-V (or -IVb), both of which play a central role in transcriptional silencing through the production of short interfering RNAs (siRNAs) (Onodera *et al*., 2005; Pikaard et al., 2008; Ream *et al*., 2009). Despite sharing a common functionality in the production of RNA from a defined DNA template, the RNA polymerases all lack DNA-sequence recognition capabilities and as such require accessory proteins to facilitate the recognition of the promoter and transcription start site of a particular induced gene. These accessory proteins are collectively termed general transcription factors (GTFs), with each RNA polymerase requiring a distinct set of GTFs to initiate site-specific transcription of their target genes. Here, I focus on the RNAP-II pathway with particular emphasis on the RNAP-II general transcription factors: TATA-binding protein (TBP), Transcription Factor IIA (TFIIA), and the TBP-like protein TRF2 (TLP). These key RNAP-II factors play a vital role in promoter recognition and the subsequent transcription of protein coding genes within all eukaryotes subsequent to the evolution of the bilateria and are, with the exception of PfTBP (McAndrew *et al*., 1993; Ruvalcaba-Salazar *et al*., 2005), not studied in *P. falciparum.*

1.2.1 The RNA polymerase II pre-initiation complex (PIC)

Human RNAP-II is composed of 12 subunits designated RPB1 to RPB12. These are generally well conserved in other eukaryotes (Young, 1991; Armache *et al*., 2003) with RPB1 and RPB2 being responsible for the protein's catalytic activity (Hampsey, 1998; Lee

A



B

TFIIA	Anti-repressor; stabilizes TBP-TATA complex; coactivator
TFIIB	Start site selection; stabilizes TBP-TATA complex; RNAP-II/TFIIF recruitment
TBP	Primary core-promoter binding factor of the TFIID complex
TFIIE	Recruits TFIIH; facilitates formation of initiation competent RNAP-II; involved in promoter clearance
TFIIF	Binds RNAP-II and may facilitate recruitment to the promoter; recruits TFIIE and TFIIH; functions with TFIIB and RNAP-II in start site selection; facilitates RNAP-II promoter escape; enhances efficiency of RNAP-II elongation
TFIIH	ATPase activity for transcription initiation and promoter clearance; helicase activity for promoter opening; transcription-coupled nucleotide excision repair; kinase activity for phosphorylating RNAP-II CTD; E3 ubiquitin ligase activity
RNAP-II	Transcription initiation, elongation, and termination; recruitment of mRNA capping enzymes; transcription-coupled recruitment of splicing and 3' end processing factors

Figure 2: Model of the eukaryotic pre-initiation complex (PIC). (A) Model of the PIC structure comprised of RNAP-II and the general transcription factors. TATA-binding protein (TBP), TFIIA, -B, -E, -F, and -H. Figure rendered using PyMOL and coordinates provided by Grünberg *et al.*, 2012. (B) Brief summary of general transcription factor functions (adapted from Thomas & Chiang, 2006).

and Young, 2000). A defining feature of RNAP-II is a C-terminal domain (CTD) located within the RPB1 subunit, containing a large number of heptapeptide repeats Tyr-Ser-Pro-Ser-Pro-Ser (YSPTSPS). Structurally, the CTD is regarded as a disordered region and therefore vulnerable to proteolytic degradation. The CTD of human RNAP-II contains 52 heat peptide repeats. However, this number does vary significantly between different species (Dahmus, 1995; Hampsey, 1998; Lee & Young, 2000). The CTD is the site of multiple post-translational modifications such as phosphorylation, glycosylation, and ubiquitination. Three separate forms of human RNAP-II (-IIO, -IIA, and -IIB) can be distinguished based on differential phosphorylation as well as the presence or absence of the CTD (Kershnar *et al*., 1998). The RNAP-IIA form is typically associated with pre-initiation complex (PIC) assembly and is marked by low levels of CTD phosphorylation (hypophosphorylation). In contrast, the RNAP-IIO state contains a hyperperphosphorylated CTD and is associated with the elongation and termination of the newly synthesised mRNA chain. Lastly, the RNAP-IIB variant lacks the CTD due to proteolytic cleavage. This form has, however, been shown to maintain its transcriptional activity on a number of different promoter constructs despite this alteration (Kim and Dahmus, 1989; Buermeyer *et al*., 1995; Zehring and Greenleaf, 1990). Glycosylation may occur on the threonine residues located at position four of the heptapeptide repeats. Glycosylation is only found on the RNAP-IIA variant and is thought to either sterically hinder protein kinases from phosphorylating the CTD or to this end alter the CTD conformation to a form that limits the interaction with protein kinases. Overall, the exact function of this modification is unclear. However, it is thought that glycosylation may help to regulate early stages of RNAP-II transcription by promoting the RNAP-A form involved in PIC assembly (Kelly *et al*., 1993; Comer & Hart, 2001). Poly-ubiquitination is the third of the well characterised RNAP-II post-translational modifications. It has been found that ubiquitination of the CTD occurs as a response to DNA damage incurred by both UV exposure and oxidative stress, which

leads to the activation of the transcription-coupled repair process and often results in the degradation of transcriptionally stalled RNAP-II by the 26S proteasome (Bregman *et al*., 1996; Inukai *et al*., 2004).

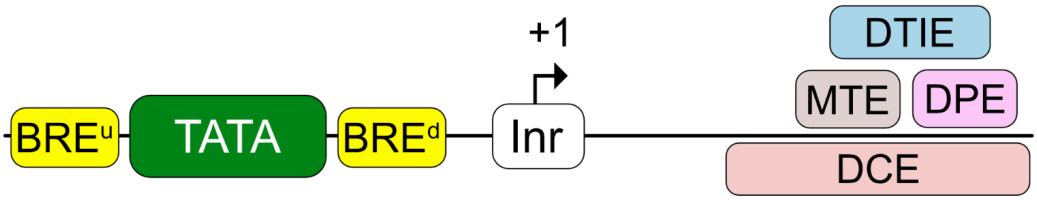
Evidence of accessory factors required for promoter-specific transcription initiation by RNAP-II was first seen when purified RNAP-II was supplemented with crude nuclear extract fractions to facilitate *in vitro* transcription from the Adenovirus type 2 major late promoter (Ad2ML) (Weil *et al*., 1979; Thomas & Chiang, 2006). These nuclear extract fractions were then further purified to isolate and to identify the factors required for promoter-specific RNAP-II transcription initiation. Collectively these factors were designated the general transcription factors (GTFs) and consist of Transcription factor IIA (TFIIA), TFIIB, TFIID, TFIIE, TFIIF, and TFIIH (Fig. 2B). Subsequent work demonstrated that the positioning of RNAP-II at the start site of transcription requires the assembly of a pre-initiation complex (PIC), a large protein mega-complex at the core promoter of an activated gene (Fig. 2A). Core promoter regions are highly variable in structure and often gene-specific (further discussed in Section 1.2.2). The text book model of PIC assembly discussed here is based on studies with TATA box-containing model promoters, such as the adenovirus type 2 major late promoter (Ad2ML). Major advances have been made in the understanding of PIC formation within the last five years with the publication of cryo-EM data for both human and yeast PIC structures (He *et al*., 2013; Murakami *et al*., 2013). A simplified order of events leading to the initiation of transcription in human cells can be described as follows (reviewed in Kandiah *et al*., 2014). PIC formation commences with binding of TATA box-binding protein (TBP) to the TATA box element (Bleichenbacher *et al*., 2003; Cianfrocco *et al*., 2013). TFIIA and TFIIB are then recruited and stabilise this interaction (Imbalzano *et al*., 1994b; Hieb *et al*., 2007; Cianfrocco *et al*., 2013). Next, RNAP-II is recruited by TFIIB in the presence of TFIIF and TFIIE. The subsequent

recruitment of TFIIH then leads to a stable TBP-TFIIA-TFIIB-RNAP-II/TFIIF-TFIIE-TFIIH transcriptionally closed PIC complex in the absence of ATP. TFIIH-mediated promoter melting, in an ATP-dependent manner, transitions the complex to a transcriptionally open state. Once in the open state, RNAP-II typically cycles through multiple rounds of abortive transcription in which many short (2-15 base) mRNAs are produced before proceeding from initiation to elongation were it transitions along the DNA template to synthesise the nascent full-length mRNA molecule (Luse & Jacob, 1987; Holstege *et al*., 1997). In most eukaryotes, transcription by RNAP-II is briefly paused after synthesising approximately 20-60bp of mRNA to allow for the disconnection of the actively transcribing elongation complex from the remaining transcriptional apparatus at the core promoter, the so-called re-initiation scaffold, in an event known as promoter escape (Dvir *et al*., 1997; Saunders et al., 2006; Hahn, 2004; Min *et al*., 2011 Adelman *et al*., 2012; Kwak *et al*., 2013). The re-initiation scaffold at the core promoter acts to increase the overall rate of transcription re-initiation as only RNAP-II, TFIIF, and the ejected TFIIB need to be re-recruited for subsequent rounds of transcription (Yudkovsky *et al*., 2000). In addition to the general transcription factors, the regulatory networks that govern gene activation involve a host of gene-specific transcription factors such as activator/repressor proteins, as well as the general cofactors that mediate the interactions between these two groups. Generally speaking there are two main classes of regulatory cofactors that govern the interactions between specific transcription factors and the general transcription factors of the RNAP-II PIC (reviewed in Thomas & Chiang, 2006).These are the TBP-associated factors (TAFs) of the TFIID complex, and a group of cofactors isolated from crude fractions of HeLa cell nuclear-extract known as the upstream stimulatory activity (USA) cofactors which include the mediator complex. With the exception of the TAFs, briefly discussed in the context of TFIID (Section 1.2.3), these regulatory networks go beyond the scope of the research presented here and as such will not be further discussed.

1.2.2 The Core Promoter Elements (CPEs)

The core promoter is typically defined as the region of DNA (50-100bp) spanning the transcription start site of a given gene. In mammals, core promoters have been categorised as either “high-GC” or “low-GC” promoters based on the relative GC-content and CpG dinucleotide frequency within these regions (reviewed in Lenhard et al, 2012). In addition, high-GC promoters, also called broad type promoters, are found to have multiple transcription start sites (TSSs), while low-GC promoters exhibit a more clearly defined TSS, usually initiating from a single nucleotide position. Genome-wide studies have since been used to further differentiate mammalian promoters into three main sub-types. Type I, are low-GC promoters that are found to be responsible for tissue-specific gene activation, while type II promoters, generally associated with housekeeping genes, typically consist of a single CpG island overlapping the TSS. Lastly, type III promoters are associated with developmentally regulated genes and are defined by the occurrence of multiple large CpG-islands. In addition to these atypically high CpG-sites, core promoters also contain one or more short DNA-motifs that are conserved with respect to both their sequence and their corresponding position relative to the TSS (Fig. 3A). These motifs are recognised and bound by components of the RNAP-II pre-initiation complex (Fig. 3B), and are collectively known as the core promoter elements (CPEs) (reviewed in Roy and Singer, 2015; Danino *et al*., 2015). There are currently eight common CPEs that are found in varying combinations within the core promoter regions of eukaryotes. The TATA box and initiator (Inr) elements are the two most abundant CPEs, and are found together or in isolation within eukaryotic promoters (Suzuki et al., 2001; Gershenzon & Ioshikhes; 2005). Both the TATA box and the Inr element provide a platform for the initial recognition and binding of the core promoter by TFIID. The TATA box is usually present 24-31 bp upstream of the transcription start site of TATA-containing genes and serves as the binding site for the

A



B

Promoter Element	Position	Consensus Sequence (5'-3')	Bound Protein
BRE^u	-38 to -32	(G/C)(G/C)(G/A)CGCC	TFIIB
TATA	-31 to -24	TATA(A/T)A(A/T)(A/G)	TBP
BRE^d	-23 to -17	(G/A)T(T/G/A)(T/G)(G/T)(T/G)(T/G)	TFIIB
Inr	-2 to +5	YYAN(T/A)YY	TAF1/TAF2
MTE	+18 to +29	C(G/C)A(A/G)C(G/C)(G/C)AACG(G/C)	unknown
DPE	+28 to +34	(A/G)G(A/T)CGTG)	TAF6/TAF9
DCE	+6 to +34	CTTC, CTGT, AGC	TAF1
DTIE	+21 to +33	G(G/C/T)(G/C/T)R(A/T/G)N(A/T/C)GG	unknown

Figure 3: Eukaryotic core promoter elements. (A) Schematic of the eukaryotic core promoter showing the relative positions of DNA-sequence motifs associated with transcription start sites. (B) Table showing consensus sequences of common core promoter elements, nucleotide positions, and basal transcription machinery commonly associated with each element. BRE^u, upstream TFIIB-recognition element; TATA, TATA box; BRE^d, downstream TFIIB-recognition element; Inr, initiator; MTE, Motif-ten element; DPE, downstream promoter element; DCE, downstream core element; DTIE, downstream transcription initiation element.

TATA-binding protein (TBP) of the TFIID complex (Bleichenbacher *et al*., 2003). When present, the Inr element spans 7bp over and around the transcription start site and is capable of initiating transcription independently of the TATA box element (Smale, & Baltimore 1989; Roy and Singer, 2015). Based on random DNA-binding site selection assays, recognition of the Inr element sequence is thought to take place through interactions with the TBP-associated factors TAF1 and TAF2 of TFIID (Chalkley & Verrijzer, 1999). The downstream promoter element (DPE) has been found to occur with the Inr element in mostly TATA-less promoters at position +28 to +34 of the transcription start site (Burke & Kadonaga, 1996, 1997). Unlike the TATA box, it is not able to function independently but rather functions cooperatively with the Inr element in the binding of TFIID through TAF6 and TAF9 (Kutach and Kadonaga, 2000; Shao *et al*., 2005). A second core promoter element found downstream of the transcription start site is the downstream core element (DCE) (Lewis *et al*., 2000; Lee *et al*., 2005). The DCE is unique in that it is made up of three subelements spanning the +6 to +34 region and which are all bound by TAF1 of TFIID (Fig. 3B) (Lee *et al*., 2005). In vitro transcription experiments have also shown that the third subelement usually located within the +30 to +34 downstream region is able to function independently of the other sub-elements and bioinformatics analysis of both the Eukaryotic Promoter Database (EPD) and Database of Transcriptional Start Sites (DTSS) (http://dbtss.hgc.jp/) indicate that its presence strongly excludes the DPE sequence motif (Lee *et al*., 2005). The motif ten element (MTE), usually found between +18 and +29, functions in conjunction with the Inr in RNAP-II mediated transcription (Ohler *et al*., 2002; Lim *et al*., 2004). The MTE can work synergistically with the TATA box and DPE in *in vitro* transcription using either *Drosophila* or HeLa cell nuclear extract. However, the transcription factors that bind the MTE are yet to be determined (Lim *et al*., 2004). More recently, a new CPE element called the downstream transcription initiation element (DTIE) was found in the context of human genes that are negative for both TATA box and

Inr regulatory sequences. Bioinformatics analysis placed the DTIE in positions spanning +21 to +33 relative to the TSS at a frequency which may exceed the prevalence of the TATA box and Inr elements within human promoters (Marbach-Bar *et al*., 2016). Finally, there are the TFIIB-recognition elements that have been designated according to their spatial orientation around the TATA box element as the upstream BRE^{u} and downstream BRE^{d} elements (Lagrange *et al*., 1998; Deng & Roberts, 2005). Although originally discovered in TATA box-containing core promoters, bioinformatics searches of the eukaryotic promoter database (EPD) and database of transcription start sites (DBTSS) have shown that the occurrence of BRE^{u} is in fact greater in TATA-less promoters (28.1% and 26.9%) than in TATA-containing promoters (11.8% and 13.8%) (Gershenzon *et al*., 2005), with similar results observed for the BRE^{d} (Deng & Roberts, 2005). These elements are contacted by TFIIB through the TFIIB helix-turn-helix DNA-binding motif (contacting BRE^{u}) and the recognition loop located between helices BH2 and BH3 (contacting BRE^{d}) in order to stabilise and orientate the PIC (Nikolov *et al*., 1995; Lagrange *et al*., 1998; Deng & Roberts, 2005; reviewed in Thomas & Chiang, 2006). While the current text book model of core promoter architecture is still evolving by way of the new observations and potential insights that stem from the evidence of widespread transcription initiation at what were previously thought of as enhancer elements (Anderson *et al*., 2014; Core et al., 2014), as well as promoter-promoter interactions between distal genes displaying context dependent enhancer-type activities (Li *et al*., 2012; Leung *et al*., 2015), it is clear that CPEs such as the TATA box, BREs and Inr play a vital role in the recruitment and positioning of the PIC through direct interactions with TBP(TFIID) and TFIIB, with TFIIA as an important modulator of TBP and TFIID DNA-binding activity.

1.2.3 Transcription factor IID (TFIID)

TFIID is a large multi-protein complex comprised of TBP and a set of 13 TBP-associated factors (TAFs) (Dynlacht *et al* 1991; Pugh & Tjian., 1991; Timmers & Sharp., 1991; reviewed in Papai *et al*., 2011; Gupta *et al*., 2016). Within TFIID, TAFs 4, 5, 6, 9 and 12 are present in two copies. Nine of the thirteen TAFs contain a common histone fold domain (HFD) that facilitates the formation of TAF heterodimers within the TFIID quaternary structure. Cryo-EM images of human TFIID displays a U-shaped three-dimensional structure comprised of three roughly equally-sized lobes that are able to undergo a drastic conformational re-arrangement in response to interactions with TFIIA at the core promoter (Cianfrocco *et al*., 2013). The TBP subunit is responsible for the recognition and binding of the TATA box while recognition of other core promoter elements such as the downstream core element (DCE), initiator (Inr), and downstream promoter element (DPE) are mediated by interactions with TAFs 1, 2, 6, and 9 respectively (Chalkley & Verrijzer., 1999; Burke & Kadonaga, 1997). These TAF-DNA interactions have proven to be extremely important with respect to promoter recognition as many human genes do not contain the TATA box element within their respective core promoters (Suzuki *et al*., 2001; Gershenzon & Ioshikhes, 2005). Following the identification of TBP as the primary TATA box-binding factor (Kao *et al*., 1990; Peterson *et al*., 1990., Hoffman *et al*., 1990), early experiments involving transcriptional assays showed that purified TBP was not the functional equivalent of TFIID and that the TAFs played an important role in activated transcription (Pugh & Tjian, 1990). It has since been determined that the primary function of the TAFs within the TFIID complex is to act as a facilitator of core-promoter recognition, responsible for the proper positioning of the RNA Pol II pre-initiation complex (PIC) at the TSS. In addition, TFIID TAFs can function as co-activators, that enhance PIC assembly though direct interactions with activator proteins bound at proximal and distal promoter elements. Finally,

some TAFs play an active role in the post-translational modification of chromatin (reviewed in Thomas & Chiang, 2006).

1.2.4 TATA-binding protein (TBP)

TBP binds DNA via a saddle shaped, highly conserved C-terminal core domain composed of two imperfect direct repeats of secondary structures that adopt quasi identical tertiary structures. Each of these repeats are made up of five anti-parallel β-strands and two α-helices (Fig. 4A) (Kim *et al*., 1993; Nikolov *et al*., 1996). The central DNA-binding surface is composed of two adjacent β-sheet regions forming a concave DNA-binding site that is flanked by two α-helices. The remaining two α-helices are located on the convex upper surface of the molecule where they act to facilitate interactions with transcriptional co-factors (Kim *et al*., 1993; Nikolov *et al*., 1996; Bleichenbacher *et al*., 2003). The DNA-binding surface also contains two pairs of highly conserved phenylalanine residues, in the human TBP ortholog residues F288/F305 and F197/F214. In the human TBP/TATA complex, F288 and F197 intercalate between the first and last DNA-base steps of the TATA box promoter element, with each interaction supported and positioned by the second phenylalanine residue, F305 and F214 respectively. Together, these insertions result in an overall widening of the DNA-minor grove and two kinks at the 5' and 3' end of the TATA box, resulting in the ≈ 80˚ bend of the DNA-helix architecture typically associated with TBP binding (Kim *et al*., 1993; Nikolov *et al*., 1996). The amino acid composition of the preceding unstructured N-terminal region is found to be highly divergent among species and does not seem to play any direct role in TFIID complex formation and function (Zhou *et al*., 1993). It has however been shown to play a role as a negative regulator of TBP/TFIID by preventing the binding of TBP to the TATA box of some RNAP-II (Zhao & Herr, 2002) and RNAP-III (Mittal & Hernandez, 1997) transcribed genes. Drastic alterations to

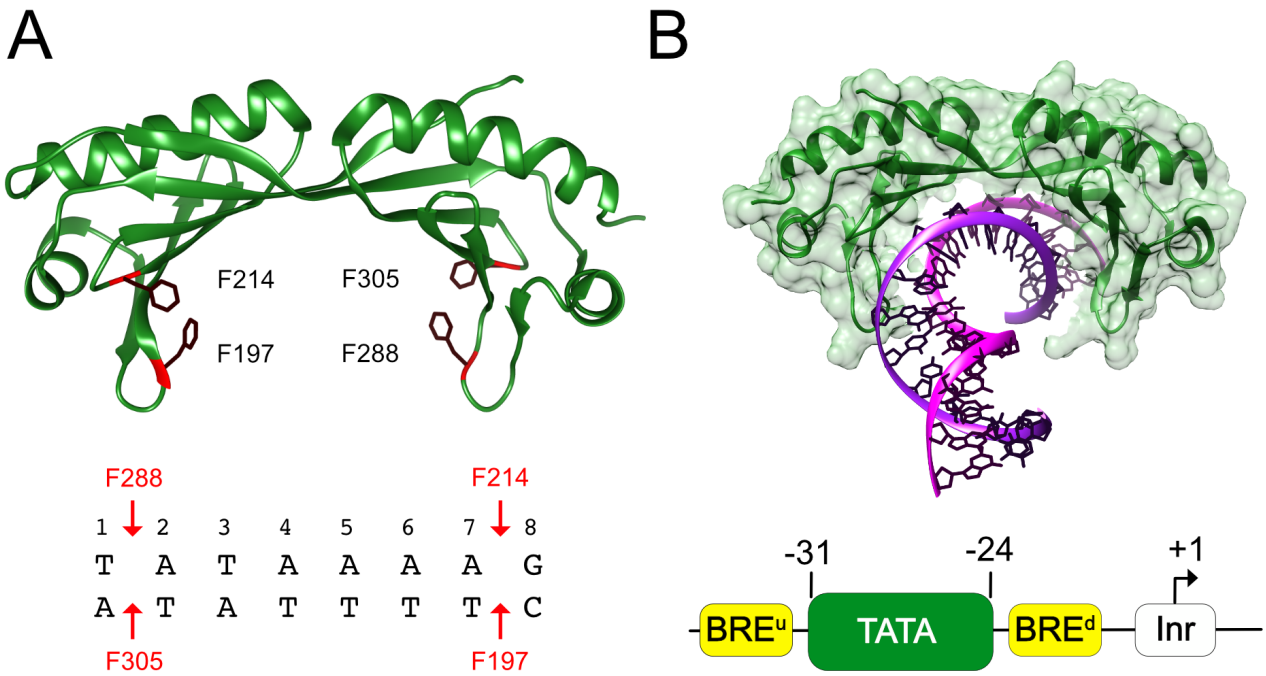


Figure 4: Structure of human TBP. Images generated from co-ordinates for the crystal structure of the human TBP-TFIIA-DNA complex (PDB ID:1NVP; Bleichenbacher *et al*., 2003) (A) Cartoon representation of the TATA-binding protein (TBP) displaying highly conserved phenylalanine residues that intercalate at the first and last base steps of the TATA-box promoter element (TATAAAAG). (B) Crystal structure of TBP-DNA complex bound to the TATA-box element (TATA). The TATA-box promoter schematic shown below illustrates the position of the Ad2ML TATA-box relative to the transcription start site (+1) flanked by up (u) and downstream (d) TFIIB recognition elements.

this N-terminal sequence in mice have also been shown to result in >90% mortality during gestation, linking it indirectly to maternal immunotolerance of the fetus during pregnancy (Hobbs *et al*., 2002). In humans it has been found that a large extension of a large poly-glutamine stretch located in the N-terminus of the TBP molecule has been linked to the occurrence of some neurological pathologies (Nakamura *et al*., 2001).

1.2.5 Regulation of TBP

The formation of TBP dimers has been observed in the crystal structures of multiple TBP orthologs (Nikolov *et al.,* 1992; Chasman *et al*., 1993; Nikolov *et al*., 1996) as well as biochemically through size exclusion chromotography and glycerol gradient sedimentation of recombinant TBP (Kato *et al*., 1994). It is thought that masking the DNA-binding surfaces through TBP homodimerization reduces the propensity for non-specific TBP/ TFIID interaction at AT-rich DNA sequences outside of the core promoter regions, which may ultimately lead to the formation of non-productive transcription complexes. The largest TFIID subunit, known as TAF 1, is also able to negatively regulate TBP binding to the TATA box through direct interaction of the TAF1 N-terminus with the TBP DNA-binding surface (Liu *et al*., 1998) as well as through competition for the TFIIA interaction sites (Kokubo *et al*., 1998) and thereby limits the stabilising effect of TFIIA on TBP-DNA binding as well as any TFIIA-mediated activation by transcription activators from distal regulatory elements. BTAF1 is a TBP-associated factor that was found to co-purify with TBP when whole-cell HeLa extracts where fractionated using a succession of column chromatography steps (Timmers *et al*., 1992). This TFIID-like complex was named B-TFIID based on the phosphocellulose (P11) B fraction from which it was further purified. BTAF1 functions by removing TBP from the TATA box in an ATP-dependent manner (Chicca *et al*., 1998). This function is thought to facilitate the clearance of TBP from non-promoter AT-rich sequences,

thereby promoting the formation of legitimate TBP-TATA complexes. As with TAF1, competition for interaction sites at the convex surface of TBP may impede any TFIIA-mediated transcriptional control (Klejman *et al*., 2005). NC2 is another example of a TBP co-factor implicated in both repression (Goppelt *et al*., 1996) and stimulation (Castaño *et al*., 2000; Lemarie *et al*., 2000; Willy *et al*., 2000) of transcriptional initiation. Crystal structure data shows that NC2 binds TBP-TATA complexes as a type of molecular clamp, interacting with both the TBP-bound DNA as well as the convex surface of the TBP molecule, thereby excluding TFIIA and TFIIB interaction and thus preventing the formation of the PIC complex (Kamada *et al.,* 2001; Innostroza *et al*., 1992; Kim *et al*., 1995; Goppelt *et al*., 1996). Chromatin immunoprecipitation studies have shown that NC2 is found complexed with promoter-bound TBP in both TATA-positive and TATA-negative genes (Gilfillan *et al*., 2005). Lastly, the NC2 complex has also been shown to stimulate transcription from TATA-negative promoters in *Drosophila* via the DPE promoter element (Burke & Kadonaga., 1997; Willy *et al*., 2000).

1.2.6 TBP-like protein (TLP, TRF2)

With regard to TBP, the repertoire of eukaryotic transcriptional machinery has been expanded through gene duplication events that have led to multiple TBP paralogs called TBP-like proteins (TLPs) or TBP-related factors (TRFs) (reviewed in Thomas & Chiang, 2006; Akhtar & Veenstra, 2011). In humans there are two such TBP paralogs, TBP-related factor 2 (TRF2; also known as TLP), and TBP-related factor 3 (TRF3). Of these, only TRF2 has an ortholog predicted amongst the hypothetical proteins annotated in the *P. falciparum* genome. Sequence alignments and 3D modelling have shown a high degree of similarity between the known human TBP and putative TLP secondary structures, with an overall sequence identity of 41% and homology of approximately 60% in the conserved C-terminal

core domain responsible for TBP-DNA interaction (Dantonel *et al*., 1999; Berk, 2000). A defining characteristic of the TLP (TRF2) is the substitution of three of the four key phenylalanine residues (F288/F305 & F197/F214) associated with the TATA box-binding activity of TBP (Section 1.2.4) (Kim *et al*., 1993; Nikolov *et al*., 1996). Phylogenetic analysis aimed at tracing the emergence of TLP(TRF2) through the examination of these phenylalanine substitutions in multiple protist, plant and animal genomes has shown that TLP(TRF2) most likely arose from a TBP-gene duplication event that occurred just prior to the evolution of the bilateria (Duttke *et al*., 2016). The authors of this study also noted that the increased rate of amino acid substitution of TLP(TRF2) relative to TBP, correlated with the loss of TLP(TRF2) DNA-binding activity, possibly due to an uncoupling from evolutionary constraints. They also noted that this increased rate of mutation slowed significantly during vertebrate evolution. The authors do not postulate a reason for this. However, it seems reasonable to assume that by this time TLP had taken on specific roles within the organism's transcriptional programs and therefore had again been tethered by positive selection. Indeed, there are abundant data to suggest that TLP functions in distinct regulatory pathways (Rabenstein *et al*., 1999; Ohbayashi et al., 2003; Chong *et al*., 2005 Wang *et al*., 2014; Kedmi *et al*., 2014). Interestingly, nearly all TFIIA and TFIIB interaction sites are conserved in TLP(TRF2), potentially allowing for TLP(TRF2) to participate in the formation of variations of the canonical RNAP-II PIC (Maldanado, 1999; Moore *et al*., 1999; Rabenstein *et al*., 1999, Teichmann *et al*., 1999). Functional data suggests that TLP (TRF2) can operate as a type of repressor protein by sequestering TFIIA away from TBP (Moore *et al*., 1999; Rabenstein *et al*., 1999, Teichmann *et al*., 1999). More recent work in *Drosophila* demonstrated that TLP functions as a positive regulator at TATA-less promoters for ribosomal protein genes under the control of the TCT sequence motif (Wang *et al*., 2014) as well as DPE- and Inr- dependent genes (Ohbayashi *et al*., 2003; Kedmi *et al*., 2014). It should be noted that the exact function or role of TLP(TRF2) is likely species

specific. RNAi knockdown experiments demonstrated embryonic lethality in TLP(TRF2)-negative *C. elegans* and *X. laevis* lines. In contrast, transgenic mice lacking TLP(TRF2) were viable but defective in male spermatogenesis (Dantonel *et al*., 2000; Kaltenbach *et al*., 2000; Veenstra at al., 2000; Martinov *et al*., 2001; Zhang *et al*., 2001). In summary, it seems clear that in metazoans, TLP(TRF2) serves important gene- and cell-type specific functions, in particular in cell differentiation and development. However, the precise mechanisms by which TLP(TRF2) exerts its influence on gene expression programs remain poorly understood.

1.2.7 Transcription factor IIA (TFIIA)

TFIIA was one of the first TFIID/TBP general transcription factors that were identified by researchers aiming to reconstitute promoter-specific RNAP-II transcription *in vitro* (Matsui *et al.,* 1980). In higher eukaryotes, the TFIIA complex is composed of three subunits: α (35kDa), β (19kDa), and γ (12kDa) (Fig. 5) (Bleichenbacher *et al*., 2003). Initial cloning experiments revealed that these three subunits were coded for by only two genes, with the α and β subunits being expressed as a single αβ peptide chain. In mammalian TFIIA, the 56kDa αβ peptide is post-translationally cleaved to yield 35kDa and 19kDa subunits (DeJong & Roeder., 1993; DeJong *et al*., 1995). Originally, this cleavage was thought to be a universal aspect of active TFIIA complexes in metazoan cells. This paradigm was, however, challenged by the discovery that TFIIA can exist in an uncleaved form complexed with TBP monomers and led to the proposition that TFIIA cleavage may have a regulatory role (Mitsiou *et al*., 2000). The subsequent identification of the cleavage site, termed the cleavage recognition sequence (CRS), provided further evidence for this notion as these cleavage sites are highly conserved among many higher eukaryotes (reviewed in Høiby *et al*., 2007). Comparison of the CRS to other known protease cleavage sites soon identified

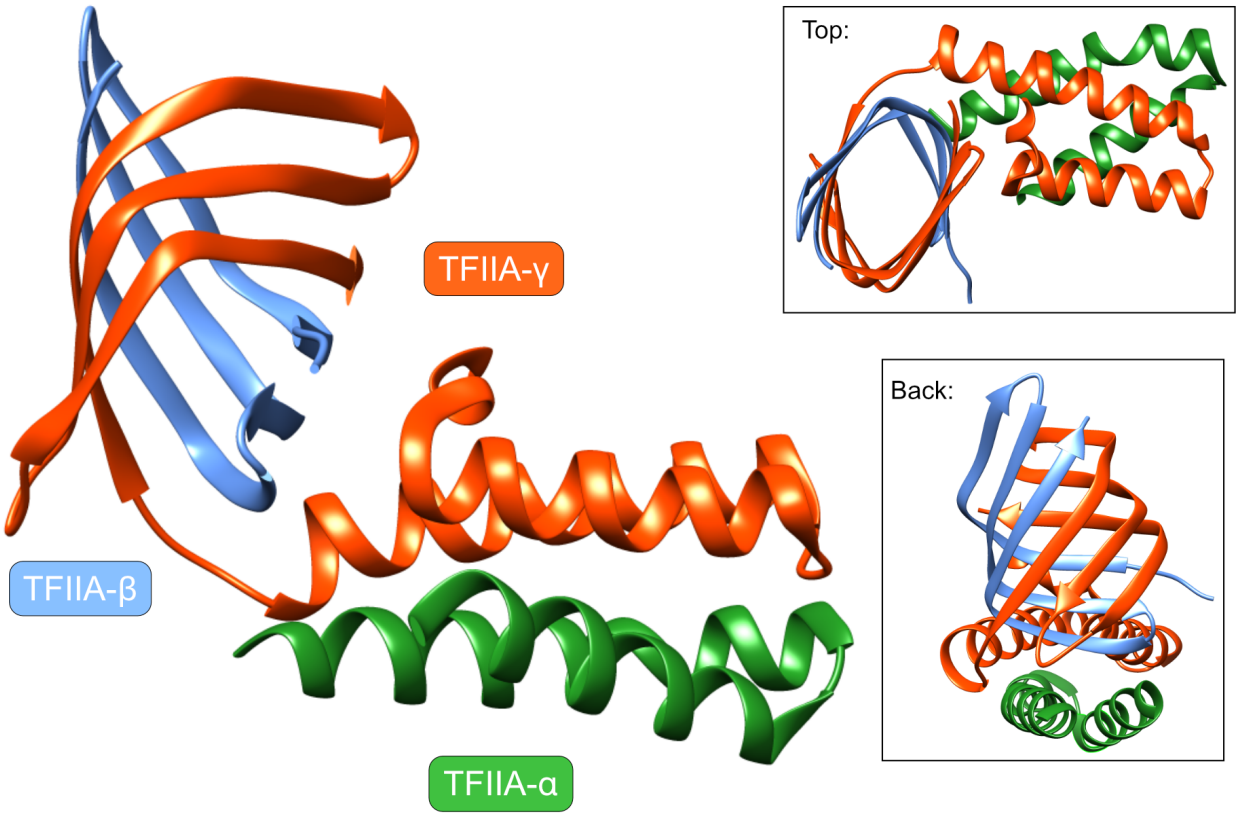


Figure 5: Model of human transcription factor IIA (TFIIA): Cartoon representation of the TFIIA structure coloured according to subunits. TFIIA-α (green) and -β (blue) are shown to form a tightly bound complex with the TFIIA-γ (orange) subunit. Image generated using PyMOL from co-ordinates for the human TBP-TFIIA-DNA crystal structure (PDB ID:1NVP; Bleichenbacher *et al*., 2003)

Taspase1 as the active protease in the processing of TFIIA (Zhou *et al*., 2006). It was found that inhibition of TFIIA cleavage through mutations in the CRS lead to a marked increase in TFIIA half-life (Høby *et al*., 2004). This vulnerability of processed αβ to proteasomal degradation indicated that Taspase1 cleavage may be the first step in a TFIIA degradation pathway and possibly facilitates the fluctuations of cellular TFIIA levels (reviewed in Høiby *et al*., 2007). More recent studies have shown Taspase1 to play a role in Inr-specific core promoter activity (Malecová *et al*., 2015). TFIID mobility-shift assays demonstrated a possible change in TFIIA-TFIID-DNA complex topology that was dependent on whether the complexed TFIIA variant was subject to Taspase1 processing. It was also found that a nucleoprotein-complex containing cleaved TFIIA was less susceptible to negative regulation by NC2. This decreased vulnerability was however dependent on the presence of the Inr element and lead to the proposal that Taspase1 cleavage promotes Inr-specific transcription initiation.

It is well established that TFIIA is able to stimulate transcription via stabilisation of the TBP-TATA complex through interactions with both TBP and specific DNA base pairs upstream of the TATA box element (Fig. 6) (Reinberg *et al*., 1987; Imbalzano *et al*., 1994b; Bleichenbacher., *et al* 2003; Hieb *et al*., 2007). Despite this effect, multiple studies have shown TFIIA to have a negligible effect on basal transcription from TATA-positive genes when using a system constituted from highly purified factors (Sawadago *et al*., 1985; Van-Dyke *et al*., 1988; Wu *et al*., 1998). These contradicting data have caused some in the field to question the classification of TFIIA as a general transcription factor (Thomas & Chiang, 2006). TFIIA has however, in many cases, been found to play a role in activator-mediated transcription, indicating that TFIIA function may be best classed as having a dual anti-repressor/co-activator role (Ozer *et al*.,1994; Yokomori *et al*., 1994; Kobayashi *et al*., 1995). The main argument for classing TFIIA as an anti-repressor is that the stimulatory

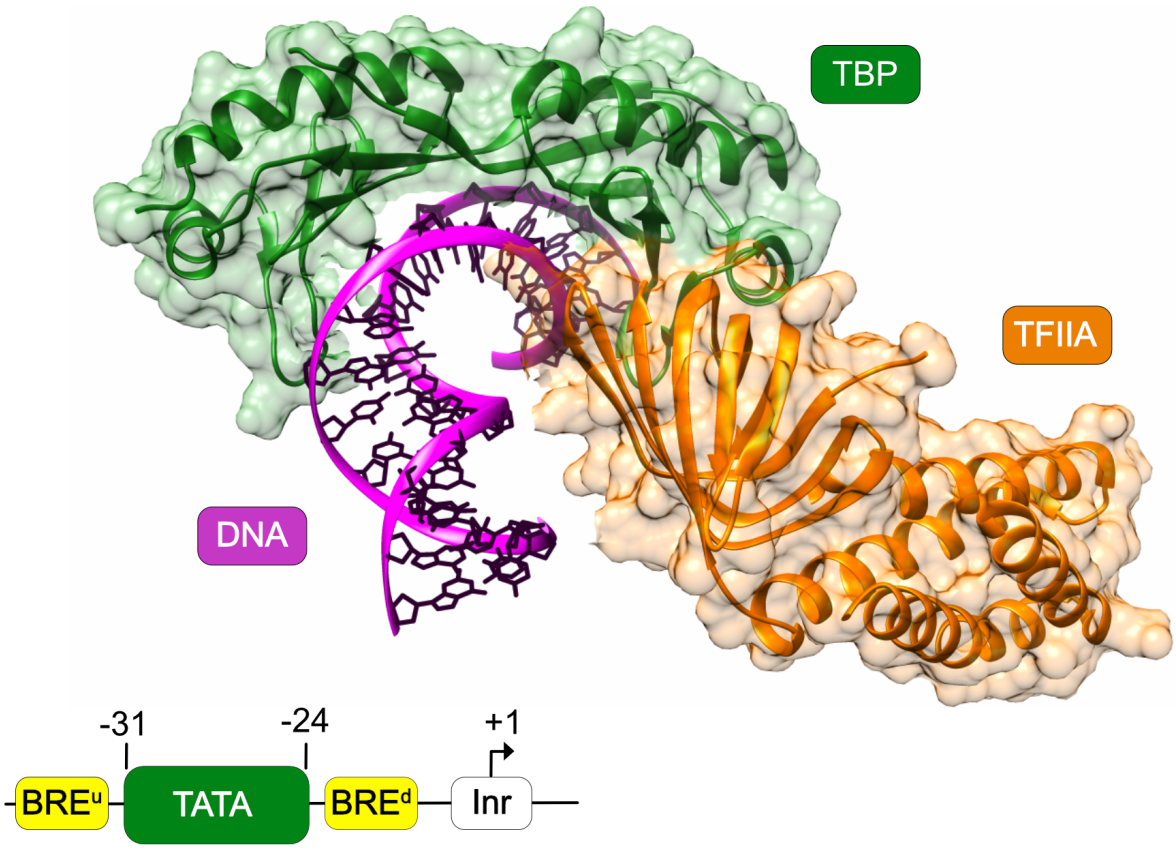


Figure 6: Model of human TBP-TFIIA complex bound to the TATA element. Image generated from PyMOL from co-ordinates for the human TBP-TFIIA-DNA crystal structure (PDB ID:1NVP; Bleichenbacher *et al*., 2003) TATA-binding protein (TBP) is bound to the TATA-box element (TATA) in complex with transcription factor IIA (TFIIA). The TATA-box is shown positioned upstream of the transcription start site (+1) flanked by the up (u) and downstream (d) TFIIB recognition elements.

effects of TFIIA on *in vitro* transcription is most notable when TBP-DNA binding conditions are such that negative regulators of TFIID/TBP may be involved *i.e.* when using partially purified TFIID nuclear fractions (Matsui *et a*l., 1980; Reinberg *et al*., 1987). TFIIA is further noted to extend the range of ionic conditions under which TBP-DNA complexes can form (Imbalzano *et al*., 1994b), and has also been shown to effect the transition from dimeric to monomeric TBP and thus promote the formation of TBP-TATA complexes (Coleman *et al*., 1999). Furthermore, crystallographic and biochemical data indicate that TFIIA interaction at the convex surface of the TBP N-terminal stirrup competes for interaction surfaces with various negative regulators of TFIID/TBP-DNA binding. These factors (all of which are likely present in the aforementioned TFIID crude fractions) include: TAF1, BTAF1, and NC2 (Kokubo *et al*., 1998; Innostroza *et al*., 1992; Kim *et al*., 1995; Goppelt *et al*., 1996; Kamada *et al.,* 2001; Bleichenbacher *et al*., 2003; Klejman *et al*., 2005). In addition, results of studies of TFIIA point mutations in yeast that focused on transcription experiments using nuclear extracts depleted of TFIIA, demonstrated that TFIIA affects transcription dependent on the promoter-type (Kang *et al*., 1995; Ozer *et al*., 1998; Chou *et al*., 1999). In particular, it was noted that there was a significant decrease in transcription from genes requiring activator-mediated initiation (Ozer *et al*., 1998). In a separate study, TFIIA interactions with Gal4 linked VP16 and Zta activation domains have also been shown to stimulate TFIIA-TFIID-DNA complex formation *in vitro* (Kobayashi *et al*., 1995). Lastly, protein-affinity chromatography and run-off transcription assays that used a combination of yeast and human transcription factors have demonstrated direct TFIIA interaction with TFIIE as well as the RAP74 (TFIIF) subunit. The stimulatory effect of TFIIA on transcription within this system was dependent on the presence of TFIIE and TFIIF indicating that it may play an auxiliary activator-role by stimulating the efficient formation of the PIC by these factors (Langelier *et al*., 2001).

1.2.8 *P. falciparum* core promoter architecture and general transcription factors

As was briefly mentioned in Section 1.1.7, putative orthologs for many of the basal RNAP-II transcriptional machinery components have been identified (Coulson *et al*., 2004; Callebaut *et al*., 2005; Bischoff *et al*., 2010). These include the subunits for RNA polymerase II, as well as TFIIA, TFIIB, TBP, TLP(TRF2), TFIIE, TFIIF, and TFIIH. In addition, TFIID subunit orthologs for TAFs 1,2,7, and 10 were also identified. It is interesting to note that all of the histone-fold domain TAFs appear to be missing from the *P. falciparum* 3D7 genome. This suggests a TFIID complex that is very different to the known structures of eukaryotic TFIID (Cianfrocco *et al*., 2013). A recent landmark study mapping RNAP-II transcription initiation sites within the *P. falciparum* genome has provided first insights into the parasite's core promoter architecture and a basis with which to compare experimental findings (Adjalley *et al*., 2016). The authors found that transcription start sites (TSSs) were generally distributed over broad promoter regions. Systematic analysis of these TSSs within each gene locus showed an average of 6 clusters of TSSs both upstream and within the gene-coding regions (GCRs). The peculiar occurrence of multiple such clusters or "TSS blocks" in individual *P. falciparum* genes is possibly indicative of multiple core-promoter regions per gene, each of which has a loosely defined TSS. Eighty-one percent (81%) of these TSS blocks were located within 1000bp of the GCRs, with more than 65% of these falling within 500bp. The majority of TSS block distribution was found only 50bp upstream of the GCRs. The close proximity of TSS blocks to the translation start codon is significant as this suggests very short 5' untranslated regions (UTRs) in the *P. falciparum* mRNA transcripts. Surprisingly, 49% of all TSS blocks were found downstream of the start codons. The vast majority (≈ 90%) of these downstream TSS were also found to be closely followed by an in-frame ATG codon, indicating that many *P. falciparum* transcription products code for truncated gene products.

Finally, 10% of the TSS blocks could not be matched to the current gene annotation, and were instead found within non-coding or very long intergenic regions. The authors speculate that a significant number of these transcripts may feature as part of the parasite's transcriptional regulatory network. Thirty-one percent (31%) of the analysed genes also showed evidence of antisense transcription initiation with 47% of those initiation events occurring in non-coding genome regions and roughly 10% overlapping with the 3' end of the sense-strand genes. It was noted that these overlaps could possibly lead to direct interference with processing and translation of sense strand mRNA. Bidirectional promoter regions were found to be relatively common in *P. falciparum,* as more than half of the annotated transcription units were present in a divergent arrangement, with the bidirectional TSS blocks usually falling within 400bp of each other. With respect to epigenetic markers, the researchers found that trimethylation of histone H3 at lysine 4 (H3K4me3) and acetylation of histone H3 at lysine 9 (H3K9ac) at nucleosomes positioned upstream of the TSSs (+1 nucleosome) were good indicators of TSS blocks. However, these markers, together with nucleosomes containing the histone variants H2A.Z and H2B.Z, were in no way good indicators of transcriptional activity. This finding is in contrast to what has typically been noted in other eukaryotes. When analysing local base content around the most active TSSs, it was found that there was a strong A/T nucleotide bias within these regions. This was found to be true for TSS blocks both upstream as well as downstream of the GCRs, confirming that internal TSS sites are likely to be true TSSs, as opposed to an experimental artefact. The dinucleotide signature T-A at the -1/+1 positions of the TSSs were found to be associated with the most dominant initiation sites. This trend was found not to correlate with relative promoter activity and, rather surprisingly, these TSS blocks also coincided with two downstream GC-islands located approximately 150bp and 210bp from the TSS. Furthermore, it was found that these GC-islands correlated with a decline in local nucleosome occupancy and were equally visible

downstream of internal TSS blocks. An increase in GC content was also found around the borders of TSS blocks, particularly at TSS blocks controlling less active genes. These deviations in nucleotide content correlated with a shift in the +1 nucleosome position and a general decrease in nucleosome-free regions, indicating a link between nucleotide base content, nucleosome positioning, and transcriptional activity. Lastly, the authors provided a detailed analysis of TSS block usage during the blood stage development of the parasite. They first looked for correlations between the relative gene positioning within the genome and TSS block selection. It was found that bi-directional initiation occurred coordinately for the vast majority of the genes found in a divergent configuration, with adjacent 5' regions located on opposite strands, suggesting that this configuration may be a marker for co-expression. Only a small minority (< 10%) of these divergent genes exhibited independent cycling behaviour, reflected in differential temporal patterns of their TSS block usage. Interestingly, the authors noted the occurrence of divergent gene pairs that exhibited distinct transcription profiles whilst maintaining uniformity in the temporal usage of their TSS blocks. This may point to the existence of additional, unidentified core promoter elements present in only one of the two gene promoters. Taken together, the results of this study suggest that bi-directional promoter regions may contain multiple regulatory elements enabling both co-ordinated and dissociated TSS usage.

While this study provides a first advance in the understanding of *P. falciparum* promoter architecture, the precise biochemical mechanisms governing PIC formation and transcription initiation in *P. falciparum* remain to be elucidated. The extreme A/T bias of the *P. falciparum* genome poses an extreme challenge for identification of *P. falciparum* core promoter elements and the identification and functional characterisation of general transcription factors. This factor, together with the conservation of chromatin modifying enzymes (Coulson *et al*., 2004) led the field of malaria research to focus largely on the

epigenetic mechanisms of *P. falciparum* gene regulation in the past (see Section 1.1.7). The general focus of the project presented here was on gene activation at the point of transcription initiation, with a specific focus on PfTBP, PfTLP and two putative PfTFIIA candidates (Callebaut *et al*., 2005) with the hope to highlight any differences to the human system that could provide a rationale for the future development of novel broad-stage antimalarial compounds.

Chapter 2

Materials and Methods

2.1 Bioinformatic analysis of PfTBP, PfTLP, and putative PfTFIIA complexes

All amino acid sequences for known and/or predicted *P. falciparum* general transcription factor orthologs were retrieved from the parasite genome database (PlasmoDB Release 29) (Aurrecoechea *et al*, 2008). Sequences for human and yeast TBP and TFIIA orthologs were retrieved from the National Centre for Biotechnology Information (NCBI) GenBank online protein database (Benson *et al*., 2013). Protein sequences and associated identifier codes are listed in Table 1.

2.1.1 Multiple sequence alignments

Sequence alignments for TBP and TFIIA orthologs were conducted using the European Bioinformatics Institute (EBI-EMBL) Clustal Omega multiple sequence alignment tool with alignment order set to “input” (Larkin *et al*., 2007; Goujon *et al*., 2010). The sequence stretches chosen for alignment and the subsequent annotations were informed by either the known crystallographic structures (Tan *et al*.,1996; Nikolov *et al*., 1996; Bleichenbacher *et al*., 2003), and/or secondary structure predictions based on hydrophobic cluster analysis (HCA) of the putative PfTFIIA subunits (Callebaut *et al*., 2005). The TFIIA-γ subunit Clustal Omega alignments were adjusted manually based on the available HCA data. Gaps were removed in the alignment of residues 1-50 of ScTFIIA-γ1 and PfTFIIA-γ1s, and the PfTFIIA-γ2s sequence (K78 to S169) was shifted four residues to the right so that residue

K78 of PfTFIIA-γ2s was aligned with residues M1 of ScTFIIA-γ1 and PfTFIIA-γ1s. A three residue gap was inserted between T56 and Q57 of ScTFIIA-γ1 and a one residue gap inserted between N57 and Q58 of PfTFIIA-γ1s. The positions of the HsTFIIA-γ residues were added by superimposing the yeast TOA1 and HsTFIIA-γ alignment data from the human and yeast TBP-TFIIA-DNA crystal structures (Bleichenbacher *et al*., 2003). The ScTFIIA-γ1 region Q103 to E122, and the PfTFIIA-γ2s region A174 to I184, were each shifted left by nine and six residues respectively, so that Q103 of ScTFIIA-γ1 was aligned with D158 of PfTFIIA-γ1s, and A174 of PfTFIIA-γ2s. The HsTFIIA-γ1 region T85 to G104 was superimposed as before. Overall, each TFIIA alignment is made up of two separate alignments, one N-terminal alignment, and one C-terminal alignment. The PfTLP alignment is adapted from a larger COBALT sequence alignment by Bing, 2015 of HsTBP and selected TRF orthologs (default settings). COBALT is a constraint based alignment program which uses conserved domain and protein motif databases together with local sequence similarity (RPS-BLAST, BLASTP, and PHI-BLAST respectively) to aid alignment of distantly related proteins (Papadopoulos & Agarwala 2007). For ease of comparison to the Clustal Omega alignments, residue similarities in the PfTLP alignment were annotated manually according to the Gonnet PAM 250 matrix parameters describe in the online EBI Clustal Omega FAQ section (http://www.ebi.ac.uk/Tools/msa/clustalo/help/faq.html). (*) indicates identical residues. (:) indicates conservation between groups of strongly similar properties scoring > 0.5 in Gonnet PAM 250 matrix. (.) indicates conservation between groups of weakly similar residues <+0.5 Gonnet PAM 250 matrix. For all alignments, residues were manually highlighted according to hydrophobicity and/or degrees of conservation. Schematics of the TFIIA secondary structure arrangements were based on the crystallographic data of the human TBP-TFIIA complex (Bleichenbacher *et al*., 2003) as well as HCA data of the putative *P. falciparum* TFIIA orthologs (Callebaut *et al*., 2005).

Database ID	Name	Amino acid sequence
CAG33057.1	HsTBP	MDQNNSLPPYAQGLASPQGAMTPGIPIFSPMMPYGTGLTPQPIQNTNSLSILEE QQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQAVAAAA VQQSTSQQATQGTSGQAPQLFHSQTLTTAPLPGTTPLYPSPMTPMTPITPATPA SESSGIVPQLQNIVSTVNLGCKLDLKTIALRARNAEYNPKRFAAVIMRIREPRTTA LIFSSGKMVCTGAKSEEQSRLAARKYARVVQKLGFPAKFLDFKIQNMVGSCDVK FPIRLEGLVLTHQQFSSYEPELFPGLIYRMIKPRIVLLIFVSGKVVLTGAKVRAEIY EAFENIYPILKGFRKTT
EDN63125.1	ScTBP	MADEERLKEFKEANKIVFDPNTRQVWENQNRDGTKPATTFQSEEDIKRAAPES EKDTSATSGIVPTLQNIVATVTLGCRLDLKTVALHARNAEYNPKRFAAVIMRIREP KTTALIFASGKMVVTGAKSEDDSKLASRKYARIIQKIGFAAKFTDFKIQNIVGSCD VKFPIRLEGLAFSHGTFSSYEPELFPGLIYRMVKPKIVLLIFVSGKIVLTGAKQREE IYQAFEAIYPVLSEFRKM
PF3D7_0506200	PfTBP	MHKEEKEKDKKKDNNYDNDYHNDYHNDYHNDYHNDYDNENDCNNNEREGSV HGDNIHAYEEQNYENDILNIINKQMDNIKSNDPIKNNSNNSYNNNFLDMNFLEQD QLFLENINQDNVVSAHYTSEYDNNEKEKSDDLKNKLVHKNISLNIHNIISSANLCI DINLRLVAVSIRNAEYNPSKINTLIIRLNKPQCTALIFKNGRIMLTGTRTKKDSIMGC KKIAKIIKIVTKDKVKFCNFKIENIIASANCNIPIRLEVLAHDHKEYCNYEPELFAGLV YRYKPTSNLKSVILIFVSGKIIITGCKSVNKLYTVFQDIYNVLIQYKN
NP_056943.1	HsTFIIA-αβ	MANSANTNTVPKLYRSVIEDVINDVRDIFLDDGVDEQVLMELKTLWENKLMQSR AVDGFHSEEQQLLLQVQQQHQPQQQQHHHHHHHQQAQPQQTVPQQAQTQQ VLIPASQQATAPQVIVPDSKLIQHMNASNMSAAATAATLALPAGVTPVQQILTNSG QLLQVVRAANGAQYIFQPQQSVVLQQQVIPQMQPGGVQAPVIQQVLAPLPGGI SPQTGVIIQPQQILFTGNKTQVIPTTVAAPTPAQAQITATGQQQPQAQPAQTQAP LVLQVDGTGDTSSEEDEDEEEDYDDDEEEDKEKDGAEDGQVEEEPLNSEDDV SDEEGQELFDTENVVVCQYDKIHRSKNKWKFHLKDGIMNLNGRDYIFSKAIGDA EW
NP_014837.1	ScTFIIA-αβ (TOA1)	MSNAEASRVYEIIVESVVNEVREDFENAGIDEQTLQDLKNIWQKKLTETKVTTFS WDNQFNEGNINGVQNDLNFNLATPGVNSSEFNIKEENTGNEGLILPNINSNNNI PHSGETNINTNTVEATNNSGATLNTNTSGNTNADVTSQPKIEVKPEIELTINNANI TTVENIDDESEKKDDEEKEEDVEKTRKEKEQIEQVKLQAKKEKRSALLDTDEVG SELDDSDDDYLISEGEEDGPDENLMLCLYDKVTRTKARWKCSLKDGVVTINRN DYTFQKAQVEAEWV
PF3D7_0716400	PfTFIIA-αβ	MSLLSDSDLKEIHKKIIDSTIKNCGQFYNARILEAIKMRWIKIYEQKLRNKELGLVS NKKENKNSKIQLSDDFDDDEFEDADVEEKEIVEIQEEVENDPAEEDELNDLDNIS ISDLSDVDPETNNIIIGVCDKISKPCGRRNVGSNWKIKLKGGLMKIDGKEMFFHG LQGELEF
NP_001307859.1	HsTFIIA-γ	MAYQLYRNTTLGNSLQESLDELIQSQQITPQLALQVLLQFDKAINAALAQRVRNR VNFRGSLNTYRFCDNVWTFVLNDVEFREVTELIKVDKVKIVACDGKNTGSNTTE
NP_012865.1	ScTFIIA-γ (TOA2)	MAVPGYYELYRRSTIGNSLVDALDTLISDGRIEASLAMRVLETFDKVVAETLKDN TQSKLTVKGNLDTYGFCDDVWTFIVKNCQVTVEDSHRDASQNGSGDSQSVISV DKLRIVACNSKKSE
PF3D7_1250700	PfTFIIA-γ1s	MEHGNFDEKFGDSILLESLKEALQQMIEEFYVEKEKGIQIYKEACMNVKKEILDN SNQLSDVHMSGQLKSYYCRNDMWTFFFKNSLFKINKNKKMKSSSKDYKNYQP LNLRVYKNFYDKKEEFLKNCVDKNNVKFFKNFPKLYSNIVHKENNEVESDDVFF YYDGLIKILCIEESTI
PF3D7_0933700	PfTFIIA-γ2s	MSDNFFIPTFEHIYEQSIPPPLIKHSRKCKNYNKKRKNIIKYTSQDICVEQRNYDK ENKEHKYVCINFHTFNKKCYHKNIRLLDHTHLDFTLGESLMYLLNVYLQNKIITKH LYNRMFYSFQLAIIEVISKLKEGKEYKWNIKGKLINSKKENNIQIVYIENAFLSFKSI VLHAPLLKIKSIDI
PF3D7_1428800.2	PfTLP	MYPPCKKKKLNNNEVTNIFLKNENNMSVHNISMNAVLCSSLNLDNIYKYFSNCIY NPREFKCMRIDVPVTLSTVQKYIKYLNNKKEKQNDDICKMKNEQVNQTKHSNN NIKEEIKTNGSNKLSDNKLLDSSNNSSSNKILTNNKSHENLINISKNNSYLDDNVN KNNFFINDNNSDNKKIDTSDIKNMDPFIESEHIINKKLIINVSIFSNGKIICTGNNSIE ACKIAMKKIEKKLKQLNFKNIKLKKITITNILAVYNVGFSIVLPLFAQYYKSVDYDPN VFPACKVKIALMNDENKSNDNKEQNDNNFAWCNAKNTIDKDKSKVDIVSASIFS TGNITLTGGKSYENLQKCINILLPYLIKSKSQH

Table 1: Primary amino acid sequences for TBP and TFIIA orthologs: The sequence regions used for the alignments presented in Ch3. are shown in red. In the instances where more than one alignment was used for a single figure, the regions for the second alignment are shown in blue.

2.2 Cloning of 6H-PfTFIIA-αβ-γ1 and 6H-PfTFIIA-αβ-γ2 expression vectors

2.2.1 pET 11d expression system (Agilent technologies)

The pET 11d expression vector contains the promoter and leader sequences from the T7 bacteriophage *gene 10.* The T7 promoter sequence is rarely found outside of the T7 phage genome and is therefore not transcribed by the host cell polymerase. Target cDNA is cloned downstream of the T7 promoter which can be activated by the IPTG induced expression of T7 polymerase from the *Lac*UV5 promoter of *E. coli* Bl21 (DE3) cells.

2.2.3 Restriction digestion and gel isolation.

Unless otherwise stated, all restriction enzymes, gel extraction kits, and PCR purification kits were obtained from Thermo Fisher Scientific (https://www.thermofisher.com). All protocols were followed as per manufacturers instructions.

2.2.4 Agarose gel electrophoresis

Agarose gels were cast with 0.8-2%(w/v) agarose dissolved in 0.5x TBE (40mM Tris-Cl; pH 8.3, 45mM Boric acid, 1mM Ethylenediaminetetraacetic acid [EDTA]). Electrophoresis was conducted at 80V in 0.5x TBE running buffer until sufficient separation was achieved.

2.2.4 Polymerase chain reactions (PCRs)

PCR reactions were performed according to standard protocol (Sambrook & Russel, 2001) using Kapa Taq PCR kits from Kapa Biosystems (07958447001) in Kapa Buffer A with a

final concentration of 1.5mM Mg^{2+}. PCR amplification of *P. falciparum* sequences were carried out with an elongation temperature of 60˚C for optimal synthesis of A/T-rich sequences (Su *et al*., 1996).

2.2.5 Primers and pET 11d vector constructs used in the cloning and sequencing of the PfTFIIA-γ1 and PfTFIIA-γ2 co-expression vectors

Primers were designed using SerialCloner (Version 2.6, SerialBasics, http://serialbasics.free.fr/Serial_Cloner.html). Melting temperature and self-complementarity analysis was carried out using the Oligonucleotide Properties Calculator available at http://biotools.nubic.northwestern.edu/OligoCalc.html. All primer sequences are listed in Table 2. The vector maps together with open reading frame (ORF) sequences and relevant plasmid features are presented in the appendix sections A1 and A2.

Procedure	Primer name	Primer sequence
Amplification of TFIIA-γ1 expression unit	pET11d-T7XCISE-AatII-FWD	TTCTTAGACGTCTAATACGACTCACTATAGG GGAATTG Tm (50mM NaCl): 63.1 °C | hairpin: no | self-complementarity: yes
Amplification of TFIIA-γ1/ γ2 expression unit	pET11d-T7XCISE-AatII-REV	TTCTTAGACGTCGGGGTTATGCTAGTTATTG CTCAG Tm (50mM NaCl): 63.5 °C | hairpin: no | self-complementarity: yes
Amplification of TFIIAγ2 expression unit	pET11d-T7XCISE-HindIII-FWD	GCATTAAAGCTTTAATACGACTCACTATAGG GGAATTG Tm (50mM NaCl): 63.1 °C | hairpin: no | self-complementarity: yes
Colony PCR TFIIA-γ1	PfIIA-SML1-NcoI-FWD	GATATACCATGGAACATGGGAACTTTGATG Tm (50mM NaCl): 58.4 °C | hairpin: yes | self-complementarity: yes
Colony PCR TFIIA-γ1/γ2	pET11d-AMP-REV	CTCATGAGCGGATACATATTTG Tm (50mM NaCl): 58.4 °C | hairpin: no | self-complementarity: no
Colony PCR TFIIA-γ2	PfIIA-SML2-PagI-FWD	GATATATCATGAGCGATAATTTTTTTATACCCA Tm (50mM NaCl): 57.6 °C | hairpin: yes | self-complementarity: yes

Table 2: Primer sequences for PfTFIIA expression vector construction: Restriction sites shown in blue.

2.2.6 Cloning strategy

Bacterial expression vectors for the 6His-tagged variants of the individual PfTFIIA subunits 6H-PfTFIIA-αβ, 6H-PfTFIIA-γ1s, and 6H-PfTFIIA-γ2s were previously cloned in collaboration with Dr Thomas Oelgeschläger (unpublished data; Milton, 2012). The expression vectors consist of a pET 11d (Promega) vector frame containing a 6His-tag inserted between *Nco*I and *Nde*I sites downstream of the T7 promoter. In an attempt to increase the solubility of the various PfTFIIA subunits (Ch3. Results, Section 3.4.1), co-expression vectors were constructed to explore the possibility of complex formation at the point of translation. Preparative restriction enzyme digests followed by 5' de-phosphorylation reactions were set up to prepare two vector frames (Frame-γ1 and Frame-γ2) from the expression plasmid pET11d-6His-PfTFIIAαβ. Frame-γ1, which was used for the PfTFIIA-γ1 co-expression vector, was prepared by digestion with *Aat*II at base position 73 (5882bp downstream of the T7 terminator sequence). Frame-γ2, which was used for the PfTFIIA-γ2 co-expression vector, was prepared by double digestion with *Hind*III at position 6195 (6122bp downstream of the T7 terminator sequence) and *Aat*II (5882bp downstream of the T7 terminator sequence). A double digest for Frame-γ1 was not possible due to the PfTFIIA-γ1s cDNA containing an internal *Eco*RI site at position 408 and an internal *Hind*III cleavage site at position 5797 of the ORF. The T7 expression units for the untagged PfTFIIA-γ1s and PfTFIIA-γ2s subunits were PCR amplified from plasmids already available in the laboratory: pET11d-6His-PfTFIIAγ1 and pET11d-6His-PfTFIIAγ2. The forward (pET11d-T7XCISE-AatII-FWD; pET11d-T7XCISE-HinDIII-FWD) and reverse (pET11d-T7XCISE-AatII-REV) primers contained the appropriate *Aat*II and *Hind*III restriction site overhangs. PCR products purified using a ThermoScientific GeneJet PCR purification kit (K0701), were digested with *Aat*II (PfTFIIA-γ1s insert) or *Aat*II/*Hind*III (PfTFIIA-γ2s insert) and purified by gel extraction. Ligation reactions with a 1:3 vector to

insert ratio were conducted and followed by transformation into *E.coli* One Shot Stbl13 Top10 cells (Invitrogen). Colonies were screened by colony PCR to identify colonies containing the plasmids with the correct inserts. Plasmid mini-preps were prepared from positive clones before being sent for sequencing analysis. The final vectors contained a T7 expression unit for the 6His-tagged PfTFIIA-αβ subunit followed downstream by a second T7 expression unit for the untagged PfTFIIA-γ1s (396bp downstream) or PfTFIIA-γ2s (290bp downstream) subunits.

2.2.7 DNA sequencing

All DNA sequencing for this project was carried out by the Central Analytical Facilities DNA Sequencing Unit at the University of Stellenbosch, South Africa (http://www.sun.ac.za/english/faculty/science/CAF).

2.3 Expression of *P. falciparum* general transcription factors.

Note: Buffer recipes listed in the Appendix

2.3.1 Bacterial expression system

Codon optimised *E. coli* BL21 CodonPlus (DE3)-RIL (Agilent technologies) was used in all instances. The frequency of usage for the codons AGA (arginine), CUA (leucine), AUA (isoleucine) in *E. Coli* are relatively low in comparison to *P. falciparum* and thus may lead to tRNA depletion through the over-expression of heterologous proteins. The BL21 CodonPlus (DE3)-RIL cells have been optimised for the expression of amino acid sequences from AT-rich eukaryotic genomes by including additional copies of the genes

argU, ileY, and leuW, all of which code for the less common tRNAs often used by *P. falciparum.*

2.3.2 Transformation of protein expression vectors into *E.coli* BL21 CodonPlus (DE3)-RIL

Cells were transformed according to manufacturers recommendations. We found that expression of *P. falciparum* general transcription factors was toxic to *E. coli* cells. For this reason, transformants were spread plated on Super Optimal Broth with Catabolite Repression medium (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) containing 34µg/ml chloramphenicol and 100µg/ml ampicillin. The presence of glucose in the medium prevents leaky expression of toxic proteins and thus selection for cells in which expression is compromised by mutation or deletion in the expression plasmid.

2.3.3 SDS-PAGE analysis and Immunoblot analysis

SDS-PAGE assays were conducted using the Mini-PROTEAN Electrophoresis System (Bio-Rad). Gels were prepared according to standard protocol (Sambrook & Russel, 2001). 5% stacking and 15% separating gels were used for PfTFIIA complexes. Samples were mixed with 4 x Laemmli Sample Buffer (Bio-Rad) and heated to 80˚C for 2min. Gels were run at a constant amperage of 45mA until completion. Gels were stained with Bio-Safe G-250 Coomassie premixed stain (Bio-Rad) according to manufacturers protocol. Precision Plus Protein All Blue Prestained Protein Standards (Bio-Rad) or Colour Prestained Protein Standard Broad Range (NEB) were used in all instances as protein molecular weight markers.

Immunoblots were carried out using a Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad) apparatus. Proteins were transferred to polyvinylidene difluoride (PVDF) membrane (Immobilon-P Membrane; EMD millipore) as per the manufacturers protocol. Following transfer, the membranes were stained for 1hr with Colloidal Gold Total Protein Stain (Bio-Rad) and rinsed with 100ml H_2O and 100ml TBST buffer. The membrane was then blocked by incubation with 5% (w/v) fat free milk powder dissolved in TBST buffer for 30min. Incubation with previously determined primary antibody dilutions (unpublished data; Fig. 13D; Bing, 2015; Talvik, 2016) was carried out in TBST buffer containing 5% (w/v) fat free milk powder overnight at 4˚C. Unbound primary antibody was removed by washing the membrane 4 x 5min in TBST buffer before incubating with a 1:3000 dilution of horseradish peroxidase conjugated anti-rabbit IgG (Sigma-Aldrich) in TBST buffer for 30min at room temperature. Immunoreactant proteins were visualised by chemiluminescence on X-ray film using Pierce ECL Western Blotting Substrate (Thermo Fischer Scientific).

All primary antibodies were produced by Biogenes (www.biogenes.de) Germany. Rabbit polyclonal antibodies were raised against peptides from the respective proteins. Peptide sequences are listed in Table 3.

Database ID	Name	Peptide sequence
PF3D7_0506200	PfTBP	C-SEYDNNEKEKSDDL
PF3D7_0716400	PfTFIIA-αβ	C-GLVSNKKENKNSKI
PF3D7_1250700	PfTFIIA-γ1s	C-INKNKKMKSSSKDY
PF3D7_0933700	PfTFIIA-γ2s	C-VEQRNYDKENKEHK
PF3D7_1428800.2	PfTLP	C-NDENKSNDNKEQND

Table 3: Peptide sequences used to raise rabbit polyclonal antibodies

2.3.4 Thrombin cleavage removal of hexahistidine tag (6H-tag)

Thrombin cleavage of the 6H-tag from the recombinant PfTFIIA complexes were carried out using immobilised thrombin (Thrombin CleanCleave Kit; Sigma). Two separate 200µl cleavage reactions were set up using 320ng of either 6H-PfTFIIA-γ1 or 6H-PfTFIIA-γ2 by combining 100µl protein sample (40mM Tris-HCl pH 7.2, 100mM KCl, 0.5% NP-40, 10mM β-EtSH, 100ng/µl BSA) with 100µl 2x reaction buffer (2x Cleancleave buffer,100mM KCl; 0.3% NP-40; 10mM β-EtSH, 100ng/µl BSA). Reactions were incubated at 21˚C with 40µl immobilised thrombin agarose in a tabletop shaker set at 150rpm. Cleaved fractions were centrifuged briefly by microcentrifuge to settle the resin before being pipetted off. Cleaved and uncleaved fractions were analysed by SDS-PAGE and visualised by silver staining (ProteoSilver Silver Staining Kit; Sigma).

2.3.5 Expression of PfTFIIA-γ1 and PfTFIIA-γ2 complexes

The bacterial expression vectors for the 6H-PfTFIIA-γ1 or 6H-PfTFIIA-γ2 complexes were transformed into *E. Coli* BL21 CodonPlus (DE3)-RIL cells and plated on selective SOC-agar plates containing 100µg/ml ampicillin (Amp) and 34µg/ml chloramphenicol (Cam). A freshly selected colony was then used to inoculate a 50ml LB starter culture containing 100µg/ml Amp, 50µg/ml Cam, and 50mM glucose. The starter culture was incubated at 37˚C overnight in an orbital shaker set at 250rpm. The following day the starter culture was harvested by centrifugation at room temperature at 3000rpm in a Beckman JA-14 rotor for 5min. The cell pellet was resuspended in 25ml LB (Amp 100µg/ml, Cam 50µg/µl) pre warmed to 37˚C in order to remove β-lactamase and expose the cells to fresh antibiotics. Expression cultures (9 x 500ml LB, Amp 100µg/ml, Cam 50µg/µl) were then inoculated with 2.5ml resuspended starter culture and incubated at 37˚C in an orbital shaker at 250

rpm until an OD_{550} reading of 0.4-0.6 was reached. Cultures were then cooled on ice to 25˚C before being incubated for a further 20 hours at 25˚C with shaking (250rpm). No IPTG was used for induction due to the toxicity of induced over expression. Instead, protein production relied solely on the leaky expression of the *Lac*UV5 promoter, leading to bursts of mRNAs coding for the T7 polymerases required for the activation of the pET 11d expression system. Saturated cultures were harvested the next day by centrifugation at 3000rpm for 10min at 4˚C in a Beckman JA-14 rotor. Cell pellets were combined and re-centrifuged for 10min to remove the remaining LB medium. A combined wet cell mass of 20g was obtained from 4.5L expression culture. The cell pellets were flash frozen in liquid N_2 and stored at -80˚C before further processing.

2.3.6 Purification of PfTFIIA-γ1 and PfTFIIA-γ2 complexes

Cell pellets were resuspended in 5ml sonication buffer (50mM Tris-HCl pH 7.2, 500mM KCl; 0.5% Nonidet P-40; 10mM β-mercaptoethanol) per gram cell mass containing 50µl/g protease inhibitor (p8340, Sigma-Aldrich), 1mg/ml lysozyme, 1µl Benzonase per 5g cell mass (Sigma-Aldrich) to reduce viscosity of the resultant cleared lysate, and incubated on ice for 30min. Suspensions were sonicated 6 x 10sec with 20sec cooling periods between bursts. Lysates were cleared by centrifugation at 10 000 rpm at 4˚C in a Beckman JA-14 rotor for 30min, transferred to a clean centrifuge tube, and centrifuged a second time. Cleared lysate (approximately 100ml total) was then passed over 300µl Nickel-Nitrilotriacetic acid (Ni-NTA) affinity resin (Sigma) in a Poly-Prep chromatography column (Bio-Rad) by gravity flow. The protein-bound resin was washed with 10ml sonication buffer followed by 10ml wash buffer (20mM Tris-HCl pH 7.2, 500mM KCl, 0.5% NP-40, 10mM β-EtOH, 20mM imidazole) and eluted with 3ml elution buffer (20mM Tris-HCl pH 7.2, 20% v/v glycerol, 500mM KCl, 0.5% NP-40, 10mM β-EtSH, 250mM imidazole). Elutions were

collected in 500µl fractions and analysed by SDS-PAGE. The protein-containing fractions were then salt adjusted to 100mM KCl using a PD Miditrap buffer exchange column (GE Healthcare). Salt-adjusted Ni-NTA eluates were further purified, either on Q sepharose (PfTFIIA-γ1) or SP sepharose (PfTFIIA-γ2) ion exchange resins (Sigma). Eluates were passed three times over 250µl resin in Poly-Prep disposable columns (Bio-Rad) by gravity flow. The columns were rinsed with 6ml BC-100 buffer (BC-0 buffer containing 100mM KCl). The bound proteins were eluted stepwise in 300µl fractions with BC-0 buffer containing increasing KCl concentrations (BC-200,-300,-400,-500,-600,-700). BSA was added to eluted fractions to a final concentration of 100ng/µl before flash freezing in liquid N_2 and storing at -80˚C. The purification scheme developed for PfTFIIA-γ1 and PfTFIIA-γ2 can be found in figure 12A.

2.3.7 Expression and purification of PfTBP and PfTLP

Expression and purification of 6His-tagged PfTBP core domain (PfTBP) that was used for DNA-binding assays was done in collaboration with Ms. Gertrud Talvik (Talvik, 2015). 6His-tagged PfTLP (PfTLP) used in DNA-binding assays was provided by Mr Steven Bing (Bing, 2015). A summary of the purification schemes developed are presented in figure 7.

2.4 Promoter sequences and DNA probe preparation for EMSA and ITA analysis

The promoter sequences chosen for this study are found upstream of genes critical to parasite pathogenicity. *gbp-130* encodes a 130kD glycophorin-binding protein found on the merozoite cell surface. This protein has been shown to be an essential mediator of erythrocyte binding and subsequent parasite invasion (Perkins, 1984). The *kahrp* gene encodes the knob-associated histidine-rich protein. Multiple studies have linked

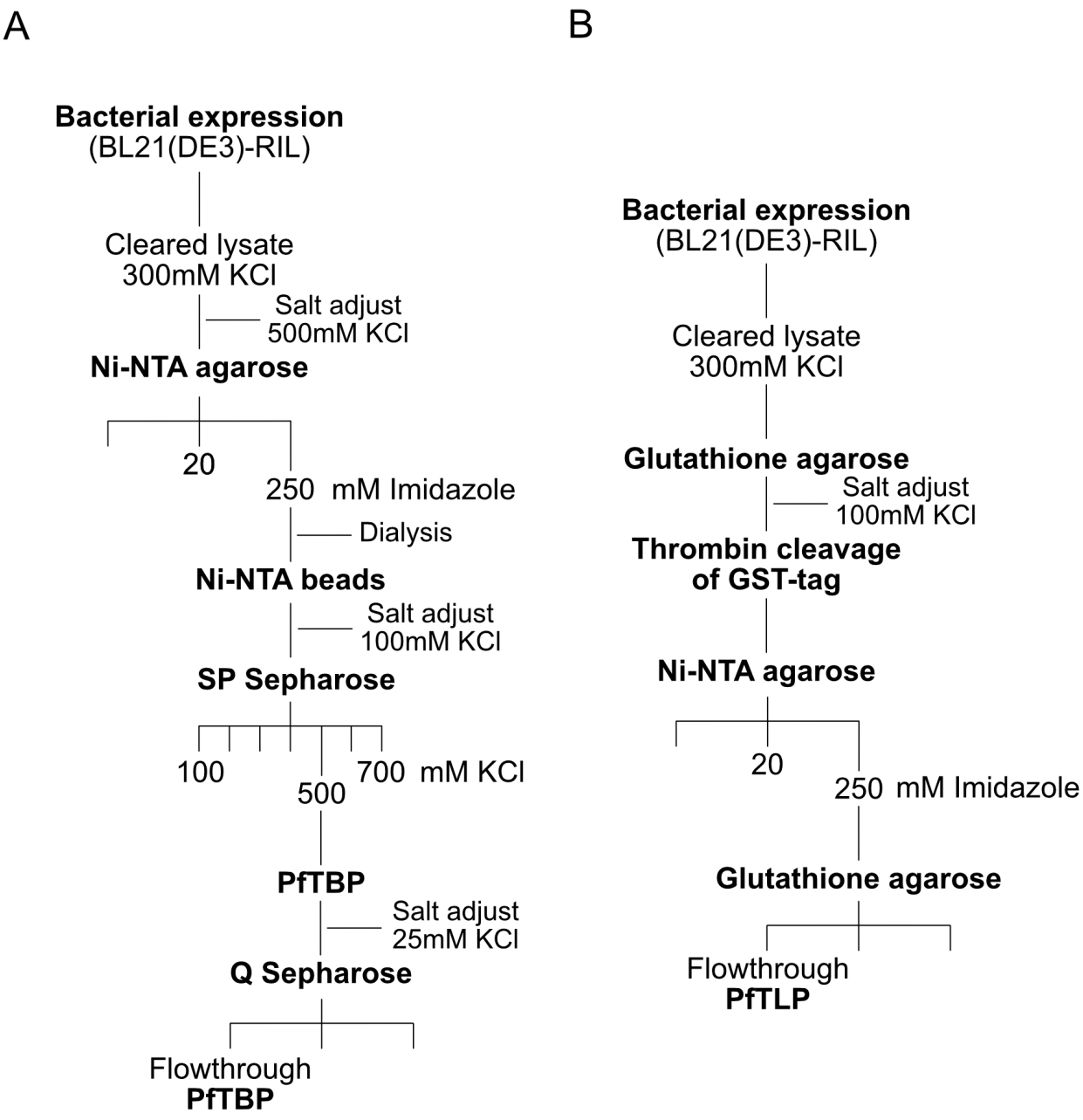


Figure 7: Summary of purification steps for 6H-PfTBP and 6H-PfTLP. (A) Purification scheme for 6H-PfTBP expressed and purified in collaboration with Talvik, 2016. (B) Purification scheme for 6H-PfTLP expressed as double epitope tagged GST-6H-PfTLP (Bing, 2015).

interactions between this protein and host RBC membrane proteins to the cytoadherence displayed by infected red blood cells (RBCs) (Pei *et al*., 2005; Weng *et al*., 2014). Transcription start sites (TSS) for both the *gbp-130* and *kahrp* genes have been mapped (Lanzer *et al*., 1992a; Lanzer *et al*., 1992b). The GC-rich TATA-less mouse *terminal deoxynucleotidyl transferase* (TdT) promoter was used for comparative purposes. The wild-type TdT sequence lacks any TATA-motif while a mutated variant (TdT [TATA+]) contains a prototypical TATA-box element TATAAAA known to promote stable binding of eukaryotic TBPs (Tan., *et al* 1993; Nikolov *et al*., 1996; Malecová *et al*., 2015).

Plasmid vectors containing the putative *P. falciparum* promoter sequences for the *gbp-130* (pTHG5gbp130) and *kahrp* (pTHG5Kharp) genes are derivatives of pGEM-7Zf (Promega) and contain promoter regions from -181 to +51 of the *gbp-130* gene and -172 to +63 of the *kahrp* gene. For EMSA probe generation, the GBP-130 and KAHRP probes were PCR amplified from these vectors using biotinylated primers Bio-pTOTdTPCR-FWD and Bio-pTOTdTPCR-REV listed in Table 4. The mouse TdT and TdT (TATA+) promoter sequences spanning the regions -41 to +33 and were amplified from plasmids pTOG5TdT(-41to+33) and pTOG5TdT(-41/TATA/to+33) respectively (Malcova *et al*., 2007). ITA probes were generated using a biotinylated forward primer (Bio-pTOPCR-US-FWD) and an unlabelled reverse primer (pTO(G5)TdTPCR-REV). Following PCR, probes were purified by polyacrylamide gel purification using the Qiagen QiAquick purification protocol for polyacrylamide gel extraction. The DNA probe sequences are listed in Table 5.

Procedure	Primer name	Primer sequence
EMSA probe generation	Bio-pTOTdTPCR-FWD	GAATTCTAGAGTCTCCGCTCG Tm (50mM NaCl):61.2˚Cl hairpin:no lSelf-complimentarity: yes
EMSA probe generation	Bio-pTO(G5)TdTPCR-REV	AACGCGTTGGGAGCTCTCC Tm (50mM NaCl):61.6˚Clhairpin:no lSelf-complimentarity: no
ITA probe generation	Bio-pTOPCR-US-FWD	AATTGGGCCCGACGTCGCA Tm (50mM NaCl):61.6˚C|hairpin:no |Self-complimentarity:no
ITA probe generation	pTO(G5)TdTPCR-REV	AACGCGTTGGGAGCTCTCC Tm (50mM NaCl):61.6˚Clhairpin: no l Self-complimentarity: no

Table 4: Primer sequences for DNA-probe generation

Promoter	Sequence (5' to 3')	Length (bp)
kahrp	CGAGCGGAGACTCTAGAATTCtgatgatatgttttatatttcttgttttaaattattagaataaaaagaaataattattatttcatggaaataatataatattattattttttttttttaattatttagtagttatgttttgtcgtttttctcatttatttattataatttacctatagtatatactatgatgtatatttactctagtatgaagaataaagttaatgtaaaatattactacactacatgcagttttaGGATCCGGAGAGCTCCCAACGCGTT	282
gbp-130	CGAGCGGAGACTCTAGaattctttgaagtacactcaaaataagttatataccatatgtttttaacatatattatatatatatatatatatataatataatataatattataattatttttaatattattaaattgaacataattattttatatcttactattatttttagaaaatttattatatatacatgcaatcataaataatgttttccctgaaccttttttcaatgaaataagttaacacaccattcctttttGGATCCGGAGAGCTCCCAACGCGTT	282
TdT	AATTGGGCCCGACGTCGCAtgctcctctagagcttgcatgcctgcaggtcggagtactgtcctccgagcggagtactgtcctccgagcggagtactgtcctccgagcggagtactgtcctccgagcggagtactgtcctccgagcggagactctagaattcgagctcggtacctatgggtctgctggtgagaggacatcagagccctcattctggagacaccacctgatggcacagacagaggatccGGAGAGCTCCCAACGCGTT	266
TdT (TATA+)	AATTGGGCCCGACGTCGCAtgctcctctagagcttgcatgcctgcaggtcggagtactgtcctccgagcggagtactgtcctccgagcggagtactgtcctccgagcggagtactgtcctccgagcggagtactgtcctccgagcggagactctagaattcggtacctatgtataaaatggtgagaggacatcagagccctcattctggagacaccacctgatggcacagacagaggatccGGAGAGCTCCCAACGCGTT	260
ITA probes: blue- primer binding sites; red- TSS		
kahrp	AATTGGGCCCGACGTCGCATGCTCCTCTAGAGCTTGCATGCCTGCAGGTCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGACTCTAGAATTCtgatgatatgttttatatttcttgttttaaattattagaataaaaagaaataattattatttcatggaaataatataatattattattttttttttttaattatttagtagttatgttttgtcgtttttctcatttatttattataatttacctatagtatatactatgatgtatatttactctagtatgaagaataaagttaatgtaaaatattactacactacatgcagttttaGGATCCGGAGAGCTCCCAACGCGTT	422
gbp-130	AATTGGGCCCGACGTCGCATGCTCCTCTAGAGCTTGCATGCCTGCAGGTCGGAGACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGACTCTAGAATTCtttgaagtacactcaaaataagttatataccatatgtttttaacatatattatatatatatatatatatataatataatataatattataattatttttaatattattaaattgaacataattattttatatcttactattatttttagaaaatttattatatatacatgcaatcataaataatgttttccctgaaccttttttcaatgaaataagttaacacaccattcctttttGGATCCGGAGAGCTCCCAACGCGTT	422

Table 5: Sequences of DNA-probes used in EMSA and ITAs: Primer binding-sites shown in blue. Transcription start sites (TSSs) shown in red. TATA box motif shown in orange.

2.4.1 Immobilisation of ITA probes to Streptavidin magnetic beads

Probes were immobilised on Dynabeads M-280 Streptavidin (Invitrogen) as per manufacturers instructions. Beads were re-suspended and stored in 1x template immobilisation wash buffer (10mM Tris-HCl pH 7.5; 0.5mM EDTA; 1M NaCl) containing 100ng/µl BSA. Binding efficacy was checked by an agarose gel comparison of known amounts of unbound probe to that of bound probe liberated from the beads through phenol-chloroform extraction and recovered by ethanol precipitation.

2.5 Electrophoretic mobility shift assays (EMSAs)

4µl template mix (50mM HEPES pH 8.0, 12.5mM $MgCl_2$, 12.5mM DTT, 250ng/µl BSA), containing 5*f*mol promoter DNA-probe and the relevant amount of poly(dG-dC) competitor DNA, were combined with a 6µl protein mix's, containing tested protein fractions diluted in BC-100 buffer (BC-0 buffer, 100mM KCl), to give a final reaction volume of 10µl (20mM HEPES pH 8.0, 5mM $MgCl_2$, 5mM DTT, 62.5mM KCl, 12.5% glycerol v/v, 25mM Tris-HCl, 100ng/µl BSA). Reactions were incubated either on ice (0˚C) or 30˚C for 45min. Protein-DNA complexes were then separated from free probe by Mg^{2+} agarose gel electrophoresis at 4˚C (1.4-1.5% w/v agarose, 0.5x TBE, 5mM $MgCl_2$). The probes were then transferred to a positively charged nylon membrane (Immobilon-Ny+, EMD Millipore), UV cross-linked three times for 1min (1200J/cm^2), and visualised on X-ray film by chemiluminescence using a Pierce Chemiluminescent Nucleic Acid Detection kit (Thermo-Scientific).

2.6 Immobilised template assays (ITAs)

20µl template mix's (50mM HEPES pH 8.0, 12.5mM $MgCl_2$, 12.5mM DTT, 250ng/µl BSA) containing 1000$f$mol immobilised probe and 400ng poly(dG-dC) competitor DNA were combined with 30µl protein mix's, containing tested protein fractions diluted in BC-100 buffer (BC-0 buffer, 100mM KCl) to give a final reaction volume of 50µl (20mM HEPES pH 8.0; 5mM $MgCl_2$; 5mM DTT, 100ng/µl BSA). Reactions were incubated at 30˚C for 45min. The probes were then magnetically separated from the binding reaction mix and washed with 100µl reaction buffer (20mM HEPES pH 8.0; 5mM $MgCl_2$; 5mM DTT, 100ng/µl BSA). Finally, bound proteins were eluted with 1x SDS-loading buffer and heated at 80˚C for 2min. These fractions were then loaded onto a 15% SDS-PAGE gel together with known quantities of protein input for immunoblot analysis.

2.7 Systematic evolution of ligands by exponential enrichment (SELEX)

SELEX is a well-established molecular biology approach that can be used to elucidate the DNA-binding specificity of proteins by making use of successive rounds of DNA-binding and amplification of selectively bound DNA by PCR to enrich a pool of random oligonucleotides for sequences that are favoured by the target protein(s). A detailed protocol is described in Ogawa & Biggin, 2012. Primers and DNA-probes that were used in the SELEX procedure are listed in Table 6. Unless otherwise stated, all restriction and/or DNA-modifying enzymes used here were obtained from New England BioLabs (www.neb.com). All PCR purification and gel extraction kits were obtained from Qiagen (www.qiagen.com). Procedures were carried out as per manufacturers instructions unless noted otherwise. A Qiagen QiAgility pipetting robot was used for all pipetting involved in the setting up of qPCR reactions.

Procedure	Primer name	Primer sequence
n16 library template	ssSELEX72-n16	GGATTTGCTGGTGCAGTACAGTGGATCC(n16)GGATCCTTAGGAGCTTGAAATCGAGCAG
SELEX selection rounds	Selex80-*Bam*HI-FWD	GGATTTGCTGGTGCAGTACA Tm (50mM NaCl):58.4˚C
SELEX selection rounds	Selex80-*Bam*HI-REV	CTGCTCGATTTCAAGCTCCT Tm (50mM NaCl):58.4˚C
SELEX concatemerization	Bio-Selex80-*Bam*HI-FWD	GGATTTGCTGGTGCAGTACA Tm (50mM NaCl):58.4˚C
SELEX concatemerization	Bio-Selex80-*Bam*HI-REV	CTGCTCGATTTCAAGCTCCT Tm (50mM NaCl): 58.4˚C
Colony PCR	M13-FWD	GTTTTCCCAGTCACGAC Tm (50mM NaCl):52.4˚C
Colony PCR and sequencing	M13-REV	CAGGAAACAGCTATGAC Tm (50mM NaCl):49.9˚C
SELEX concatemerization II	Bio-Plus*Eco*RI-FWD	ACTGGCGAATTCGGATTTGCTGGTGCAGTACA // Tm (50mM NaCl):73.8˚C
SELEX concatemerization II	Bio-Plus*Eco*RI-REV	ACTGGCGAATTCCTGCTCGATTTCAAGCTCCT // Tm (50mM NaCl):73.8˚C

Table 6: Primer and template sequences used in the SELEX procedure: Restriction enzyme sites shown in blue. The SELEX concatemerization II formed part of a strategy used to identify truncated SELEX products formed in the original SELEX enrichment rounds (Ch3. Results, Section 3.6.3)

2.7.1 Generation of SELEX input DNA library (n16 library)

Single stranded 72bp DNA molecules (ssSELEX72-n16) containing a central randomised 16bp sequence flanked by 28bp primer binding sites were synthesised and purified by the MCB oligonucleotide synthesis laboratory. A double stranded DNA library (dsSELEX72-n16) was generated by a single-cycle polymerase reaction using the ssSELEX72-n16 template and the reverse primer (SELEX80-*Bam*HI-REV). 15 x 50µl single cycle reactions were carried out with the following parameters: 95˚C 5min, 95˚C 30sec, 64˚C 1min, 72˚C 9min using a two step reaction profile. The first step consisted of a 95˚C 5min denaturing step followed by a cycle of 95˚C denaturation for 30sec, annealing at 64˚C for 1min, and

elongation at 72˚C for 1min. The reactions were then pooled and gel purified using 15% PAGE and the user developed protocol for polyacrylamide gel extraction published in the Qiagen QIAEX II handbook. The concentration of the final preparation was determined by nano-drop and adjusted to 10ng/µl. The quality of the preparation was assessed by polyacrylamide gel electrophoresis analysis. In order to confirm that there is no inherent sequence bias within the commercially synthesised pool of random oligonucleotides, the SELEX procedure was carried out in the absence of the protein DNA-binding steps by using a dilution of the random sequence library (n16 library) as an “eluted” fraction with which to follow through with the subsequent amplification, cloning, and sequencing procedures (Fig. 8) (Ogawa & Biggin, 2012). A large amount of biotinylated n16 library sequences were produced by PCR amplification with biotinylated primers (Bio-Selex80-BamHI-FWD, Bio-Selex80-BamHI-REV; Table 6) using a two step PCR reaction profile. The first step consisted of a 5min 95˚C denaturation step followed by 20 cycles of 95˚C denaturation for 30sec, annealing at 64˚C for 1min, and elongation at 72˚C 1min. After the initial 20 cycles a 20µl supplement mix containing fresh Kapa Taq polymerase and primers was added to each reaction (now 40µl in total) followed by a final single reaction step to ensure maximum yield of double stranded product with the following parameters: denaturation at 95˚C for 30sec, annealing at 64˚C for 5min, and elongation at 72˚C for 10min. The resultant product was gel purified and analysed by native polyacrylamide gel electrophoresis. 10µg of the biotinylated sequences were digested with 1000U *Bam*HI for 3hrs in NEB Buffer 3.1 to remove the flanking sequences. This was followed by incubation with 40µl packed Streptavidin sepharose (GE healthcare) for 3hrs in 500µl 0.3M NaOAc pH 5.20 to remove flanking sequences from the cleaved 16bp random sequences (n16). A 20µl ligation reaction containing the *Bam*HI digested fragments in 1x T4 ligase buffer was incubated overnight at 4˚C with 10U T4 ligase. The resultant concatemers (250bp-1000bp) were isolated by agarose gel extraction. The 3’ and 5’ overhangs were filled in by

incubation in 1x NEB buffer 2 with 9U T4 DNA Polymerase and 15U Klenow fragment (Lrg) in a 50µl reaction incubated at 21˚C for 1hr. The resultant mixture of concatemers were separated by 1.5% agarose gel electrophoresis resulting in a smear of DNA within the gel lane. Concatemers of the desired sizes were isolated using the molecular weight markers as a guide to identify the correct regions of the gel to be excised. Once isolated, the concatemers were gel purified using a GeneJet gel purification kit (ThermoScientific), followed by phenol–chloroform extraction and ethanol precipitation to remove any remaining salts from the extraction procedure. Purified concatemers were blunt-end cloned in a 10µl ligation reaction containing 24ng *Sma*I digested pUC19 sequencing vector and 5U T4 ligase (Fermentas) in 1x T4 ligase buffer (Fermentas) incubated overnight at 16˚C. The resultant ligation reaction was then purified by phenol–chloroform extraction and ethanol precipitation, with the DNA being resuspeded in 2µl H_2O. 1µl pUC19-ligated concatemers were then transformation by electroporation (2.0Kv 200Ω 25µF) into ElectroMAX DH5α-E competent cells (Invitrogen). Transformants were plated on LB agar plates containing 50mg/ml X-gal, 100mM IPTG, and 150µg/ml ampicillin. Recombinant transformants were identified by blue/white selection (Sambrook & Russel, 2001) and the resultant clones were further screened for inserts by colony PCR. Plasmids containing inserts > 300bp were then amplified and isolated by plasmid mini-prep before being sent for sequencing. The 16bp sequences were screened for any apparent sequence bias using the online Multiple Em for Motif Elicitation (MEME) application of the MEME-suite (http://meme-suite.org/; Bailey *et al*., 2009).

2.7.2 Immobilisation of 6H-PfTBP on Ni-NTA magnetic beads

100ng 6H-PfTBP was immobilised to 10µl nickel magnetic beads (Merck Millipore;) in a 50µl binding reaction containing 20mM HEPES pH 8.0, 5mM $MgCl_2$, 1mM DTT, 20% v/v

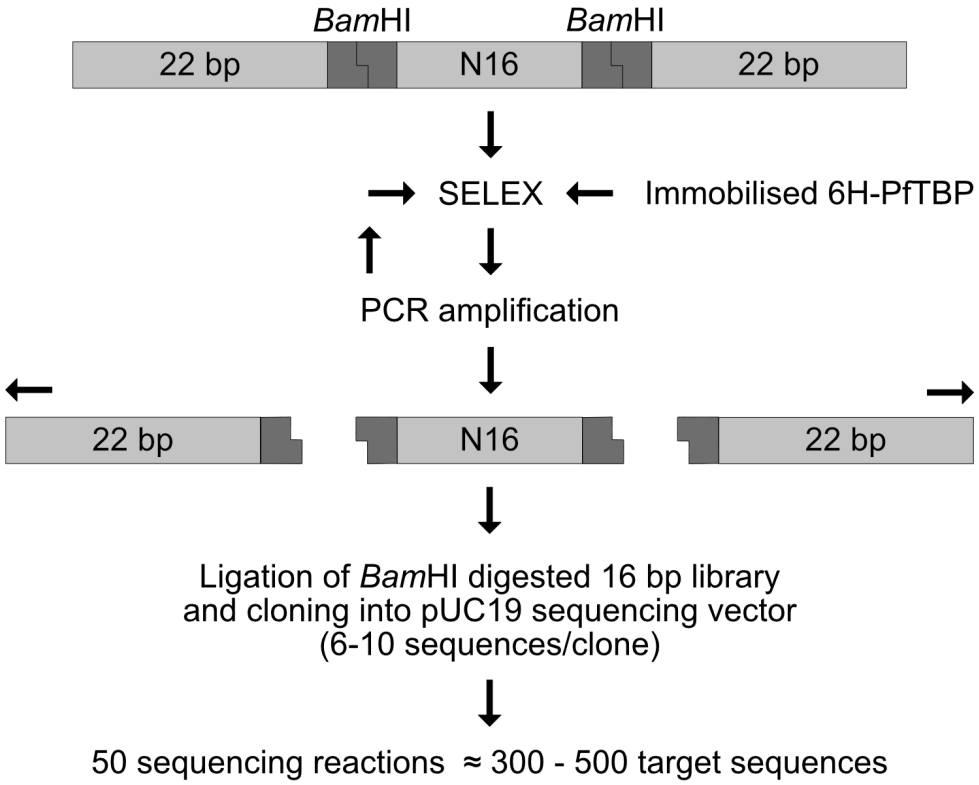


Figure 8: SELEX workflow. 20µl DNA-binding reactions containing 10ng random 72bp oligonucleotide probes (n16) were incubated at 21˚C in the presence of 400ng poly(dG-dC) with 12ng 6H-PfTBP immobilised to Ni^{2+}-magnetic beads. Bound DNA oligonucleotides were separated from the unbound fraction using a magnetic-separator and washed with 1000µl SELEX wash buffer followed by elution in high salt (500mM NaCl). The bound fraction was PCR amplified and used as input DNA for the next round of selection. Once several rounds of SELEX have been completed the enriched oligonucleotide library was PCR amplified with Biotin-tagged primers followed by digestion with *Bam*HI and incubation with Streptavidin sepharose to remove the flanking sequences. The 16bp probes were ligated and cloned into pUC19 for sequencing analysis.

glycerol, and 100ng/µl BSA. Reactions were incubated at 21˚C for 1hr with shaking in Eppendorf LoBind microcentrifuge tubes. Beads were washed with 2 x 100µl reaction buffer and resuspended in 120µl BC-100 complete (BC-O, 100mM KCl, 100ng/µl BSA, 0.5% NP-40).

2.7.3 Selective enrichment of PfTBP DNA-binding motifs by SELEX using 72bp dsSELEX-n16 input library.

Quantitative PCR (qPCR) using a Rotor-Gene 6000 (Qiagen) and the KAPA SYBR FAST qPCR Kit (Kapa biosystems) was used to determine the amount of immobilised 6H-PfTBP needed in a 20µl binding reaction (20mM Hepes pH 8.4, 5mM $MgCl_2$ 1mM DTT ,60mM KCl 100ng/µl BSA; incubated at 21˚C) to bind 0.1% of 10ng input DNA in the presence of 400ng poly(dG-dC) competitor DNA. 100ng 6H-PfTBP was immobilised and resuspended in BC-100 complete (Section 2.7.2). Dilutions were then made to set up for binding reactions containing 0ng, 5ng, 10ng, and 15ng 6H-PfTBP. Bound oligonucleotides were separated from the reaction mixture by magnetic pulldown of 6H-PfTBP, washed twice with 1000µl aliquots of SELEX wash buffer (1xBC-0, 10mM HEPES pH 8.0, 60mM KCl, 5mM $MgCl_2$, 5mM DTT, 100ng/µl BSA) and eluted with 200µl SELEX elution buffer (500mM NaCl,1x BC-0, 10mM HEPES pH 8.0, 60mM KCl, 5mM $MgCl_2$, 5mM DTT, 100ng/µl BSA). Oligonucleotides were purified by phenol-chloroform extraction and ethanol precipitation and quantified relative to a standard curve of n16 library dilutions.

Chapter 3

Results

3.1 Bioinformatics analysis of the *Plasmodium falciparum* TATA box binding Protein (PfTBP).

A bioinformatics analysis of the PfTBP amino acid sequence from the parasite genome database (Gene entry PF3D7_0506200; PlasmoDB Release 29) (Aurrecoechea *et al.*, 2008) was conducted in order to consolidate current data with the original gene characterisation and prediction of the *P. falciparum* TBP ortholog (McAndrew *et al.*, 1993), and with structural data from the human and yeast crystallographic models of TBP-DNA and TBP-TFIIA-DNA interactions (Tan *et al*., 1996; Bleichenbacher *et al*., 2003). This was done to examine the degree of conservation of secondary structure elements as well as of residues shown to be important for TBP interaction with DNA and TFIIA. The amino acid sequences of the highly conserved C-terminal core domains responsible for DNA-binding (reviewed in Thomas and Chiang, 2006) from the human and yeast (*S. cerevisiae*) orthologs were aligned with the PfTBP core domain using the EBI Clustal Omega multiple sequence alignment tool and default alignment parameters (Fig. 9). The PfTBP ortholog was originally predicted to be a 228 amino acid (26kDa) protein which, as with orthologs from other species, is comprised of a highly conserved C-terminal domain preceded by a variable unstructured N-terminal sequence of 45 amino acids (McAndrew *et al*., 1993). Current gene annotations based on the whole genome sequencing data (Gardner *et al.*, 2002; Aurrecoechea *et al*., 2008) have extended this N-terminal region, predicting a wild-type ortholog with a much longer N-terminal domain, comprised of 327 amino

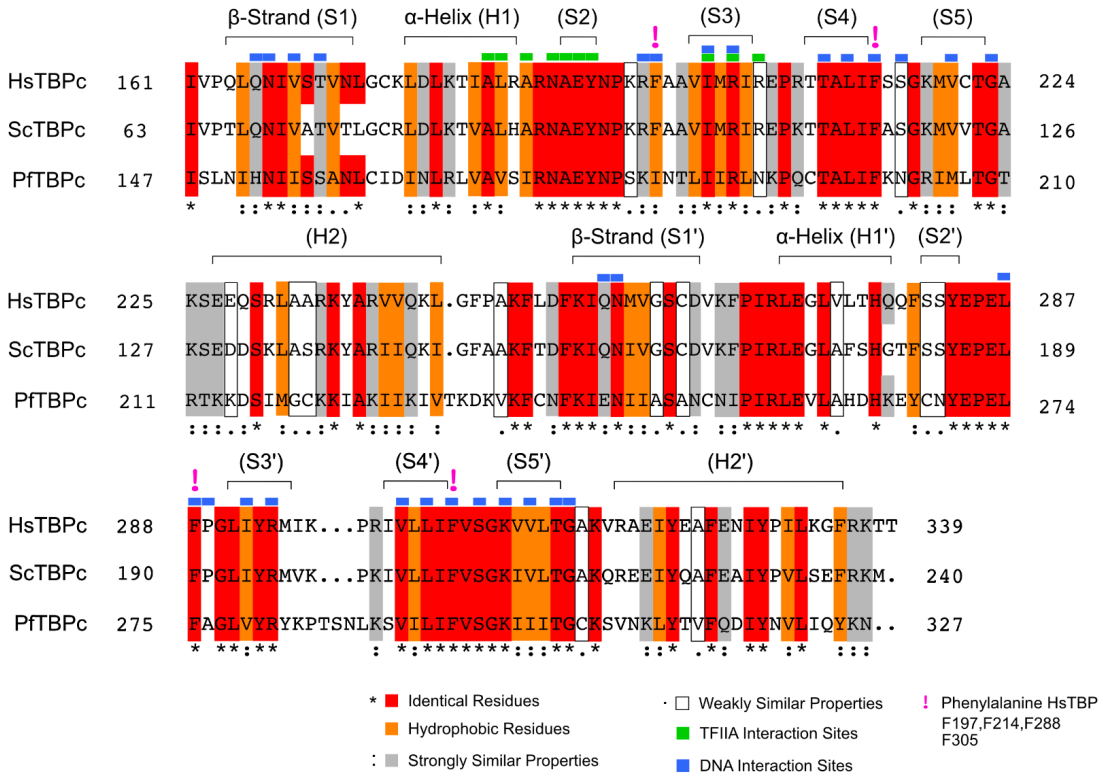


Figure 9: Sequence alignment of human, yeast (*S. cerevisiae*), and *P. falciparum* TBP-core domain orthologs. Amino acid sequences from *P. falciparum* (Pf), *S. cerevisiae* (Sc), and human (Hs) TBP orthologs were aligned using the Clustal Omega sequence alignment tool (Sievers *et al.*, 2011). Positions with sequence identity are highlighted red and marked with an asterisk (*) in order to signify regions of high sequence conservation. Conserved hydrophobic positions are highlighted orange to illustrate the conservation of hydropathy patterns previously determined (McAndrew *et al.*, 1993). Non-hydrophobic residues considered strongly similar (denoted ":") or weakly similar (denoted ".") properties are shown in grey (strongly similar) or boxed (weakly similar). The relative positions of secondary structures determined from the human TFIIA-TBP-DNA crystallographic data are indicated above the alignment (Bleichenbacher *et al*, 2003). Residues shown to be important for DNA and/or TFIIA interaction are indicated by blue (DNA) and green (TFIIA) box markers (Tan., 1996 *et al*; Bleichenbacher *et al.*, 2003).

acids (38.1 kDa). The exact function of the non-conserved N-terminal region is not fully understood. Analyses of TBP mutants lacking the N-terminal domain have shown that this region is dispensable for TATA box binding and nucleation of PIC assembly.

Despite a very low overall level of primary sequence identity between the PfTBP core domain and the human TBP core domain ortholog (39%) relative to the levels of sequence conservation seen among other eukaryotes, for example between human TBP and TBP of S. cerevisiae (79%), there are still stretches of identical residues interspersed with residues determined to have strongly similar properties (scoring > 0.5 in the Gonnet PAM 250 matrix) which generally fall within the regions associated with secondary structure (Fig. 9). 27 of 29 (93%) residues shown to make contact with the minor groove of DNA in the yeast TFIIA-TBP-DNA crystal structure (Tan *et al*., 1996) through a combination of hydrogen bonding and hydrophobic interactions are either conserved or substituted with a highly similar amino acid residue. The two substitutions are serine S216 (HsTBP) to asparagine N202 (PfTBP), both of which have polar uncharged side chains and thus scored as weakly similar, and proline P289 (HsTBP) to alanine A276 (PfTBP), scored divergent. The most notable substitution is that of phenylalanine F197 in HsTBP corresponding to ScTBP F99, which is replaced by isoleucine I183 in PfTBP and scored as highly similar. Although both residues are hydrophobic and scored as highly similar, the aromatic nature of the F197 R-group is considered to be critical for TBP-DNA interaction as it is one of the phenylalanine residues intercalating between base pairs to stabilise the bent DNA helix structure associated with TBP binding to the TATA box element (Nikolov *et al*., 1996). Indeed, mutation of conserved phenylalanine residues has been shown to severely compromise TBP-DNA binding activity *in vitro* and *in vivo* (Klejman *et al*., 2005). Based on these data it seems reasonable to predict that substitution of phenylalanine F197 (HsTBP) with isoleucine I183 (PfTBP) would minimally specifically affect DNA

bending by PfTBP and hence the trajectory of DNA in the PfTBP-DNA complex and would highly diminish overall PfTBP DNA-binding activity (Nikolov *et al*., 1996). Another potentially significant point regarding PfTBP-DNA interaction is that the spatial positioning of some residues known to interact with DNA has been altered by an insertion of 3 amino acids (TSN, positions 285-297 PfTBP). In yeast the motifs RXXXXXXTALIF and RXXXXXXVLLIF spanning the S3 to S4 and S3' to S4' regions have been shown by deletion studies to be essential for TBP-DNA interaction (Reddy & Hahn, 1991). In *Plasmodium falciparum* the second motif is altered by the addition of the three amino acids between R281 (PfTBP) and V291 (PfTBP), possibly leading to an alteration in the DNA sequence specificity of the PfTBP ortholog (McAndrew *et al.*, 1993). Eight out of ten (80%) of the residues shown to interact with TFIIA when bound to promoter DNA in the human TFIIA-TBP-DNA crystal structure (Bleichenbacher *et al*, 2003) have been conserved in the PfTBP ortholog. The two exceptions being substitutions of alanine A186 (HsTBP) to isoleucine I173 (PfTBP) (scored divergent) and arginine R204 (HsTBP) to asparagine N191 (PfTBP) (scored weakly similar). Lastly, a conserved proline, P163 in HsTBP and P65 in ScTBP, which has been shown to be essential for RNA Pol II transcription in yeast (Schultz *et al.*, 1992; MacAndrew *et al.*, 1993) has been substituted to leucine L149 (PfTBP) (scored divergent). The exact function of this residue is yet to be determined.

In light of these observations it is likely that although PfTBP may have conserved function based on its secondary structure homology and TFIIA interaction sites, its interaction with DNA may well be altered with regard to overall DNA-binding activity, sequence specificity, and the structure/trajectory of DNA in the PfTBP-DNA nucleoprotein complex.

3.2 Bioinformatic analysis of putative *Plasmodium falciparum* transcription factor IIA orthologs.

Hydrophobic cluster analysis (HCA) and profile-based homology search methods such as PSI-Blast were used to identify *P. falciparum* general transcription factors known to be associated with RNAP-II transcription in metazoans (Callebaut *et al.*, 2005). Among the hypothetical GTFs found in the *P. falciparum* genome, were candidates for the TFIIA αβ (Gene entry PF3D7_0716400; PlasmoDB Release 29) and TFIIA γ subunits, which associate to form the quaternary structure of transcription factor IIA (TFIIA). Interestingly, two potential candidates for the γ subunit were identified, termed here PfTFIIA-γ1s (Gene entry PF3D7_1250700; PlasmoDB Release 29) and PfTFIIA-γ2s (Gene entry PF3D7_0933700; PlasmoDB Release 29) (Aurrecoechea *et al.*, 2008). This observation gives rise to the possibility that *P. falciparum* possesses two distinct TFIIA complexes containing different γ subunits and with possibly distinct function. This may be a feature unique to *P. falciparum* and may therefore provide opportunity for the elucidation of unique *P. falciparum* RNAP-II transcription initiation mechanistic insights.

In order to consolidate data from the original homology predictions with structural data from the human and yeast (*S. cerevisiae*) crystallographic models of TBP-TFIIA-DNA interactions (Tan *et al*., 1996; Bleichenbacher *et al*., 2003), the EBI Clustal Omega multiple sequence alignment tool (default alignment parameters) was used to perform multiple sequence alignments between the hypothetical *Plasmodium* proteins and their respective human and yeast orthologs for those regions predicted by HCA to contain conserved secondary structure elements (Callebaut *et al*., 2005) (Figs. 10B, 11A).

3.2.1 Alignment of PfTFIIA-αβ subunits

The schematic presented in Fig. 10A shows the overall structure of the TFIIA-αβ subunit. TFIIA-αβ is composed of two structural domains, an N-terminal α-helical domain and a C-terminal domain made up of β-strands. One significant difference between TFIIA-αβ orthologs is the length of unstructured species-variable sequences bridging these two regions (Fig. 10A) (Callebaut *et al.*, 2005). It is important to note that the TFIIA-αβ subunit of *P. falciparum* does not contain a Taspase1 cleavage site, and is therefore likely to exist in an uncleaved form as is seen in yeast (*S. cerevisiae*). Multiple sequence alignments of the two structural domains show a significant lack of conservation of primary sequence identity between *P. falciparum* (Pf), human (Hs), and yeast (Sc) TFIIA-αβ (Fig. 10B). Indeed, the predicted PfTFIIA-αβ subunit is either matched with a high E value when using the human ortholog (E = 0.088), or not matched at all when using the full-length yeast TOA1 sequence as a BLASTp query sequence in the PlasmoDB database. The *Plasmodium* ortholog was originally identified by PSI-BLAST results showing significant C-terminal similarities with the TOA1 sequence (residues 214 to 286) after the second iteration ($E = 3 \times 10^{-4}$). HCA analysis then revealed the presence of the N-terminal helical domain which could be aligned with the yeast and human orthologs (Callebaut *et al.*, 2005). Despite the lack of primary sequence conservation in both the N- and C-terminal regions, patterns of conservation are discernible when examining the hydrophobic residue positions (Fig. 10B). The DNA interaction site (residues 343-346 HRSK) located in the HsTFIIA-αβ β-strand 2 (S2) region (Fig. 10B, blue box marker), which has been shown to interact with the DNA backbone both within and upstream of the TATA box major groove in the human TFIIA-TBP-DNA structure has been substantially altered in the PfTFIIA-αβ ortholog. The S2 histidine H343 (HsTFIIA-αβ) is deleted and the S2 arginine R344 (HsTFIIA-αβ) is substituted for isoleucine I134 (PfTFIIA-αβ) (scored divergent).

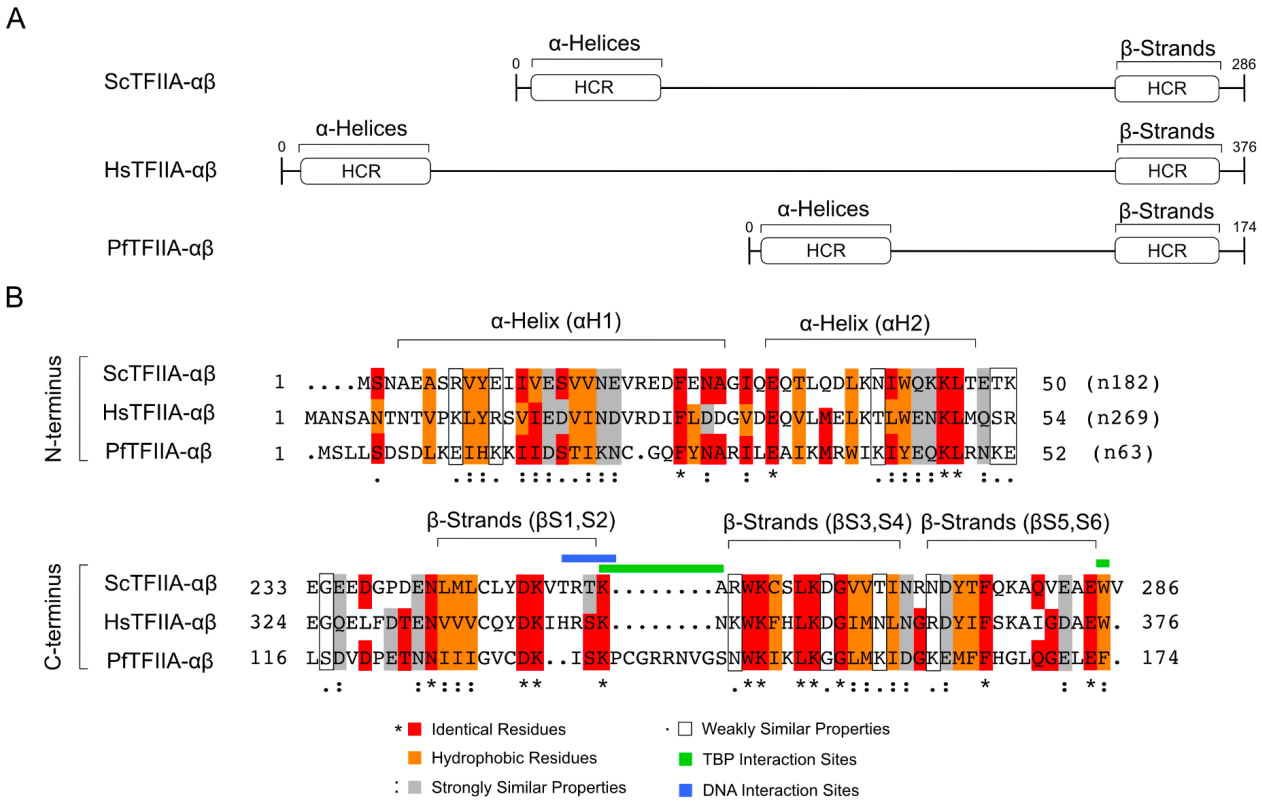


Figure 10: Sequence alignment of human, yeast (*S. cerevisiae*), and *P. falciparum* TFIIA-αβ subunit orthologs. (A) Schematic representations of TFIIA-αβ subunits from *P. falciparum* (Pf), *S. cerevisiae* (Sc), and human (Hs) TFIIA orthologs showing the relative spatial orientation of known structural regions together with predicted hydrophobic cluster regions (HCR) (Bleichenbacher *et al.*, 2003; Callebaut *et al.*, 2005). HCRs are regions of secondary structure prediction based on patterns of conserved hydrophobic residues (Gaboriaud *et al.*, 1987). (B) Amino acid sequences were aligned using the Clustal Omega sequence alignment tool (Sievers *et al.*, 2011). Positions with sequence identity are marked with an asterisk (*) and highlighted red to signify regions of high sequence conservation. Conserved hydrophobic positions highlighted in orange illustrate the conservation of hydropathy patterns previously determined (Callebaut *et al.*, 2005). Non-hydrophobic residues, which have either strongly similar (denoted ":") or weakly similar (denoted ".") properties are shown in grey (strongly similar) or boxed (weakly similar). The relative positions of secondary structures determined from the human TFIIA-TBP-DNA crystal structure are indicated above the alignments (Bleichenbacher *et al.*, 2003). Residues shown to be important for DNA and/or TBP interaction as determined from structural data are indicated by blue (DNA) and green (TBP) box markers.

Furthermore, the TBP interaction sites in HsTFIIA-αβ have also not been conserved in PfTFIIA-αβ (Fig. 10B, green box markers). The loop between S2 and S3 that docks alongside tryptophan W376 in HsTFIIA-αβ near the N-terminal stirrup region of HsTBP forming a hydrophobic pocket is extended by an insertion of 8 amino acids directly after the conserved lysine K136 in PfTFIIA-αβ. Tryptophan W376 in HsTFIIA-αβ, which is inserted within this hydrophobic region, has also been substituted to phenylalanine F174 in PfTFIIA-αβ (scored strongly similar). In light of the human TFIIA-TBP-DNA structural data these changes to the DNA interaction sites suggest that the ability of PfTFIIA to stabilise binding of TBP to the TATA box is likely altered. Deletion of H343 in HsTFIIA-αβ results in the loss of a hydrogen bond to the phosphate backbone of the DNA sense-strand upstream of the TATA box element. In addition, substitution of R344 in HsTFIIA-αβ for isoleucine I134 in PfTFIIA-αβ results in the loss of another two hydrogen bonds located within the TATA box element on the anti-sense strand diagonally across from the H343 in HsTFIIA-αβ (Bleichenbacher *et al*., 2003). TFIIA stabilises the TBP-DNA interaction by altering the conformation of the nucleoprotein complex, demonstrated by a TFIIA-mediated reduction in the bending angle of TBP-bound DNA (Hieb *et al*., 2007). It is easy to imagine that the loss of direct TFIIA-DNA contacts at several positions spanning the width of the TBP-DNA interaction surface may prevent this conformational re-arrangement. With respect to TBP-TFIIA interaction, some of the aforementioned changes seem unlikely to abolish potential co-operation between the two transcription factors. Extension of the loop between β-strands S2 and S3 does not necessarily preclude this region from functioning in its original capacity as a shield for the hydrophobic pocket located near the TBP N-terminal stirrup. Similarly, the effect of tryptophan W376 in HsTFIIA-αβ being substituted for phenylalanine F174 in PfTFIIA-αβ is possibly small as both residues are hydrophobic in nature and have aromatic side chains. This substitution does however result in a loss of hydrogen bonding between tryptophan W376 in HsTFIIA-αβ and asparagine N189 in

HsTBP, which may be significant as N189 forms part of a highly conserved TBP region in both the human and *P. falciparum* orthologs (section 3.1). In summary, the TFIIA-αβ subunit of metazoan TFIIA is the primary site of interaction between TFIIA and DNA in the TFIIA-TBP-DNA nucleoprotein complex. Although secondary structure elements between human, yeast, and *P. falciparum* PfTFIIA-αβ orthologs appear to be conserved there are substantial alterations to the region involved in DNA interactions. These alterations likely lead to a loss in the ability of PfTFIIA to stabilise interactions between TBP and the TATA box promoter element. In addition, there are clear changes in the TBP-TFIIA interaction sites.

3.2.2 Alignment of PfTFIIA-γ subunits

The schematic presented in Fig. 11B shows the overall structure of the human TFIIA-γ subunit (HsTFIIA-γ), which is made up of a α-helical domain followed directly by a β-sheet domain (Fig. 11B). The yeast TFIIA-γ (ScTFIIA-γ) and *P. falciparum* (PfTFIIA-γ1s) orthologs differ from this in that they both contain insertions of an unstructured amino acid region that is located within their respective β-sheet domains (Fig. 11B) (Callebaut *et al.*, 2005; Bleichenbacher *et al.*, 2003). An unstructured region is also present in the second *P. falciparum* ortholog (PfTFIIA-γ2s). However, this region precedes the N-terminal α-helical domain (Fig. 11B). Multiple sequence alignments of the human and yeast orthologs with the structural domain regions of the PfTFIIA-γ1s and PfTFIIA-γ2s subunits predicted by HCA (Callebaut *et al.*, 2005) show a substantial divergence in the primary amino acid sequences of these orthologs relative to the human and yeast variants as well as in respect to one another (Fig. 11A). It should be noted that minor adjustments to the Clustal Omega alignment had to be made at the extreme C-terminus corresponding to amino acid

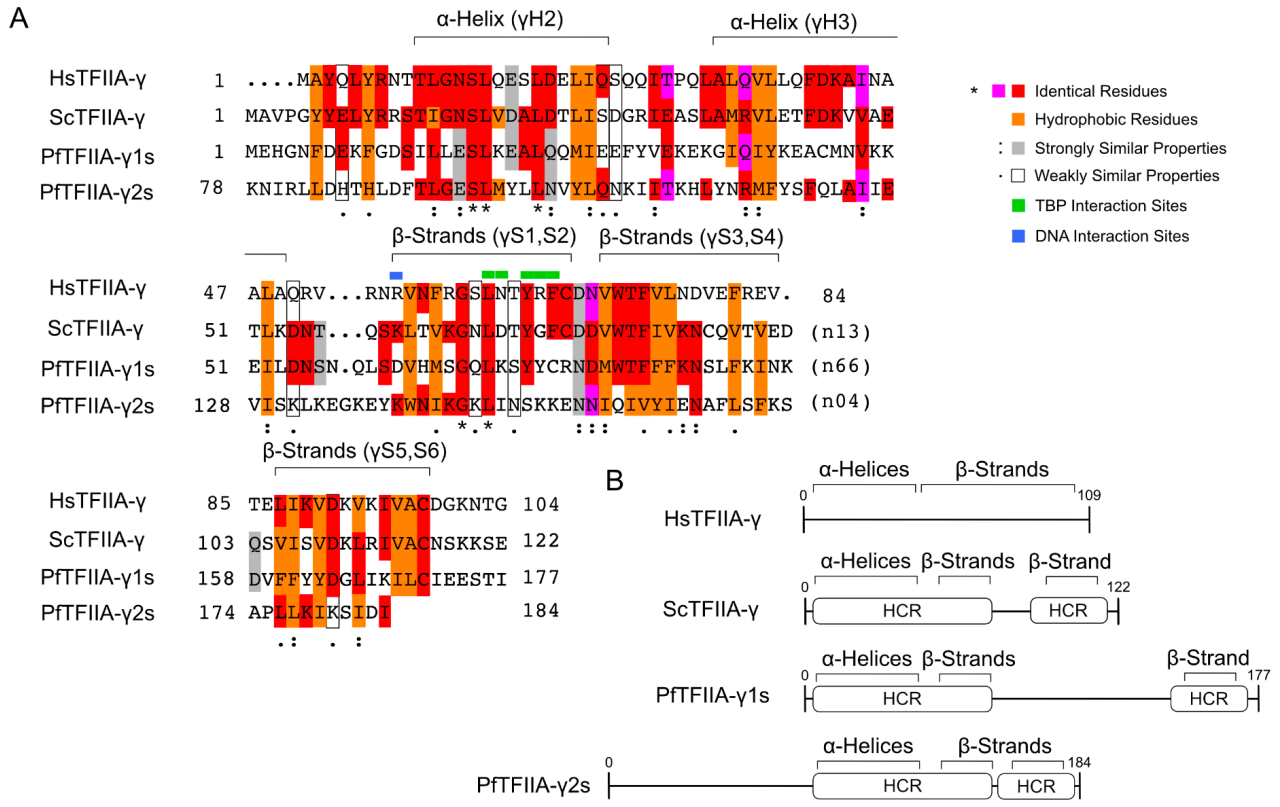


Figure 11: Sequence alignment of human, yeast (*S. cerevisiae*), and two putative *P. falciparum* TFIIA-γ subunit orthologs. (A) Amino acid sequences of the TFIIA-γ subunits from *P. falciparum* (Pf), *S. cerevisiae* (Sc), and human (Hs) TFIIA orthologs were aligned using the Clustal Omega sequence alignment tool (Sievers *et al.*, 2011). Positions with sequence identity are marked with an asterisk (*) and highlighted red, or magenta, in order to signify regions of high sequence conservation. Conserved hydrophobic positions are highlighted orange to illustrate the conservation of hydropathy patterns previously determined (Callebaut *et al.*, 2005). Non-hydrophobic residues which have either strongly similar (denoted ":") or weakly similar (denoted ".") properties are shown in grey (strongly similar) or boxed (weakly similar). The relative positions of secondary structures determined from the human TFIIA-TBP-DNA crystal structure (Bleichenbacher *et al*, 2003) are indicated above the alignments. Residues shown to be important for DNA and/or TBP interaction as determined from structural data are indicated by blue (DNA) and green (TBP) box markers. (B) Schematic representations of TFIIA-γ subunit orthologs showing the relative spatial orientation of known structural regions together with predicted hydrophobic cluster regions (HCR) (Bleichenbacher *et al.*, 2003; Callebaut *et al.*, 2005). HCRs are regions of secondary structure prediction based on patterns of conserved hydrophobic residues (Gaboriaud *et al.*, 1987)

positions 85-104 of HsTFIIA-γ in order to better align the subunits according to the HCA plots previously described (Callebaut *et al.*, 2005). As with the PfTFIIAαβ subunit there is some conservation of hydrophobic positions but overall, the sequences lack large regions of conserved homology (Fig. 11A).

The DNA interaction site R55 located in HsTFIIA-γ (Fig. 11A, blue box marker), which provides electrostatic interaction with the DNA in the crystallographic data (Bleichenbacher *et al.*, 2003), is somewhat conserved both in the ScTFIIA-γ and PfTFIIA-γ2s orthologs, where this residue has been substituted for another positively charged polar amino acid, lysine K139 in PfTFIIA-γ2s, lysine K59 in ScTFIIA-γ. PfTFIIA-γ1s however shows a more radical substitution with arginine R55 in HsTFIIA-γ1 being replaced with the negatively charged aspartic acid D61. The main TBP-TFIIA interaction sites in the human TFIIA-TBP-DNA complex (Bleichenbacher *et al*, 2003) are parallel β-strand interactions made up of a H-bond network between Y65 and F67 in HsTFIIA-γ, and in the N189, A190, and Y192 residues within the HsTBP S2 β-sheet domain, extended further by an indirect interaction between HsTBP N189 and HsTFIIA-γ N63 via a water molecule. A second H-bond network is formed between HsTFIIA-γ R66 and the A184, L185, and A187 residues in the the C-terminus of the HsTBP H1 α-helical domain. Finally the R205 side chain in HsTBP forms direct contacts with HsTFIIA-γ L62 and HsTFIIA-γ N63 via hydrogen bonding to the main-chain carbonyl groups (Bleichenbacher *et al.*, 2003). With the exception of the lysine residue L62 in HsTFIIA-γ, there is no conservation of TBP interaction sites found in PfTFIIA-γ1s or PfTFIIA-γ2s. N63 of HsTFIIA-γ is substituted with K69 in PfTFIIA-γ1s, and I147 in PfTFIIA-γ2s. Y65 of HsTFIIA-γ is conserved in PfTFIIA-γ1s, but has been substituted for cysteine C149 in PfTFIIA-γ2s (scored divergent). Arginine R66 of HsTFIIA-γ is also not conserved, being substituted for Y72 in PfTFIIA-γ1s and K150 in PfTFIIA-γ2s. Lastly, phenylalanine F67 in HsTFIIA-γ is substituted for cysteine C73 in PfTFIIA-γ1s and

lysine K151 in the PfTFIIA-γ2s ortholog. It is important to note that two of these TBP interaction sites, N63 and R66 in HsTFIIA-γ, are also not conserved in the yeast ScTFIIA-γ, and therefore, the differences seen at these positions in the PfTFIIA orthologs may not be of much significance to TBP-TFIIA interaction in *P. falciparum*.

In summary, the bioinformatics analysis of the two putative PfTFIIA-γ subunits provides good evidence, based on secondary structure prediction, that PfTFIIA-γ1s and PfTFIIA-γ2s are indeed legitimate PfTFIIA orthologs. However, there is an extreme divergence in primary sequence identity with respect to one another as well as to the human ortholog (Fig. 11A). In addition, there are alterations to the spatial arrangement of the proposed secondary structure features in the PfTFIIA-γ1s and -γ2s orthologs relative to the known arrangement seen in the crystal structure of HsTFIIA (Fig. 11B) (Bleichenbacher *et al.*, 2003). As there is also almost no conservation of the main TBP interaction sites present in the human ortholog, located primarily on the TFIIA-γ subunit, one would expect the two different PfTFIIA complexes to posses biochemical characteristics that are very different from the human TFIIA and from each other (Bleichenbacher *et al.*, 2003).

3.3 Bioinformatic analysis of the *Plasmodium falciparum* TBP-like protein (PfTLP)

The *P. falciparum* TBP-Like protein (PfTLP) (Gene entry PF3D7_1428800.2 ; PlasmoDB Release 29) (Aurrecoechea et al., 2008) is a putative TBP paralog predicted to be part of the set of general transcription factors associated with the parasite's RNAP-II mediated transcription (Coulson *et al*., 2004; Bischoff *et al*., 2010). Vertebrates have evolved a range of TBP paralogs through gene duplication events leading to an expansion of the available TATA-binding factors able to promote basal transcription initiation (reviewed in Thomas & Chiang, 2006; Akhtar & Veenstra, 2011). This is thought to have given rise to novel control

mechanisms through the preferential interaction of these paralogs with distinct core promoter sub-types as evidenced by TBP-related factor (TRF2) mediated control of TCT- and DPE- dependent TATA-less promoters (Wang *et al*., 2014; Kedmi *et al*., 2014; Duttke *et al*., 2016). Three distinct TBP family proteins have thus far been identified: TBP-related factor (TRF1), discovered in *Drosophila* and later found to be specific to insect cells, TBP-like factor (TRF2 or TLP), found in all metazoans, and finally the most recent vertebrate-specific paralog, TBP2 (or TRF3) (Crowley *et al*., 1993; Dantonel *et al*., 1999; Bártfai et al., 2004; reviewed in Akhtar & Veenstra, 2011). Multiple sequence alignment and phylogenetic analysis showed that of the three TBP paralogs, PfTLP is evolutionarily most closely related to TRF2 (TLP) (Bing, 2015). Secondary-structure modelling revealed that PfTLP is likely to have a structure very similar to the that of the conserved C-terminal core domain typically associated with TBP family proteins (Bing, 2015). Figure 12 shows an adaption of a COBALT sequence alignment by Bing, 2015 of HsTBP and selected TRF orthologs. It has been included as a point of reference for later discussions regarding data obtained from functional assays. For purposes of comparison to the Clustal Omega alignments of PfTBP and PfTFIIA, residue similarities in the PfTLP alignment are indicated according to the Gonnet PAM 250 matrix parameters (Ch2. Materials & Methods, Section 2.1.1). As with preceding alignments, Fig. 12 has been annotated according to structural data for the human and yeast TBP-DNA (Nikolov *et al*., 1996; Tan *et al*., 1996) and TBP-TFIIA-DNA (Bleichenbacher *et al*., 2003) complexes. PfTLP shows a substantial divergence in primary sequence identity relative to HsTBP with the most notable difference being the insertion of large unstructured regions in the S4 and S4' β-strand regions (131 and 31 amino acid residues respectively), as well as a very short N-terminal region (Fig. 12). Examination of the DNA-interaction sites show that only 16 of 29 (55%) residues is HsTBP known to interact with the DNA minor groove (Tan *et al*., 1996; Bleichenbacher *et al*., 2003) have been either conserved or substituted with highly similar (scoring > 0.5 in

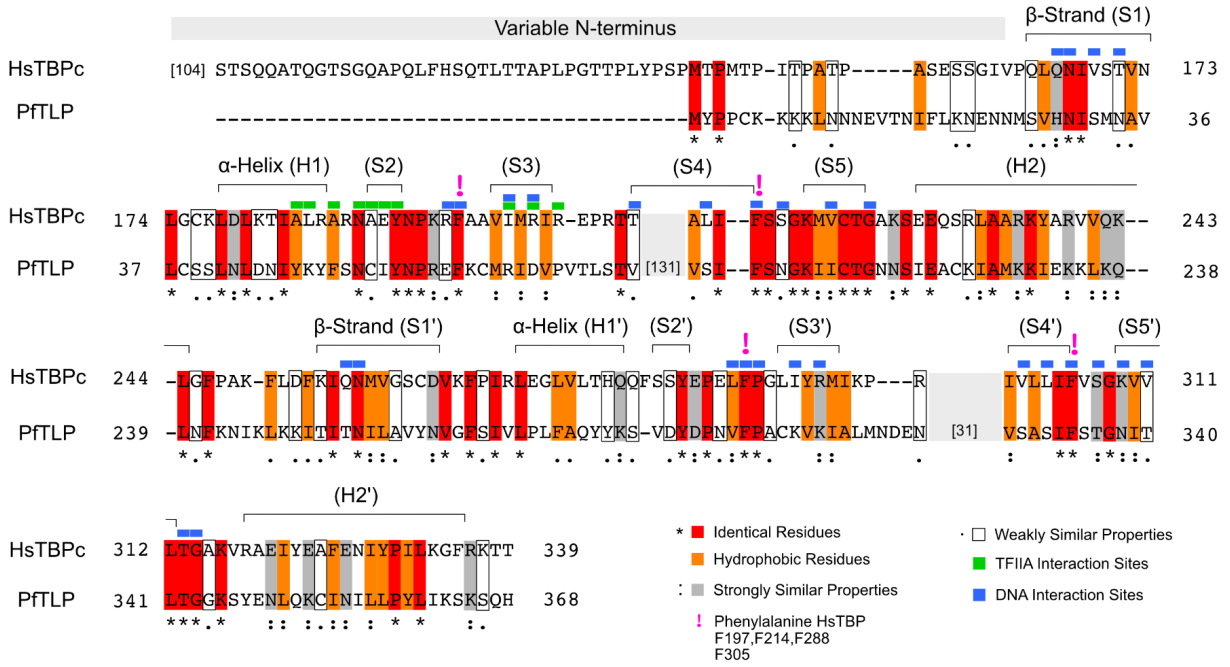


Figure 12: Sequence alignment of human TBP-core domain and *P. falciparum* TBP-Like protein (PfTLP). Figure adapted from a COBALT multiple sequence alignment by Bing, 2015. Residue similarities were annotated manually according to the Gonnet PAM 250 matrix parameters (Ch2. Materials & Methods, Section 2.1.1). Positions with 100% sequence identity are marked with an (*) and highlighted red in order to signify regions of high sequence conservation. Conserved hydrophobic positions are highlighted orange to illustrate conservation of hydropathy. Non-hydrophobic residues which have either strongly similar (denoted ":"), or weakly similar (denoted ".") properties are shown coloured grey (strongly similar) or boxed (weakly similar). The relative positions of secondary structures predicted from TBP structural data are indicated above the alignment (Nikolov *et al.*, 1992; Bleichenbacher *et al*., 2003). Residues predicted to be important for DNA and/or TFIIA interaction from structural analysis of the TBP-TFIIA-DNA complex are indicated by blue (DNA) and green (TFIIA) box markers (Tan *et al*., 1996; Bleichenbacher *et al*., 2003).

the Gonnet PAM 250 matrix) amino acid residues (Fig. 12). In addition, it should also be noted that the two unstructured loop regions (131 and 31 amino acid residues) are inserted into the RXXXXXXTALIF and RXXXXXXVLLIF motifs, which have been shown in yeast TBP to contain residues vital for DNA-binding activity (Reddy & Hahn, 1991; discussed in Section 3.1). Surprisingly, all four of the highly conserved phenylalanine residues F197, F214, F288, and F305 in HsTBP, essential for stable TBP-DNA interaction at the TATA box (Nikolov *et al*., 1996; Klejman *et al*., 2005), are conserved in PfTLP. This is a characteristic that is unique among TRF2-type TBP paralogs, which in other eukaryotes lack one or more of these residues. The lack of conserved phenylalanine residues in TRF2-type TBP paralogs is thought to be the reason for their lack of TATA-binding activity (reviewed in Thomas & Chiang, 2005; Duttke *et al*., 2016). Similarly, only 2 of 10 (20%) residues associated with TFIIA interaction in HsTBP (Bleichenbacher *et al*., 2003) have been conserved in PfTLP, with all other residues showing substitution to either divergent or weakly similar (scoring < 0.5 in the Gonnet PAM 250 matrix) residues. In addition, the analysis of associated DNA-interaction sites (Bleichenbacher *et al*., 2003) seem to indicate that, despite the relatively low conservation (55%) of important DNA-interaction sites compared to PfTBP (97%), key residues typically associated with TATA-binding activity are conserved (Fig. 12) (Nikolov *et al*., 1996). It is also evident, based on the low level of TFIIA interaction site conservation, that any traditional relationship to TFIIA would likely necessitate an interaction partner that has a biochemical surface distinct from that found on human TFIIA. In summary, *P. falciparum* TLP (PfTLP) is unique compared to currently described metazoan TLP (TRF2) orthologs in that it possesses all four of the highly conserved phenylalanine residues associated with the TATA-binding activity of TBP. This is a very surprising feature in that it seems to indicate DNA-binding activity, and possibly TATA-binding activity for PfTLP, a characteristic not seen with other TLP orthologs. However, aside from this key similarity with TBP, the overall conservation of DNA

interaction sites outside of these four phenylalanine residues is relatively low. In addition, the insertions of large unstructured loop regions within the C-terminal domain of PfTLP adds precedence to the notion that PfTLP may have an altered DNA-sequence specificity With respect to TFIIA interaction, the levels of TFIIA interaction site conservation is very low, indicating the altered surface chemistry of PfTLP suggests at the very least a very different mode of interaction with TFIIA or perhaps an inability to interact with TFIIA all together.

3.4 Co-expression and purification of two putative PfTFIIA complexes

Transcription factor IIA (TFIIA) was one of the earliest general RNAP-II transcription factors (GTFs) shown by *in vitro* transcription assays to be necessary for site-specific transcription initiation by RNAP-II (Weil *et al*. 1979; Matsui *et al*., 1980; Samuels *et al*., 1982). Human TFIIA is a multi-subunit complex made up of three polypeptide chains: α (35kDa), β (19 kDa), and γ (12 kDa) which are initially transcribed from two genes, with the αβ chain being translated as a single unit and subject to post-translational cleavage by Taspase 1 (reviewed in Høiby 2007; Malecová *et al*., 2015). The TBP-TFIIA-DNA crystal structure shows a tightly bound and intertwined α-β-γ ternary complex forming contacts with both DNA and promoter-bound TBP (Bleichenbacher *et al*., 2003). Bioinformatics studies identified an ortholog for the αβ subunit and two putative orthologs for the γ subunit, raising the possibility that *P. falciparum* contains two distinct PfTFIIA complexes with different gamma subunits.

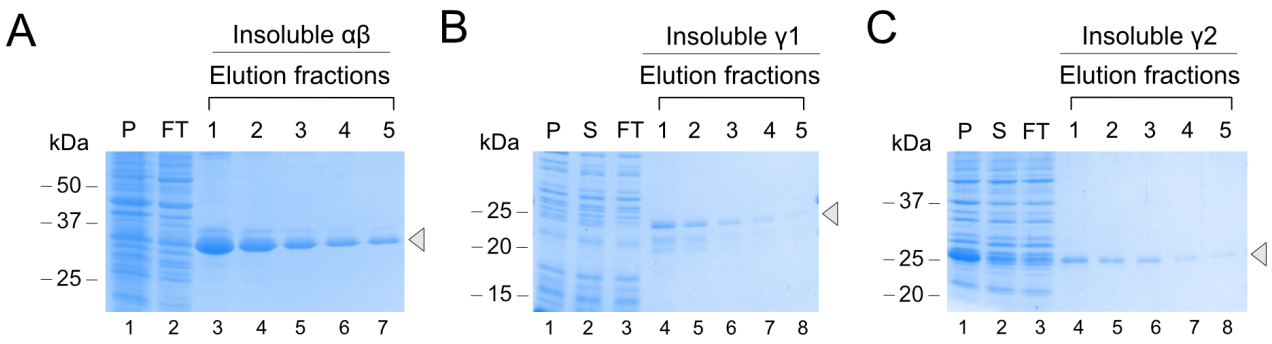


Figure 13: SDS-PAGE analysis of 6His-tagged putative PfTFIIA subunits purified from the insoluble fraction of of *E. Coli* BL21 (DE3)-RIL cell lysate by Ni-NTA affinity chromatography. Bacterial pellets were resuspended in 8M urea buffer. 10µl samples of 50µl elution fractions were loaded alongside equivalent amounts of solubilised bacterial pellet (P), column flowthrough (FT), and soluble fractions (S) for comparison. The positions of the specific 6H-tagged proteins are indicated by grey arrowheads.

3.4.1 Cloning and expression of isolated 6His-tagged PfTFIIA-αβ, PfTFIIA-γ1s, and PfTFIIA-γ2s subunits

Initial efforts focused on the expression and purification of the individual subunits in isolation with the view of generating PfTFIIA complexes through previously established techniques of controlled denaturation/renaturation, which achieved human TFIIA complex formation with individual subunits during the re-folding process (Sun *et al*.,1994; Dejong *et al*., 1995; Ozer *et al*., 1994; Malecova *et al*., 2007). Bacterial expression vectors for 6His-tagged variants of the putative PfTFIIA subunits PfTFIIA-αβ, PfTFIIA-γ1s, and PfTFIIA-γ2s were previously cloned in collaboration with Dr Thomas Oelgeschläger (Milton, 2012; unpublished data). Vectors were transformed into the codon optimised *E.coli* strain BL21-DE3-RIL and expression induced by the addition of IPTG. SDS-PAGE analysis of whole-cell lysates before and after induction followed by Ni-NTA affinity purification of cleared lysate prepared under native conditions (QIAexpressionist, 2002) failed to show enrichment of these proteins (data not shown). To investigate whether these proteins were

expressed but insoluble, bacterial pellets were resuspended in 8M urea sonication-buffer. After sonication and centrifugation, the cleared solubilised pellet was passed over a Ni-NTA column and elution fractions were analysed by SDS-PAGE (Fig. 13). The 6H-PfTFIIA-αβ subunit has a calculated molecular mass of 22kDa but migrated as a single enriched band with an apparent molecular mass of about 30kDa in the SDS-PAGE gel. This unusually low mobility of the PfTFIIA-αβ subunit in SDS-PAGE can be explained by its very low pI (4.84). The resultant negative charge is known to hamper the association of SDS molecules, thereby reducing the net negative charge per unit length of the denatured protein subunit and thus its mobility in the electric field (Fig 13A). The two γ subunit candidates PfTFIIA-γ1s (23.4 kDa) and γ2s (24.3 kDa) can be seen migrating around the 25kDa markers of Figures 13B and 13C respectively, consistent with the calculated molecular mass of the 6H-tagged variants. Low molecular mass peptides visible around the 20kDa marker in lanes 4 + 5 of figure 13B are evidence of possible proteolytic degradation, which may account for the comparatively low expression levels of the PfTFIIA-γ1s and γ2s subunits relative to the PfTFIIA-αβ subunit. Multiple variations of these initial expression conditions were employed in an attempt to alleviate the protein solubility problem. Changes included variations in expression temperature, IPTG concentration, cell density at point of induction, as well as the use of commercially available auto-induction media. Ultimately these attempts were unsuccessful and did not result in the production of appreciable levels of soluble protein (data not shown).

3.4.2 Co-expression of the PfTFIIA-γ1s and PfTFIIA-γ2s containing PfTFIIA complexes

The heterologous co-expression of protein complex subunits is a well-established method to circumvent protein aggregation of individual subunits and the subsequent solubility problems that are often associated with high levels of expression in *E.coli* cells (Tolia *et al*.,

2006). It is thought that the exposure of hydrophobic surfaces, which would otherwise be shielded by the relevant interaction partner, may lead to the formation of insoluble protein aggregates known as inclusion bodies (Kane & Hartley, 1988; Carrio & Villaverde, 2002). Simultaneous expression of subunits from one or several expression vectors within the same cell has in many cases been shown to alleviate this problem through complex formation at the point of translation. The crystal structure of the ternary TBP-IIA-DNA nucleoprotein complex (Bleichenbacher *et al*., 2003) shows that interaction between the TFIIA αβ and γ subunits leads to a very intertwined and tightly associated structure that likely explains the difficulties experienced when expressing the isolated subunits. In addition to possibly alleviating solubility problems, co-expression also provides a means to test if the γ-subunits can stably interact with PfTFIIA-αβ. This would be evidenced by co-purification of untagged putative γ subunits with an epitope-tagged PfTFIIA-αβ subunit (i.e 6H-PfTFIIA-αβ), and would consequently also provide evidence to suggest that these are indeed bona fide PfTFIIA complexes. Co-expression vectors were constructed through the insertion of untagged PfTFIIA-γ1s and PfTFIIA-γ2s cDNA expression units (T7 promoter + ORF + T7 terminator) downstream of the expression unit for 6His-tagged PfTFIIA-αβ (Ch2, Materials & Methods 2.2). Complexes were expressed after cold shock at 25˚C without the use of IPTG, relying solely on leaky expression from the LacUV5 promoter of the host strain plasmid. Figure 14 shows the SDS-PAGE analysis of the Ni-NTA column and subsequent ion-exchange chromatography steps used to purify the PfTFIIA-γ1 (Figs 14A, B) and PfTFIIA-γ2 (Figs 14C, D) complexes following co-expression. 6His-tagged complexes were purified on Ni-NTA columns under high salt (500mM KCl) and eluted with 250mM imidazole buffer. Early elution fractions (Figs 14A, C; lanes 3-5) were analysed by SDS-PAGE and clearly show the presence of the 6H-PfTFIIA-αβ subunit in roughly equivalent amounts. Analysis of the PfTFIIA-γ1s and γ2s subunits clearly shows the co-purification of the PfTFIIA-γ1s subunit (Fig 14A; lane 3), enriched relative to the input

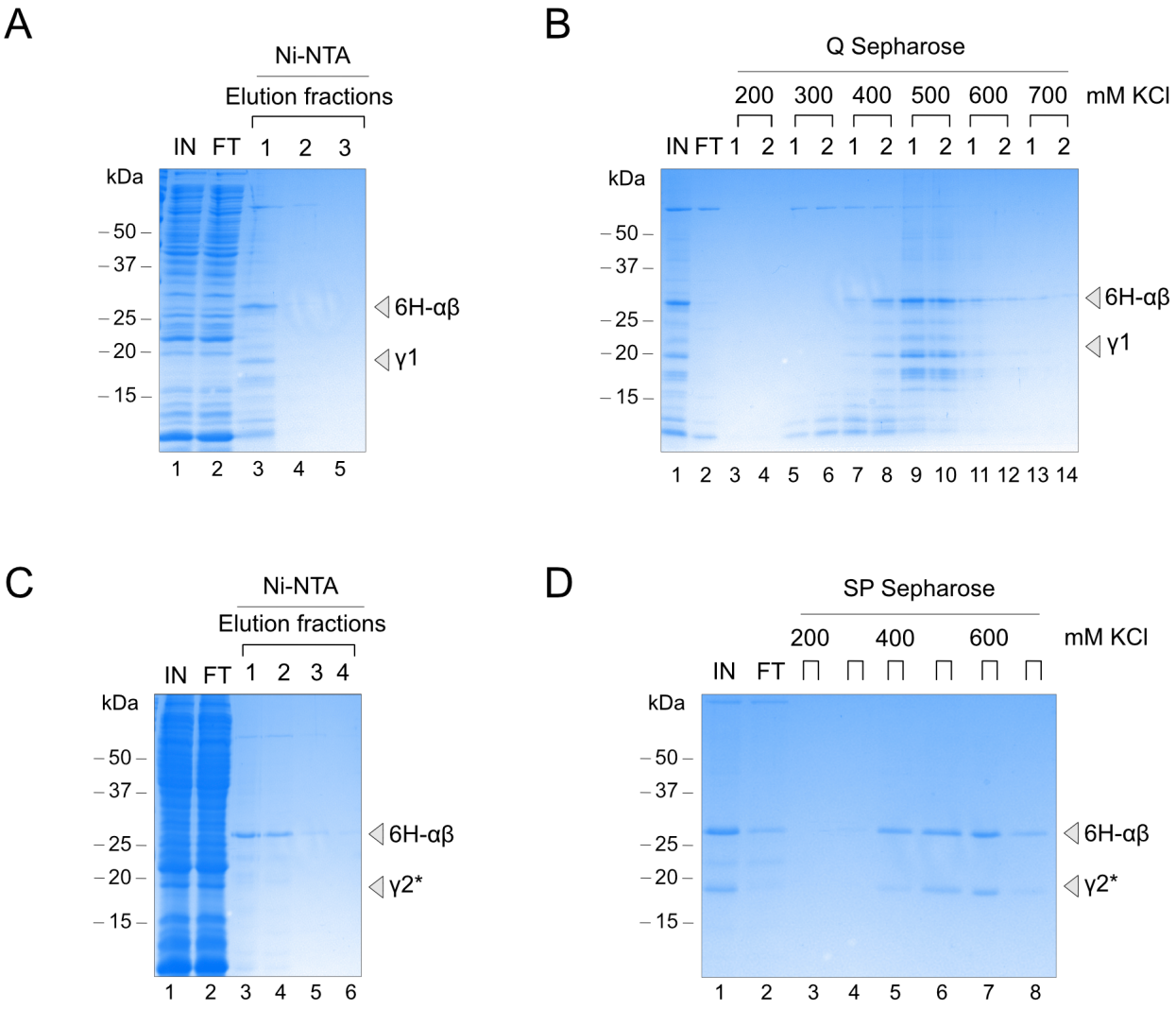


Figure 14: SDS-PAGE analysis of 6H-PfTFIIA-γ1 and 6H-PfTFIIA-γ2 complexes purified from *E. Coli* BL21 (DE3)-RIL cell lysate by Ni-NTA affinity and ion-exchange chromatography. (A,C) Whole cell lysates were cleared by centrifugation and purified by Ni-NTA columns. Proteins were eluted with 250mM imidazole buffer. 10µl samples of 500µl elution fractions were loaded for analysis alongside input (IN) and flowthrough (FT) fractions. Pooled elution fractions containing PfTFIIA complexes were subsequently passed over either Q-Sepharose (B; PfTFIIA-γ1) or SP-Sepharose (D; PfTFIIA-γ2) columns. Proteins were eluted in a step-wise fashion by the addition of 1 column volume (300µl) elution buffer containing increasing concentrations of KCl as indicated. * SDS-PAGE gels were stained with Coomassie stain; the γ2 subunit is stained poorly below a certain threshold concentration.

fraction and present in a roughly 1:1 ratio relative to the PfTFIIA-αβ subunit. In contrast, the PfTFIIA-γ2s subunit could initially not be detected (Fig 14C). The presence of contaminating proteins in the Ni-NTA purifications (Figs 14A, C) can be attributed to the low expression levels of PfTFIIA complexes. Following Ni-NTA purification PfTFIIA-containing fractions were adjusted to 100mM KCl (Ch2, Materials and Methods 2.3.6) and subjected to ion-exchange chromatography (Figs 14B, D). The combined subunits of the PfTFIIA-γ1 complex have a pI of 5.38, and was therefore further purified on a Q-sepharose anion exchange resin (Fig. 14B). Fractions were eluted by increasing the KCl concentration of the elution buffer from 200mM to 700mM in a stepwise manner. The PfTFIIA-γ1 complex was eluted in the 400-500mM KCl range (Fig 14B; lanes 8-10) with some low molecular weight (<15 kDa) contaminants being excluded in the earlier elutions (Fig 14B; lanes 5-10). The PfTFIIA-γ2 complex has a combined pI of 7.69 and was therefore purified on a SP-sepharose cation exchange resin (Fig. 14D). Interestingly, the PfTFIIA-γ2s subunit was now found to be visible after SDS-PAGE and coomassie staining. This suggests that higher concentrations of the complex were present on the gel and that the subunit stains poorly below a certain threshold amount on the gel (Fig. 14D). Overall, the elution profile of the PfTFIIA-γ2 complex was similar to that of PfTFIIA-γ1, with the complex eluting gradually over the 400mM to 700mM KCl range (Fig 14D; lanes 5-8). Figure 15 shows a summary of the purification protocols developed for the two putative PfTFIIA complexes. The fact that both untagged PfTFIIA-γ1s and PfTFIIA-γ2s subunits co-purified with 6H-tagged PfTFIIA-αβ subunit over two separate purification steps, provides strong evidence for stable association with PfTFIIA-αβ and the formation of two distinct bona fide PfTFIIA complexes. Two independent preparations of both PfTFIIA complexes were expressed and purified (Fig 16A). Western blot analysis was then carried out using PfTFIIA preparation 1 to confirm the identity of PfTFIIA-αβ, PfTFIIA-γ1s, and PfTFIIA-γ2s and to establish the appropriate antibody dilutions for detection in immobilised template

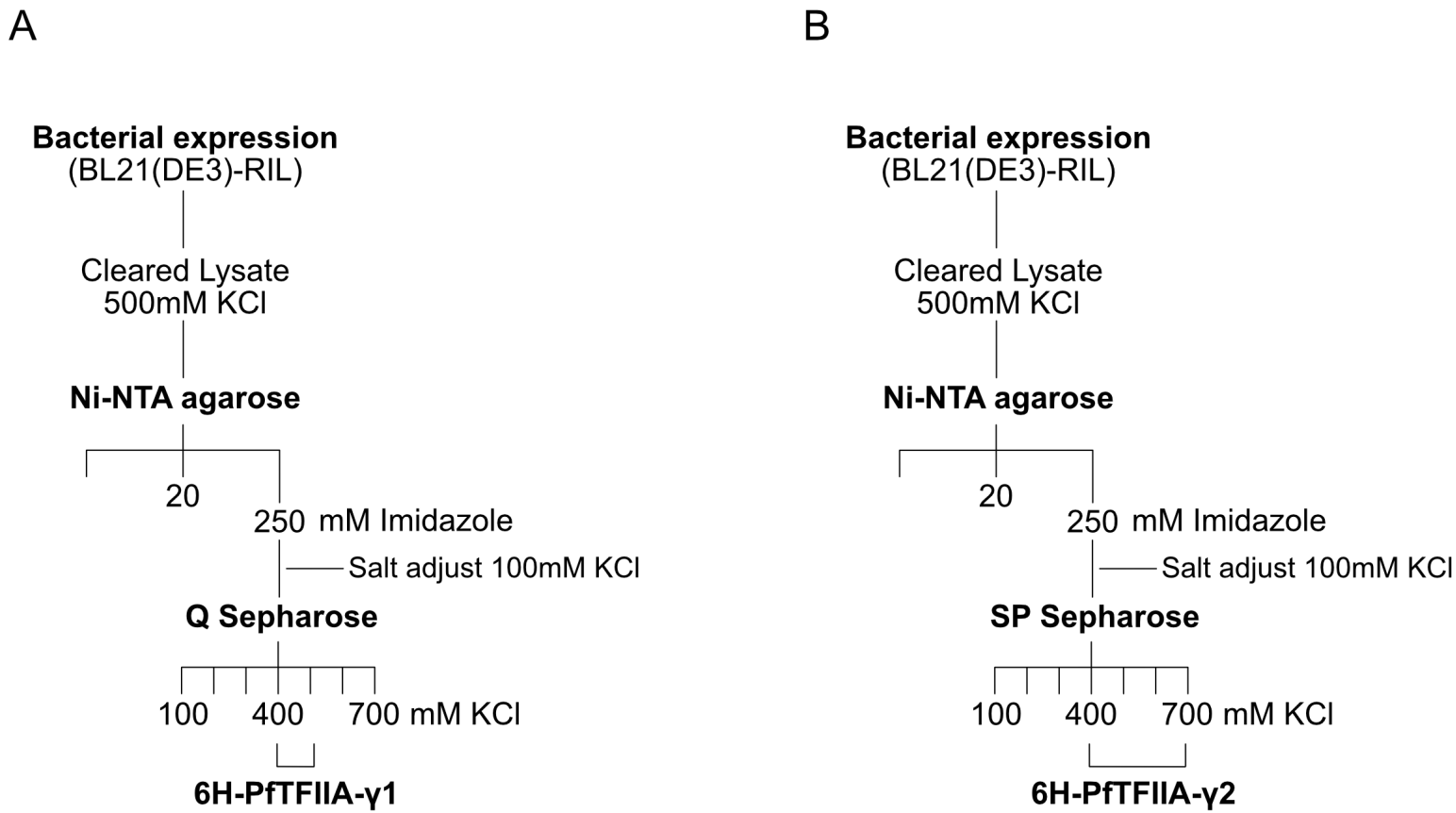


Figure 15: Purification of 6H-PfTFIIA-γ1 and 6H-PfTFIIA-γ2 complexes. (A) PfTFIIA-γ1 was purified by Ni-NTA affinity chromatography followed by ion-exchange chromatography on Q-Sepharose resin. (B) PfTFIIA-γ2 was purified by Ni-NTA affinity chromatography followed by ion-exchange chromatography on SP-Sepharose resin.

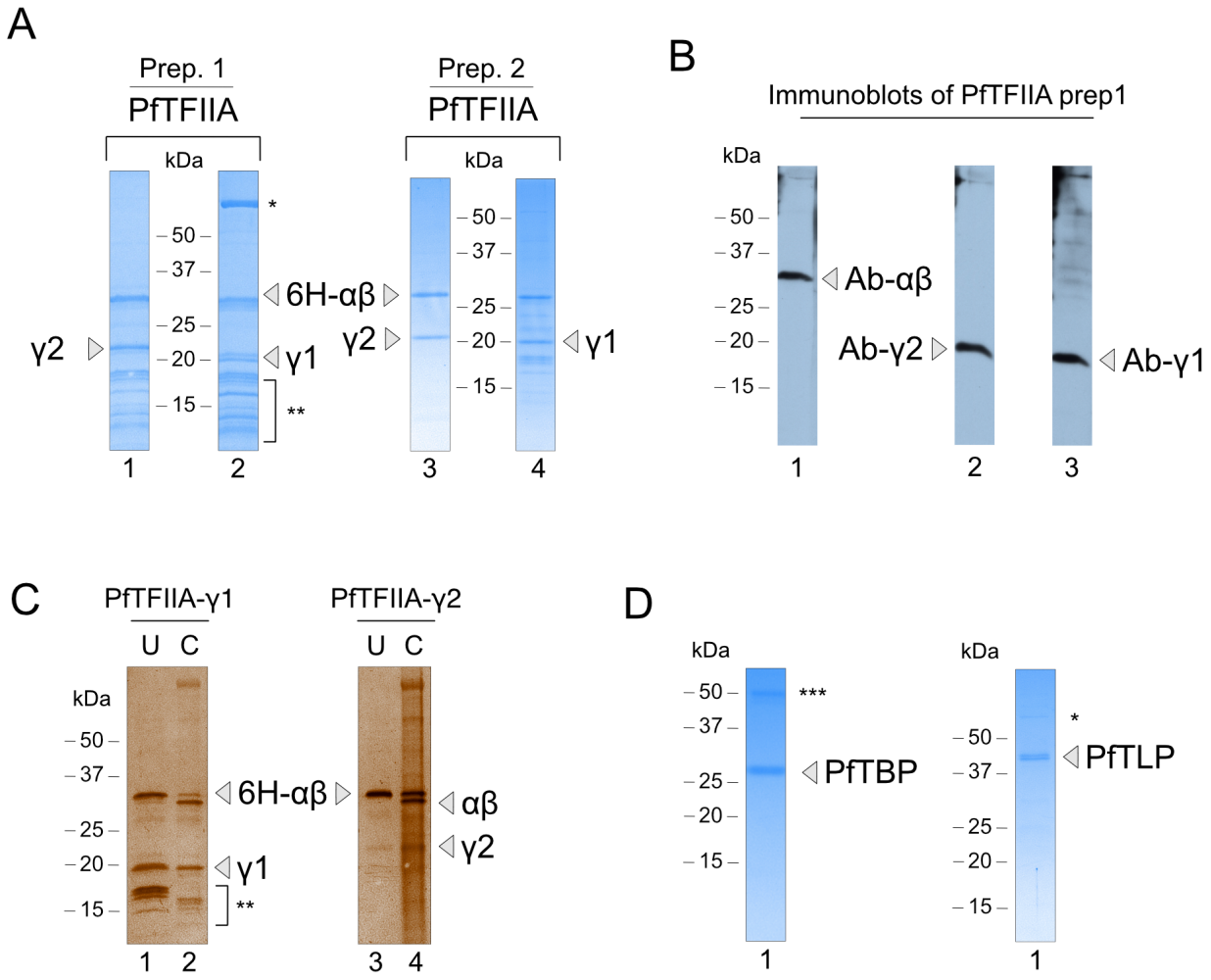


Figure 16: SDS-PAGE analysis of 6His-tagged recombinant *P. falciparum* transcription factors used for functional characterisation experiments. A: Putative PfTFIIA preparations, 6H-PfTFIIA-γ1 (lanes 2 + 4), 6H-PfTFIIA-γ2 (lanes 1 + 3), * BSA, ** low MW contaminants. B: Immunoblots of PfTFIIA prep 1: 6H-PfTFIIA-γ2 loaded in lanes 1 + 2 and probed with anti-Pf-IIA-αβ (lane 1) or anti-Pf-IIA-γ2s (lane 2). 6H-PfTFIIA-γ1 loaded in lane 3 and probed with anti-Pf-IIA-γ1s. C: Thrombin cleavage removal of the 6His-tag from the PfTFIIAαβ subunit of PfTFIIA preparation 2 showing altered mobility of cleaved PfTFIIAαβ subunit (6H-αβ) as well as low molecular weight C-terminus truncations, (u: uncleaved c: cleaved) D: 6H-PfTBP (lane 1) and 6H-PfTLP (lane 2) preparations used in protein DNA binding assays, *** high MW *E. coli* contaminant , * BSA.

assays which were later used in the initial functional characterisation of the two PfTFIIA orthologs (Fig. 16B) (Section 3.7). Thrombin cleavage experiments were also carried out to remove the 6H-tag from the PfTFIIA complexes (preparation 2) in anticipation of possibly using these proteins in conjunction with PfTBP to examine PfTBP DNA-binding specificity using a modified SELEX assay (further discussed in section 3.6). Partially successful cleavage of these tags from the complexes were indicated by changes in the mobility of the αβ subunits in both PfTFIIA-γ1 (Fig. 16C, compare lanes 1 and 2) and PfTFIIA-γ2 (Fig. 16C, compare lanes 3 and 4) complexes. No new bands became visible in the SDS-PAGE analysis of the cleaved fractions, indicating that the integrity of the two PfTFIIA complexes remain unaffected by thrombin cleavage. Interestingly, it is noted that the mobility of the low molecular weight contaminants are also altered by the thrombin digestion (Fig. 16C, lanes 2 and 4), indicating that these contaminants are truncated 6H-tagged PfTFIIA-αβ subunits. The incomplete cleavage seen in these preparations is likely due to aggregation brought about by the relatively low salt concentration of the cleavage reactions (100mM KCl) compared to the conditions under which the complexes were initially purified (500mM KCl). It was thought that these uncleaved aggregates would not necessarily pose a problem in later assays as they will likely not be able to form complexes with PfTBP and thus not effect the site selection of PfTBP in the SELEX experiment. At this stage, the effects of PfTFIIA on PfTBP DNA-binding specificity has not been addressed due to time constraints on the project however this data does provide a starting point from which this question can further be explored.

3.5 Initial characterisation of PfTBP and PfTLP

TBP and at least some of its TLP paralogs are the only sequence-specific DNA-binding proteins in the eukaryotic RNAP-II general transcription machinery (reviewed in Thomas & Chiang, 2006). Their structure is primarily made up of a conserved C-terminal domain responsible for the recognition and stable binding of the core promoter in order to direct the assembly of the RNAP-II PIC (Tan *et al*., 1993; Kim *et al*., 1993; Nikolov *et al*., 1996; Duttke *et al*., 2016; reviewed in Thomas & Chiang, 2006). *P. falciparum* orthologs for both TBP and TLP were identified through bioinformatics analysis several years ago but their DNA-binding properties have so far not been sufficiently characterised (McAndrew *et al*., 1993; Ruvalcaba-salazar *et al*., 2005; Bischoff *et al*., 2010). Initial work in our lab on the characterisation of these proteins was done in collaboration with two MSc students, Steven Bing and Gertrud Talvik. This work provided first evidence for stable binding of both PfTBP and PfTLP to previously characterised putative *P. falciparum* promoter regions of the genes *gbp-130* and *kahrp* (Ruvalcaba-Salazar *et al*., 2005; Bing, 2015; Talvik, 2016).

EMSAs are a powerful and effective way of analysing protein-DNA interactions *in vitro* (Garner & Revzin., 1981; Hellman & Fried., 2007). In these assays proteins are incubated with nucleic acid probes and stable protein-DNA complexes are separated from unbound DNA by non-denaturing gel electrophoresis. Protein-DNA interaction is detected by reduced mobility of the DNA probe when bound to protein (a mobility shift or gel shift). Traditionally, EMSA assays are performed using native polyacrylamide gels. However, results from our early attempts to perform EMSAs showed that protein-DNA complexes formed with PfTBP or PfTLP were quantitatively retained in the loading wells (Bing, 2015; Talvik, 2016). This suggested aggregation of PfTBP/PfTLP nucleoprotein complexes or the occurrence of multiple PfTBP/PfTLP binding events to a single DNA-probe leading to the

formation of large multi-protein complexes that were unable to migrate into the gel. This problem was solved by switching to a Mg^{2+} agarose-gel system as the increased pore size of the agarose-gel matrix is able to accommodate larger protein-DNA assemblies (Zerby & Lieberman, 1997; Malecová *et al*., 2005). In these EMSAs it was found that binding of putative *P. falciparum* promoters with increasing amounts of PfTBP or PfTLP resulted in a gradual decrease of probe mobility that corresponded to a decrease in free-probe signal until quantitative binding was reached (Talvik, 2016). These data were consistent with the polyacrylamide EMSA results and confirmed binding of multiple PfTBP or PfTLP molecules to the DNA probes under the conditions used. To determine whether PfTBP binding is site-selective, further experiments were carried out using poly(dG-dC) as competitor DNA. Because of the extremely high A/T-base content of the *P. falciparum* genome (Gardner *et al*., 2002), a G/C-rich competitor such as poly(dG-dC) was deemed to be best suited to discriminate specific and non-specific protein-DNA interactions. The results of representative experiments are shown in Fig. 17.

3.5.1 PfTBP forms stable DNA complexes at A/T-rich sequences distinct from a consensus TATA box

In EMSA competition experiments, GBP-130 probes (5*f*mol) were first incubated with saturating amounts of PfTBP in the absence of competitor DNA. This resulted in a quantitative shift of the free-probe (Fig. 17, compare lanes 1 and 2). The mobility of the PfTBP-DNA complex is relatively high when compared to Mg^{2+} agarose EMSA shifts seen with HsTBP under similar reaction conditions (Malecová *et al*., 2015). In EMSAs, the movement of a protein-DNA complex relative to that of free-probe is largely influenced by three factors: the molecular weight of the bound protein, the tertiary structure of the bound protein, and the structure of the DNA-protein complex. Stable binding of HsTBP to the

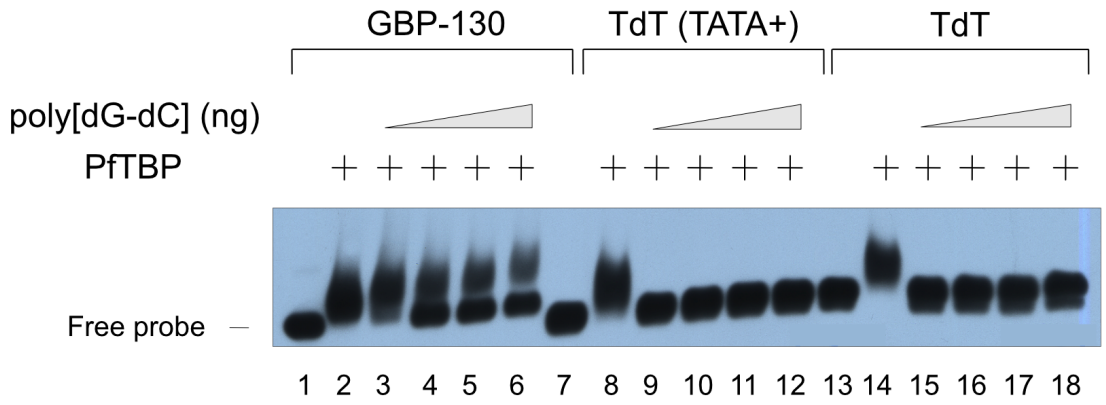


Figure 17: PfTBP binding to DNA is site-selective. Binding reactions (10µl) contained 12ng 6H-PfTBP and 5 *f*mol 282bp GBP-130, 260 bp TdT (TATA+) or 266 bp TdT biotinylated DNA-probes as indicated. Reactions were incubated on ice for 45min in standard transcription buffer (Ch2. Materials & Methods, Section 2.5). PfTBP binding was challenged with 40ng, 100ng, 200ng, and 400ng poly(dG-dC) as indicated. Protein-DNA complexes were separated from free-probe on a 1.4% Mg^{2+} agarose gel, transferred to a positively charged nylon membrane, and visualised by chemiluminescence on X-ray film.

---

minor groove of TATA box containing DNA induces a severe bend in the DNA-helix (> 80˚) through the insertion of two phenylalanine side chains into the first and last base steps of the the TATA box element (F288 & F197) (Nikolov *et al*., 1996; Hieb *et al*., 2007). This bend causes an unusually slow migration of TBP-DNA complexes in EMSAs (Lieberman *et al*., 1991; Zhao & Herr, 2002). Considering the similarities in both size and structure of the PfTBP and HsTBP C-terminal regions, together with the absence of one of the conserved phenylalanines in PfTBP (see Section 3.1), it seems likely that the relatively high mobility of the PfTBP-DNA complex seen in our experiments indicates the absence of, or a much lesser degree of, DNA bending in the PfTBP-DNA complex. Competition with 100ng poly(dG-dC) resulted in a partial disruption of PfTBP complexes leading to roughly 50% recovery of the free-probe signal (Fig. 17, compare lanes 2 and 4). Interestingly, further increasing the amount of competitor DNA to 400ng poly(dG-dC) did not further decrease the amount of PfTBP-DNA complexes (Fig. 17, compare lanes 4 and 6). The resistance of

the remaining PfTBP-DNA complexes to poly(dG-dC) challenge can be interpreted as follows: Binding of PfTBP to the DNA-probe occurs at multiple sites. A sub-population of these sites is sensitive to poly(dG-dC) challenge, suggesting that the relative affinity of PfTBP for these sites is low and likely due to non-specific protein-DNA interaction. A distinct population of sites is bound with comparatively higher affinity and is resistant to poly(dG-dC) challenge, suggesting that these sites mediate sequence-specific interactions, and thus contain sequence motifs recognised by PfTBP. The data therefore suggest that the formation of stable PfTBP-DNA complexes is sequence-specific.

To test whether a consensus TATA box can mediate the formation of stable PfTBP-DNA complexes, protein-DNA binding experiments were repeated using two variants of the GC-rich TATA-less mouse terminal deoxynucleotidyl transferase (TdT) promoter (Malecová *et al*., 2007), the wild-type TdT sequence lacking the TATA-motif and a mutated variant containing the prototypical TATA box element TATAAAA known to promote stable binding of eukaryotic TBPs (Tan., *et al* 1993; Nikolov *et al*., 1996; Malecová *et al*., 2015). As with GBP-130, addition of saturating amounts of PfTBP resulted in quantitative binding of both TdT-probe variants (Fig. 17, compare lanes 1 and 2 with lanes 7 and 8, and 13 and 14). However, in contrast to the putative GBP-130 promoter region addition of just 40ng poly(dG-dC) was sufficient to completely abolish the PfTBP binding to the TdT promoter sequence (Fig. 17, compare lanes 3 and 15). Furthermore, and most significantly, the presence of a TATA box element did not increase the resistance of PfTBP-TdT complexes to poly(dG-dC) challenge (Fig. 17, compare lanes 9 and 15). In summary, these data suggest that PfTBP binds in a sequence-specific manner to A/T-rich DNA probes such as the P. falciparum *gbp-130* and *Kahrp* promoter sequences. However, PfTBP differs from other eukaryotic TBPs in that it does not appear to form stable complexes with a consensus TATA-sequence.

3.5.2 Temperature dependence of PfTBP DNA binding

We initially used reaction conditions established for human and yeast TBP in our DNA-binding assays. Eukaryotic TBPs have been shown to bind DNA in a two stage process in which the DNA is at first bound loosely, followed by a transition to a more stable complex through the bending of the DNA-helix (Hoopes *et al*., 1992; Zhao & Herr, 2002; Masters *et al*., 2003; Delgadillo *et al*., 2009). Incubation at 30˚C thermodynamically favours the transition to a bent TBP-DNA complex and is therefore generally used as the standard condition for TBP DNA-binding experiments (Petri *et al*., 1995; Lieberman *et al.,* 1991; Lee *et al.,* 1992). In our initial experiments, reactions were placed on ice after incubation at 30˚C, before being loaded for electrophoresis. However, we noticed a high degree of variability in the efficiency and stability of protein-DNA complex formation by PfTBP in Mg^{2+} agarose EMSAs that appeared to correlate with the amount of time reaction samples were left on ice before electrophoresis (Talvik, 2015). This prompted investigations into the effects of incubation temperature on PfTBP-DNA complex formation. The results of these experiments are summarised in Fig. 18. In this experiment, GBP-130 DNA-probe was incubated with increasing amounts of PfTBP either on ice (0˚C) or at 30˚C. In order to limit PfTBP interactions to specific sites, non-specific interactions were suppressed by the addition of 400ng poly(dG-dC) to the binding reaction (Fig. 18). The results in Fig. 18 clearly demonstrate increased PfTBP DNA binding activity at 0˚C compared to 30˚C. Thus, in contrast to human and yeast TBP, PfTBP does not require higher incubation temperatures to facilitate stable complex formation. Rather, higher incubation temperatures appear to diminish stable PfTBP-DNA complex formation. This observation is consistent with the notion that stable PfTBP DNA binding does not involve severe DNA bending, indicated by the relatively high mobility of PfTBP-DNA complexes in Mg^{2+} agarose EMSA experiments.

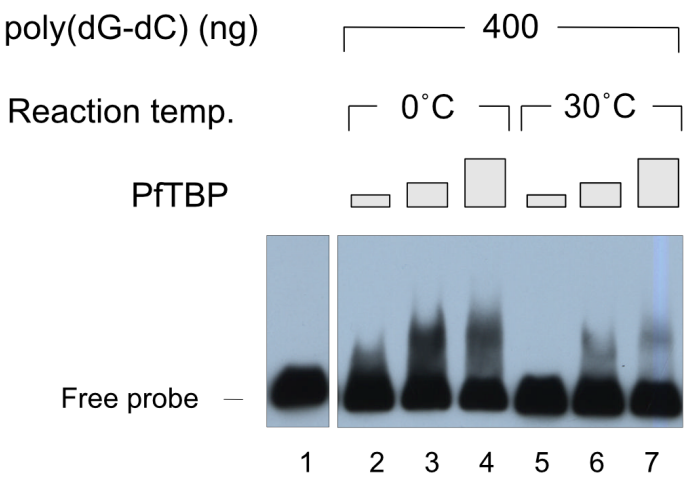


Figure 18: Effect of reaction temperature on PfTBP DNA binding. Binding reactions (10µl) contained 3ng, 6ng, and 12ng 6H-PfTBP incubated in the presence of 5 *f*mol biotinylated GBP-130 DNA-probe (282 bp) and 400ng poly(dG-dC) competitor DNA. Reactions were incubated on ice or at 30˚C for 45min in standard transcription buffer (Ch2. Materials & Methods, Section 2.5). Protein-DNA complexes were separated from free-probe on a 1.4% Mg^{2+} agarose gel, transferred to a positively charged nylon membrane, and visualised by chemiluminescence on X-ray film.

3.5.3 DNA binding properties of PfTLP are distinct from those seen in PfTBP

Preliminary Mg^{2+} agarose EMSA analysis of PfTLP demonstrated that it had the ability to form stable complexes with the putative *P. falciparum* promoters KAHRP and GBP-130 (Bing, 2015; unpublished data). Additionally, it was shown to form complexes with the TATA-containing Adenovirus type 2 major late (Ad2ML) promoter sequence in the absence of competitor DNA (Bing, 2015). Differences in PfTLP binding affinity for these probes was not immediately apparent as in all three cases DNA binding reactions presented very similar EMSA results. The PfTLP mediated mobility-shifts were characterised by the appearance of a predominant fast-migrating shift that was followed by several slower and comparatively faint shifts. The formation of multiple PfTLP-dependent shifts were attributed

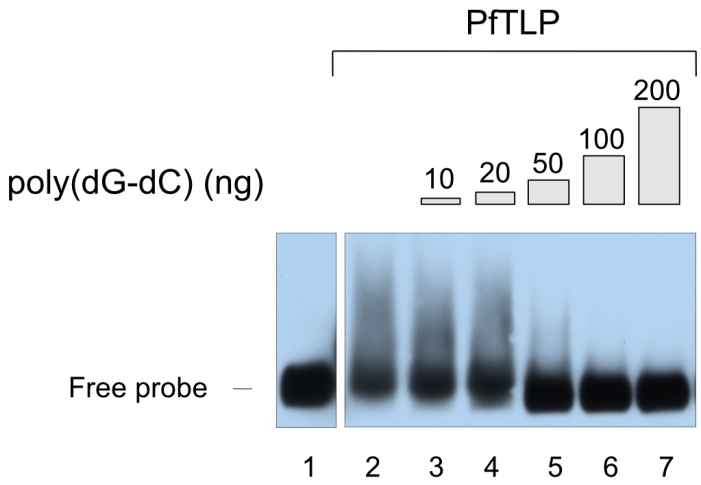


Figure 19: PfTLP protein-DNA complexes are highly sensitive to poly(dG-dC) competitor challenge. Binding reactions (10µl) containing 120ng 6H-PfTLP and 5 *f*mol biotinylated GBP-130 DNA-probe (282 bp) were incubated on ice for 45min in standard transcription buffer (Ch2. Materials & Methods, Section 2.5). PfTLP binding was challenged with 10ng, 20ng, 50ng, 100ng and 200ng poly(dG-dC) as indicated. Protein-DNA complexes were separated from free-probe on a 1.4% Mg^{2+} agarose gel, transferred to a positively charged nylon membrane, and visualised by chemiluminescence on X-ray film.

---

to two possible causes: the first being that several PfTLP molecules would bind to the same DNA-probe leading to PfTLP-DNA complexes of varying sizes, and the second being that single binding events at different positions along the probes lead to PfTLP-DNA complexes with dissimilar conformations. This second explanation stemmed largely from the observation that all four of the conserved phenylalanine residues responsible for DNA-bending in eukaryotic TBPs are conserved in PfTLP (Ch1. Introduction, Section 1.2.4 & 1.2.6) (Bing, 2015). The DNA-binding activity of recombinant PfTLP is illustrated by a representative experiment shown in Fig. 19. In this experiment 60ng PfTLP was incubated with 5*f*mol GBP-130 probe at 0˚C for 45min. This resulted in a PfTLP-mediated shift and a corresponding reduction in the free probe signal (Fig 19, compare lanes 1 and 2). The

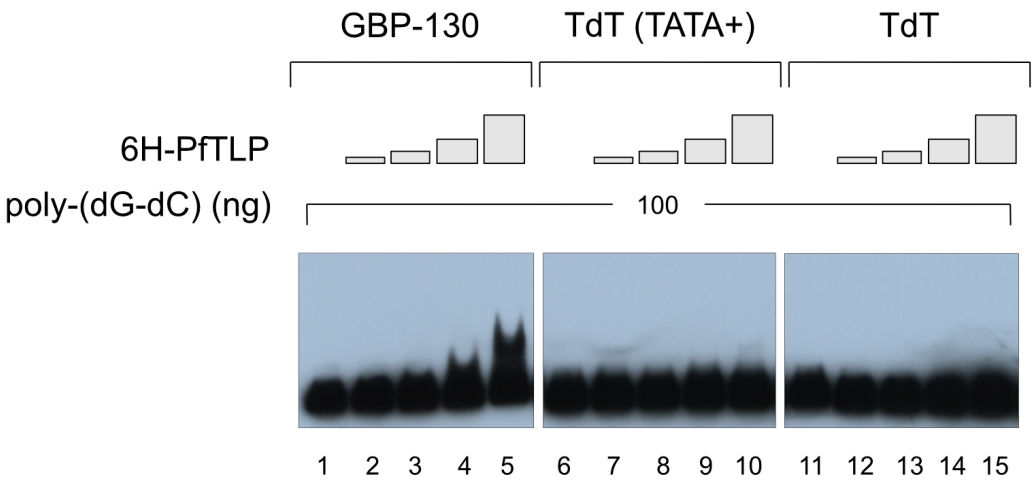


Figure 20: PfTLP binding to putative *P. falciparum* promoter is site-selective. Binding reactions (10µl) contained 10ng, 20ng, 40ng, and 80ng 6H-PfTLP and 5 *f*mol 282bp putative GBP-130, 260 bp TdT (TATA+) or 266 bp TdT biotinylated DNA-probes as indicated. Non-specific DNA interactions were suppressed by the addition of 100ng poly(dG-dC) competitor DNA as indicated. Reactions were incubated on ice for 45min in standard transcription buffer (Ch2. Materials & Methods, Section 2.5). Protein-DNA complexes were separated from free-probe on a 1.5% Mg^{2+} agarose gel, transferred to a positively charged nylon membrane, and visualised by chemiluminescence on X-ray film.

---

resolution of PfTLP-DNA shifts were at first very poor. The diffuse appearance of shifts was suggests dissociation of complexes during electrophoresis (Fig. 19, lanes 2, 3 and 4) (Bing, 2015). Through assay optimisation experiments it was found that increasing the agarose-gel percentage by just 0.1% drastically improved the visualisation of these PfTLP-DNA complexes (see Fig. 20 and Fig. 28). Competition experiments revealed that PfTLP binding to A/T-rich putative *P. falciparum* promoter regions was abolished by the presence more than 20ng poly(dGdC) in the binding reaction (Fig. 19; compare lanes 2, 4 and 5). These findings indicate a very high sensitivity of PfTLP-DNA interaction to poly(dG-dC) relative to what was seen with PfTBP under similar conditions when using the same DNA-probe (Fig. 17; lane 3). As with PfTBP, addressing the possibility of site-specific PfTLP binding was done through additional DNA-binding experiments using the putative

GBP-130 promoter and mouse TdT and TdT(TATA+) promoter sequences. A summary of these results is shown in Figure 20. In these experiments, increasing amounts of PfTLP were first added to protein-DNA binding reactions containing 5 *f*mol of the GBP-130 DNA-probe and 100ng poly(dG-dC). The addition of increasing amounts of PfTLP to the GBP-130 binding reactions resulted in the gradual emergence of a mobility-shift seen above the free-probe signal (Fig. 20, compare lanes 1,2,3,4 and 5). Interestingly, based on the relative mobility of the complex, PfTLP appears to form an unbent complex with the GBP-130 promoter despite the conservation of the phenylalanine residues responsible for the intercalation into the DNA minor groove (Nickolov *et al*., 1996). To further test the observed PfTLP-Ad2ML shifts (Bing, 2015) and to provide a means of direct comparison to the DNA-binding activity of PfTBP seen in Fig. 16, the experiment was repeated using the mouse TdT and TdT(TATA+) probes. Even after the addition of 40ng PfTLP, stable PfTLP-DNA complexes with the TdT promoter could not be detected in the presence of 100ng poly(dG-dC) (Fig. 20, compare lanes 11 and 15). Furthermore, no TATA box mediated binding of PfTLP was found under similar conditions when using the TdT(TATA+) DNA-probe (Fig. 20, compare lanes 6 and 10). Taken together, these data indicate that as seen with PfTBP, there appears to be a sequence-specific recognition of binding sites on the GBP-130 probe by PfTLP. In addition, these sites appear to be absent in the mouse TdT promoter sequence. Thirdly, as with the human TLP ortholog (TRF2), PfTLP does not appear to recognise or form stable-complexes with the TATA box element (Duttke *et al*., 2016). There is also evidence of both quantitative and qualitative differences between the respective complexes formed by PfTBP and PfTLP with the GBP-130 DNA-probe. PfTLP binding appears to be less efficient considering the relative amounts of protein needed to produce a shift signal, approximately 200 *f*mol/µl PfTLP versus approximately 40 *f*mol/µl PfTBP. Furthermore, the PfTLP-DNA complexes appear to be less stable as evidenced by

their increased sensitivity to poly(dG-dC) challenge and the dissociation of the protein-DNA complexes seen as a smear below the PfTLP-DNA complexes.

3.6 Characterisation of PfTBP sequence specificity by Systematic Evolution of Ligands by Exponential Enrichment (SELEX)

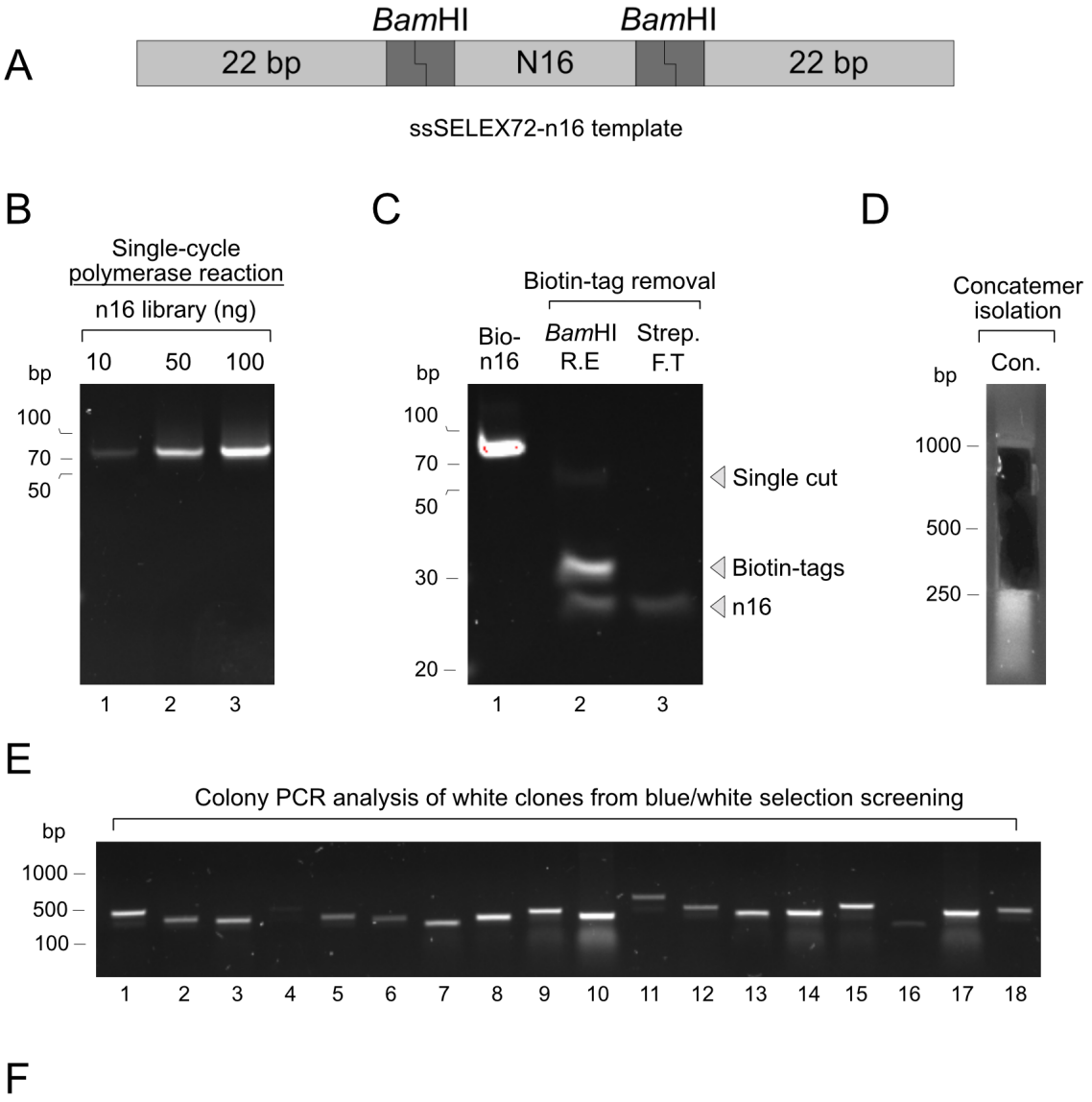
The positioning of the RNAP-II PIC is achieved primarily by the recognition of short DNA-sequence motifs known as core promoter elements (CPEs). This process is mediated by several RNAP-II general transcription factors, often acting in conjunction with gene-specific transcriptional activators bound to distal promoter elements (reviewed Smale & Kadonaga, 2003; Gross & Oelgeschläger, 2006; Thomas & Chiang, 2006). Studies of eukaryotic promoter architecture have identified several such sequences found in varying combinations within approximately 50bp of most transcription start sites. Typically, CPEs are identified using *in vitro* transcription in conjunction with DNA-binding assays such as DNase I footprinting (Burke & Kadonaga, 1996; Marbach-Bar *et al*., 2015). The data generated from these assays are then used in bioinformatics searches to locate the given sequence elements in the promoter regions of other genes. The extremely high A/T-base content of *P. falciparum* intergenic regions (often > 90% A/T) has however made the *in silico* prediction of CPEs very difficult (Gardner *et al*., 2002). This extreme base content bias and the dearth in functional data regarding the parasites general transcription machinery are the main reasons why the structure of core promoters in *P. falciparum* could not be determined thus far. Motivated by the observation that DNA binding by PfTBP is sequence-specific, and that this specificity is not mediated by the TATA box element (Section 3.7), a systematic evolution of ligands by exponential enrichment (SELEX) approach (Ch2. Materials & Methods, Section 2.7) was adapted to try elucidate sequence motifs recognised by PfTBP. The SELEX assay is a PCR-based method for the enrichment

of DNA sequences from a random DNA sequence pool based on their potential to be bound by a protein of interest (Ogawa & Biggin, 2012).

For the experiments described here, a DNA library containing 16 nucleotide random sequences (n16 library) was prepared (Ch2. Materials & Methods, Section 2.7.1). Briefly, 6H-PfTBP was immobilised on Ni^{2+}-magnetic beads and incubated with a pool of random 16bp target sequences flanked by primer-binding sites. Bound oligonucleotides were isolated by magnetic pull-down of 6H-PfTBP and eluted with high salt. The eluted sequences were then PCR amplified, purified, and used for another round of SELEX selection. After seven cycles of enrichment, the primer sites were removed by enzyme restriction digestion. The enriched DNA pool was ligated to form concatemers containing 10-15 copies of 16bp target sequences. These concatemers were then cloned into a pUC19 sequencing vector and analysed. The resultant datasets were scanned for reoccurring sequence motifs using the online Multiple Em for Motif Elicitation (MEME) application of the MEME-suite (Bailey *et al*., 2009; http://meme-suite.org/).

3.6.1 Construction and verification of a 16 nucleotide random-sequence DNA library (n16 Library)

The success of the SELEX assay is largely predicated on the input DNA being truly random. Any nucleotide and/or sequence bias brought about either by way of the initial chemical synthesis procedure and/or the various amplification and cloning steps would obviously lead to a skewed result. The results of this procedure are summarised in Fig. 21. In order to test for any possible intrinsic sequence biases, the SELEX procedure was carried out in the absence of PfTBP selection and subjected to the respective PCR amplification and cloning steps. The n16 library was constructed using single stranded



Meme input	Sequence count	Sequence motifs detected	AT-base content
n16 library	480	none	48.6%

Figure 21: Construction of a random n16 72bp SELEX input library and verification of non-bias. DNA products were analysed by 15% PAGE (panels B and C) or 2% agarose gel electrophoresis (panels D and E) and visualised by ethidium bromide staining. (A) Schematic of ssSELEX72-n16 template DNA showing a central randomised 16bp sequence flanked by 22bp primer binding sites. (B) Single-stranded ssSELEX72-n16 oligonucleotides were converted to the double- stranded n16 library by a single-cycle polymerase reaction. (C) The n16 library was amplified by PCR using biotinylated primers (lane 1) and flanking sequences removed by *Bam*HI digestion followed by incubation with streptavidin sepharose (lanes 2, 3). *Bam*HI digested n16 library fragments were ligated and the resultant concatemers (250bp-1000bp) isolated by agarose gel extraction (D), cloned into a pUC19 sequencing vector and transformed by electroporation into *E.coli* DH5α cells. (E) White colonies were screened for inserts by colony PCR. Plasmids containing inserts > 300bp were amplified and sequenced. (F) A total of 480 16bp sequences were analysed for any apparent sequence bias.

oligonucleotides (ssSELEX72-n16), 72 nucleotides in length, comprised of a 16bp random DNA-sequence flanked by 22bp PCR-primer binding sites and containing *Bam*HI recognition sequences (Fig. 21A). Single-cycle polymerase reactions were carried out using the ssSELEX72-n16 templates and the reverse SELEX80-*Bam*HI primer (Ch2. Materials & Methods, Table 5) to produce the double stranded n16 library. Reactions were pooled and gel-purified before being analysed on a native poly-acrylamide gel (Fig. 21B). A single sharp 72bp band reactive to ethidium bromide staining was seen, confirming the completion of the dsDNA synthesis (Fig. 21B). The next step in the SELEX procedure requires PCR amplification of the enriched n16 library using biotinylated primers (Fig. 21C, lane 1). This is done to facilitate the removal of the flanking sequences. *Bam*HI digestion resulted in a mixture of fully digested n16 target sequences, together with biotin-linked flanking sequences and any remaining undigested or partially digested oligonucleotides (Fig. 21C, lane 2), which were removed by incubation with streptavidin-coated beads (Fig. 21C, lane 3). Following gel purification, the n16 sequences were ligated to form concatemers, which in agarose gel electrophoresis, appeared as a long smear due to the variation in concatemer lengths (Fig. 21D). Concatemers ranging from approximately 250bp to 1000bp in length were cut out from the agarose gel, purified, cloned into pUC19 vector, and transformed by electroporation into *E.coli* DH5-α cells. Positive clones were selected by blue/white screening on X-gal-containing agar plates (Sambrook & Russell, 2001), and white colonies were then analysed by colony PCR to select clones with an insert > 350bp (Fig. 21E). The plasmid DNA from these clones was subsequently isolated and sent for sequencing analysis. The raw sequence data of the n16 sequences was manually curated to exclude any truncated sequences or cloning artefacts prior to MEME analysis. The results of the MEME analysis are shown in Fig. 21F. No significant evidence of sequence motifs could be detected in a total pool of 480 sequences (Appendix A3).

Furthermore, the A/T:G/C ratio was found to be roughly equal to a 1:1 ratio, indicating that there was no significant nucleotide bias in the n16 library.

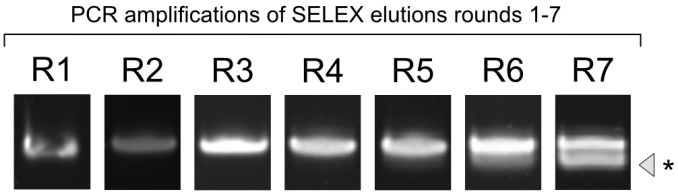
3.6.2 n16 library enrichment by PfTBP selection

In order to elucidate potential PfTBP binding motifs from the verified library, seven consecutive rounds of SELEX selection were carried out. The results of this procedure are summarised in Fig. 22. Quantitative PCR (qPCR) was used to determine the amount of immobilised 6H-PfTBP that would be needed in a 20µl SELEX binding reaction to bind approximately 1% (0.1ng) of the 10ng input DNA in the presence of 400ng poly(dG-dC) competitor DNA. These conditions favour isolation of high affinity binding sites over sites bound in a non-specific manner (Ogawa & Biggin, 2012). The amount of competitor DNA was chosen based on previous EMSA results showing that 400ng poly(dG-dC) was adequate to suppress non-specific PfTBP-DNA interactions (Section 3.7.1). Four binding reactions containing 0ng, 5ng, 10ng, and 15ng immobilised 6H-PfTBP were incubated with 10ng n16 DNA library at 21˚C for 45min in 20µl standard transcription buffer (Ch2. Materials & Methods, Section 2.5). The bound oligonucleotides were separated together with PfTBP from the sequence pool and extensively washed before elution. Quantification of the elutions by qPCR indicated that 10ng PfTBP was sufficient to bind 0.1ng of the input DNA under the given reaction conditions (Fig. 22A). The seven rounds of SELEX selection were then carried out with the bound DNA from each enrichment cycle being PCR amplified and analysed by native poly-acrylamide gel electrophoresis (Fig. 22B). It was found that, in addition to the specific 72bp PCR product, a second shorter product with higher mobility was formed between enrichment cycles 5 and 7 (Fig. 22B, compare R5, R6, and R7). Further processing of the enriched n16 library after seven rounds of selection (R7) by PCR biotinylation (Fig. 22C, lane 1) and *Bam*HI cleavage, revealed that this lower

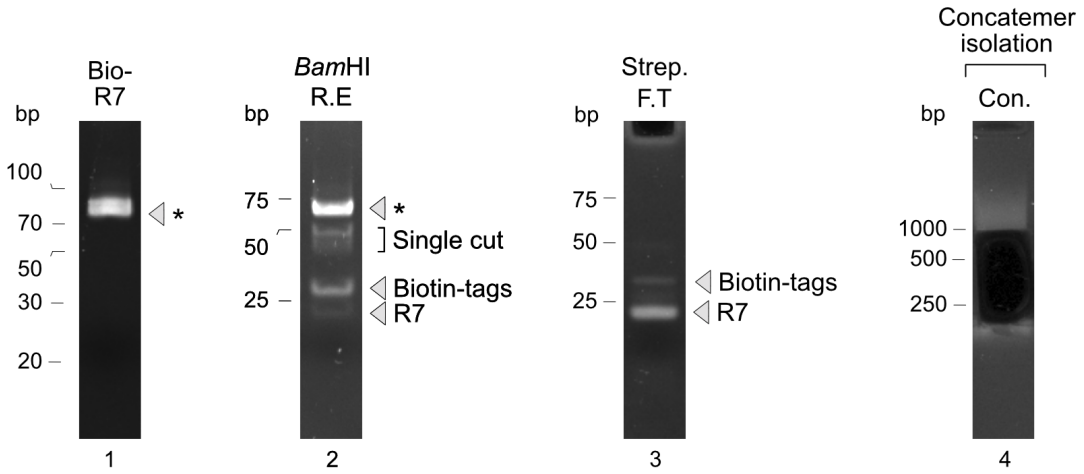
A

6H-PfTBP (ng)	Input DNA bound (ng)
15	0.1109
10	0.0902
5	0.0492
0	0.0044

B



C



D

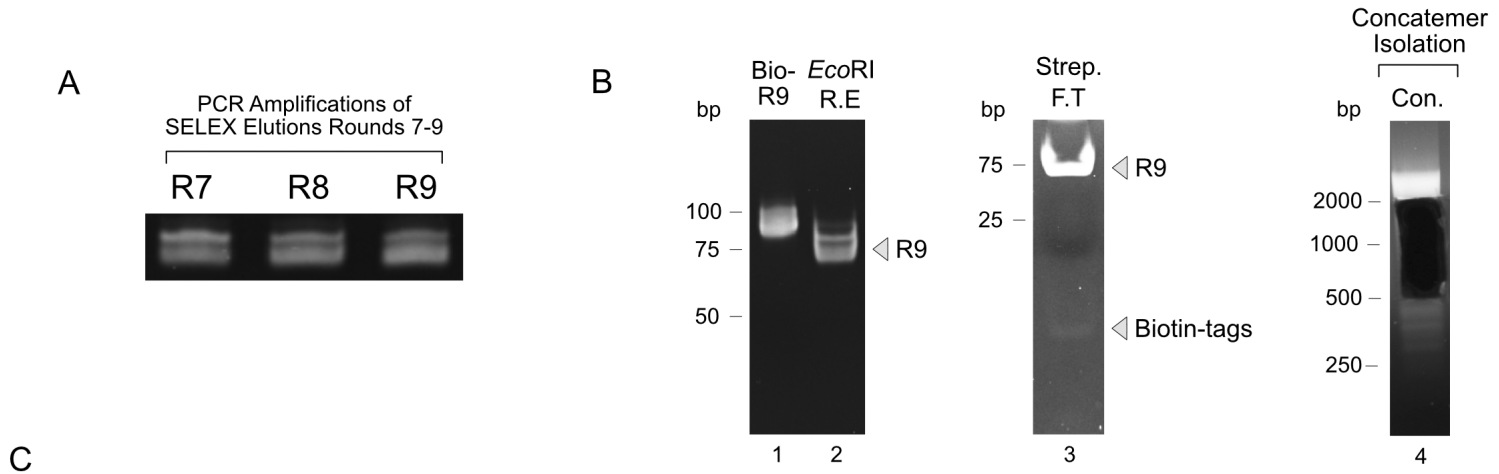
Meme Input	Sequence count	Sequence motifs detected	AT base content
n16 Library	480	none	48.6%
Round 7	397	none	61.0%

Figure 22: SELEX enrichment of DNA sequences from a 72bp n16 library bound by PfTBP. DNA products were analysed by 15% PAGE (panels B and C) or 2% agarose gel electrophoresis (panel C, lane 4) and visualised by ethidium bromide staining. (A) Quantitative PCR was used to determine the amount of immobilised 6H-PfTBP needed in a 20µl binding reaction to bind 1% of 10ng input DNA (n16 library) in the presence of 400ng poly(dG-dC). (B) 15% polyacrylamide gel electrophoresis of PCR-amplified SELEX elutions. 50ng library DNA from selection rounds 1 to 7 (R1-R7) was loaded. The position of a second PCR product band with a higher mobility appearing in selection rounds 5 to 7 is indicated by an asterisk. (C) After the seventh round of selection the enriched library was biotinylated by amplification with biotin-tagged primers before being digested with *Bam*HI to allow for ligation of multiple enriched sequences into concatemer chains (panels 1 + 2). The cleaved primer binding sites originally flanking the 16bp target sequences were removed together with any undigested sequences by incubation with streptavidin sepharose (panel 3). Following ligation the resultant concatemers ranging from 250-1000bp were isolated by agarose gel extraction (panel 4). (D) Concatemers were cloned into the pUC19 sequencing vector and analysed for motif enrichment using the online Multiple Em for Motif Elicitation (MEME) application of the MEME-suit (Bailey *et al*., 2009; http://meme-suite.org/).

band was resistant to *Bam*HI cleavage (Fig. 22C, lane 2). This resulted in its removal from the sequence pool in the streptavidin sepharose pulldown (Fig. 22C, lane 2). The n16 library R7 sequence pool that was responsive to *Bam*HI digestion was concatemerised, cloned, and analysed (Fig. 22C, lane 4). Motif-search analysis of 397 sequences of the n16 enriched DNA sequence pool (R7), did not reveal any significant motifs. However, notably, the overall A/T content of enriched sequences was 61% (Fig. 22D), compared to 48.6% in the original library (Fig. 22E). Thus, although no definite consensus motif for PfTBP binding could be discerned, binding of PfTBP clearly enriched A/T-containing DNA sequences, consistent with the observed preference of PfTBP in the EMSA DNA-binding experiments (Section 3.7).

3.6.3 Analysis of a truncated SELEX product enriched in selection rounds 5-7

Two explanations were posited to explain the enrichment of a truncated SELEX product seen between the R5 and R7 enrichment cycles (Fig. 22D). The first theory was that this DNA product was a PCR artefact which preferentially amplified over other library sequences during consecutive selection cycles. To eliminate this problem the correct DNA band from cycle 5 (R5) was excised from polyacrylamide gels and purified prior to being used as input DNA for continued SELEX enrichment. This approach was however unsuccessful as the truncated SELEX product re-emerged within two cycles (data not shown). This result suggested that the truncated DNA sequences were indeed selected by PfTBP. In order to determine the precise reason why these DNA-sequences could not be digested with *Bam*HI and displayed increased mobility in native polyacrylamide gels, a new set of extended biotinylated primers complimentary to the 5' ends of the original SELEX primer sequences were synthesised, each of which contained a 3' *Eco*RI restriction site. This strategy allowed for the inclusion of the truncated PCR product into



C

*Bam*HI

CCCATATATGSATGGCYCGCARGAYAACTCTAGAGGATCCCCGCTCCTAGGATCCATCATTAATTATCACTGTACTGCAC

*Eco*RI

CAGCAAATCCGAATTCGGATTTGCTGGTGCAGTACAGTGGAAAACAAGTAGTGGATCTTTAGGAGCTTGAAATCGAGCAG

*Eco*RI *Eco*

GAATTCGGATTTGCTGGTGCAGTACAGTGGACCCGGGACGTTACAAAAGGGATCCTTAGGAGCTTGAAATCGAGCAGGAA

RI *Eco*RI

TTCGGATTTGCTGGTGCAGTACACCTGCTCTTTCACCATATGGTAGGAGCTTGAAATCGAGCAGGAATTCCTGCTGCATT

TCAAGCTCCTACGATYCGGTGMAAGMGCAGGTGGGATSCACYGGACTGTMCGAATTCCYGCYAGATTTCATGCTTCCG

Figure 23: Identification of a truncated SELEX PCR product formed between SELEX rounds 5-7. (A) SELEX was continued for a further two rounds R8 + R9. (B) Biotinylated primers containing *Eco*RI restriction sites flanking the original SELEX primer sequences were used to amplify the 9th round enrichment library. The library was digested with *Eco*RI followed by purification with streptavidin sepharose (panel 3), ligated and cloned into pUC19. (C) Example of a clone sequence containing concatemerised truncated SELEX PCR products (shaded in grey). M13-sequencing primer sites coloured green. Original SELEX primer binding sites coloured blue.

the sequence pool that would later be analysed for motif enrichment. The results of these experiments are summarised in Fig. 23. The SELEX enrichment procedure was continued for a further two cycles using the R7 library containing both PCR products as input DNA (Fig. 23A). The resultant R9 library sequences were then amplified and biotinylated using the primers containing the *Eco*RI sites (Fig. 23B, lane 1). Analysis of the *Eco*RI digestion of the biotinylated PCR products prior to concatemerization confirmed that both the correct and truncated SELEX products were now included in the analysis (Fig. 23B, lane 2). After removal of *Eco*RI cleaved 3' and 5' ends as before with streptavidin sepharose (Fig. 23B, lane 3), target sequences were concatemerised, isolated by agarose gel extraction (Fig. 23B, lane 4), and cloned into the pUC19 vector. Sequencing of positive clones revealed that the aberrant SELEX PCR products were composed primarily of truncated 16bp target sequences bordered by defective *Bam*HI sites (Fig. 23C). It seems reasonable to assume that these sequences originated initially from aberrant PCR amplification of PfTBP-enriched DNA during the SELEX procedure. The fact that these sequences were strongly enriched through PfTBP binding suggests that they may contain high affinity PfTBP binding sites. However, as with full length enriched SELEX-selected DNA sequences, MEME analysis revealed no specific DNA motifs enriched in the truncated enriched PCR products.

3.7 Functional characterization of PfTFIIA-γ1 and PfTFIIA-γ2 complexes

Transcription factor IIA has in itself no inherent DNA binding capabilities. It has however been shown to act as a positive regulator of TBP for both basal and activated transcription *in vitro* (Reinberg *et al*., 1987; Imbalzano *et al*., 1994b; Maldonado *et al*., 1990; Lieberman *et al*., 1997; reviewed in Thomas & Chiang, 2006). This stimulatory effect can broadly be attributed to three main functional roles of TFIIA. In the first instance, TFIIA mediates a conformational change to the TBP-DNA complex through the alteration of the DNA bending-angle within the TBP-TATA nucleoprotein complex, and thereby increases the stability of the TBP-DNA interaction (Hieb *et al*., 2007). TFIIA also acts as an antagonist to negative regulators of TBP such as NC2 by competing for overlapping regions of interaction on the convex surface of TBP (Kim *et al*., 1995; Malecová *et al*., 2015; reviewed in Thomas & Chiang, 2006). Lastly, TFIIA is thought to play a co-activator role by facilitating interactions between transcriptional activator proteins and the TFIID complex in order to promote the assembly of the RNAP-II PIC (Kobayashi *et al*., 1995; Kang *et al*., 1995; Ozer *et al*., 1998; Chou *et al*., 1999; reviewed in Thomas & Chiang, 2006). In addition to these well established TBP-related functions, TFIIA has also been shown to form stable complexes with TBP-like protein (TLP or TRF2) and appears to facilitate the localisation of TLP to the cytoplasm (Teichman *et al*., 1999; Nakadai *et al*., 2004). A comprehensive understanding of TLPs involvement in transcriptional regulation is yet to be elucidated. However, there is strong evidence indicating that it functions in the recognition of promoter subtypes lacking the TATA-box element (Wang *et al*., 2014; Kedmi *et al.,* 2014; Duttke *et al*., 2016). Having co-expressed and purified two distinct *P. falciparum* TFIIA complexes, the PfTFIIA complexes were tested in regard to their ability to (i) bind to DNA complexes formed with either PfTBP or PfTLP, and (ii) to affect the previously established DNA-binding activity of PfTBP or PfTLP. In order to answer the question as to whether the

PfTFIIA-γ1 and PfTFIIA-γ2 can indeed form ternary complexes with PfTBP or PfTLP, immobilised template assays were conducted. Immobilised-template assays serve as a direct measure of protein-DNA interaction. Biotinylated DNA-probes immobilised to streptavidin-coated magnetic beads are used as binding targets in protein-DNA binding reactions that contain the protein(s) of interest. The unbound fraction is washed from the beads and the bound proteins are eluted and analysed by SDS-PAGE and by immunoblotting. Experimental parameters such as template and protein concentration in the binding reaction for the analysis of PfTBP and PfTLP DNA-binding using immobilised template assays (ITAs) were established early on in the laboratory by myself and Steven Bing, an MSc. student (Bing, 2015).

3.7.1 PfTBP dependent recruitment of PfTFIIA to putative GBP-130 promoter DNA

To test whether PfTBP can recruit PfTFIIA-γ1 or PfTFIIA-γ2 complexes to DNA, proteins were incubated with 1pmol immobilised GBP-130 probe in the presence of 500ng poly(dG-dC) competitor DNA to suppress unspecific PfTBP-DNA interaction (see Fig. 17, Section 3.5.1). Bound-proteins were eluted, separated by SDS-PAGE, and analysed by immunoblot together with the PfTBP and PfTFIIA-γ1 input fractions using anti-PfTBP and anti-PfTFIIA-γ1 antibodies developed in the laboratory. As shown in Fig. 24A, PfTBP bound quantitatively to the GBP-130 DNA template under these conditions (Fig. 24A, compare lanes 1 and 3). In contrast, there was no visible binding of the isolated PfTFIIA-γ1 complex to the immobilised DNA beads (Fig. 24A, compare lanes 2 and 4). In the presence of PfTFIIA-γ1, binding by PfTBP to the putative promoter was clearly diminished (Fig. 24, compare lanes 3 and 5). However, PfTFIIA-γ1 binding could not be detected (Fig. 24A, lane 5). In contrast, the presence of the PfTFIIA-γ2 complex did not affect PfTBP DNA binding (Fig. 24B, compare lanes 1 and 3). However, the addition of both PfTBP and

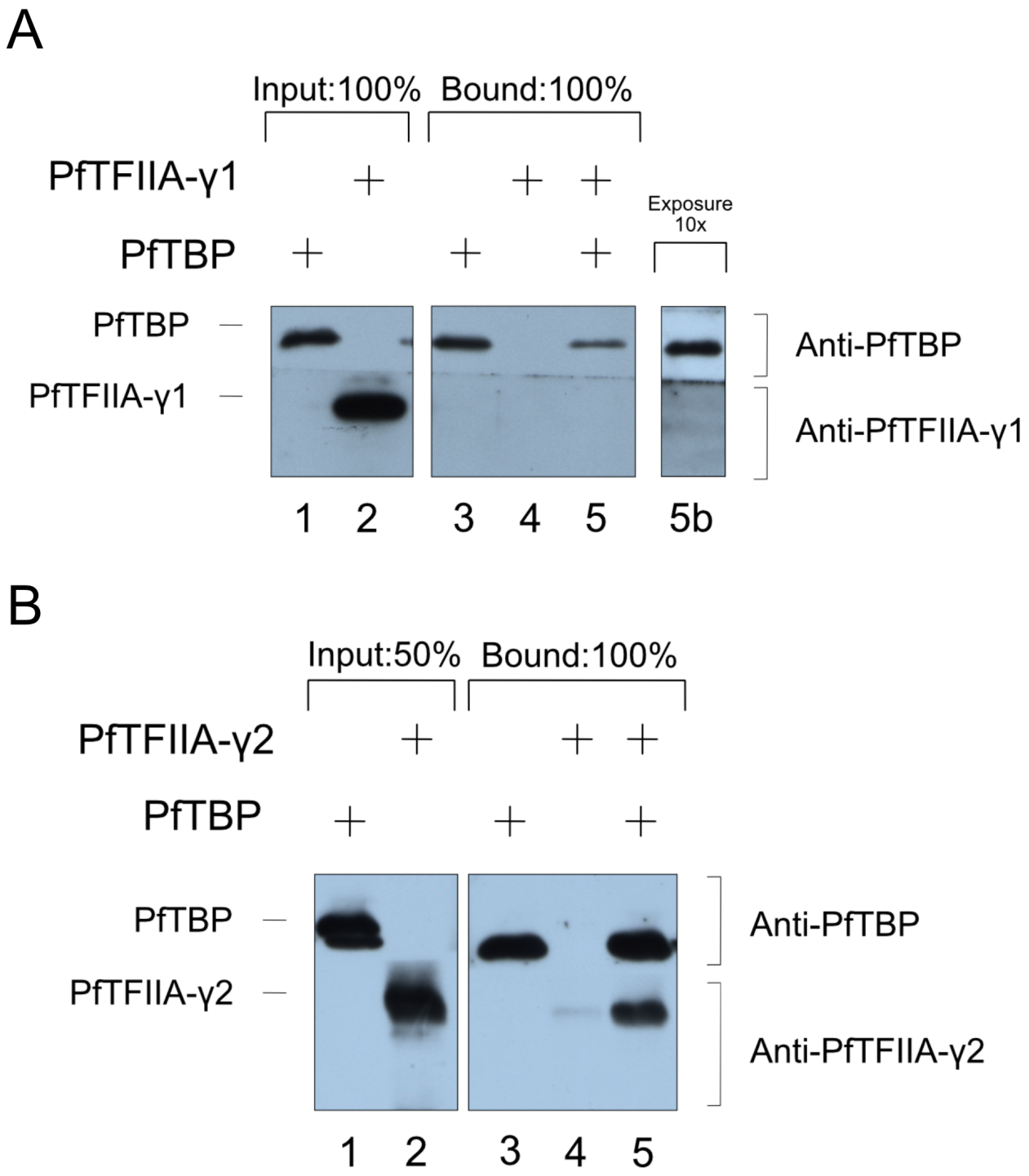


Figure 24: PfTBP dependent recruitment of PfTFIIA-γ2 to GBP-130 promoter DNA. Binding reactions (50µl) contained 50ng 6H-PfTBP, 800ng 6H-PfTFIIA-γ1 (panel A), or 160ng 6H-PfTFIIA-γ2 (panel B) as indicated, 1pmol 422bp immobilised GBP-130 DNA-probe and 500ng poly(dG-dC) competitor DNA. Protein-DNA binding reactions were incubated at 30˚C for 45min in standard transcription buffer (Ch2. Materials & Methods 2.6). Bound proteins were eluted in SDS-PAGE loading buffer, separated by SDS-PAGE, and analysed by immunoblotting. The protein-bound membrane was cut into strips and incubated with rabbit polyclonal anti-PfTBP and anti-PfTFIIA-γ1s or anti-PfTFIIA-γ2s antibodies as indicated.

PfTFIIA-γ2 to the binding reaction resulted in the recruitment of the PfTFIIA-γ2 complex to the GBP-130 DNA-probe (Fig. 24B, compare lanes 4 and 5). Overall, the data presented in Fig. 24 demonstrate that PfTFIIA-γ2 is able to form a stable protein-DNA ternary complex on putative GBP-130 promoter DNA in the presence of PfTBP. This presented data provide strong evidence for the identification of PfTFIIA-γ2 as a functional *P. falciparum* TFIIA ortholog. PfTFIIA-γ1 appears to act as a negative regulator of PfTBP binding, however, binding of PfTFIIA-γ1 to PfTBP could not be detected under these conditions.

3.7.2 PfTFIIA-γ2, but not PfTFIIA-γ1, stimulates PfTBP-DNA binding activity

The ability of the putative PfTFIIA complexes to stimulate DNA-binding by PfTBP was tested in Mg^{2+} agarose gel EMSAs using previously established reaction conditions as a guide (see Section 3.7; Bing, 2015; Talvik, 2016). The binding reactions were set up as follows. GBP-130 DNA-probes (5*f*mol) were first incubated with saturating amounts of PfTBP. This resulted in a quantitative shift of the free probe (Fig. 25, compare lanes 1 and 2). PfTBP binding was suppressed by the addition of 100ng and 400ng poly(dG-dC), resulting in the recovery of approximately 80% of the free probe (Fig. 25, compare lanes 2, 3 and 4). Increasing amounts of either PfTFIIA-γ2 or PfTFIIA-γ1 were then added to the PfTBP DNA binding reactions (Fig. 25, lanes 5 to 7 and 8 to 10 respectively). PfTFIIA-γ2 restored PfTBP DNA binding in a dose-dependent manner (Fig. 25, compare lanes 4, 5, 6, and 7). No such effect was seen with the PfTFIIA-γ1 complex. Importantly, no mobility shifts were seen with either PfTFIIA ortholog in the absence of PfTBP (Fig. 25, lanes 11 and 12). These results are consistent with the results of the immobilised template assays and confirm that recombinant PfTFIIA-γ2, but not PfTFIIA-γ1, is able to interact with PfTBP to form a ternary PfTBP/PfTFIIA-γ2/DNA complex (Fig. 24). Importantly, the results also demonstrate that PfTFIIA-γ2 interactions stimulate PfTBP DNA-binding activity.

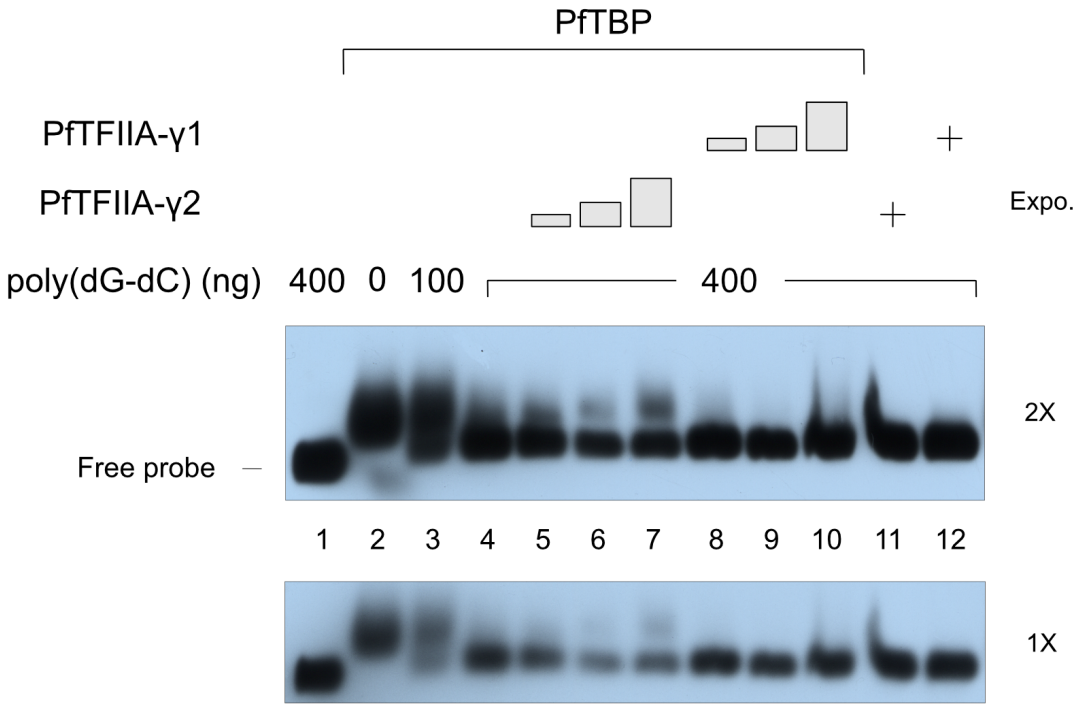


Figure 25: PfTFIIA-γ2, but not PfTFIIA-γ1, stimulates PfTBP DNA-binding activity. Binding reactions (10 µl) contained 12ng 6H-PfTBP and 5 *f*mol 282bp GBP-130 biotinylated DNA probe. 6H-PfTBP binding was suppressed by the addition of 400ng poly(dG-dC) followed by the addition of 10ng, 20ng, and 40ng PfTFIIA-γ1, or 8ng, 16ng, and 32ng PfTFIIA-γ2 as indicated. Reactions were incubated on ice for 45min in standard transcription buffer (Ch2. Materials & Methods 2.5). Protein-DNA complexes were separated from free probe on a 1.4% Mg^{2+} agarose gel, transferred to a positively charged nylon membrane, and visualised by chemiluminescence on an X-ray film. The results of two different exposure times (Expo. 1x, 2x) are shown.

3.7.3 PfTLP dependent recruitment of PfTFIIA to putative GBP-130 promoter DNA

The possibility of *P. falciparum* having two distinct PfTFIIA complexes led to the interesting question as to whether the two different putative PfTFIIA orthologs may specifically interact with either PfTBP or PfTLP. The demonstrated ability of PfTLP to stably bind putative *P. falciparum* promoter DNA in both EMSAs and ITAs (Figs. 19 and 20, Section 3.5.3; Bing, 2015) provided a means by which to test for, and characterise, PfTLP-PfTFIIA interaction. ITAs were carried out as described with PfTBP, however, these reactions contained only

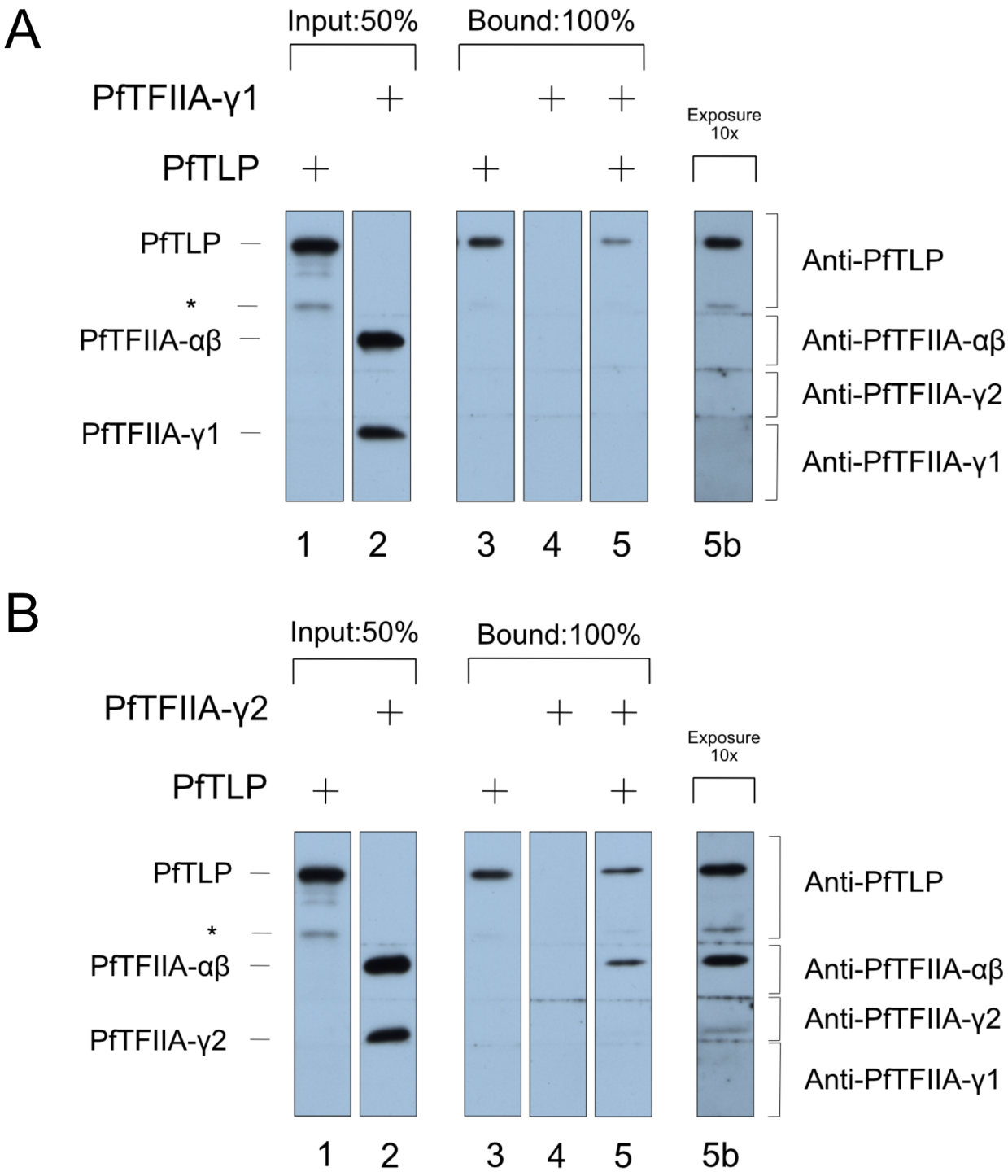


Figure 26: PfTLP-dependent recruitment of PfTFIIA-γ2 to GBP-130 promoter DNA. Binding reactions (50µl) contained 300ng 6H-PfTLP, 195ng 6H-PfTFIIA-γ1 (panel A) or 200ng 6H-PfTFIIA-γ2 (panel B) as indicated in the presence of 1 pmol 422bp immobilised GBP-130 DNA-probe and 100ng poly(dG-dC) competitor DNA. Protein-DNA binding reactions were incubated at 30˚C for 45min in standard transcription buffer (Ch2. Material & Methods 2.6). Bound proteins were eluted in SDS-PAGE loading buffer, separated by SDS-PAGE, and analysed by immunoblotting. The protein-bound membrane was cut into strips and incubated with rabbit polyclonal anti-PfTLP, anti-PfTFIIA-αβ, and anti-PfTFIIA-γ1s or anti-PfTFIIA-γ2s antibody as indicated. The position of a possible PfTLP breakdown product is marked with an asterisk.

100ng poly(dG-dC), which is sufficient to suppress non-specific PfTLP and PfTFIIA DNA binding (Fig. 19, Section 3.5.3). Under these conditions PfTLP was able to bind to the immobilised template but neither of the PfTFIIA complexes were able to bind the promoter template in isolation (Fig. 26, lanes 3 and 4). As seen with PfTBP, addition of PfTFIIA-γ1 complex resulted in a diminished PfTLP DNA binding, but no PfTLP dependent recruitment of the PfTFIIA-γ1 complex was detected. However, as seen with PfTBP, the PfTLP dependent recruitment of PfTFIIA-γ2 was detected (Fig. 26B, lanes 5 and 5b). Overall, these results are very similar to what was seen in ITAs using the PfTBP ortholog (Fig. 24), with the overall decrease of PfTLP and PfTBP DNA-binding in the presence of PfTFIIA-γ1 suggesting a possible role for PfTFIIA-γ1 as a negative regulator of PfTBP and PfTLP DNA-binding activity.

3.7.4 PfTFIIA-γ2, but not PfTFIIA-γ1, stimulates PfTLP-DNA complex formation

In order to further investigate the effect of PfTFIIA on PfTLP DNA-binding, Mg^{2+} agarose EMSAs were carried out. A representative result is shown in Fig. 27. Again, the stimulatory effect on PfTLP DNA-binding was investigated under conditions where PfTLP DNA-binding was suppressed by the addition of poly(dG-dC) competitor DNA to the binding reactions. Binding reactions containing PfTLP and 100ng poly(dG-dC) showed little evidence of a PfTLP-mediated mobility shift (Fig. 27, compare lanes 1 and 2). The addition of the PfTFIIA-γ2 to the binding reaction resulted in a quantitative shift of the KAHRP probe (Fig. 27, compare lanes 2 and 3), whereas the addition of PfTFIIA-γ1 to binding reactions had no noticeable effect on PfTLP-DNA complex formation (Fig. 27, lane 4). Interestingly, conducting these experiments in the absence of competitor DNA demonstrated a change in the mobility of the PfTFIIA-PfTLP-DNA complexes relative to the mobility shift seen with PfTLP alone. Binding reactions containing saturating amounts of PfTLP resulted in

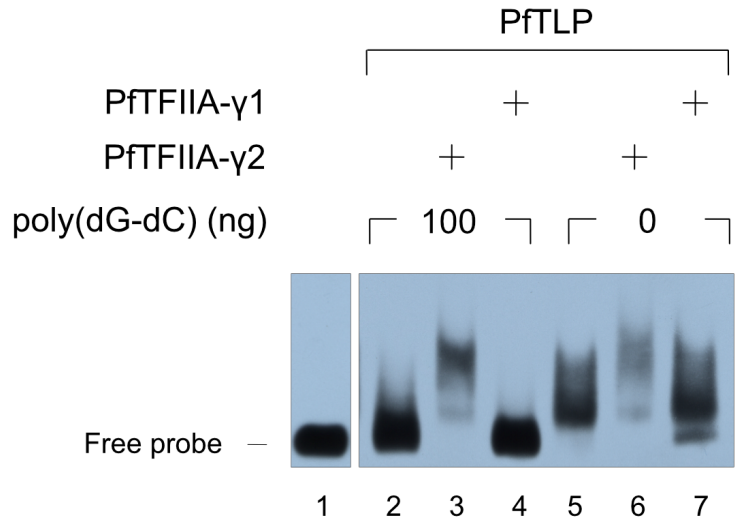


Figure 27: PfTFIIA-γ2, but not PfTFIIA-γ1, stimulates PfTLP DNA-binding activity. Binding reactions (10µl) contained 80ng 6H-PfTLP and 5 *f*mol 282bp putative KAHRP biotinylated DNA-probes in the presence of either 0ng or 100ng poly(dG-dC) competitor DNA. 16ng 6H-PfTFIIA-γ2 or 80ng 6H-PfTFIIA-γ1 were added to binding reactions as indicated. Reactions were incubated on ice for 45min in standard transcription buffer (Ch2. Materials & Methods 2.5). Protein-DNA complexes were separated from free probe on a 1.4% Mg^{2+} agarose gel, transferred to a positively charged nylon membrane, and visualised by chemiluminescence on an X-ray film.

---

quantitative binding of the free probe (Fig. 27, compare lanes 1 and 5). The addition of PfTFIIA-γ2 resulted in a further shift of the bound DNA-probe to a higher position which corresponds to the PfTFIIA-γ2 mediated shift seen in the presence of the competitor DNA (Fig. 27, compare lanes 5 and 6, with lane 3). This result further corroborates the conclusion that PfTFIIA-γ2 forms a stable ternary complex with PfTLP on DNA. Overall, these data clearly demonstrate that PfTFIIA-γ2 binds to PfTLP-DNA complexes and stimulates DNA-binding by PfTLP.

3.7.5 Effect of reaction temperature on stimulation of PfTLP binding of PfTFIIA-γ2

In vitro PfTBP DNA-binding was found to strongly favour low (0 to 4˚C) reaction temperatures (see Fig. 19, Section 3.5.2; Talvik, 2016). In contrast, DNA binding by PfTLP was found to be enhanced at 30˚C compared to 0˚C (Fig. 28). Interestingly, the addition of PfTFIIA-γ2 resulted in an increase in PfTLP DNA-binding at 0˚C (Fig. 28, compare lanes 3 and 6), but not at 30˚C, at which the formation of PfTLP DNA complexes is greatly stimulated (Fig. 28, compare lanes 3 and 7, and 7 and 10). Based on what is known about TFIIA function in the context of eukaryotic TBP-TATA-box binding energetics (Ch1. Sections 1.2.3 to 1.2.7) (Hoops *et al*., 1992; Starr *et al*., 1995; Nikolov *et al*., 1996; Tan *et al*., 1996; Masters *et al*, 2003; Hieb *et al*., 2007; Delgadillo *et al*., 2009), the increased stimulatory effects of PfTFIIA-γ2 on PfTLP DNA-binding at low temperature can most likely be attributed to thermodynamic considerations. PfTFIIA-γ2 may contribute to the binding energy required to form the IIA-TLP-DNA nucleoprotein complex, either by inducing a conformational change when complexed with PfTLP, which in turn leads to an increase in favourable PfTLP-DNA interactions, or else through direct interactions with the bound DNA in the context of the IIA-PfTLP-DNA complex.

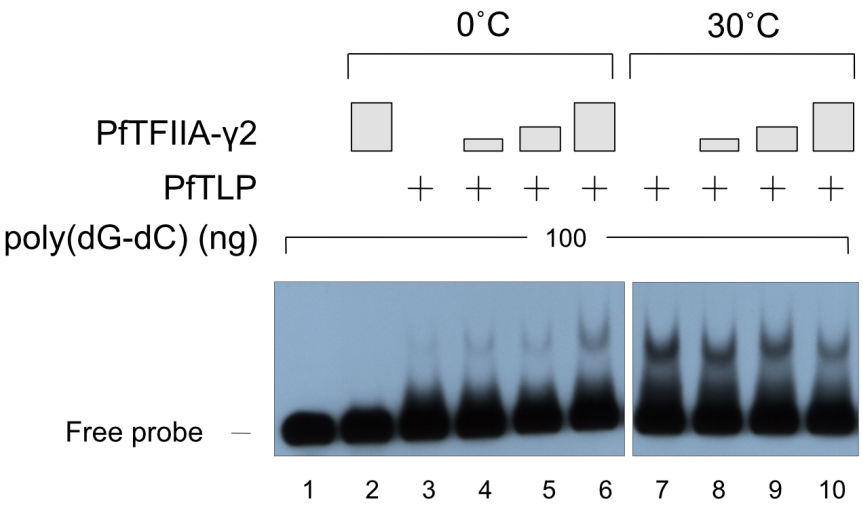


Figure 28: Effect of reaction temperature on PfTFIIA stimulation of PfTLP DNA binding activity: Binding reactions (10µl) contained 80ng 6H-PfTLP, 5 *f*mol 282bp GBP-130 biotinylated DNA-probes, 100ng poly(dG-dC) competitor DNA, and increasing amounts of 6H-PfTFIIA-γ2 (20ng, 40ng, and 80ng) as indicated. Reactions were incubated on ice or at 30˚C for 45min in standard transcription buffer (Ch2. Materials & Methods 2.5). Protein-DNA complexes were separated from free-probe on a 1.5% Mg^{2+} agarose gel, transferred to a positively charged nylon membrane, and visualised by chemiluminescence on X-ray film.

Chapter 4

Discussion

4.1 Bioinformatic analysis of the *Plasmodium falciparum* TATA box binding protein (PfTBP).

The original identification and analysis of the gene coding for *P. falciparum* TBP (McAndrew *et al*., 1993) had followed on from the publication of similar studies in *S. cerevisiae* (Horikoshi *et al*., 1989), *S. pombe* (Fikes *et al*., 1990), *A. thaliana* (Gasch *et al*., 1990), *D. melanogaster* (Hoey *et al*., 1990), and *H. sapiens* (Hoffman *et al*., 1990). By making use of highly conserved motifs in the C-terminal core domain of the known TBP orthologs, together with the knowledge of *P. falciparum* codon usage (Hyde *et al*., 1989), McAndrew and colleagues were able to sequentially amplify sections of the PfTBP open reading frame (ORF) and thereby construct a consensus sequence for the full length gene. By using a combination of multiple sequence alignments and hydrophobicity plots, the group demonstrated that, relative to other eukaryotes, PfTBP had a highly divergent primary amino acid sequence but still likely maintained a conserved secondary and tertiary structure profile. McAndrew also pointed out that many of the residues that were at that time known to be important for TBP-DNA interaction (Reddy & Hahn, 1991) had also been conserved in PfTBP and that the spatial orientation of some of these residues had been altered by a small insertion of several amino acids in the *P. falciparum* ortholog. McAndrew postulated that these changes may reflect a loss of reliance on the TATA box element by *P. falciparum* and that this was possibly necessitated by the extremely high A/T base content of the parasites intergenic sequences. Finally, the author went on to state that any

abnormalities seen in the transcriptional mechanisms of *P. falciparum* may offer insight into novel stratagems for the development of anti-malarial chemotherapeutics. These data and predictions from the McAndrew group were used as a starting point in this thesis for a more in-depth bioinformatics analysis of PfTBP that utilised the now known crystallographic structures of the human and yeast TBP orthologs in complex with the TATA box (Nikolov *et al*., 1996; Tan *et al*., 1996; Bleichenbacher *et al*., 2003), as well as the available *P. falciparum* (3D7) genome sequence (Gardner *et al*., 2002). Initial examination of the annotated PfTBP gene in the PlasmoDB database (PF3D7_0506200) showed that the ORF actually extended upstream of the ATG start codon originally predicted by McAndrew *et al.,* 1993. The current annotation for the full-length PfTBP-gene codes for a 38.1 kDa protein composed of 327 amino acids. At present there appears to be some confusion in the literature regarding which ATG is actually used as the start codon for PfTBP translation and the actual size of the full-length PfTBP protein. The reasons for this become quite clear when carefully analysing published western-blot data (Ruvalcaba-Salazar *et al*., 2005, 2006; Gopalakrishnan *et al*., 2009). In a 2005 paper entitled "Recombinant and native *Plasmodium falciparum* TATA-binding protein binds to a specific TATA box element in promoter regions", Ruvalcaba-Salazar and colleagues expressed a 26kDa PfTBP core domain protein fused to a GST-tag (total MW of 52kDa) for which they produced a polyclonal anti-GST-PfTBP antibody that is specific to the PfTBP portion of the fusion protein (Ruvalcaba-Salazar *et al*., 2005). The authors then go on to present western-blot data using this anti-PfTBP antibody to probe *P. falciparum* nuclear extract (PfNE) (Ruvalcaba-Salazar *et al*., 2005; Fig.1) and to show the specific detection of a protein band in PfNE migrating close to the 26kDa protein molecular weight marker. This result is clearly contrary to the expected size of 38.1kDa of PfTBP predicted in the most recent iteration of PlasmoDB (38.1kDa) (Gene entry PF3D7_0506200; PlasmoDB Release 29). A second set of western-blot data is then presented (Ruvalcaba-Salazar *et al*., 2005;

Fig. 8) showing that a polyclonal antibody raised against human TBP (anti-HsTBP) is also able to specifically detect PfTBP in parasite nuclear extract, in order to justify the use of anti-HsTBP in chromatin-immunoprecipitation (ChIP) experiments. No explanation was given as to why the author chose not to use their anti-PfTBP antibody in these experiments. It seems likely that the anti-PfTBP antibody turned out not to be suitable for ChIP experiments. Shortly after the publication of this initial paper, the author released a second paper entitled “Preparation and Characterization of a Monoclonal Antibody Specific to Plasmodium falciparum TATA Binding Protein”, (Ruvalcaba-Salazar *et al*., 2006). Here the authors clearly re-use the image of the western-blot probed with the anti-HsTBP antibody in the 2005 paper (Ruvalcaba-Salazar *et al*., 2005; Fig. 8), and present this data as a western blot probed with their new monoclonal anti-PfTBP antibody (Ruvalcaba-Salazar *et al*., 2006; Fig. 2). In addition, western-blot data published by Gopalakrishnan and co-workers in 2009 contradict the Ruvalcaba-Salazar western-blot data. Here, the authors probe *P. falciparum* nuclear extract with a custom-made polyclonal anti-PfTBP antibody (Washington Biotechnology). This experiment detected a specific interaction with a protein band that migrated between the 32kDa and 47kDa protein molecular weight markers, consistent with the calculated molecular weight of PfTBP (38.1kDa) in the current PlasmoDB release. It should be noted that Gopalakrishnan *et al*. claimed that the full-length PfTBP band seen in their western-blot data migrated corresponding to a molecular mass of 42kDa, a value stated to be in accordance with the calculated mass of PfTBP in the genome annotation at the time of their publication. However, it appears to be difficult to distinguish between these two values (38.1kDa and 42kDa) in their data figure based on the migration of the molecular weight markers (Gopalakrishnan *et al*., 2009, Fig. 1A). Added to this point, Gopalakrishnan *et al*. stripped and re-probed the same immunoblot membrane with anti-HsTBP, and the resulting PfTBP band is then shown in Fig. 1B to migrate alongside HsTBP (calculated MW 37.6 kDa). The anti-HsTBP antibody used here

was a polyclonal antibody able to detect TBP orthologs from multiple species; likely due to the highly conserved nature of TBP C-terminus secondary structure (https://www.scbt.com/scbt/product/tfiid-antibody-si-1). It may be possible that the PfTBP identified in the nuclear extracts of the Ruvalcaba-Salazar papers was a distinct PfTBP isoform produced from the ATG transcription start site originally proposed by McAndrew, 1993. However, in light of the apparent duplication of images in the two papers published by Ruvalcaba-Salazar, as well as the contradictory findings of Gopalakrishnan *et al*., 2009, there is some doubt as to the veracity of these western-blot data. The conclusion here is that the 38.1 kDa annotation, confirmed by Gopalakrishnan *et al*., 2009, is likely the correct molecular weight of full-length PfTBP.

The alignment and subsequent annotation of the TBP-core sequences in Fig. 9 (Section 3.1) confirmed the extremely low degree of primary sequence identity in the C-terminal domain of PfTBP (39%) relative to the human ortholog, a value much lower than that seen when comparing the HsTBP sequence to the *S. cerevisiae* ortholog (79%). However, Fig. 9 illustrates the overall conservation of hydrophobic positions between the three orthologs, a feature implying conserved secondary structure motifs (Gaboriaud *et al*., 1987; MacAndrew *et al*., 1993). Furthermore, a detailed examination of the residues involved in TBP-DNA interaction in the yeast TFIIA-TBP-DNA crystal structure (Tan *et al*., 1996), shows that 27 of 29 (93%) residues that are known to contact the DNA minor groove are either conserved or substituted with a highly similar amino acid. However, a key point to remember when considering the DNA-binding characteristics of PfTBP, is that the substitution of F197 in HsTBP (F99 in ScTBP) for isoleucine I183 in PfTBP is likely to result in a change in the the topology of the PfTBP-DNA complex through a loss of DNA kinking at the last base step in the TATA box motif. This is due to the elimination of the DNA-helix distortion brought about by the intercalation of this phenylalanine residue

(Fig. 4) (Kim *et al*., 1993; Nikolov *et al*., 1996). Additionally, the relative spatial positioning of residues involved in TBP-DNA interaction have been altered by a small insertion between β-strands S3' and S4' of PfTBP. In summary, the substitution of F197 together with the insertion of amino acids TSN at positions 285-297 of PfTBP predict a loss of TATA-specific PfTBP DNA binding activity as well as a PfTBP-DNA nucleoprotein architecture that is distinct from that which is seen in other eukaryotic systems. Examination of the known TFIIA interaction sites in the human and yeast TBP orthologs shows that 8 of 10 (80%) of these sites are conserved, suggesting a retention in the ability of PfTBP to interact with TFIIA. Lastly, the conserved proline P163 in HsTBP P163 (P65 in ScTBP), shown to be essential for RNAP-II transcription in yeast (Schultz *et al*., 1992), and also found to be conserved in TBP orthologs from *A. thaliana,* and *D. malanogaster* (Hoffman *et al*., 1992), has been substituted to leucine L149 in PfTBP (Fig. 9). These data, together with the absence of nearly all the histone-fold containing TAFs in the parasites genome sequence (Bischoff *et al*., 2010), make a strong case for a functionally divergent *P. falciparum* TFIID complex, which can thus be considered as a highly interesting target for the development of potent anti-malarial drugs aimed at lethality across all stages of the parasite life cycle.

4.2 Bioinformatic analysis of the putative *P. falciparum* transcription factor IIA (PfTFIIA) orthologs.

Owing to the extremely high A/T base content of the parasite genome, together with the resultant divergence in amino acid profiles of *P. falciparum* proteins, the putative PfTFIIA subunits were among the last of the *P. falciparum* general transcription factor orthologs to be identified (Callebaut *et al*., 2005). Interestingly, these *in silico* studies predicted two possible variants of the TFIIA-γ subunit but only one TFIIA-αβ subunit. This led to the

possibility of two distinct PfTFIIA complexes, each containing one of two γ subunit variants (denoted: PfTFIIA-γ1s and PfTFIIA-γ2s) complexed with the common αβ ortholog (PfTFIIA-αβ). In this study, multiple-sequence alignments were used to analyse the predicted secondary structure motifs (Callebaut *et al*., 2005) as well as the levels of conservation of TBP and DNA interaction sites (Bleichenbacher *et al*., 2003) within each of the putative subunits (Figs. 10 and 11). Examination of the primary amino acid sequences and, in particular, the clustering of hydrophobic residues, indicated the likely conservation of secondary structure profiles with respect to the human and yeast orthologs (Figs. 10 and 11) (Callebaut et al., 2005). It should however be noted that the spatial arrangement of the predicted β-strand secondary structures within the two γ subunit variants are altered with respect to one another, and as such, would likely form dissimilar three-dimensional structures when complexed with αβ. This is an important point when considering the relation of three-dimensional protein structure to the specificity of protein-DNA interactions (Fig. 11B; Murphy & Churchill, 2000; Garvie & Wolberger, 2001; Slattery *et al*., 2014). With regard to PfTFIIA-αβ, it was also found that there are substantial alterations to the main points of contact seen between HsTFIIA and DNA in the TFIIA-β chain of the human TFIIA crystal structure (Fig. 10) (Bleichenbacher *et al*., 2003). These changes to the common αβ subunit potentially point to a loss in the ability of the two PfTFIIA complexes to stabilise interactions between TBP and the TATA box element (Bleichenbacher *et al*., 2003; Hieb *et al*., 2007). Furthermore, there is almost no conservation of the residues involved in TBP interaction on either of the PfTFIIA-γ subunits, the γ subunit being the primary site of TBP interaction in the human and yeast orthologs (Bleichenbacher *et al*., 2003). This is a very peculiar feature of PfTFIIA as most (80%) of the TFIIA interaction sites on PfTBP are conserved (Fig. 9) (Fig. 11A). However, when considering the overlaps seen in the interaction sites of TFIIA with that of other transcription associated factors such as TAF1, NC2, and BTAF1 (Ch1. Introduction, Section 1.2.5), it is entirely possible that these

regions are conserved within PfTBP for reasons other than TFIIA interaction, and that the PfTFIIA interactions seen with PfTBP in the mobility shift and immobilised template data (Ch3. Results, Section 3.7; discussed in Section 4.7) is in fact brought about in a different way.

Interestingly, RNA-seq analysis of erythrocytic stage steady state mRNA levels for PfTFIIA-γ1s (PF3D7_1250700) and PfTFIIA-γ2s (PF3D7_0933700) (Appendix A7), shows different expression profiles for the two PfTFIIA-γ subunits. The mRNA levels of PfTFIIA-γ1s remain relatively constant in the early stages of erythrocyte infection, around 2-4 fold higher than mRNA levels of PfTFIIA-γ2s, and increase 1.5 to 2 fold 30-35 hours post-invasion. In contrast, PfTFIIA-γ2s mRNA levels increase 6 to 10-fold by hours 32-40, raising transcript levels to a number roughly equal to PfTFIIA-γ1s. Furthermore, the expression profiles of the antisense transcripts in these regions are such that antisense-PfTFIIA-γ1 transcripts are only produced in late blood stage while the antisense-PfTFIIA-γ2 transcripts mimic the PfTFIIA-γ2 sense strand profile. It is now widely accepted that antisense transcripts can play a significant regulatory role in the transcription, translation, and mRNA degradation pathways of sense strand transcripts (reviewed in Pelechano & Steinmetz, 2013). As such, the differences in antisense mRNA expression patterns for the two subunits may be taken as further evidence for the independent regulation and presumably distinct functional roles of the two PfTFIIA complexes analysed in this study. In addition, available transctiptomics data for male and female gametocytes, show that PfTFIIA-γ2s transcription is down-regulated in female gametocytes. Assuming similar mechanisms of direct PfTFIIA-DNA interaction as seen in the human and yeast systems, and considering the overall differences in PfTFIIA-γ1 and PfTFIIA-γ2 complex topologies brought about by the distinct PfTFIIA-γ subunit architectures, as well as the fluctuations in mRNA levels for both the sense and antisense transcripts coding for the two PfTFIIA-γ

candidates, it seems reasonable to postulate distinct developmentally linked activities for the two putative TFIIA complexes, presumably at distinct *P. falciparum* promoter subtypes. Indeed, this hypothesis is in line with the different abilities of the PfTFIIA-γ1 and PfTFIIA-γ2 complexes to stimulate PfTBP binding to GBP-130 promoter DNA in *in vitro* DNA binding experiments (discussed in Section 4.7). Lastly, in addition to differential stimulation of PfTBP or PfTLP DNA-binding activity, the two different PfTFIIA complexes could differentially affect activator or repressor function at different core promoters.

4.3 Bioinformatic analysis of *Plasmodium falciparum* TBP-like protein (PfTLP)

An adapted version of a larger COBALT sequence alignment originally carried out by Steven Bing (Bing, 2015) was used to illustrate the conservation of the known DNA and TFIIA interaction sites usually present in the conserved C-terminal domain of eukaryotic TBPs (Fig. 12; Tan *et al*., 1996; Bleichenbacher *et al*., 2003). This was included as a point of reference to aid discussion of the immobilised template and mobility shift data described in sections 3.5 and 3.7 of Chapter 3. PfTLP was among the first *P. falciparum* general transcription factors identified (Coulson et al., 2004) and is, to my knowledge, the only *Apicomplexan* TBP-like protein to have been analysed beyond the point of *in silico* prediction. As would be expected, the degree of divergence of PfTLP from typical eukaryotic TBP model orthologs is even greater than that seen in the already divergent PfTBP. In particular, the conservation of DNA-interaction sites drop from 93% in PfTBP to 53% in PfTLP. In addition, there are large insertions of unstructured low-complexity regions located in the S4 (131 amino acids) and S4' (31 amino acids) β-strand regions, as well as a shortened N-terminal region (Fig. 12). Furthermore, the degree of conservation of TFIIA interaction sites also drops significantly, from 80% in PfTBP to 20% in PfTLP. Despite these substantial alterations to the primary amino acid sequence of the PfTLP paralog,

there is one distinct TBP characteristic that is conserved in PfTLP but not in PfTBP. All four phenylalanine residues that are vital for the TATA box-binding activity of TBP are conserved in PfTLP. This feature stands in contrast to what is seen in the human and *Drosophila* TLP orthologs where three of these four residues have been substituted for either tryptophan, histidine, asparagine, or valine (Duttke *et al*., 2016). Taken together, these observations are somewhat paradoxical in that they both seemingly support and contradict the potential of PfTLP having TATA box-binding capabilities. The important point to note is that whatever the effect of the conservation and/or substitution of the various residues in PfTLP shown to interact with DNA in the canonical TBP core domain is, the data overall suggest that PfTBP and PfTLP have distinct DNA-binding specificity, and as such, support the notion that PfTLP potentially plays a role in specific developmentally regulated gene transcriptional programs as postulated for other eukaryotic TLPs (Duttke *et al*., 2016)

4.4 Expression and purification of putative PfTFIIA-γ1 and PfTFIIA-γ2 complexes

The heterologous expression of soluble *Plasmodium* proteins in *E.coli* is known to be an extremely challenging endeavour (reviewed in Birkholtz *et al*., 2008), with the extremely low success rate being a major hurdle to the commencement of structural and/or functional studies aimed at the development of much needed anti-malarial drugs (Aguiar *et al*., 2004; Mehlin *et al*., 2006). There are currently around 650 *P. falciparum* protein structures listed among the 130 000 macromolecular assemblies available in the Protein Data Bank (PDB) (http://www.rcsb.org/pdb/static.do?p=general_information/about_pdb/index.html). This is a surprisingly high number considering the total number of *P. falciparum* genes is approximately 5300, and reflects a sustained effort from both the public and private research sectors over the last decade despite the considerable difficulties experienced by

many earlier large-scale expression efforts (Aguiar *et al*., 2004; Mehlin *et al*., 2006; Vedadi *et al*., 2007). The challenges associated with *P. falciparum* protein expression in *E.coli* stem mainly from cloning artefacts and solubility problems brought about by a number of variables. The extremely A/T-rich genome sequence can severely compromise basic cloning procedures such as restriction digestion, ligation reactions, and bacterial transformations, resulting in the production of re-arranged recombinant clones (unpublished observations; Triglia & Kemp, 1991). If viable clones can be obtained, protein expression is often compromised by diminished cell growth and cell death due to protein toxicity (Cinquin *et al*., 2001; Bing, 2015). A large-scale study found correlations between expression problems and the relatively high molecular mass of *P. falciparum* orthologs which are typically 50% larger than their yeast ortholog (Aravind *et al*., 2003; Mehlin *et al*., 2006), as well as with the abundance of low-complexity disordered regions in the *P. falciparum* proteome (Pizzi & Frontali, 2001). In addition, proteins which lack homology to *E.coli* proteins, and, in particular, those which exhibit a very basic pI, are strongly correlated with solubility problems, whereas a very low pI typically results in no observable expression at all (Mehlin *et al*., 2006). Lastly, size exclusion chromatography experiments demonstrated that a significant number (51%) of soluble expressed *P. falciparum* proteins tend to form aggregates through self-association. It is important to note that a number of eukaryotic expression systems are available that have been successfully used in the heterologous expression of some *Plasmodial* proteins (Birkholtz *et al*., 2008). These include yeast, baculovirus, and mammalian cell systems. Ultimately however, low levels protein glycosylation in *P. falciparum* (Gowda & Davidson, 1999), and the time and costing advantages afforded by the *E.coli* system, led to the decision to endure the inherent difficulties of this approach.

In this study, two putative untagged TFIIA-γ subunits, PfTFIIA-γ1s and PfTFIIA-γ2s, were each co-expressed with a 6His-tagged variant of the putative PfTFIIA-αβ subunit in order to overcome the solubility problems experienced when expressing the proteins in isolation. Furthermore, toxicity problems brought about by over-expression of these proteins were mitigated by eliminating IPTG induction, relying instead on the leaky expression of the LacUV5 promoter together with a lowered incubation temperature (25˚C as opposed to 37˚C) to slow down cellular processes. The amount of soluble PfTFIIA proteins obtained under these conditions was very low, typically 5-10μg/L cell culture, but sufficient to conduct functional studies. In light of the toxicity problems experienced, using a cell-free expression system (Shimizu *et al*., 2006; Mudeppa *et al*., 2007; Tsuboi *et al*., 2008) may be a possible alternative strategy if one were to pursue structural studies of these orthologs, which require much higher amounts of recombinant protein.

Purification of PfTFIIA complexes was achieved through a two step process starting with a Ni-NTA affinity chromatography step at high salt (500mM KCl cleared lysate) in order to minimise any masking of the 6His-tag by protein aggregation. A subsequent buffer exchange to 100mM KCl was followed by ion exchange chromatography using Q-sepharose anion exchange resin for PfTFIIA-γ1 (pI 5.38) or SP-sepharose cation exchange resin for PfTFIIA-γ2 (pI 7.69), with the two complexes eluting between 400mM-500mM KCl and 400mM-700mM KCl, respectively. The purity of the resultant preparation was deemed sufficient for subsequent assays. Notably, based on the tightly intertwined structure of human and yeast TFIIA structures (Fig. 5) (Bleichenbacher *et al*., 2003), the stable association of the two untagged PfTFIIA-γ subunits with 6H-PfTFIIA-αβ, as evidenced by co-purification over two separate purification steps, provides strong evidence that both proteins are legitimate PfTFIIA γ-subunits, and as such, provide convincing evidence for the existence of two distinct TFIIA complexes in *P. falciparum*.

4.5 Initial characterisation of PfTBP & PfTLP

Electrophoretic mobility shift assays (EMSAs) were carried out in order to characterise the DNA-binding activity of PfTBP and PfTLP and to test predictions made in the bioinformatics analysis of these key plasmodial transcription initiation factors (Ch3, Results, Section 3.5), and to establish systems in which to functionally characterise the two putative PfTFIIA complexes. In protein-DNA binding reactions, three modes of DNA-protein interactions must be considered: specific, non-specific, and unspecific interactions. Specific binding occurs when a protein recognises and binds to a specific DNA sequence. Non-specific binding occurs when that same protein binds a DNA sequence that is different from its ideal binding sequence. Lastly, unspecific binding refers to protein-DNA interaction with no inherent sequence preference (reviewed in Murphy & Churchill, 2000; Garvie & Wolberger, 2001). Structural biology has long provided key insights into the molecular mechanisms by which proteins interact with specific DNA sequences through the analysis of hundreds of co-crystal structures of DNA-bound protein complexes. Two main mechanisms of DNA sequence recognition can be distinguished, termed base and shape readout (reviewed in Slattery *et al*., 2014). Base readout broadly refers to direct interactions between a proteins amino acid side chains and the chemical signatures of specific DNA base pairs in the major and, to a much lesser extent, minor grooves of the DNA molecule. These interactions are comprised of direct and indirect (water mediated) H-bonds as well as van der Waals interactions. Shape readout refers to the recognition of sequence-specific structural features at a given target site. These include features such as the propensity for DNA-bending, the formation of kinks, and the unwinding of the double-helix, which can be brought about by local nucleotide composition (Slattery *et al*., 2014; Hieb *et al*., 2007). Examples of both base and shape readout can be seen in the co-crystal

structures of TBP-core domains bound to the TATA box promoter element (Kim & Burley., 1994; Tan *et al*., 1996; Nikolov *et al*., 1996; Bleichenbacher *et al*., 2003).

TBP recognises and binds to the TATA box through the minor grove of the DNA double helix by an induced-fit mechanism (reviewed in Nikolov & Burley, 1997). Intercalation of phenylalanine residues into the first and last base steps of the 8bp recognition sequence causes two kinks at the 5' and 3' ends of the TATA box resulting in a partial unwinding of the DNA-helix between these two points. This unwinding causes a widening of the minor grove and facilitates multiple H-bond and hydrophobic interactions between TBP and the TATA box nucleotides (Nikolov *et al*., 1996). The alterations to the helix structure and trajectory at the point of binding ultimately cause a significant bend in the DNA-helix that can average in excess of 100˚ (Nikolov *et al*., 1996; Hieb *et al*., 2007). The TBP-DNA interactions can be further stabilised by accessory proteins, such as TFIIA and TFIIB. TFIIA interaction is mediated by a number of protein-protein contacts at the N-terminal stirrup of the TBP core domain as well as direct contacts with DNA base pairs upstream of the TATA box (Bleichenbacher *et al*., 2003). These interactions result in a conformational change in the TBP-TATA complex reducing the bend angle to 80˚, thereby significantly increasing the kinetic stability of the complex under physiological conditions (Hieb *et al*., 2017).

The key findings regarding the DNA binding activity of PfTBP from the mobility shift assays presented in chapter 3 were that PfTBP binding of putative GBP-130 promoter DNA is possibly sequence-specific. However, as predicted in the sequence analysis data of Section 3.1, which predicted a loss in TATA-specific DNA binding activity due the substitution of F197 (HsTBP) for isoleucine I183 (PfTBP), as well as the spatial alterations

observed among key amino acid residues involved in TBP-DNA interaction, PfTBP does not appear to recognise and bind to a consensus TATA box element.

Consistent with the lack of conventional TATA box-binding activity, the production of a PfTBP-DNA complex with high mobility relative to other eukaryotic TBP-DNA complexes in Mg^{2+} agarose EMSAs, can be taken as evidence for a lack of (or a much reduced) bending angle in the PfTBP-DNA nucleoprotein complex. This notion is further supported by the observation that PfTBP-DNA complex formation and stability is greatly enhanced at low incubation temperatures. This temperature preference for DNA binding is incompatible with the 30˚C preference seen in other eukaryotic TBPs, which require a higher energy system to facilitate DNA-bending through the intercalation of phenylalanine residues into the TATA box promoter element (Kim *et al*., 1993; Nikolov *et al*., 1996).

Similar to PfTBP, PfTLP was also shown to exhibit possible sequence-specific DNA binding activity. However, as with TBP, this sequence-specific binding does not appear to be mediated by a consensus TATA box element despite the conservation of all four of the phenylalanine residues typically associated with TATA-binding activity (Thomas & Chiang, 2006; Duttke *et al*., 2016). It is also important to note that, although both PfTBP and PfTLP exhibited binding activity when using probes generated from the same putative *P. falciparum* promoter sequence (GBP-130), there are clear quantitative differences in these binding activities, with PfTBP-DNA complexes displaying a much higher stability when challenged with non-specific poly(dG-dC) competitor sequences.

Overall, the demonstration of PfTBP and PfTLP DNA-binding activity, together with the observation that both of these transcription factors are able to form stable complexes with PfTFIIA-γ2, suggest that both PfTBP and PfTLP are involved in promoter recognition. More

work needs to be done to describe in detail the precise differences in PfTBP and PfTLP DNA sequence specificity and to determine which of these is the pre-dominant factor involved in nucleating the RNAP-II pre-initiation complex.

4.6 Characterisation of PfTBP sequence specificity by Systematic Evolution of Ligands by Exponential Enrichment (SELEX)

Motivated by our observation that PfTBP DNA-binding is sequence specific, but that sequences distinct from the typical TATA box core promoter element (CPE) are required, a PCR-based method for the enrichment of short DNA sequences from a pool of randomly generated oligonucleotides (SELEX) was used to try elucidate any putative CPEs recognised by PfTBP. After seven rounds of selection, no significant motifs were detected. Rather, there was a marked increase in the A/T content of the nucleotide library. This may indicate that sequence specificity seen in our Mg^{2+} EMSA data purely relates to the high A/T content of the GBP-130 probe (>80%) relative to that of the TdT probes (<50%) as opposed to any defined CPE. This notion is further supported by the lack of PfTBP DNA-binding seen using DNA-probes containing a consensus TATA box element (Ch3. Results, Section 4.5). Significantly, these results are in agreement with a recent seminal study analysing *P. falciparum* transcription start sites by way of 5' cap RNA sequencing (Adjalley *et al*., 2016). Here the authors assembled a comprehensive map of RNAP-II transcription initiation sites giving new and much needed insight into the parasites transcriptome and core promoter sequence context. Examination of the nucleotide content around the most active transcription start sites indicated that the preference for parasite transcription initiation sites is governed by a locally increased A/T nucleotide base composition. It may indeed however be possible that the the limitations of this assay, brought about by the experimental artefacts that begin to emerge after five rounds of SELEX, have prevented

the elucidation of a motif which may have only arisen in successive rounds. A possible solution to this would be to start the SELEX procedure with a random library that contains an A/T base content that is more representative of what is found in the parasites intergenic regions, i.e > 80% A/T. This would remove any “wasted rounds” filtering out G/C-rich sequences from the pool and thereby minimise the total number of rounds needed. Alternatively, it may be that TFIIA and TFIIB are required in the binding reactions in order to observe any real sequence specificity in the PfTBP binding events.

4.7 Initial characterization of two putative PfTFIIA orthologs

In order to test for the interaction with, and effects of, the two putative TFIIA complexes with respect to PfTBP and PfTLP DNA-binding, a series of immobilised template assays (ITAs) and electrophoretic mobility shift assays (EMSAs) were carried out. In both instances, it was found by ITA that the second of the two orthologs, PfTFIIA-γ2, was able to be stably recruited to putative GBP-130 promoter DNA in a PfTBP- and PfTLP-dependent manner. It was also noted that overall DNA-binding of PfTBP and PfTLP was decreased in the presence of PfTFIIA-γ1. This may well indicate a function of PfTFIIA-γ1 in reducing non-specific PfTBP and PfTLP DNA interactions. Consistent with these results, it was demonstrated by EMSA that the presence of PfTFIIA-γ2 was able to help both PfTBP and PfTLP to overcome DNA binding suppression with non-specific poly(dG-dC) competitor, presumably by promoting specific interactions with promoter DNA over non-specific interactions with the competitor DNA. Despite a lack of detectable PfTFIIA-γ1 interaction with both PfTBP and PfTLP in these assays, it would be unwise to discount this complex as a valid part of the *P. falciparum* general transcription machinery. It may well be necessary that additional transcription factors, both general and specific, be present for PfTFIIA-γ1 incorporation into the pre-initiation complex. In particular, PfTFIIB has been

shown through bioinformatics analysis to contain conserved TFIIA interaction sites (Bing, 2015) and could likely play a role in PfTFIIA-γ1 association with TBP/TFIID. It is also important to note that nuclear TBP is complexed with multiple TBP-associated factors (TAFs), some of which could play a key role in PfTFIIA-γ1 recruitment. Recent studies have demonstrated that in mouse testis, TFIIA forms part of a large 280kDa complex containing both TLP and TFIIA-like factor (ALF) together with a heat-shock protein chaperone. The studies suggest, that TLP (TRF2), together with the alternative TFIIA complex, TFIIA-like factor (ALF), and TFIIA, from a type of "TFIIA super complex" in haploid spermatid cells that is recruited to the RNAP-II PIC (Martianov *et al*., 2016). Whether PfTBP or PfTLP can interact with both PfTFIIA-γ1 and PfTFIIA-γ2 simultaneously is a interesting possibility that still needs to be tested.

Finally, a paper published at the time of writing this thesis reported the identification of a novel core promoter element, the TFIIA recognition element (IIARE), located upstream of the TATA box, which is recognised by TFIIA (Wang *et al*., 2017). The authors show that IIARE is able to enhance the activity of TATA-containing core promoters by stimulating the recruitment of TFIIA, RNAP-II, TAF4 and the coactivator p300. Furthermore, it was found that the presence of IIARE was also able to activate or repress transcription at TATA-less promoters. These findings support the interesting possibility that the two different PfTFIIA complexes characterised in this work may have important roles in directing PfTBP and PfTLP to distinct core promoter sequences.

4.8 Summary and Outlook

The data presented here provide strong evidence for the existence of two *P. falciparum* TFIIA complexes. With respect to PfTBP, it was found that PfTBP DNA-binding very likely

leads to “unbent” PfTBP-DNA nucleoprotein complexes, dissimilar from those seen in other eukaryotic systems. Furthermore, mobility shift and SELEX data indicate that DNA-binding activity, although sequence-specific, is not dependent on a TATA box element but rather on local nucleotide content. Similarly, PfTLP has also been shown here to have sequence specific DNA-binding activity that is distinct from that of PfTBP. Despite some structural indications of TATA box binding activity, PfTLP DNA-binding activity was also found not to be positively influenced by the presence of a TATA box element. Overall, this work provides a first important basis for more detailed studies on the *in vitro* DNA-binding activities of PfTBP and PfTLP together with two newly confirmed TFIIA complexes.

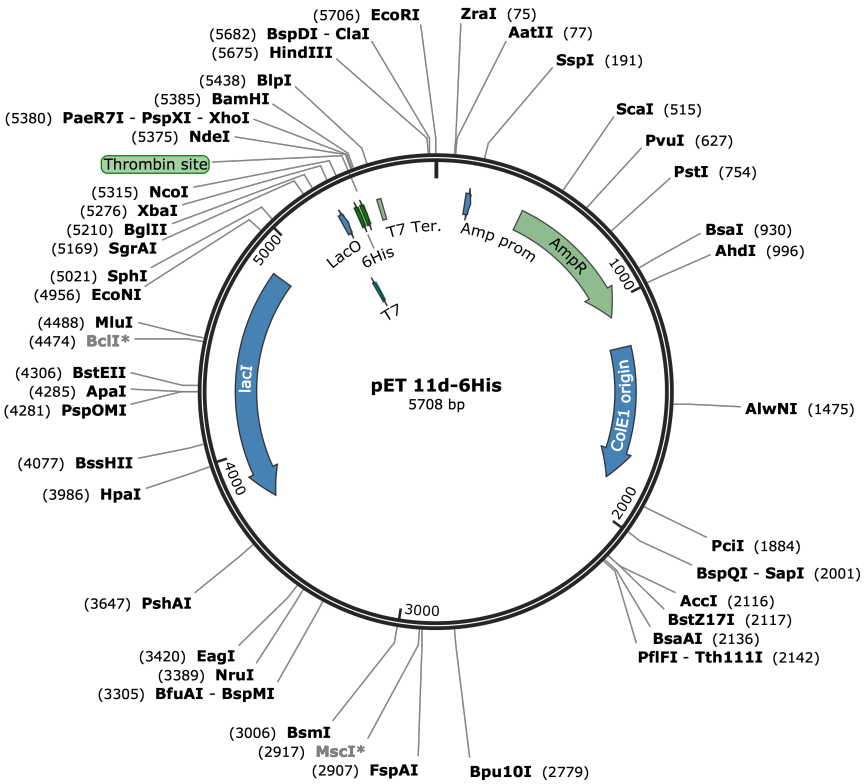
The demonstrated interactions between PfTFIIA-γ2 and both PfTBP and PfTLP may well provide a novel stratagem for antimalarial small molecule development. It is widely accepted that a large majority of the approximately 3000 druggable proteins in the human system function as complexes within various interaction networks and not as isolated effectors (Hopkins & Groom, 2002), a characteristic that is likely shared by *P. falciparum* (LaCount *et al*., 2005) and indeed all other metazoans. This prevailing paradigm has led to the research and development of a class of small organic molecules known as protein-protein interaction disruptors (PPIDs) (Fontaine *et al*., 2015). Based on the high levels of divergence seen in the bioinformatics analysis of PfTBP and the PfTFIIA orthologs with respect to their human counterparts, the key role of these proteins in parasite transcription, and the demonstrated interactions of PfTFIIA-γ2 with both PfTBP and PfTLP, it seems reasonable to suggest that a PPID targeting PfTFIIA-γ2 interactions with PfTBP/PfTLP may well be the source of an effective antimalarial drug treatment.

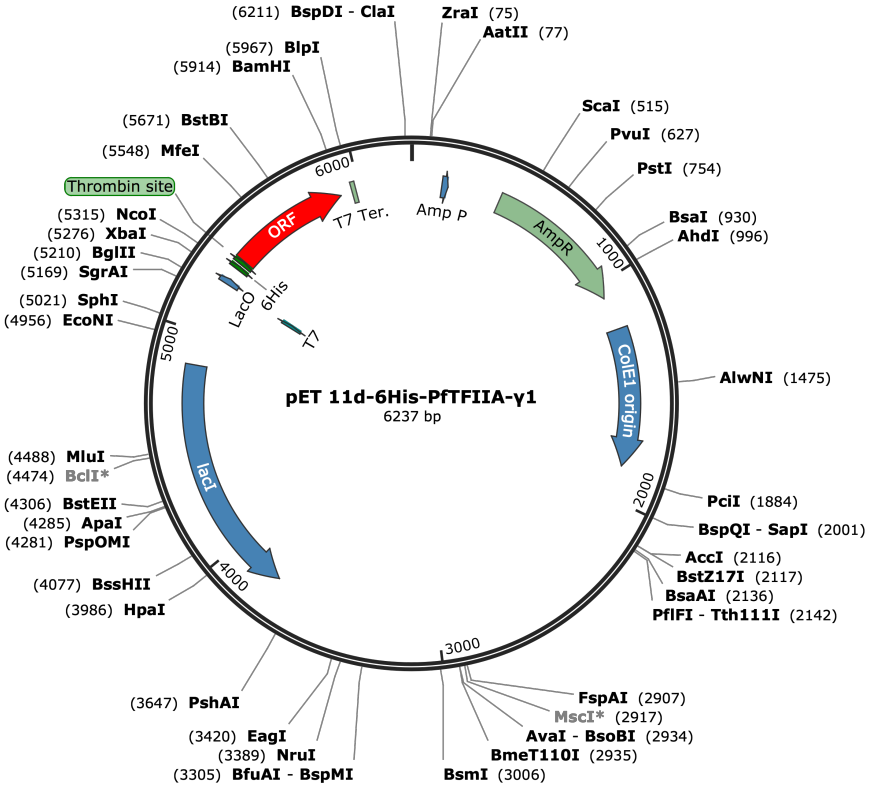
Appendix

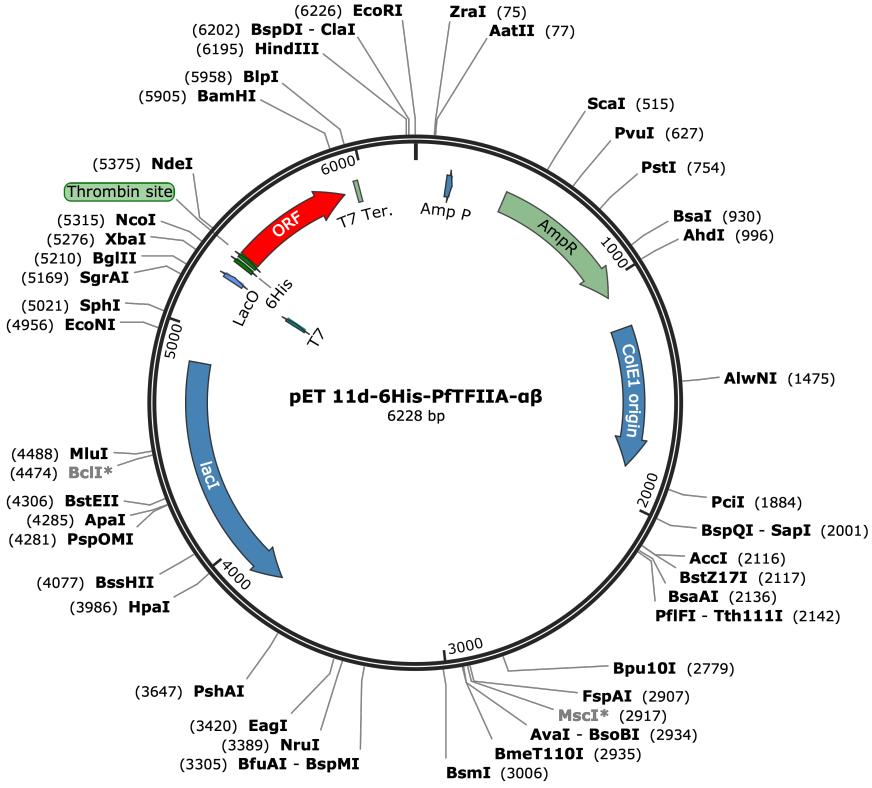
A.1 PfTFIIA expression vectors and pET 11d vector maps

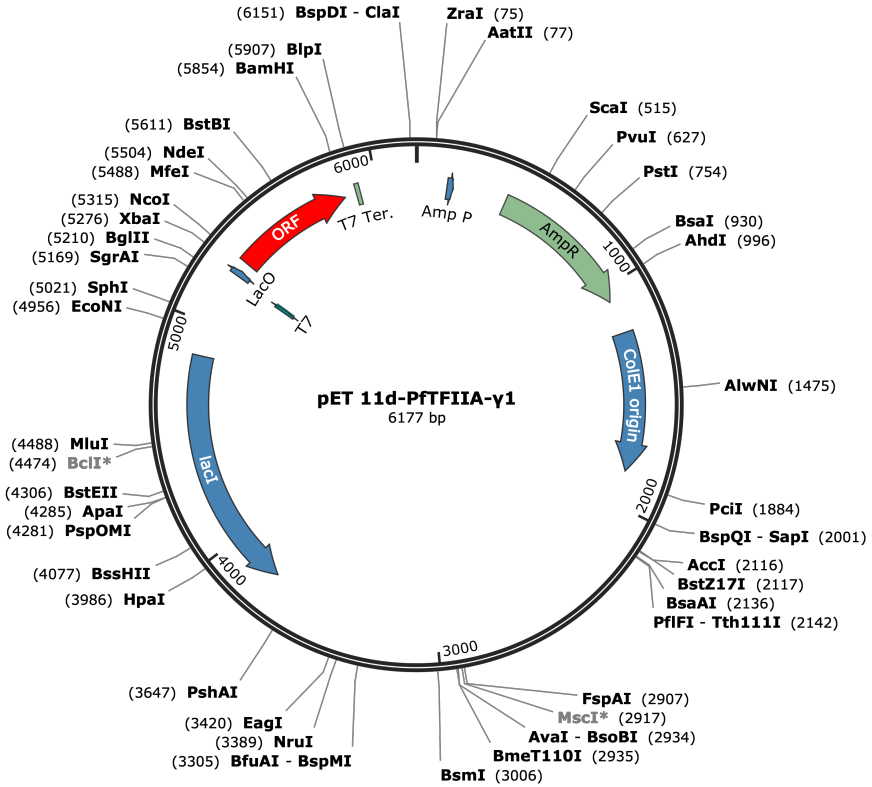
Vector name	Coding	Purification tag	Size (bp)
pET 11d-6His	N/A	6His-	5708
pET11d-6His-PfTFIIA-αβ	PfTFIIA-αβ subunit	6His-	6228
pET11d-6His-PfTFIIA-γ1	PfTFIIA-γ1 subunit	6His-	6237
pET11d-6His-PfTFIIA-γ2	PfTFIIA-γ2 subunit	6His-	6258
pET11d-PfTFIIA-γ1	PfTFIIA-γ1 subunit	non	6177
pET11d-PfTFIIA-γ2	PfTFIIA-γ1 subunit	non	6198
pET11d-6His-PfTFIIA-αβ-γ1	PfTFIIA-αβ-γ1 complex	6His-	6935
pET11d-6His-PfTFIIA-αβ-γ2	PfTFIIA-αβ-γ2 complex	6His-	6850

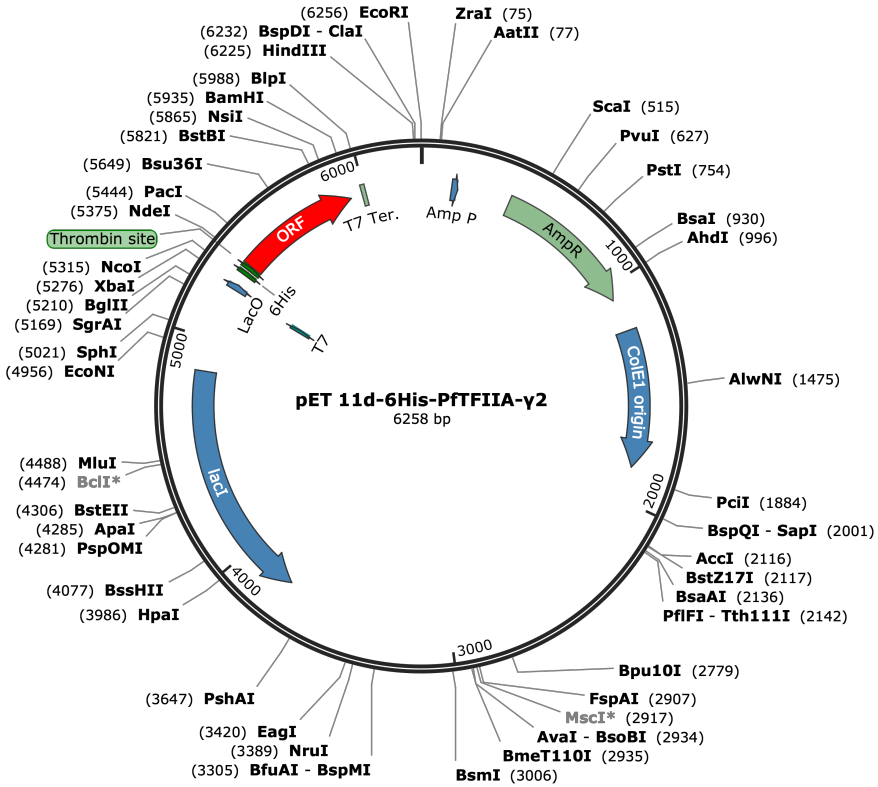
Table A1 : PfTFIIA expression vectors

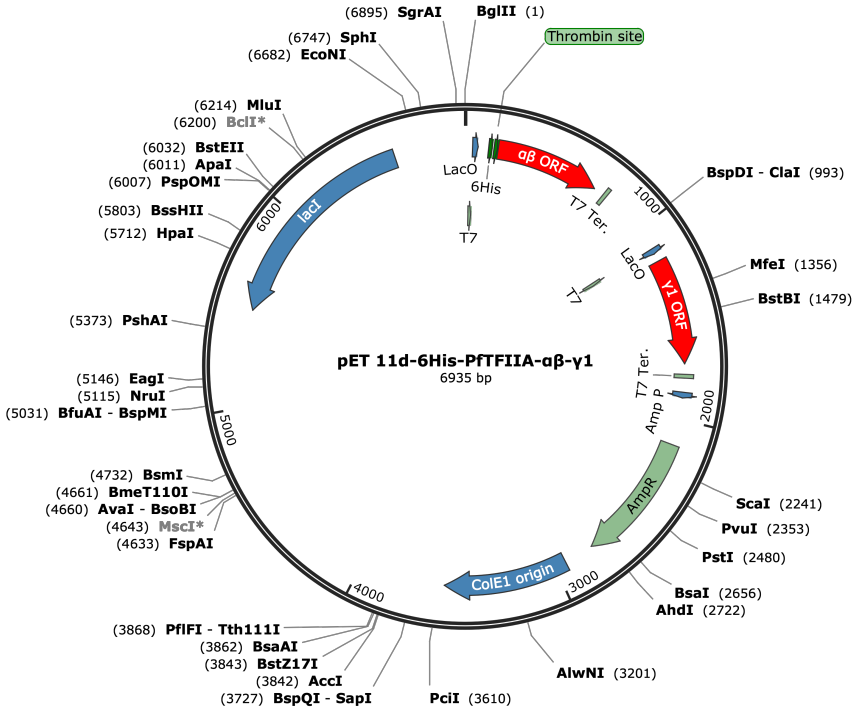


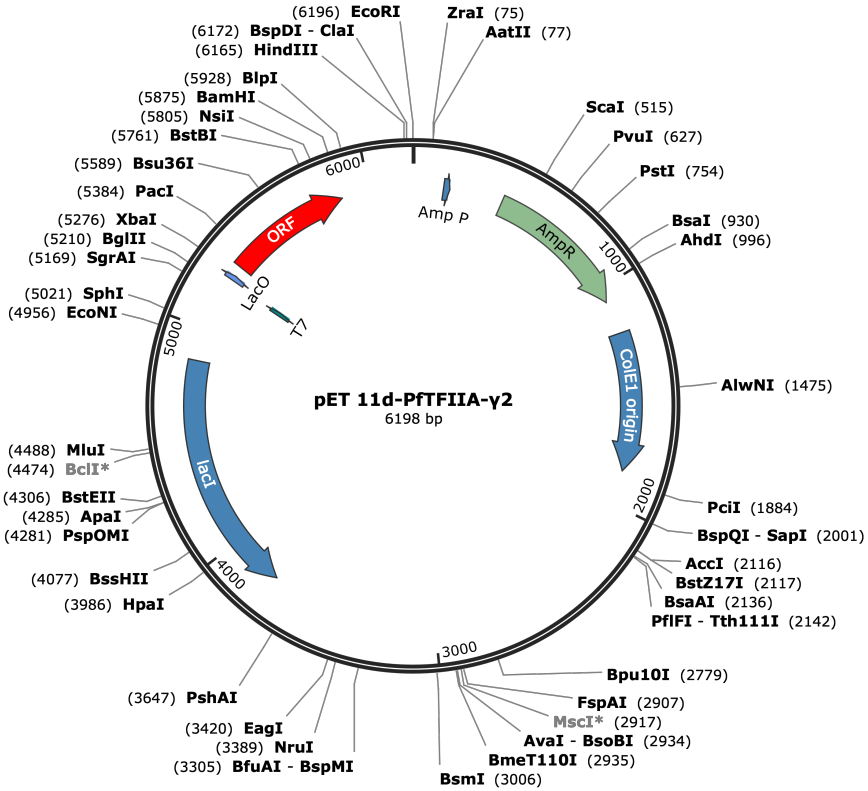

pET 11d-6His-PfTFIIA-γ1
6237 bp
(6211) BspDI - ClaI
ZraI (75)
AatII (77)
(5967) BlpI
(5914) BamHI
ScaI (515)
PvuI (627)
(5671) BstBI
(5548) MfeI
PstI (754)
Thrombin site
BsaI (930)
(5315) NcoI
AhdI (996)
(5276) XbaI
(5210) BglII
(5169) SgrAI
(5021) SphI
(4956) EcoNI
AlwNI (1475)
(4488) MluI
(4474) BclI*
PciI (1884)
(4306) BstEII
BspQI - SapI (2001)
(4285) ApaI
(4281) PspOMI
AccI (2116)
BstZ17I (2117)
(4077) BssHII
BsaAI (2136)
(3986) HpaI
PflFI - Tth111I (2142)
(3647) PshAI
FspAI (2907)
MscI* (2917)
(3420) EagI
AvaI - BsoBI (2934)
(3389) NruI
BmeT110I (2935)
(3305) BfuAI - BspMI
BsmI (3006)
ORF
T7 Ter.
Amp P
AmpR
ColE1 origin
lacI
LacO
6His
T7
1000
2000
3000
4000
5000
6000

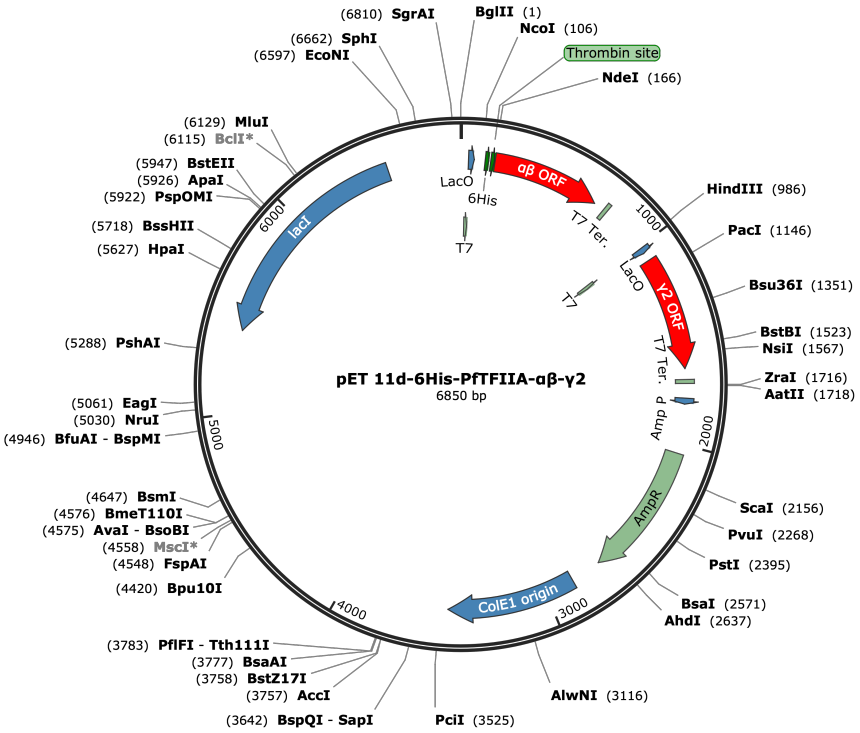

pET 11d-6His-PfTFIIA-αβ
6228 bp
(6226) EcoRI
ZraI (75)
(6202) BspDI - ClaI
AatII (77)
(6195) HindIII
(5958) BlpI
(5905) BamHI
ScaI (515)
PvuI (627)
(5375) NdeI
PstI (754)
Thrombin site
BsaI (930)
(5315) NcoI
AhdI (996)
(5276) XbaI
(5210) BglII
(5169) SgrAI
(5021) SphI
(4956) EcoNI
AlwNI (1475)
(4488) MluI
(4474) BclI*
PciI (1884)
(4306) BstEII
BspQI - SapI (2001)
(4285) ApaI
(4281) PspOMI
AccI (2116)
BstZ17I (2117)
(4077) BssHII
BsaAI (2136)
(3986) HpaI
PflFI - Tth111I (2142)
Bpu10I (2779)
(3647) PshAI
FspAI (2907)
MscI* (2917)
(3420) EagI
AvaI - BsoBI (2934)
(3389) NruI
BmeT110I (2935)
(3305) BfuAI - BspMI
BsmI (3006)
ORF
T7 Ter.
Amp P
AmpR
ColE1 origin
lacI
LacO
6His
T7
1000
2000
3000
4000
5000
6000


pET 11d-PfTFIIA-γ1
6177 bp
ORF
T7 Ter.
Amp P
AmpR
ColE1 origin
lacI
LacO
T7
1000
2000
3000
4000
5000
6000
(6151) BspDI - ClaI
ZraI (75)
AatII (77)
(5907) BlpI
(5854) BamHI
(5611) BstBI
(5504) NdeI
(5488) MfeI
(5315) NcoI
(5276) XbaI
(5210) BglII
(5169) SgrAI
(5021) SphI
(4956) EcoNI
ScaI (515)
PvuI (627)
PstI (754)
BsaI (930)
AhdI (996)
AlwNI (1475)
PciI (1884)
BspQI - SapI (2001)
AccI (2116)
BstZ17I (2117)
BsaAI (2136)
PflFI - Tth111I (2142)
FspAI (2907)
MscI* (2917)
AvaI - BsoBI (2934)
BmeT110I (2935)
BsmI (3006)
(3647) PshAI
(3420) EagI
(3389) NruI
(3305) BfuAI - BspMI
(4077) BssHII
(3986) HpaI
(4488) MluI
(4474) BclI*
(4306) BstEII
(4285) ApaI
(4281) PspOMI


pET 11d-6His-PfTFIIA-γ2
6258 bp
ORF
T7 Ter.
Amp P
AmpR
ColE1 origin
lacI
LacO
6His
T7
Thrombin site
1000
2000
3000
4000
5000
6000
(6256) EcoRI
ZraI (75)
AatII (77)
(6232) BspDI - ClaI
(6225) HindIII
(5988) BlpI
(5935) BamHI
(5865) NsiI
(5821) BstBI
(5649) Bsu36I
(5444) PacI
(5375) NdeI
(5315) NcoI
(5276) XbaI
(5210) BglII
(5169) SgrAI
(5021) SphI
(4956) EcoNI
ScaI (515)
PvuI (627)
PstI (754)
BsaI (930)
AhdI (996)
AlwNI (1475)
PciI (1884)
BspQI - SapI (2001)
AccI (2116)
BstZ17I (2117)
BsaAI (2136)
PflFI - Tth111I (2142)
Bpu10I (2779)
FspAI (2907)
MscI* (2917)
AvaI - BsoBI (2934)
BmeT110I (2935)
BsmI (3006)
(3647) PshAI
(3420) EagI
(3389) NruI
(3305) BfuAI - BspMI
(4077) BssHII
(3986) HpaI
(4488) MluI
(4474) BclI*
(4306) BstEII
(4285) ApaI
(4281) PspOMI


pET 11d-6His-PfTFIIA-αβ-γ1
6935 bp
(6895) SgrAI
BglII (1)
Thrombin site
(6747) SphI
(6682) EcoNI
(6214) MluI
(6200) BclI*
(6032) BstEII
(6011) ApaI
(6007) PspOMI
(5803) BssHII
(5712) HpaI
(5373) PshAI
(5146) EagI
(5115) NruI
(5031) BfuAI - BspMI
(4732) BsmI
(4661) BmeT110I
(4660) AvaI - BsoBI
(4643) MscI*
(4633) FspAI
(3868) PflFI - Tth111I
(3862) BsaAI
(3843) BstZ17I
(3842) AccI
(3727) BspQI - SapI
PciI (3610)
AlwNI (3201)
AhdI (2722)
BsaI (2656)
PstI (2480)
PvuI (2353)
ScaI (2241)
MfeI (1356)
BstBI (1479)
BspDI - ClaI (993)
αβ ORF
LacO
6His
T7
T7 Ter.
LacO
T7
Y1 ORF
T7 Ter.
Amp P
AmpR
ColE1 origin
lacI
1000
2000
3000
4000
5000
6000


pET 11d-PfTFIIA-γ2
6198 bp
(6196) EcoRI
ZraI (75)
AatII (77)
(6172) BspDI - ClaI
(6165) HindIII
(5928) BlpI
(5875) BamHI
(5805) NsiI
(5761) BstBI
(5589) Bsu36I
(5384) PacI
(5276) XbaI
(5210) BglII
(5169) SgrAI
(5021) SphI
(4956) EcoNI
(4488) MluI
(4474) BclI*
(4306) BstEII
(4285) ApaI
(4281) PspOMI
(4077) BssHII
(3986) HpaI
(3647) PshAI
(3420) EagI
(3389) NruI
(3305) BfuAI - BspMI
BsmI (3006)
BmeT110I (2935)
AvaI - BsoBI (2934)
MscI* (2917)
FspAI (2907)
Bpu10I (2779)
PflFI - Tth111I (2142)
BsaAI (2136)
BstZ17I (2117)
AccI (2116)
BspQI - SapI (2001)
PciI (1884)
AlwNI (1475)
ScaI (515)
PvuI (627)
PstI (754)
BsaI (930)
AhdI (996)
ORF
LacO
T7
Amp P
AmpR
ColE1 origin
lacI
1000
2000
3000
4000
5000
6000



A.2 Expression vector ORFs and surrounding sequences

Blue- T7 promoter; purple- 6His tag; red- thrombin cleavage site; lower case letters- open reading frame; orange- T7 terminator.

pET11d-6His:

GAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAAT

AATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCATCATCATC

ACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCTCGAGGATCCGGCTGCTAAC

AAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAA

CCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATAT

CCGGATATCCCGCAAGAGGCCCGGCAGTACCGGCATAACCAAGCCTATGCCTACAGC

ATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGATTTCATACACGGT

GCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTATCGATGATA

pET11d-6His-PfTFIIA-αβ:

AGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGA
GCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCAT
GGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCC
ATatgtctttattgagtgattctgatttaaaggaaattcataaaaagattattgacagcacaataaaaaattgtggtcagttttataa
tgcaagaatattggaagcaataaaaatgaggtggataaaaatatatgagcagaaattaagaaacaaggaacttggattagt
ctcgaacaaaaaggaaaacaaaaacagcaaaatacaattaagtgacgattttgatgatgacgaatttgaagatgccgatgtt
gaagaaaaagaaatagtagaaattcaagaagaagttgaaaatgatccagccgaagaagatgaattaaatgacctggaca
atatttctataagcgatttaagtgacgttgacccagaaacgaataatattataataggagtttgtgataagatatctaaaccttgtg
gtagaagaaatgttggaagtaattggaaaattaaactgaaaggaggacttatgaagattgatgggaaggaaatgttttttcacg
gtttacaaggagaattggaattttaatagGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGT
TGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGG
TCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATCCCGCAAGAGGCCC
GGCAGTACCGGCATAACCAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGG
ATGACGATGAGCGCATTGTTAGATTTCATACACGGTGCCTGACTGCGTTAGCAATTTAA
CTGTGATAAACTACCGCATTAAAGCTTATCGATGATAAGCTGTCAAACATGAGAA

pET11d-6His-PfTFIIA-γ1:

GAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAAT
AATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCATCATCATC
ACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATatggaacatgggaactttgatgagaagtttggtgatt
cgatattattggaatctttaaaagaggcccttcaacaaatgattgaagagttttatgtggaaaaggagaaaggaatacaaatat
ataaggaggcatgtatgaatgtgaaaaaagaaatattagacaattcaaatcaattgtcagatgttcatatgagtggacaactta
aaagttattactgtcggaatgatatgtggacatttttttttaagaattctctgtttaaaattaataagaataaaaaaatgaaaagttctt
cgaaggattacaaaaattatcaacctttaaatttacgagtatataaaaatttttatgacaaaaaggaagagtttcttaaaaattgtg
ttgacaaaaataatgttaaatttttttaaaaatttccctaagctttattcaaatatcgttcataaggaaaacaatgaggttgaaagtga

tgatgttttttttattatgatggtttaatcaaaatattatgtatagaagaatctaccatataatagGGATCCGGCTGCTAAC
AAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAA
CCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATAT
CCGGATATCCCGCAAGAGGCCCGGCAGTACCGGCATAACCAAGCCTATGCCTACAGC
ATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGATTTCATACACGGT
GCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTATCGATGATA
AGCTGTCAAACATGAGAA

pET11d-6His-PfTFIIA-γ2:

GAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAAT
AATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCATCATCATC
ACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATatgagcgataatttttttatacccacctttgaacatatc
tatgaacaaagcatccccccgcctttaattaaacattcaagaaaatgcaaaaattataataagaaaagaaaaaatattatcaa
atacacatctcaagatatttgtgtagaacaaagaaattacgacaaagaaaataaggaacataagtatgtatgtataaattttcat
acgttcaacaaaaaatgttatcataaaaatatccgactattagaccatacacatttggatttcaccttaggggaatctctaatgtat
cttcttaatgtatatttacaaaataaaattataactaagcatttgtataatagaatgttctattcgtttcagctggcaatcattgaagta
ataagtaaactcaaagaaggaaaagaatataagtggaatataaaagggaaattaataaattcgaaaaaagaaaataatat
acaaattgtttatatagaaaatgcatttctgtcctttaaaagtatagtattacatgccccttgttaaaaattaaatctatagatatata
atagGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGC
TGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG
CTGAAAGGAGGAACTATATCCGGATATCCCGCAAGAGGCCCGGCAGTACCGGCATAAC
CAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATT
GTTAGATTTCATACACGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGC
ATTAAAGCTTATCGATGATAAGCTGTCAAACATGAGAA

pET11d-6His-PfTFIIA-αβ-γ1:

GAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAAT

AATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCATCATCATC

ACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATatgtctttattgagtgattctgatttaaaggaaattcat

aaaaagattattgacagcacaataaaaaattgtggtcagttttataatgcaagaatattggaagcaataaaaatgaggtggat

aaaaatatatgagcagaaattaagaaacaaggaacttggattagtctcgaacaaaaaggaaaacaaaaacagcaaaata

caattaagtgacgattttgatgatgacgaatttgaagatgccgatgttgaagaaaaagaaatagtagaaattcaagaagaagt

tgaaaatgatccagccgaagaagatgaattaaatgacctggacaatatttctataagcgatttaagtgacgttgacccagaaa

cgaataatattataataggagtttgtgataagatatctaaaccttgtggtagaagaaatgttggaagtaattggaaaattaaactg

aaaggaggacttatgaagattgatgggaaggaaatgttttttcacggtttacaaggagaattggaattttaatagGGATCCG

GCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAA

CTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAG

GAACTATATCCGGATATCCCGCAAGAGGCCCGGCAGTACCGGCATAACCAAGCCTATG

CCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGATTTCA

TACACGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTAT

CGATGATAAGCTGTCAAACATGAGAATTCTTGAAGACGAAAGGGCCTCGTGATACGCC

TATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCTAATACGACTCACTAT

AGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAA

GGAGATATACCatggaacatgggaactttgatgagaagtttggtgattcgatattattggaatctttaaaagaggcccttca

acaaatgattgaagagttttatgtggaaaaggagaaaggaatacaaatatataaggaggcatgtatgaatgtgaaaaaaga

aatattagacaattcaaatcaattgtcagatgttcatatgagtggacaacttaaaagttattactgtcggaatgatatgtggacattt

ttttttaagaattctctgtttaaaattaataagaataaaaaaatgaaaagttcttcgaaggattacaaaaattatcaacctttaaattt

acgagtatataaaaatttttatgacaaaaaggaagagtttcttaaaaattgtgttgacaaaaataatgttaaatttttttaaaaatttc

cctaagctttattcaaatatcgttcataaggaaaacaatgaggttgaaagtgatgatgttttttttttattatgatggtttaatcaaaatat

tatgtatagaagaatctaccatataatagGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAG

pET11d-6His-PfTFIIA-αβ-γ2:

GAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAAT

AATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCATCATCATC

ACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATatgtctttattgagtgattctgatttaaaggaaattcat

aaaaagattattgacagcacaataaaaaattgtggtcagttttataatgcaagaatattggaagcaataaaaatgaggtggat

aaaaatatatgagcagaaattaagaaacaaggaacttggattagtctcgaacaaaaaggaaaacaaaaacagcaaaata

caattaagtgacgattttgatgatgacgaatttgaagatgccgatgttgaagaaaaagaaatagtagaaattcaagaagaagt

tgaaaatgatccagccgaagaagatgaattaaatgacctggacaatatttctataagcgatttaagtgacgttgacccagaaa

cgaataatattataataggagtttgtgataagatatctaaaccttgtggtagaagaaatgttggaagtaattggaaaattaaactg

aaaggaggacttatgaagattgatgggaaggaaatgttttttcacggtttacaaggagaattggaattttaatagGGATCCG

GCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAA

CTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAG

GAACTATATCCGGATATCCCGCAAGAGGCCCGGCAGTACCGGCATAACCAAGCCTATG

CCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGATTTCA

TACACGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTT

AATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTT

GTTTAACTTTAAGAAGGAGATATACCatgagcgataatttttttatacccacctttgaacatatctatgaacaaag

catccccccgcctttaattaaacattcaagaaaatgcaaaaattataataagaaaagaaaaaatattatcaaatacacatctc

aagatatttgtgtagaacaaagaaattacgacaaagaaaataaggaacataagtatgtatgtataaattttcatacgttcaaca

aaaaatgttatcataaaaatatccgactattagaccatacacatttggatttcaccttaggggaatctctaatgtatcttcttaatgta

tatttacaaaataaaattataactaagcatttgtataatagaatgttctattcgtttcagctggcaatcattgaagtaataagtaaact

caaagaaggaaaagaatataagtggaatataaaagggaaattaataattcgaaaaaagaaaataatatacaaattgtttat

atagaaaatgcatttctgtcctttaaaagtatagtattacatgccccttttgttaaaaattaaatctatagatatataatagGGATC

CGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAAT

AACTAGCATAACCCCGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCC

A.3 SELEX n16 library verification: MEME results

Inputs & Settings

Sequences

Source	Alphabet	Sequence Count
sequences.fa	DNA	480

Background

Name	Freq.	Bg.				Bg.	Freq.	Name
Adenine	0.243	0.243	A	~	T	0.243	0.243	Thymine
Cytosine	0.257	0.257	C	~	G	0.257	0.257	Guanine

Other Settings

Motif Site Distribution	ZOOPS: Zero or one site per sequence
Site Strand Handling	Sites may be on either strand
Maximum Number of Motifs	10
Motif E-value Threshold	no limit
Minimum Motif Width	6
Maximum Motif Width	16
Minimum Sites per Motif	2
Maximum Sites per Motif	480
Bias on Number of Sites	0.8
Sequence Prior	Simple Dirichlet
Sequence Prior Strength	0.01
EM Starting Point Source	From substrings in input sequences
EM Starting Point Map Type	Uniform
EM Starting Point Fuzz	0.5
EM Maximum Iterations	50
EM Improvement Threshold	0.00001
Trim Gap Open Cost	11
Trim Gap Extend Cost	1
End Gap Treatment	Same cost as other gaps

Hide Advanced Settings

Discovered Motifs

	Logo	E-value	Sites	Width	More	Submit/Download
1.		1.0e+005	2	12		
2.		1.1e+005	8	11		
3.		1.2e+005	4	11		
4.		1.5e+005	2	15		
5.		1.8e+005	2	12		
6.		2.0e+005	4	15		
7.		3.6e+004	7	16		
8.		3.8e+004	7	15		
9.		3.1e+004	7	16		
10.		2.2e+005	2	8		

Stopped because requested number of motifs (10) found.

n16 Library verification sequences n1-480

>1
TTTCTCGCTGATTTCG
>2
CCTTCAAAATTTAGCA
>3
GCGGTCAGCGGTGCAC
>4
TTGAAGGCCACAAGGA
>5
TAGGTTCCTTATAGTG
>6
GATGAGGCTGGTGGTC
>7
ATACTAGATCAGGTCT
>8
ACTTGCTGGTTTTTGC
>9
TACTGTGTGTTCTTCT
>10
AGCATTTGCGCGCATA
>11
AAGCCATGTTGCCGCG
>12
GGTGCACCGGTCAACG
>13
GACGAGGGAAATGCAG
>14
CCTCTAGAACGTAGTG
>15
TGTGGCGTTCTAGATA
>16
GGCTCAACCGGAAACG
>17
CACTTACTGCAAATGC
>18
TCCAGATACCAGGACC
>19
TCTAGGTTTAACAACG
>20
CCAACGCAAACTGGAG
>21
ATAGCTAAGTAGACGT
>22
TGTGTGTGACAGAGAG
>23
TCGGAGCGCCTTGGAG
>24
ATCCCGCAACTCCCGG
>25
TATCAGCATTGTGTGA
>26
TGGGCCCGGTATAGAC
>27
TCCATTAACAGGTAGT
>28
CTGACCTCGCTGGGAG
>29
ACTAATAACGCGACCC
>30
CTTGGTCTGGGTCAGA
>31
TTTTGCATTCTACCTC
>32
CCTCATCGTTATCGTC
>33
CCCTGCACGCCTATGC
>34
GAGTTTAGTTTACTTC
>35
CACCTTCCCCGCGGAA
>36
TATGGTGGACGGCCTG
>37
CCGCAGCAGAACCGCG
>38
CCGAAGTAAAGTGTGA
>39
GATATCTGACAGTAGT
>40
TCAGTTCACCGCGTGG
>41
CCCCACCAATTGCACG
>42
ATCAGAGTCCTGAATC
>43
ATTTGACCTAATTTGT
>44
GGGGTTGAGTCCTCAA
>45
AGACCATTTTCCTGGC
>46
ACAGTCCATTTGTACT
>47
TTGTGCGTGGTATGCC
>48
AACCCATACTATGACG
>49
CAATTCGTATTGTGGC
>50
CAAACACCACATCACG
>51
CGGTTTGCGATCTGAA
>52
CGTGAAGGGGGGGTGC
>53
CATACGAACGAAAGTT
>54
AGAAGTGACTATACGT
>55
GATGACAGTAGCTCCC
>56
AAGTGCCTGGGTGCTG
>57
TGGCGCTACGATAAAC
>58
TGGCCTTGCGAGGCAC
>59
CGCAAATACTTCAACG
>60
CTTCGTCCGCGCACCA
>61
CGTGTTAGTCGATCCC
>62
TGAACACCGTAGATTA
>63
AGCGTCGTTGCTTGTT
>64
ATTTTTCCTCTGGCAG
>65
AGTCCAGCCATGAAGC
>66
CACACTACCAACATAA
>67
TTAAGCACACTCCACT
>68
GCCACGCACCAATAGG
>69
GTCACGAAGTGCTAAA
>70
CCAATAATTACTGTCC
>71
AACTGGGGGGGTGGCG
>72
CCCACGATCGGTAGTC
>73
GACACGGAAGAGCCCG
>74
AGCGTCATGGAGGCAT
>75
CACCAACACAAGATGT
>76
CTCAGGCTTTAGTGGA
>77
GATTGGAGTAATGCCA
>78
TATTGAGTCACTAATA
>79
GCTCAATTGTTCGTAG
>80
CGATGGTTTGTATCGA
>81
CAAGATGGTCGCCCGC
>82
TATGACTTATCCCGGA
>83
GGGGCGTTCGCACGTG
>84
AAATGGAGGTATTTTC
>85
CCAACTACTATTAAAC
>86
GTGGAACGCCTCCTCT
>87
GACAACATGCTCGATG
>88
TAGTTTTATGTGCGTC
>89
CTACTTTGCATAGCAG
>90
GGGTTGCAGGTTCTTG
>91
TCGTAATCTTGAATAG
>92
ATAGTAAGACGCATAG
>93
TAGAAGAGTTTAGTTC
>94
CAACTGTCGATAATTA
>95
ACCGCTCGTGTTGTCT
>96
GACCCACAAGAGGTGA
>97
ATGCTGAACGGGGATA
>98

ACACATGGCTCGGTCA
>99
CGCCTGTAATGTCGCT
>100
GACAATCTAGCCTCGC
>101
ACACCACTGATCCTAC
>102
ACATTAAGAGGTGCCC
>103
CCCACATCCTCACCAA
>104
AGGCTCGGAAGTCTCA
>105
TCGCCTAAGAAAAGAT
>106
CACAGGATAGAGTAGC
>107
AGTCGAATCCGTGAAG
>108
GGACTGAAATTTTTCC
>109
GAGACGAACGCTTAGA
>110
TCGTGTCCCATGTTTG
>111
GATTTCAAGCTCCTAA
>112
TTTGCATCGCATACTG
>113
TTTGTAGGCAGGTACT
>114
TATTTTGGGGAGCGTG
>115
GTACCTGTGACACCAG
>116
CTTGCATGGAGTGGTG
>117
TAATGCCCGGCGTGTT
>118
CGGGCACTCCGTAACT
>119
ATTCGCCCTCCATTCA
>120
GCTGGTGCAGTACAGT
>121
CCCGGGAGCGTCCCAC
>122
GCCTGCTAATAATTTG
>123
TGTGGGGTCAGTTAGA
>124
CGTGAAATTTAAACTG
>125
TGGGGACCAACCCAAG
>126
AGACTACTTATCACCA
>127
AGTAGCCATTGATTCC
>128
AGTAGAGGACCGTTAC
>129
AAACTGAATGATTTTG
>130
CTGCTATTCGACGACC
>131
GTACTGAAAAATACAA
>132
ACTAATTTACGCTACC
>133
AGTCCCAAAGCCGATT
>134
AGGCTGGGAGATCCTA
>135
CAGACCCCTGGCTCTA
>136
TGTCCCCAAACAGCTC
>137
TACTGGGATACCACAC
>138
CTGTGATAACAGGACT
>139
GCCACTTGCTAATCGT
>140
ATGCGTTTGTGTGGAG
>141
GTCTGCCCGTCTGGGC
>142
CGGGGCAGGCGTACTT
>143
GAGGGAACTAGTCTGC
>144
CGCAGAACCGGTCAGA
>145
GGTTGAGTAGACCATA
>146
GTTATTTCACCCACAG
>147
CTATTTACACTGCACC
>148
CACTATGTCGCTGACC
>149
CTACAAGGGCACGTCG
>150
CCGCAGGAATTAGTGA
>151
ACTCGATACATCGATT
>152
GAAGACGGGAGGACAT
>153
GTCCAGGTCTTGTTTT
>154
AAAGGTGCCCGGGTCA
>155
GCTTCCATCTCGCGAT
>156
TAAGCTCATCAGCTAT
>157
CAACTGTCCGAATCTC
>158
AGCAAGAGACCGACAC
>159
TTGGGTGAGCTTTCAT
>160
CATGGTTGGCATTGGG
>161
ACCGATTTCGGGGCCT
>162
TGAGTTGCCCAAGGGG
>163
AGCAGGGTCAGCGTTT
>164
ATATTGGCGCAGCGCC
>165
GTGAACGGTGACATCA
>166
TCACGCTAATGTATCC
>167
CGGTCGCGTCATTTAA
>168
AATGGTGCGGTCTTCC
>169
TCAGCCGTGCCTTGCT
>170
ACGCCTCGCCTAAGCT
>171
GGGTTTGCTCTGGCAG
>172
CCACCTAGTTCTTATC
>173
TAGATAGATCTTAACG
>174
ACGGTTGGTCCGGTCC
>175
GCAAATCACTCGTCAA
>176
GTGGGGATGACTATCC
>177
TGTAGGGGCCTTACCC
>178
ATAGACATGACGCGGG
>179
GGTGCTCCGTAACTTT
>180
CTTCCCATGGATAAAG
>181
TCGGCTTTGTCCGGCA
>182
TAGTGGATTGTATTAG
>183
CTGTTCCCGCCCTGAT
>184
AACCCGCGCTCAAAGT
>185
TTCCGAATTTTCCATG
>186
CGCCTTCACGTTGTTG
>187
GCATGTCTGGAACCGT
>188
ATCGGATGTGAACGCC
>189
AAGCTACGGCCGTGGG
>190
TTTGCCCATTGCGGAT
>191
TTTGGCCAAGTACCGG
>192
ATTCGTACTTGAGATG
>193
AATTAGGCCAGGGGTC
>194
GGCGAGCAATAAATTT
>195
GAAAAACCCGCATTGC
>196
GGGGCATGGGTCAGGC
>197

TAGAATGTGGATGTAG
>198
CGCTACGGATCATATA
>199
TCTATAATGCATCACC
>200
TCCTAATGCATAACGC
>201
AAGGGGCCGTATAGAG
>202
AATCGAGAATGTTCTG
>203
CTGTCCCGGTACTTTT
>204
TGCAGCGCGGGTTTTG
>205
CTTACGCAGGGAAGCC
>206
GTCTCCGGGGAGTCTA
>207
CTACAACATGACCTTG
>208
CGTCGCGTAGTGCCAG
>209
TAGTTCGTGTCTTGAC
>210
CTCGCCAAGCTAAGCT
>211
CCGTTACACTCTAAAT
>212
GTCCTGATCAGGTCAG
>213
CTCATAATGTGTGGCT
>214
AGATGTAAGCTGAATC
>215
AAAATAGGGCGCAGTG
>216
CCTCACGGATGCCAGC
>217
TTACTTACCGGGTAGC
>218
TGTGTGGGTCTGTGAC
>219
ATAGGTCACCTCTTAC
>220
CCTACAGCTGTAGTTA
>221
TTGACGAGCTCTTCGT
>222
TGTTCACAGATAGGCA
>223
GTGTGTCTCGCCGTTG
>224
GTTCGCGCCGCCCGCC
>225
TTCCCACAGGTAGTTA
>226
AGGATTTCCAGCATGA
>227
TAATAAACGGTCTAAG
>228
GAGATAATTGAGAACG
>229
GGTACACTCCACCAAA
>230
TAACCACAGGCCTACC
>231
CGTTATCAATCTGAGA
>232
TTGGAGCGACATGTTC
>233
TTAGGGGAAAACGAAC
>234
CGTAGCACACTTGCAG
>235
GATCTGTACATTCCAG
>236
CTTGATGGGACGAGGT
>237
TTGCCCCCCAAGCTCT
>238
ATAAGCCCGAGCTAAA
>239
GTTCACCACGAAACCT
>240
ATAGGTCGACACAAGA
>241
GACAATACCTGGTCCT
>242
CCGATCCAGGTAAGTA
>243
GCACACCTAAGGGCTT
>244
GCCGACCCCTCACAGT
>245
TATGTGGGTACCCTCG
>246
GGTCCCCGGGCGTCTG
>247
AGACTGAAAGTAGAAC
>248
GATCTAGTACGCTGCG
>249
TCTGCGGGCTAATCAG
>250
CGGTCGATCAAGCCAC
>251
TCGCATGAGCGCTTAG
>252
TAGATAAGTTCCCAAG
>253
TGCTGATAAATTTGGA
>254
GGCATGGCCGGTGTGT
>255
CTTTGGCACGGTGGAC
>256
GTAAACGACCAGCTAA
>257
TTGGGCTATGAAATTT
>258
ATAGTTGAACTCTTTC
>259
AGATAATTGATTAGTC
>260
TACTCTGGGCATGCGG
>261
AAGACAAGCTGACCTT
>262
CGTTGATCTCCATGAC
>263
GCAGGATGCCTATCTC
>264
TTGCTAGTCCGCTCCA
>265
GATCTAAAGCAGCCCC
>266
TTGCTGGTTGTTACAG
>267
ATGTCCGACCGCTCAT
>268
GCACCGAGCAGTGCTG
>269
AACGGATAACGGTTAC
>270
CCAGCCCCAAGCACCG
>271
TTGTGCGCCGAAATTT
>272
GCAAAACCTAAGATAC
>273
TAACCGCGTTACCTAA
>274
TGGCACTAGCTCATCG
>275
ACATGGAGGCGCGCAT
>276
GGTAGAGCCCCCTCCC
>277
ACCAAATAGTGGACGC
>278
CGCGCACGCGCAGATG
>279
CTCTGGGTCAGGAACG
>280
TGTGCTTTGTATCGCT
>281
AGACAAGGTTGGTGCA
>282
CAATGATGTTCTGCCC
>283
GTGGCGCCTAAAGGCT
>284
GATAGGTAACATTGGC
>285
GGTTAAGGCTCGTAGT
>286
TAAAATGGAGCTACAA
>287
CTTGCCGTTCAGGATC
>288
GGGGAGCCTCATTGTC
>289
CGTGGTTTTCAACGTA
>290
TACGTCGCGCAGTCAC
>291
TTGTGGCATGAGAGAG
>292
TACTTTGGCCGATATT
>293
TGCCGGTTAGCCAATA
>294
TTCGAATCTACTGAGT
>295
GGATACGACTAGCACT
>296

TCGGGACGGGGTGTGG
>297
ACTGCCAACTCCGCCG
>298
ATATGCCAGCCGCTGG
>299
CACAGGATTCATGTCG
>300
GGGAACTGCGTGAAAC
>301
GCCCGTGGGTCGAGAC
>302
CTCAAACCAGACCTGA
>303
CGTCAGAAATATAGCT
>304
CTTGGGCTAATTGTAC
>305
AAAGTACATTTCAATG
>306
GCTAAAGGTTATCTAG
>307
CCTAAGAATGCCATGC
>308
GCTGGCAGACTCCCCG
>309
AGACAACGCTGACAGG
>310
AGTCCGTGAAGTCGTA
>311
CGTCATTTTTTGTAAT
>312
AATTCCCTATCTGAGC
>313
TTCTCTCATCAATTCA
>314
CCGAGGTCGGCGCAGT
>315
CAAGCTATGTGAATTT
>316
TTTTCAGGTTTGTTCC
>317
AGAGGAATGCGTATAC
>318
AGAACCGGCGGGCTAA
>319
GCGTGCGCCGACCTGG
>320
GTTGTCTTATCGTACA
>321
GGGAACAGGAGGGTCT
>322
ACCGTAGGATAACATA
>323
TGACAGGTCACTTACT
>324
GGAATTGCCGCATGAG
>325
ACGTTCATGGCGAAAC
>326
AAATGTATAAACAGCT
>327
CCGCGCGGCACACGAA
>328
ACAGCACGCATAGAAG
>329
GCGGACACAACCCACC
>330
ATGTCAACTTGATTGC
>331
CCATGAGGGCAACCGC
>332
TCGAACTAGCAACAGT
>333
AAGCTTGTCGTAAAGC
>334
TCAAGACACATTCACC
>335
ATAATGCGGGCTTCCC
>336
TACAGGTCCTCTTCCA
>337
ATAACTTGCGCCCAGC
>338
GTACATAAGAAACCCA
>339
GCTGTTGCCCGTCACG
>340
TTACTGCCAGTGGGCA
>341
TACCACAGGCAACCAT
>342
TAAGTTATCGAAGTTG
>343
ATCGGTTGGGGGCACC
>344
AGAAGGCCGTGACTAA
>345
CATGAGAAGCCCTAGT
>346
AGTTTCCCTTTCTGCG
>347
TAGTAGCGACTCAACC
>348
CCAACTGTGATAAGTA
>349
CTGGTGACATCATACC
>350
TGAGCCTAGAACACGT
>351
TTGTGTAAAAAGACCC
>352
CTGATGGGTGGATTCA
>353
ACCCCTCCCGAAAGGG
>354
GACGGGCACACTAATT
>355
GAGCATCCGATGCCCG
>356
TCGTAAGGACCCGTAA
>357
ACTTAACTAAAGCTCC
>358
CCACCTATGACTGACT
>359
GCGCAGTATTTACCGC
>360
TTGTGGGGAAATGGGT
>361
TAAAAGCTGAGTCGCC
>362
GTATGAGGCGGAATCA
>363
CCAGCGTATGGAACTG
>364
CCTAGTTACGGGGTTT
>365
ACCGAGTCACATCACA
>366
GCTAAAAAGTTTTCTC
>367
AGGCACATGCTTGCCC
>368
AAATATAGTGCCCTCG
>369
GTCGCTCTTCGCTCCG
>370
AAACCCGGAATCTATT
>371
TATCCGGCCGTGTATG
>372
CATCGGAGTACGATTT
>373
CCAGATCTCGGTGCCG
>374
CTGTGCTAAGACGTTC
>375
CGGCAACACGGYAGTA
>376
TAGTCTACCCGTACTG
>377
GGGTTAGCTCGCCATA
>378
TTCGCACTTTTCGGCA
>379
CTTTCTTTTCAGCATA
>380
AATCTTGGGTTTGCCG
>381
TCTGCTGCGTTGAGGT
>382
GTGTGGAAGCTTACAC
>383
AAAGGATTTGCCCGCC
>384
ATAGGTAAAAATGTTG
>385
GGTAAAGAAGATGACA
>386
AAAGGCGGAGCCAGTC
>387
CGCAGGGCGGAATATA
>388
CGCTGTGAACCTCGTA
>389
GGTTTAGACTAATCGC
>390
CATATCTTAGACCATT
>391
TGCCGCGTCTTCAGCA
>392
TCAGTATCCCGCCCCG
>393
CCATAAAATCAAAATC
>394
CTGGCCGCTTTCGCCG
>395

GACCACTCCATCAGGA
>396
CCGTTAGGAGCACTCG
>397
AGAGACTATCCACTAT
>398
GCTGGCAGGTCGGGTT
>399
AGTATTCAGACTGGGG
>400
GGGGGACATGACATCA
>401
CCGGTGAAACTTACTA
>402
GCAGGTCCGGACGTTC
>403
GACTAGTCGGGGGCCC
>404
AACCGTGTAGGGACTG
>405
GCTCCAACAGTTCAGG
>406
GTGCGGACAATAAGAT
>407
ACTGACTTAAGTCACC
>408
CACCTAAGAGCTTCCG
>409
ATTTTCGACGGTAGCA
>410
TCGACGAGTAAGTCCG
>411
GAGTGGCACACGCAGC
>412
CCCACCTCAATTACAA
>413
TTTAACGCGTCGCCTG
>414
CCACCCAATATCTTGC
>415
CATAAAGCCAGGCTGT
>416
TTGCTCGCGGAAGCTA
>417
TTGTGGAGGTCGGTTT
>418
GCTCTAACAACTGCCG
>419
TGGTAGGAGTCGTGCG
>420
GTACAACACAGCTACG
>421
CACAGGCTCCAGCCTG
>422
CGATGGCCGGGGCATG
>423
AGCCGGCATTTATCCG
>424
GAAACGTGAACTGTAC
>425
GGGTGATGTTAATTTC
>426
CTTTGTTCTCCCTTAG
>427
GGGGAGCGGGCCTCAA
>428
TATATGTCCACAACGG
>429
GGCATGTGTACATACA
>430
CTTAGGTGTGCCCTCT
>431
GAACTTTTGACCAGGA
>432
GTAGTGGTACCGTTTC
>433
ACAAAGGACACGTAGC
>434
GGGGGGGCGGCATGTC
>435
CTTGTGTCTCGTGGGG
>436
CGTACGGTAGCCGACC
>437
GGCACCGCTTGGTGGA
>438
GTGCGCCTCGGGAGCC
>439
GGCATCCTTTAACCGT
>440
TGAATTCCTTGCGGGG
>441
GCTACCTGGGAGCCCA
>442
GCTCCTCACACGCCTC
>443
GCTTCTTCTTTAGTTG
>444
CAAAATACCAGGTCGA
>445
CTTCGTGGCTGTTCAC
>446
TGCCAACAGAGTTGGT
>447
GATACTCAAGGGCTTC
>448
ACTTGTTTGATGCCCT
>449
CTCCTGGGTAAGTTTT
>450
TTAATGGTCCCTACTC
>451
AATATCGTAACACGAG
>452
AGTGCATGCCTGGGGC
>453
TCAGCGCCCTCGTCTC
>454
GGTGCTAACTACGGGA
>455
GTTTCAATAAATGTAG
>456
TGCACTAGTCGTCGTC
>457
TGTAGGTCAAGTTGGC
>458
AACGGCCTTGGCTGGA
>459
GGTTGCGATTTGGACG
>460
TGAATGGTATACAAAG
>461
CTTGTTGGTACCAGTC
>462
AAGTGTGTGACTTGCT
>463
GACGTATGGAACTGGA
>464
GTAGTTTTGGTAAACG
>465
TTAGATACTCGCTAGC
>466
ATCGGCACATTGTACT
>467
CTAGGACTATGGCCCT
>468
AAGCCATGTTTGCCTT
>469
CTGCGAACCCAACTTT
>470
ATCGTCAATAACGACG
>471
TGTCGATCGCAGGCGT
>472
GTCCACTGGCGCGTGA
>473
TTGCGAGCATGGCAAG
>474
CGGGGTAGTGCAGCAT
>475
ATGTTTCTGCTATTCG
>476
CCGCCGGGGACCTAAT
>477
ACCCACCGACGAATCG
>478
AAGATTGTAGTAACTT
>479
AAGTGCTGGCCGGATA
>480
GCATTATGTGACGGTA

A.4 SELEX Round 7 (R7) final analysis: MEME results

Inputs & Settings

Sequences

Source	Alphabet	Sequence Count
sequences.fa	DNA	397

Background

Name	Freq.	Bg.				Bg.	Freq.	Name
Adenine	0.305	0.305	A	~	T	0.305	0.305	Thymine
Cytosine	0.195	0.195	C	~	G	0.195	0.195	Guanine

Other Settings

Motif Site Distribution	ZOOPS: Zero or one site per sequence
Site Strand Handling	Sites may be on either strand
Maximum Number of Motifs	10
Motif E-value Threshold	no limit
Minimum Motif Width	6
Maximum Motif Width	16
Minimum Sites per Motif	2
Maximum Sites per Motif	397
Bias on Number of Sites	0.8
Sequence Prior	Simple Dirichlet
Sequence Prior Strength	0.01
EM Starting Point Source	From substrings in input sequences
EM Starting Point Map Type	Uniform
EM Starting Point Fuzz	0.5
EM Maximum Iterations	50
EM Improvement Threshold	0.00001
Trim Gap Open Cost	11
Trim Gap Extend Cost	1
End Gap Treatment	Same cost as other gaps

Hide Advanced Settings

Discovered Motifs

	Logo	E-value	Sites	Width	More	Submit/Download
1.		9.3e+000	7	15		
2.		2.2e+003	3	10		
3.		2.3e+003	11	15		
4.		1.8e+004	2	14		
5.		1.6e+004	10	7		
6.		2.5e+004	2	9		
7.		3.7e+004	2	9		
8.		4.7e+004	5	11		
9.		6.0e+004	5	16		
10.		7.6e+004	2	11		

Stopped because requested number of motifs (10) found.

Selex R7 Sequences n1-397

>1
AGTGTGATGATCGCTC
>2
TAACCACCACTAACAA
>3
TATTGTGTAGTCAAGT
>4
TGTGCTGCGGGGTGCC
>5
TGAAGCAAACAAATGC
>6
CCCGCGATGGTGATTC
>7
TTTCGTTAACAAACTC
>8
AAAAATTATATCAAAG
>9
CTCTTTTATTGTGAGT
>10
GTAATCACTCGCAAGC
>11
AAACACCTCACCTTGC
>12
AAAATGTTGCTTAGAG
>13
CTAACTTCCAAACCAA
>14
GATTTTCGTCAATGCG
>15
AGTGAGGCAAAATGCA
>16
CATCTGCCCTGATGGC
>17
TAATATGGAGTTGGAA
>18
GATTTCTTCAGCCCCT
>19
CTTAAACCCCTACACA
>20
CAGCGTACGTTTCCGA
>21
CATGGATTGGGTTTAC
>22
CACTACACACATTAGA
>23
AAGTCTGGATTTTGAT
>24
TTTAATTTTTTTTAAT
>25
CTATCATTAATAAAAT
>26
ATGAAGTTCGAATAAT
>27
AGAAAGCAGCATAATT
>28
GAGTTTGTAAAATTTT
>29
ATCCCAAAAAACACTA
>30
TTCCGTGTTGCCCGCT
>31
GTCGGGCAAAATTGTG
>32
AATTGAATCCAGCTAG
>33
GGTCACCGGCCGGCTC
>34
TTTGCATGGATACGCT
>35
TACACGACCTTCGTTA
>36
ACCACGAGTTATATCA
>37
GAAACAACTCAATGCT
>38
GTACGGATTTTGAAGC
>39
TAGCGTATTTTCCGAA
>40
GTTACCTGGCTATACT
>41
AGCGCATTGGGTCGGA
>42
GTTGTATTTTCTACTT
>43
TTTTATAAAATCCAAC
>44
TCCCTGTGGGAATAAT
>45
ACATTTGTTGATTTAG
>46
TGAGACAGTTTACTTC
>47
AAAAATGAGCTTGAAA
>48
TTGCACGACTCCATTC
>49
AGGCTCGAGTAATTTG
>50
TGTAAAGTAGGAGTGA
>51
ACTAGCGCCGAATCTG
>52
TATTATCACTTAAAGC
>53
ATCTAGATTTTATACG
>54
TCTTATTGTTTTGACA
>55
ACTTTGGAAAAGGATC
>56
ACGCGCTTTTTCGTGT
>57
TCTATACGGGTGGGCT
>58
GCATTACCTGCATTTT
>59
TTTTCGTATGTTTTAG
>60
ATGGTTTTCATTCTTA
>61
ATGGTTAACCGGGCTG
>62
ATTTGATCACACGTAC
>63
ACGTAAATAGCCGTAG
>64
GTTCATGTTTTACCTT
>65
GTCAATGCACAAAATC
>66
AAGATTAAATTGAACG
>67
TATGAACAATTCTGTT
>68
TATAAGAATATGTAAC
>69
TCCCGCACACGGGGAC
>70
TTTTTCCGCTTATCGG
>71
TCTGCGATTTTATCTT
>72
AGCGATAGACCTGGTA
>73
CCCAAAGACGCATTGC
>74
ACATATATCATCACAC
>75
ACAGTAAAGGCGCATT
>76
ACCCAGTTTATCATTT
>77
GTTCACAAACCCAGTA
>78
GATGTTGAGGACATTA
>79
TCTTTGTGGATGCAAA
>80
GCTCGTGTATTTCTGC
>81
AAAATCTCCTACCAGC
>82
ACCCCATGAATTGACG
>83
GCGAAGCACGCGTCTG
>84
CATTGCCATGGCCCTA
>85
CAGACGAGATCTAAAA
>86
GTATAAGAACCTGACG
>87
ATGGATTAGGGACTTA
>88
TTGGCTATTGATGTGT
>89
CGACTGGTAGCATCAC
>90
GATTACGTCTTTTTGC
>91
CTCGACGGGAATAATC
>92
GACTTATGTTGTACAA
>93
GGCGAAGACAGTATTA
>94
TGGCCACACTCATTAT
>95
ATGTTTTTTTTCTTT
>96
AGCAAGCGGAGATAAC
>97
TATGCGCCAGGGTTCC
>98
AAAAGAGAGCCCCATA

>99
ACGTAGAGTGGGTATC
>100
GGGTAACGCCCTCACA
>101
GTAATAGGCCCGAGTA
>102
TCGCCACGGTTAGCCT
>103
AATACATATTAACTTA
>104
GTGTGTTGCTATGAGA
>105
CGTTGTGAGGTATGAT
>106
CTGGTTTGCACATGTT
>107
CGTAACGAAATTGGGT
>108
TTCAAAAATATTTTCC
>109
AAAAATACTTCTAGCA
>110
TGTCTATTGCCGGGTC
>111
TATCCGCTCAAATTAG
>112
ATATCGTTACGTAAAC
>113
CAACTAAACAACTGTA
>114
AGCATGTTAAAACCCG
>115
TCTTTTTTGCCGCCTA
>116
GCGGCTGTATCATTTA
>117
AAAAACGGCCATAAAA
>118
AAAGGCAGTAAAAGAT
>119
AGAATCTGGTCCTAAT
>120
AAGCCTGCCTGATAAA
>121
CTCTGTTCACCTGTTC
>122
CGAGTAATTTCATTTG
>123
TAAATCACAGGTAATC
>124
TTTTGGGAAGAAATGA
>125
TACAGAAAATTAATTT
>126
CCTTTCAAGCTCCTAA
>127
TATAATTAGTCTGCAT
>128
GTATATTGGGTACGTA
>129
ATCTCCAGTACCGATT
>130
TTTCTATGGGAAACAT
>131
CCAGTATGCGTATGGT
>132
AAAAAATGTGTTTGGG
>133
AAGGAGCAAAAAAAGT
>134
TTTCTGATGATGTCTT
>135
CAATTAATGGCTTCTT
>136
TCGTTCAGACCAGGGC
>137
GATACATTGTATTTTG
>138
CCAGCCGGCACGAAAA
>139
ACCAGATCCGTGCGCT
>140
GTGGGTGAGCGCAGCT
>141
TGACGGCACCAACCTT
>142
TTCATCATAATAGCTT
>143
CAGCCCCAACGGAGGC
>144
TGTCGTTTTCACATTT
>145
TATTTGGCTAGGGTCT
>146
TTTGGTCGAGCGTAAT
>147
CCCTTCATCGGTTTAA
>148
TGATTCTACCCGGACA
>149
AATATCCTCCGTGAAT
>150
CGTTGGGTAAGTCCTT
>151
TGACTCTTACTCCTCG
>152
AGTCTAAAAACCCTTA
>153
TTACTATATACCACTC
>154
CGTGCCAATATTGTTG
>155
CATATTTCGAGGCAAG
>156
ATAGTGAACCGCATCT
>157
GCGGGTTTTTTTCCAT
>158
CTGTGTAATAATTCGT
>159
AATGTAGAAAACTGGA
>160
TGTTTTCACTTTTGGT
>161
TGGTTCAAGTTTGTCT
>162
TTCCGACGGGTGGATC
>163
ATGTGGGGCAACCCTG
>164
CTAAGCCCAAATAAAA
>165
TTTGGGGGTATACGAT
>166
TACACAACATATACCA
>167
CTTAGGACTTTCTTGT
>168
AAAATCAGAAAATTCA
>169
CATARAGAACAAAACC
>170
GTAGGTGTAAGTGCTT
>171
AAGCGGCCAAAATCGA
>172
TGCGTGTGCGGGGTTA
>173
TATTTTCACGATTTAT
>174
CAGAATGGATAAGCAC
>175
GGTATTAACCCTTTCT
>176
GTTAACGATGAATATT
>177
AAAATAACAAAAACTA
>178
AATGTGGTTCAATAAA
>179
AATGTAAAGCCTTATA
>180
GTATATAACCCTACCT
>181
TACATCATTAGAAATC
>182
GAGCAAACAACAAACG
>183
TTAATTGTGGCATATA
>184
TACAAACATTGGAAAT
>185
AATAGGTAAAAAGGAC
>186
TTTTACTATTTTTTCC
>187
CTAAAATATAAAACGA
>188
ACGGTTTCCGAATGTC
>189
CTTTCTAAGACGCTAA
>190
AACATAAAGGCCATAC
>191
TTTAATATTTTTTGGT
>192
CATCTGGAGAATAAAT
>193
CAGTCGATCAAGCAAC
>194
AAACCAATTGGAAGGC
>195
TAAAAGTAATATAGAT
>196
CTAACATGGCACAAAC
>197
TTCTTTTATATGTGGT

>198
ATGTTACCTATTTTAA
>199
ACTCACCTTCCGAATC
>200
AATATTTCGTTGGTTA
>201
TCGATTCACGCAAAAC
>202
GGGGGCATGGAAATTG
>203
GGCAGCTTTTTTATAG
>204
GCGAGCCTTTTTATCT
>205
ACGTCAACGCTTGTTA
>206
ATCTTGAACGCAATCG
>207
TAGAATGCATACGTTC
>208
GATCGCATTTTCGGAA
>209
GTAAGGAAATGAACGA
>210
TTGAGACATAAAGTGG
>211
GGACCCAATTTCAACG
>212
ACTAAATCGGGCATTG
>213
CTTAGGAGTATCGAGT
>214
AGCCTAATTATGTAAG
>215
ACTTAAGGCCCATCAC
>216
CAATGTGATAAGGAAT
>217
TTAATTAACCTTCGAC
>218
TATGACACGTAAAACA
>219
CAGAGTCTAAAAACTC
>220
TTCTTAGTGTGTTTTA
>221
TAACTTATTTTATAAG
>222
GTTTGTTGTGTAATAA
>223
AAAACGTTTCCAAAGA
>224
GATCCAACCCTATACC
>225
TGGTGTACACTCATTA
>226
CGATTAAAAAAATATA
>227
AGGCCATTCAATGAGG
>228
ATATGGAAACCACGAT
>229
TCAAAATGAAAGACCC
>230
AATATACAGTAACCCA
>231
AAAAAATTAACATCCA
>232
TATCGTGATATACATG
>233
ATCACACTTTGATTAT
>234
ACGGATAATACGTTAT
>235
TCGTACCGTAGCTTTT
>236
TTCCTATATTTCGCAT
>237
AAAAATGTTAGAACAT
>238
ATGGGGGGGGGGATAC
>239
ATACTAAAAGCACAGT
>240
TTTGTGTAGTTTTGGG
>241
GTTTCACTTTTCCAGC
>242
CTTCTCGCACACGCTA
>243
CAATGGTATCTGACGA
>244
ATCGCTCATTGTGGTA
>245
ATAATTCTCCCCGCTC
>246
CAGGGGAACAGTACAT
>247
CATATATGCCCCAATA
>248
CAGGTTTCGAGATGTA
>249
CACTCTTCTTGCTTTG
>250
CAGTGTATTCGTGAAC
>251
AAGTGTTACAGAAATC
>252
TTTCTCTTTCTGCCGC
>253
TTGTCCCGGATGTGTG
>254
TTTCCAAAAGTACCCG
>255
AGAAATGAGAGAACCT
>256
ATTATTACAGGTCAAG
>257
ATTCGCATTACTGGCG
>258
ACCATACGTTAGAAGC
>259
ACGTTCGATCACAAAA
>260
ATAATTTTTACATTGA
>261
GAATCATTCAATATTA
>262
TAAGATCTTGAGAATG
>263
AATAATGTATGGTCGT
>264
TTCGGAAGGAGTCGAG
>265
AATACTGGCCAGGGGC
>266
TTTCTTAGTTATGCCC
>267
GCGCGATTTCTTGGTT
>268
GAAAGACCGTGATAGT
>269
CTCTATTGATGGCTCT
>270
GCAGTTCATACGTCGC
>271
AAGAAGTTTGGTACCA
>272
AAAAAATGATCCTTCA
>273
GTAATACAAATAAATT
>274
CGCAGTTTTGGATAAA
>275
GTCAACAATAATCCGC
>276
GGATAACTCCGGCCAT
>277
CGAATGCTTCTACCTT
>278
GAGATGTCGGAAAGAT
>279
TTGCCGTCAGTGTGAT
>280
AAAAAGTTTGGCAAGA
>281
GAACCAAGTAATTAGA
>282
AGATTAATCAATTCAA
>283
TTTTATATTAATCCTA
>284
TTTTCTTGTTTTATAT
>285
CTTGTTGTTTTTTTCT
>286
TGTTGTAACTGTCATG
>287
CGAAGGACGTAACTAT
>288
ACTCTGATATACGCTG
>289
TGAAATACGATACCAG
>290
CGTAATAGACGTCCAT
>291
GGTTGTGGTTGCCTTA
>292
TATTCCCAGGTTTCCC
>293
GGAGAGATATTGAGTT
>294
ACTTGACCCTACCTTC
>295
TCGGTCAAATTTCGCT
>296
GAAAAAGACAGAACTT

>297
CATTTCTTCATGAAAA
>298
CAATACCCAATGTCTC
>299
GACACGGCTTTCTTAG
>300
GGTAAGGGGTGCACCT
>301
CGATCTCATCTAACAA
>302
CAAACGAGTCTTGAAC
>303
ACAATAATACAAATTA
>304
CGATTAGTACCACAAA
>305
AATAATACTAAATGTA
>306
AGGTTGTTGCCGCGTT
>307
TGAGTTCATGGAATTT
>308
ATTTGTCTCTTTTCAC
>309
GCATTTTAATAACTAT
>310
AACTAACCAAACATTT
>311
GCAAACAAAAGAACAA
>312
ATAGAGACTATTTTTG
>313
GCTTTCTTTCTGTGTT
>314
TAGCACCCTTCGAATG
>315
AGGTGCCAACCACAGA
>316
TGATTCTGATGAGGTG
>317
TATAATCTTGCGATGT
>318
AATAAACTACCAAAAA
>319
GAATTAGATAAAACAA
>320
CGTACAACCCAATGCA
>321
GGGTAGGCTCCGTTAA
>322
AATCCCCTTCTGTGGA
>323
AGTTTTTTTCCAGCGC
>324
ATGTTGGCTTAGGCAA
>325
ACCTTGATTGTTGCAT
>326
GCGAATAAAGATATAC

>1.7.1-1
GAGTAATAGTTCACAA
>1.7.1-2
AAACAACCTAAGAAAA
>1.7.1-3
GGTATAGAGGCATCCG
>1.7.1-4
TGAATGGTCAACAGTT
>1.7.2-2
CGATCCTTAATAATGC
>1.7.2-3
TATGTATCTGATCAAC
>1.7.2-4
CCCGCAGAGACATCCC
>1.7.2-5
CTCAACCTGTGGACCA
>1.7.2-6
GCACAACAATAAATAC
>1.7.3-3
TAAAACTGGACGGGCG
>1.7.3-5
AACAATTGTTGATGTC
>1.7.3-6
GATCACTAATGAAAAT
>1.7.3-7
AGTTATGATGCAGCCG
>1.7.3-8
CTAGAGCTTTTTATGT
>1.7.4-2
AACATATCGATCAAAT
>1.7.4-3
CACCCAGACCGTTCAA
>1.7.4-4
GAAATCCTGATTCGTA
>1.7.4-5
TGAAAAGTTGATACGA
>1.7.4-6
AGTTGGGTAGGATGTC
>1.7.4-7
CAGCCTAATCCCCGGA
>1.7.4-8
GCATAAATCTTCCTTT
>1.7.4-9
AGTACCTCCACGATGT
>1.7.4-10
AAGTCGGACCATTGAG
>1.7.4-11
GCTTATACCCTTACTT
>1.7.4-12
AACACACTGTGATATC
>1.7.4-13
GCGGCGCGCGAATTTT
>1.7.4-15
AACAATAGATAGTAAA
>1.7.5-2
AATAGGCATAACAAGT
>1.7.5-3
AATGGCGTCGAAATCA
>1.7.5-5
TTTCACACAGCGTAAC
>1.7.6-2
TTCTTCCGATCCGATG
>1.7.6-3
CATAAATTCACAGGTC
>1.7.6-5
GCGAGGCTGGTTCATG
>1.7.6-6
ATAGGTTCTCATCTTG
>1.7.6-10
TTTGAATTACCCAACG
>1.7.6-12
CTCAATGAAAAAACTT
>1.7.6-13
ACGTTGCCGATTCTTG
>1.7.6-14
CGAAAAAGCGAGGGTC
>1.7.7-1
CAATATGATTGTATAC
>1.7.7-2
AAAAAAAGGACATTCA
>1.7.7-3
TCCTTTTGGGTGACGT
>1.7.7-4
TTATCCCAACACGAGC
>1.7.7-5
AATAAATATCAGGGTA
>1.7.8-2
TGCAGCGTGGTTAATT
>1.7.8-4
AGTAAAATAAGGTTAA
>1.7.8-5
CCGCCTGCAAAAATAA
>1.7.8-6
GTTTTAGACTGAAAGA
>1.7.8-7
TTGTGATATATTATAG
>1.7.9-1
TTAAACCAAGATAAAC
>1.7.9-2
CACTAGCTGTTAATGT
>1.7.9-3
AATTTAGTGCAAGACA
>1.7.9-4
ACCCGCGCCCTTCCGT
>1.7.9-5
CTAAACGACATCACTT
>1.7.9-6
CAACGTTAGCCGTCCA
>1.7.9-7
GTAACGTTCCGAGGGA
>1.7.9-8
GCTTAAACTATTTGAT
>1.7.9-9
CATATTCTTTTCGAGT
>1.7.10-2
GTTCACTCCTTGAGAT
>1.7.10-3
CCAAACAATTCTGACA
>1.7.11-1
ATACCAAGATAGAAAA
>1.7.11-2
TAGTTGCTAGTTTTAT
>1.7.11-3
ATGTATAGTAATGCGC
>1.7.11-5
GCATGAATTTTTTAAA
>1.7.11-6
ATGCACAACCGAAGGA
>1.7.11-7

```
GAAATCACGCGGCTAA
>1.7.11-8
ATAAGTTTAGGGATTT
>1.7.12-1
GTTGTTACTTTCTCAG
>1.7.12-3
AACCTCCGAATAAGTG
>1.7.12-5
TATTCGTTTAGTTTTT
>1.7.13-2
GGAGCTATTGTTTCGT
>1.7.13-3
AGATGGTCGGCGTTTG
```

A.5 Stock buffers: Oelgeschläger Lab Protocols (2012)

Unless otherwise stated all reagents were obtained from Sigma-Aldrich

Ampicillin

- 100mg/ml in $_{R_o}H_2O$.
- Store at -20˚C
- Use 100µg/ml for selection on agar plates/liquid medium

Chloramphenicol

- 34 mg/ml in ethanol.
- Store at -20°C.
- Use 34 µg/ml for selection on agar plates and 50 µg/ml in liquid medium

EDTA 0.5M, pH 8.0

For 500ml:

• Add 93.06g EDTA disodium (MW 372.24) to 400ml H_2O.

• Stir vigorously on a magnetic stirrer.

• Adjust the pH to 8.0 with NaOH pellets (~ 10g).

note: the disodium salt of EDTA will not go into solution until the pH approaches

~ pH 8.0

- Sterilize by autoclaving.
- Store at RT.

Ethidium Bromide (EtBr) 10mg/ml

• Add 1g ethidium bromide to 100ml H_2O.

Note: ethidium bromide is a known mutagen. Handle with caution.

• Stir vigorously on a magnetic stirrer to ensure that the dye has dissolved.

• Wrap the bottle in aluminium foil.

- Store at RT.
- Use 0.5µg/ml in agarose gel during electrophoresis.

LB medium (Luria-Bertani medium)

Composition	Stock Solutions	for 2L
Tryptone (Merck)	n/a	20g
Yeast extract (Merck)	n/a	10g
171mM NaCl	n/a	20g

- Add to 1900ml R_oH_2O
- Shake/stir until solutes are completely dissolved.
- Adjust pH to 7.0 with 5N NaOH (~ 0.2 ml)
- Adjust to a final volume of 2L
- Sterilize by autoclaving
- Store at RT.

SOC medium (Super Optimal Broth with Catabolite repression)

Composition	Stock Solutions	for 2L
Tryptone (Merck)	n/a	40g
Yeast extract (Merck)	n/a	10g
8.6mM NaCl	n/a	1g

- Add to 1900ml $_{Ro}H_2O$
- Shake/stir until solutes are completely dissolved.
- Adjust pH to 7.0 with 5N NaOH (~ 0.2 ml)
- Add 10 ml of 250mM KCl
- Adjust to a final volume of 2L
- Sterilize by autoclaving
- Store at RT.

IMPORTANT: Just before use add 5ml of sterile 2M $MgCl_2$ (10mM final) and 20 ml of 1M glucose.

5x TBE Buffer

Composition (1x)	Stock Solutions	for 2L
40 mM Tris•HCl	n/a	108.0g
45mM Boric acid	n/a	55g
1mM EDTA pH 8.0	0.5M	40ml

- Add to 1800ml $_{Ro}H_2O$
- Shake/stir until solutes are completely dissolved.
- Make up to 2L with $_{Ro}H_2O$
- store at RT

5x Tris-glycine electrophoresis buffer

Composition (1x)	Stock Solutions	for 2L
25 mM Tris•HCl	n/a	30.2g
190mM Glycine	n/a	188g
0.1% SDS	10%	100ml

- Dissolve 30.2 g Tris base and 188g of glycine in 1800ml $_{Ro}H_2O$.
- Add 100ml of 10% SDS stock solution and adjust volume to 2L.
- Store at RT.

A.6 Protein purification and western blot buffers: Oelgeschläger Lab Protocols

Unless otherwise stated all reagents were obtained from Sigma-Aldrich.

Note: All proteins were flash frozen and Stored at -80˚C. Aliquots were prepared to avoid repeated freeze/thaw cycles.

Lysozyme

• Prepare a stock solution at a concentration of 50 mg/ml in water.

• Dispense into aliquots and store at -20°C.

• Discard each aliquot after use.

2x BC-0

Composition	Stock Solutions	for 2L
40 mM Tris•HCl pH 7.37 at RT	1M	80ml
40% Glycerol	100%	800ml
0.4 mM EDTA pH 8.0	500mM	1.6ml

• adjust volume to 2L with $_{Ro}H2O$.

• sterilize by autoclaving.

• store at 4°C

Sonication and wash buffer

Composition	Stock Solutions	for 100ml
50 mM Tris•HCl pH 7.37 at RT	1M	80ml
500mM KCl	2M	20ml
20mM Imidazole	2M (pH 7.4)	1ml
0.5% NP40	10%	5ml

• adjust volume to 100ml with $_{Ro}$H2O.

• keep on ice or store at 4°C.

• add immediately before use:

70 µl β-EtSH (14.3 M), final concentration: 10mM

Elution Buffer

Composition	Stock Solutions	for 20ml
BC-0	2x	10ml
500mM KCl	2M	5ml
250mM Imidazole	2M (pH 7.4)	2.5ml
0.5% NP40	10%	1ml

• adjust volume to 20ml with Ro H2O.

• keep on ice or store at 4°C.

• add immediately before use:

14µl β-EtSH (14.3M), final concentration: 10mM

40µl protease inhibitor cocktail for purification of His:tagged proteins

(Sigma; P8849)

5x Transfer buffer

Composition (1x)	Stock Solutions	for 2L
25mM Tris-HCl	n/a	30.28g
192mM Glycine	n/a	144.13g

- Dissolve 30.28g Tris-HCl and 144.13g Glycine in 1800ml $_{Ro}H_2O$
- Make up to 2L with $_{Ro}H_2O$
- For 2L 1x buffer, add 400ml 100% Methanol (20% final conc.) to 400ml 5x buffer. Make up to 2L with $_{Ro}H_2O$

10x TBST

Composition (1x)	Stock Solutions	for 2L
50mM Tris-HCl	n/a	121.14g
150mM NaCl	n/a	175.32g
0.1% Tween 20	20%	n/a

- Dissolve 30.28g Tris-HCl and 144.13g Glycine in 1400ml $_{Ro}H_2O$
- For 1x buffer, add Tween 20 to 0.1% v/v
- Make up to 2L with $_{Ro}H_2O$

A7. Blood stage mRNA levels of PfTFIIA-γ1s and PfTFIIA-γ2s.

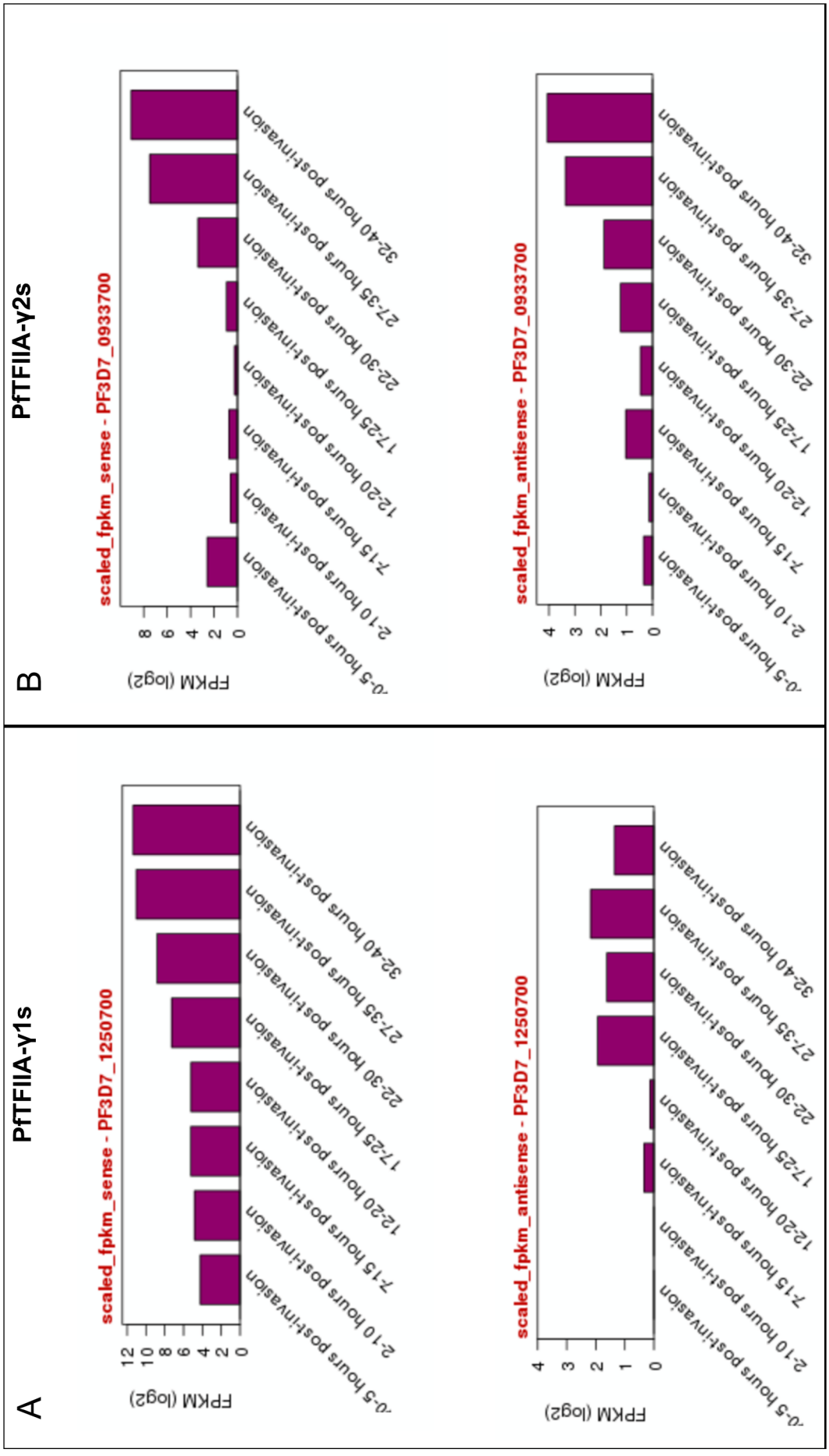


Fig. A8: mRNA profiles of sense and anti-sense transcripts for PfTFIIA-γ1s and PfTFIIA-γ2s. A. Blood stage mRNA levels for PfTFIIA-γ1s sense strand (top) and anti-sense strand (bottom). B. Blood stage mRNA levels for PfTFIIA-γ2s sense strand (top) and anti-sense strand (bottom). (http://plasmodb.org/plasmo/)

References

Adelman, K. and Lis, J.T., 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature Reviews Genetics*, 13(10), pp.720-731.

Adjalley, S.H., Chabbert, C.D., Klaus, B., Pelechano, V. and Steinmetz, L.M., 2016.Landscape and dynamics of transcription initiation in the malaria parasite Plasmodium falciparum. *Cell reports*, 14(10), pp.2463-2475.
Vancouver

Aguiar, J.C., LaBaer, J., Blair, P.L., Shamailova, V.Y., Koundinya, M., Russell, J.A., Huang, F., Mar, W., Anthony, R.M., Witney, A. and Caruana, S.R., 2004. High-throughput generation of P. falciparum functional molecules by recombinational cloning. *Genome research*, 14(10b), pp.2076-2082.

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

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T. and Ntini, E., 2014. An atlas of active enhancers across human cell types and tissues. *Nature*, 507(7493), pp. 455-461.

Andersson, R., Sandelin, A. and Danko, C.G., 2015. A unified architecture of transcriptional regulatory elements. *Trends in Genetics*, 31(8), pp.426-433.

Aravind, L., Iyer, L.M., Wellems, T.E. and Miller, L.H., 2003. Plasmodium biology: genomic gleanings. *Cell*, 115(7), pp. 771-785.

Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A.C., Khim, N., Kim, S., Duru, V., Bouchier, C., Ma, L. and Lim, P., 2014. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. *Nature*, 505(7481), pp.50-55.

Armache, K.-J., Kettenberger, H., and Cramer, P. 2003 Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc Natl Acad Sci USA* (100) 6964–6968.

Ashley, E.A., Recht, J. and White, N.J., 2014. Primaquine: the risks and the benefits. *Malaria journal*, 13(1), p.418.

Aurrecoechea, C., Brestelli, J., Brunk, B.P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O.S. and Heiges, M., 2009. PlasmoDB: a functional genomic database for malaria parasites. *Nucleic acids research*, 37(suppl 1), pp.D539-D543.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S., 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic acids research*, p.gkp335.

Baragaña, B., Hallyburton, I., Lee, M.C., Norcross, N.R., Grimaldi, R., Otto, T.D., Proto, W.R., Blagborough, A.M., Meister, S., Wirjanata, G. and Ruecker, A., 2015. A novel multiple-stage antimalarial agent that inhibits protein synthesis. *Nature*, 522(7556), pp.315-320.

Baum, J., Papenfuss, A.T., Baum, B., Speed, T.P. and Cowman, A.F., 2006. Regulation of apicomplexan actin-based motility. *Nature Reviews Microbiology*, 4(8), pp.621-628.

Bártfai, R., Balduf, C., Hilton, T., Rathmann, Y., Hadzhiev, Y., Tora, L., Orbán, L. and Müller, F., 2004. TBP2, a vertebrate-specific member of the TBP family, is required in embryonic development of zebrafish. *Current Biology*, 14(7), pp. 593-598.

Beier, J.C., 1998. Malaria parasite development in mosquitoes. *Annual review of entomology*, 43(1), pp.519-543.
Vancouver

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Sayers, E.W., 2013. GenBank. *Nucleic acids research*, 41(D1), pp.D36-D42.

Berk, A.J., 2000. TBP-like factors come into focus. *Cell*, 103(1), pp.5-8.

Buermeyer, A.B., Strasheim, L.A., McMahon, S.L., and Farnham, P.J. 1995.
Identification of cis-acting elements that can obviate a requirement for the C-terminal domain of RNA polymerase II. J *Biol Chem,* 270 pp.6798–6807.

Bing, S., 2015. Expression and initial characterisation of the Plasmodium falciparum general transcription factors TFIIB and TLP (Masters dissertation, University of Cape Town).

Birkholtz, L.M., Blatch, G., Coetzer, T.L., Hoppe, H.C., Human, E., Morris, E.J., Ngcete, Z., Oldfield, L., Roth, R., Shonhai, A. and Stephens, L., 2008. Heterologous expression of plasmodial proteins for structural studies and functional annotation. *Malaria journal*, 7(1), pp.197.

Bischoff, E. and 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(1), p.34.
Bleichenbacher, M. et al., 2003. Novel Interactions Between the Components of Human and Yeast TFIIA/TBP/DNA Complexes. *J. Mol. Biol.* (332), pp. 783–793.

Bozdech, Z., Llinás, M., Pulliam, B.L., Wong, E.D., Zhu, J. and DeRisi, J.L., 2003. The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. *PLoS Biol*, 1(1), pp.e5.

Bregman, D.B., Halaban, R., van Gool, A.J., Henning, K.A., Friedberg, E.C., and Warren, S.L. 1996. UV-induced ubiquitination of RNA polymerase II: a novel modification deficient in Cockayne syndrome cells. *Proc Natl Acad Sci USA*, (93), pp. 11586–11590.

Burke, T.W. and Kadonaga, J.T. 1996. Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes & Development*, (10), pp.711–724.

Burke, T.W. and Kadonaga, J.T. 1997. The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. *Genes & Development*, (11), pp.3020–3031.

Calderwood, M.S., Gannoun-Zaki, L., Wellems, T.E. and Deitsch, K.W., 2003. Plasmodium falciparum var genes are regulated by two regions with separate promoters, one upstream of the coding region and a second within the intron. *Journal of Biological Chemistry*, 278(36), pp. 34125-34132.

Callebaut, I., Prat, K., Meurice, E., Mornon, J.P. and Tomavo, S., 2005. Prediction of the general transcription factors associated with RNA polymerase II in Plasmodium falciparum: conserved features and differences relative to other eukaryotes. *BMC genomics*, 6(1), p.100.

Carrio, M.M. and Villaverde, A., 2002. Construction and deconstruction of bacterial inclusion bodies. *Journal of biotechnology*, 96(1), pp. 3-12.

Castaño, E., Gross, P., Wang, Z., Roeder, R.G. and Oelgeschläger, T., 2000. The C-terminal domain-phosphorylated IIO form of RNA polymerase II is associated with the transcription repressor NC2 (Dr1/DRAP1) and is required for transcription activation in human nuclear extracts. *Proceedings of the National Academy of Sciences*, 97(13), pp. 7184-7189.

Chalkley, G.E. and Verrijzer, C.P. 1999. DNA binding site selection by RNA polymerase II TAFs: a TAFII250-TAFII150 complex recognizes the initiator. *EMBO J,* 18, pp. 4835–4845.

Charman, S.A., Arbe-Barnes, S., Bathurst, I.C., Brun, R., Campbell, M., Charman, W.N., Chiu, F.C., Chollet, J., Craft, J.C., Creek, D.J. and Dong, Y., 2011. Synthetic ozonide drug candidate OZ439 offers new hope for a single-dose cure of uncomplicated malaria. *Proc. of the Nat. Aca. of Sci*, 108(11), pp.4400-4405.

Chasman, D.I., Flaherty, K.M., Sharp, P.A., and Kornberg, R.D. 1993. Crystal structure of yeast TATA-binding protein and model for interaction with DNA. *Proc. of the Nat. Aca. of Sci*, 90, pp.8174–8178.

Chicca, J.J., II, Auble, D.T., and Pugh, B.F. 1998. Cloning and biochemical characterization of TAF-172, a human homolog of yeast Mot1. *Mol. Cell. Biol*. 18 (3), pp. 1701–1710.

Chong, J.A., Moran, M.M., Teichmann, M., Kaczmarek, J.S., Roeder, R. and Clapham, D.E., 2005. TATA-binding protein (TBP)-like factor (TLF) is a functional regulator of transcription: reciprocal regulation of the neurofibromatosis type 1 and c-fos genes by TLF/TRF2 and TBP. *Mol. Cell. Biol*, 25(7), pp.2632-2643.

Chou, S., Chatterjee, S., Lee, M. and Struhl, K., 1999. Transcriptional activation in yeast cells lacking transcription factor IIA. *Genetics*, 153(4), pp.1573-1581.

Church, J. and Maitland, K., 2014. Invasive bacterial co-infection in African children with Plasmodium falciparum malaria: a systematic review. *BMC medicine*, 12(1), p.31.

Cianfrocco MA, Kassavetis GA, Grob P, Fang J, Juven-Gershon T, Kadonaga JT, Nogales E: Human., 2013 TFIID binds to core promoter DNA in a reorganized structural state. *Cell*, 152, pp.120-131.

Cinquin, O., Christopherson, R.I. and Menz, R.I., 2001. A hybrid plasmid for expression of toxic malarial proteins in Escherichia coli. *Molecular and biochemical parasitology*, 117(2), pp.245-247.

Cohen, S., McGregor, I.A. and Carrington, S., 1961. Gamma-globulin and acquired immunity to human malaria. *Nature*, 192, pp.733-7.

Cohen, J., Nussenzweig, V., Vekemans, J. and Leach, A., 2010. From the circumsporozoite protein to the RTS, S/AS candidate vaccine. *Human vaccines*, 6(1), pp.90-96.

Coleman, R.A., Taggart, A.K.P., Burma, S., Chicca, J.J., II, and Pugh, B.F. 1999. TFIIA regulates TBP and TFIID dimers. *Mol. Cell Biol,* 4, pp.451–457.

Collart, M.A. 1996. The NOT, SPT3, and MOT1 genes functionally interact to regulate transcription at core promoters. *Mol Cell Biol*, 16, pp.6668–6676.

Collins, C.R., Hackett, F., Strath, M., Penzo, M., Withers-Martinez, C., Baker, D.A. and Blackman, M.J., 2013. Malaria parasite cGMP-dependent protein kinase regulates blood stage merozoite secretory organelle discharge and egress. *PLoS Pathog*, 9(5), p.e1003344.

Comer, F.I. and Hart, G.W. 2001. Reciprocity between O-GlcNAc and Ophosphate on the carboxyl terminal domain of RNA polymerase II. *Biochemistry,* 40, pp. 7845–7852.

Cooper, G.M. and Hausman, R.E., 2000. *The Cell*, pp. 258-263. Sunderland: Sinauer Associates.

Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A. and Lis, J.T., 2014. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature genetics*, 46(12), pp.1311-1320.

Coteron, J.M., Marco, M., Esquivias, J., Deng, X., White, K.L., White, J., Koltun, M., El Mazouni, F., Kokkonda, S., Katneni, K. and Bhamidipati, R., 2011. Structure-guided lead optimization of triazolopyrimidine-ring substituents identifies potent Plasmodium falciparum dihydroorotate dehydrogenase inhibitors with clinical candidate potential. *Journal of medicinal chemistry*, 54(15), pp.5540-5561.

Coulson, R.M., Hall, N. and Ouzounis, C.A., 2004. Comparative genomics of transcriptional control in the human malaria parasite Plasmodium falciparum. *Genome research*, 14(8), pp.1548-1554.

Crowley, T.E., Hoey, T., Liu, J.K., Jan, Y.N., Jan, L.Y. and Tjian, R., 1993. A new factor related to TATA-binding protein has highly restricted expression patterns in Drosophila. *Nature,* 361(6412) pp.557-561

Cui, L. and Miao, J., 2010. Chromatin-mediated epigenetic regulation in the malaria parasite Plasmodium falciparum. *Eukaryotic cell*, 9(8), pp.1138-1149.

Dahmus, M.E. 1995. Phosphorylation of the C-terminal domain of RNA
polymerase II. *Biochim Biophys Acta,* 1261, pp.171–182.

Dantonel, J.C., Wurtz, J.M., Poch, O., Moras, D. and Tora, L., 1999. The TBP-like factor: an alternative transcription factor in metazoa?. *Trends in biochemical sciences,* 24(9), pp.335-339.

Dantonel, J.C., Quintin, S., Lakatos, L., Labouesse, M. and Tora, L., 2000. TBP-like factor is required for embryonic RNA polymerase II transcription in C. elegans. *Molecular cell*, 6(3), pp.715-722.

Das, S., Hertrich, N., Perrin, A.J., Withers-Martinez, C., Collins, C.R., Jones, M.L., Watermeyer, J.M., Fobes, E.T., Martin, S.R., Saibil, H.R. and Wright, G.J., 2015. Processing of Plasmodium falciparum merozoite surface protein msp1 activates a spectrin-binding function enabling parasite egress from RBCs. *Cell host & microbe*, 18(4), pp.433-444.

Decherf, G., Egée, S., Staines, H.M., Ellory, J.C. and Thomas, S.L., 2004. Anionic channels in malaria-infected human red blood cells. *Blood Cells, Molecules and Diseases*, 32(3), pp.366-371.

Deitsch, K.W., Calderwood, M.S. and Wellems, T.E., 2001. Malaria: Cooperative silencing elements in var genes. *Nature*, 412(6850), pp.875-876.

Deitsch, K., Duraisingh, M., Dzikowski, R., Gunasekera, A., Khan, S., Le Roch, K., Llinás, M., Mair, G., McGovern, V., Roos, D. and Shock, J., 2007. Mechanisms of gene regulation in Plasmodium. *The American journal of tropical medicine and hygiene*, 77(2), p.201.

DeJong, J., Bernstein, R. and Roeder, R.G., 1995. Human general transcription factor TFIIA: characterization of a cDNA encoding the small subunit and requirement for basal and activated transcription. *Proc. of the Nat. Acad. of Sci*, 92(8), pp.3313-3317.

DeJong, J. and Roeder, R.G., 1993. A single cDNA, hTFIIA/alpha, encodes both the p35 and p19 subunits of human TFIIA. *Genes & Development*, 7(11), pp.2220-2234.

Delgadillo, R.F., Whittington, J.E., Parkhurst, L.K. and Parkhurst, L.J., 2009. The TATA-binding protein core domain in solution variably bends TATA sequences via a three-step binding mechanism. *Biochemistry*, 48(8), pp.1801-1809.

Deng, W. and Roberts, S.G., 2005. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes & Development*, 19(20), pp.2418-2423.

Desai, S.A. and Rosenberg, R.L., 1997. Pore size of the malaria parasite's nutrient channel. *Proc. of the Nat. Acad. of Sci*, 94(5), pp.2045-2049.

Desai, S.A., Bezrukov, S.M. and Zimmerberg, J., 2000. A voltage-dependent channel involved in nutrient uptake by red blood cells infected with the malaria parasite. *Nature*, 406(6799), pp.1001-1005.

De Silva, E.K., Gehrke, A.R., Olszewski, K., Leo´ n, I., Chahal, J.S., Bulyk, M.L.,and Llina´ s, M. (2008). Specific DNA-binding by apicomplexan AP2 transcription factors. *Proc. Natl. Acad. Sci. USA* 105, 8393–8398.

Donati, D., Zhang, L.P., Chen, Q., Chêne, A., Flick, K., Nyström, M., Wahlgren, M. and Bejarano, M.T., 2004. Identification of a polyclonal B-cell activator in Plasmodium falciparum. *Infection and immunity*, 72(9), pp.5412-5418.

Dondorp, A.M., Ince, C., Charunwatthana, P., Hanson, J., Van Kuijen, A., Faiz, M.A., Rahman, M.R., Hasan, M., Yunus, E.B., Ghose, A. and Ruangveerayut, R., 2008. Direct in vivo assessment of microcirculatory dysfunction in severe falciparum malaria. J*ournal of Infectious Diseases*, 197(1), pp.79-84.

Dondorp, A.M., Nosten, F., Yi, P., Das, D., Phyo, A.P., Tarning, J., Lwin, K.M., Ariey, F., Hanpithakpong, W., Lee, S.J. and Ringwald, P., 2009. Artemisinin resistance in Plasmodium falciparum malaria. *New England Journal of Medicine*, 361(5), pp.455-467.

Douglas, A.D., Baldeviano, G.C., Lucas, C.M., Lugo-Roman, L.A., Crosnier, C., Bartholdson, S.J., Diouf, A., Miura, K., Lambert, L.E., Ventocilla, J.A. and Leiva, K.P., 2015. A PfRH5-based vaccine is efficacious against heterologous strain blood-stage Plasmodium falciparum infection in Aotus monkeys. *Cell host & microbe*, 17(1), pp.130-139.

Duraisingh, M.T., Triglia, T., Ralph, S.A., Rayner, J.C., Barnwell, J.W., McFadden, G.I. and Cowman, A.F., 2003. Phenotypic variation of Plasmodium falciparum merozoite proteins directs receptor targeting for invasion of human erythrocytes. *The EMBO journal*, 22(5), pp.1047-1057.
Vancouver

Duttke, S.H., Doolittle, R.F., Wang, Y.L. and Kadonaga, J.T., 2014. TRF2 and the evolution of the bilateria. *Genes & Development*, 28(19), pp.2071-2076.

Dvir, Arik, Siyuan Tan, Joan Weliky Conaway, and Ronald C. Conaway., 1997. Promoter Escape by RNA Polymerase II formation of an escape-competent transcriptional intermediate is a prerequisite for exit of polymerase from the promoter. *Journal of Biological Chemistry,* 272 (45) pp.28175-28178.

Dvorin, J.D., Martyn, D.C., Patel, S.D., Grimley, J.S., Collins, C.R., Hopp, C.S., Bright, A.T., Westenberger, S., Winzeler, E., Blackman, M.J. and Baker, D.A., 2010. A plant-like kinase in Plasmodium falciparum regulates parasite egress from erythrocytes. *Science*, 328(5980), pp.910-912.

Dynlacht, B.D., Hoey, T., and Tjian, R. 1991. Isolation of coactivators associated with the TATA-binding protein that mediate transcriptionalactivation. *Cell*, 66:563–576.

Dzikowski, R., Frank, M. and Deitsch, K., 2006. Mutually exclusive expression of virulence genes by malaria parasites is regulated independently of antigen production. *PLoS Pathog*, 2(3), p.e22.

Dzikowski, R., Li, F., Amulic, B., Eisberg, A., Frank, M., Patel, S., Wellems, T.E. and Deitsch, K.W., 2007. Mechanisms underlying mutually exclusive expression of virulence genes by malaria parasites. *EMBO reports*, 8(10), pp.959-965.

Fang, J. and McCutchan, T.F., 2002. Malaria: Thermoregulation in a parasite's life cycle. *Nature*, 418(6899), pp.742-742.

Fang, J., Zhou, H., Rathore, D., Sullivan, M., Su, X.Z. and McCutchan, T.F., 2004. Ambient glucose concentration and gene expression in Plasmodium falciparum. *Molecular and biochemical parasitology*, 133(1), pp.125-129.

Felgner, P.L., Roestenberg, M., Liang, L., Hung, C., Jain, A., Pablo, J., Nakajima-Sasaki, R., Molina, D., Teelen, K., Hermsen, C.C. and Sauerwein, R., 2013. Pre-erythrocytic antibody profiles induced by controlled human malaria infections in healthy volunteers under chloroquine prophylaxis. *Scientific reports*, 3, pp.3549-3549.

Fikes, J.D., Becker, D.M., Winston, F. and Guarente, L., 1990. Striking conservation of TFIID in Schizosaccharomyces pombe and Saccharomyces cerevisiae. *Nature*, 346(6281), p.291.

Florens, L., Washburn, M.P., Raine, J.D., Anthony, R.M., Grainger, M., Haynes, J.D., Moch, J.K., Muster, N., Sacci, J.B., Tabb, D.L. and Witney, A.A., 2002. A proteomic view of the Plasmodium falciparum life cycle. *Nature*, 419(6906), pp. 520-526.

Fontaine, F., Overman, J. and François, M., 2015. Pharmacological manipulation of transcription factor protein-protein interactions: opportunities and obstacles. *Cell Regeneration*, 4(1), p.2.

Freitas-Junior, L.H., Bottius, E., Pirrit, L.A., Deitsch, K.W., Scheidig, C., Guinet, F., Nehrbass, U., Wellems, T.E. and Scherf, A., 2000. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of P. falciparum. *Nature*, 407(6807), pp.1018-1022.

Gaboriaud, C., Bissery, V., Benchetrit, T. and Mornon, J.P., 1987. Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. *FEBS letters*, 224(1), pp.149-155.

Gamo, F.J., Sanz, L.M., Vidal, J., de Cozar, C., Alvarez, E., Lavandera, J.L., Vanderwall, D.E., Green, D.V., Kumar, V., Hasan, S. and Brown, J.R., 2010. Thousands of chemical starting points for antimalarial lead identification. *Nature*, 465(7296), pp.305-310.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S. and Paulsen, I.T., 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*, 419(6906).

Garner, M.M. and 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.

Garvie, C.W. and Wolberger, C., 2001. Recognition of specific DNA sequences. *Molecular cell*, 8(5), pp.937-946.

Gasch, A., Hoffmann, A., Horikoshi, M., Roeder, R.G. and Chua, N.H., 1990. Arabidopsis thaliana contains two genes for TFIID. *Nature*, 346(6282), p.390.

Ge, H., Martinez, E., Chiang, C.M. and Roeder, R.G., 1996. Activator-dependent transcription by mammalian RNA polymerase II: in vitro reconstitution with general transcription factors and cofactors. *Methods in enzymology*, 274, pp. 57-71.

Geiger, J.H., Hahn, S., Lee, S., and Sigler, P.B. 1996. Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science* 272, pp.830–836.

Gershenzon, N.I. and Ioshikhes, I.P. 2005. Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics,* 21, pp.1295–1300.

Gilbert, S.C., Plebanski, M., Gupta, S., Morris, J., Cox, M., Aidoo, M., Kwiatkowski, D., Greenwood, B.M., Whittle, H.C. and Hill, A.V., 1998. Association of malaria parasite population structure, HLA, and immunological antagonism. *Science*, 279(5354), pp.1173-1177.

Gilfillan, S., Stelzer, G., Piaia, E., Hofmann, M.G., and Meisterernst, M. 2005. Efficient binding of NC2 · TATA-binding protein to DNA in the absence of TATA. *J Biol Chem* 280, pp.6222–6230.

Gómez-Pérez, G.P., Van Bruggen, R., Grobusch, M.P. and Dobaño, C., 2014. Plasmodium falciparum malaria and invasive bacterial co-infection in young African children: the dysfunctional spleen hypothesis. *Malaria journal*, 13(1), p. 335.

Goppelt, A., Stelzer, G., Lottspeich, F., and Meisterernst, M. 1996. A mechanism for repression of class II gene transcription through specific binding of NC2 to TBP-promoter complexes via heterodimeric histone fold domains. *EMBO J*, 15, pp.3105–3116.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. and Lopez, R., 2010. A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic acids research*, 38(suppl 2), pp.W695-W699.

Gowda, D.C. and Davidson, E.A., 1999. Protein glycosylation in the malaria parasite. *Parasitology today*, 15(4), pp. 147-152.

Greenwood, B. and Doumbo, O.K., 2016. Implementation of the malaria candidate vaccine RTS, S/AS01. *The Lancet*, 387(10016), pp.318-319.

Gross, P. and Oelgeschläger, T., 2006, January. Core promoter-selective RNA polymerase II transcription. *Biochemical Society Symposia* (Vol. 73, pp. 225-236). Portland Press Limited.

Grünberg, Sebastian, Linda Warfield, and Steven Hahn., (2012). Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nature structural & molecular biology,* 19(8), pp.788-796.

Guiguemde, W.A., Shelat, A.A., Bouck, D., Duffy, S., Crowther, G.J., Davis, P.H., Smithson, D.C., Connelly, M., Clark, J., Zhu, F. and Jiménez-Díaz, M.B., 2010. Chemical genetics of Plasmodium falciparum. *Nature*, 465(7296), pp.311-315.

Gupta, A.P., Chin, W.H., Zhu, L., Mok, S., Luah, Y.H., Lim, E.H. and Bozdech, Z., 2013. Dynamic epigenetic regulation of gene expression during the life cycle of malaria parasite Plasmodium falciparum. *PLoS Pathog*, 9(2), p.e1003170.

Gupta, K., Sari-Ak, D., Haffke, M., Trowitzsch, S. and Berger, I., 2016. Zooming in on Transcription Preinitiation. *Journal of molecular biology*, 428(12), pp.2581-2591.

Hahn, S., 2004. Structure and mechanism of the RNA polymerase II transcription machinery. *Nature structural & molecular biology*, 11(5), p.394.

Hampsey, M. 1998. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev,* 62, pp.465–503.

He, Y., Fang, J., Taatjes, D.J. and Nogales, E., 2013. Structural visualization of key steps in human transcription initiation. *Nature*, 495(7442), pp.481-486.

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.

Hieb, A.R., Halsey, W.A., Betterton, M.D., Perkins, T.T., Kugel, J.F. and Goodrich, J.A., 2007. TFIIA changes the conformation of the DNA in TBP/TATA complexes and increases their kinetic stability. *Journal of molecular biology*, 372(3), pp.619-632.

Høiby, T., Zhou, H., Mitsiou, D.J. and Stunnenberg, H.G., 2007. A facelift for the general transcription factor TFIIA. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression,* 1769(7), pp.429-436.

Hobbs, N.K., Bondareva, A.A., Barnett, S., Capecchi, M.R. and Schmidt, E.E., 2002. Removing the vertebrate-specific TBP N terminus disrupts placental β2m-dependent interactions with the maternal immune system. *Cell*, 110(1), pp.43-54.

Hoey, T., Dynlacht, B.D., Peterson, M.G., Pugh, B.F. and Tjian, R., 1990. Isolation and characterization of the Drosophila gene encoding the TATA box binding protein, TFIID. *Cell*, 61(7), pp.1179-1186.

Hoffman, A., Sinn, E., Yamamoto, T., Wang, J., Roy, A., Horikoshi, M., and Roeder, R.G. 1990b. Highly conserved core domain and unique N terminus with presumptive regulatory motifs in a human TATA factor (TFIID). *Nature,* 346, 387–390.

Holstege, F.C., Fiedler, U. and Timmers, H.T.M., 1997. Three transitions in the RNA polymerase II transcription complex during initiation. *The EMBO journal*, 16(24), pp.7468-7480.

Hoopes, B.C., LeBlanc, J.F. and Hawley, D.K., 1992. Kinetic analysis of yeast TFIID-TATA box complex formation suggests a multi-step pathway. *Journal of Biological Chemistry*, 267(16), pp.11539-11547.

Hopkins, A.L. and Groom, C.R., 2002. The druggable genome. *Nature reviews Drug discovery*, 1(9), pp.727-730.

Horikoshi, M., Wang, C.K., Fujii, H., Cromlish, J.A., Weil, P.A. and Roeder, R.G., 1989. Cloning and structure of a yeast gene encoding a general transcription initiation factor TFIID that binds to the TATA box. *Nature*, 341(6240), p.299.

Horrocks, P. and Lanzer, M., 1999. Differences in nucleosome organization over episomally located plasmids coincides with aberrant promoter activity in P. falciparum. *Parasitology international*, 48(1), pp.55-61.

Horrocks, P., Wong, E., Russell, K. and Emes, R.D., 2009. Control of gene expression in Plasmodium falciparum–ten years on. *Molecular and biochemical parasitology*, 164(1), pp.9-25.

Hviid, L. and Jensen, A.T., 2015. Chapter Two-PfEMP1–A Parasite Protein Family of Key Importance in Plasmodium falciparum Malaria Immunity and Pathogenesis. *Advances in parasitology*, 88, pp.51-84.

Hyde, J.E., Kelly, S.L., Holloway, S.P., Snewin, V.A. and Sims, P.F., 1989. A general approach to isolating Plasmodium falciparum genes using non-redundant oligonucleotides inferred from protein sequences of other organisms. *Molecular and biochemical parasitology*, 32(2-3), pp.247-261.

Imbalzano, A.N., Kwon, H., Green, M.R. and Kingston, R.E., 1994. Facilitated binding of TATA-binding protein to nucleosomal DNA. *Nature*, 370(6489), p.481.

Imbalzano, A.N., Zaret, K.S. and Kingston, R.E., 1994b. 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.

Inostroza, J.A., Mermelstein, F.H., Ha, I., Lane, W.S., and Reinberg, D. 1992. Dr1, a TATA-binding protein-associated phosphoprotein and inhibitor of class II gene transcription. *Cell,* 70, pp.477–489.

Johnson, K.M., Wang, J., Smallwood, A. and Carey, M., 2004. The immobilized template assay for measuring cooperativity in eukaryotic transcription complex assembly. *Methods in enzymology*, 380, p.207.

Joice, R., Nilsson, S.K., Montgomery, J., Dankwa, S., Egan, E., Morahan, B., Seydel, K.B., Bertuccini, L., Alano, P., Williamson, K.C. and Duraisingh, M.T., 2014. Plasmodium falciparum transmission stages accumulate in the human bone marrow. *Science translational medicine*, 6(244), pp.244re5-244re5.

Kaltenbach, L., Horner, M.A., Rothman, J.H. and Mango, S.E., 2000. The TBP-like factor CeTLF is required to activate RNA polymerase II transcription during C. elegans embryogenesis. *Molecular cell*, 6(3), pp.705-713.

Kamada, K., Shu, F., Chen, H., Malik, S., Stelzer, G., Roeder, R.G., Meisterernst, M., and Burley, S.K. 2001. Crystal structure of negative cofactor 2 recognizing the TBP-DNA transcription complex. *Cell,* 106, pp.71–81.

Kandiah, E., Trowitzsch, S., Gupta, K., Haffke, M. and Berger, I., 2014. More pieces to the puzzle: recent structural insights into class II transcription initiation. *Current opinion in structural biology*, 24, pp.91-97.

Kane, J.F. and Hartley, D.L., 1988. Formation of recombinant protein inclusion bodies in Escherichia coli. *Trends in Biotechnology*, 6(5), pp.95-101.

Kang, J.J., Auble, D.T., Ranish, J.A., and Hahn, S. 1995. Analysis of yeast transcription factor TFIIA: distinct functional regions and a polymerase II-specific role in basal and activated transcription. *Mol Cell Biol* 15, pp.1234–1243.

Kantele, A. and Jokiranta, T.S., 2011. Review of cases with the emerging fifth human malaria parasite, Plasmodium knowlesi. *Clinical infectious diseases*, 52(11), pp.1356-1362.

Kao, C.C., Lieberman, P.M., Schmidt, M.C., Zhou, Q., Pei, R., and Berk, A.J. 1990. Cloning of a transcriptionally active human TATA binding factor. *Science,* 248, pp.1646–1650.

Kato, K., Makino, Y., Kishimoto, T., Yamauchi, J., Kato, S., Muramatsu, M. and Tamura, T.A., 1994. Multimerization of the mouse TATA-binding protein (TBP) driven by its C-terminal conserved domain. *Nucleic acids research*, 22(7), pp. 1179-1185.

Kedmi, A., Zehavi, Y., Glick, Y., Orenstein, Y., Ideses, D., Wachtel, C., Doniger, T., Ben-Asher, H.W., Muster, N., Thompson, J. and Anderson, S., 2014. Drosophila TRF2 is a preferential core promoter regulator. *Genes & Development*, 28(19), pp.2163-2174.

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

Kershnar, E.,Wu, S.-Y., and Chiang, C.-M. 1998. Immunoaffinity purification and functional characterization of human transcription factor IIH and RNA polymerase II from clonal cell lines that conditionally express epitope-tagged subunits of the multiprotein complexes. *J Biol Chem* 273, pp. 34444–34453.

Khunrae, P., Dahlbäck, M., Nielsen, M.A., Andersen, G., Ditlev, S.B., Resende, M., Pinto, V.V., Theander, T.G., Higgins, M.K. and Salanti, A., 2010. Full-length recombinant Plasmodium falciparum VAR2CSA binds specifically to CSPG and induces potent parasite adhesion-blocking antibodies. *Journal of molecular biology*, 397(3), pp.826-834.

Kim, J.L., Nikolov, D.B. and Burley, S.K., 1993. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, 365(6446), p.520.

Kim, J.L. and Burley, S.K., 1994. 1.9 Å resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nature Structural & Molecular Biology*, 1(9), pp.638-653.

Kim, T.K., Zhao,Y., Ge, H., Bernstein, R., and Roeder, R.G. 1995. TATA binding protein residues implicated in a functional interplay between negative cofactor NC2 (Dr1) and general factors TFIIA and TFIIB. *J Biol Chem,* 270, pp.10976–10981.

Kim, W.-Y. and Dahmus, M.E. 1989. The major late promoter of adenovirus-2 is accurately transcribed by RNA polymerase IIO, IIA, and IIB. *J Biol Chem,* 264, pp.3169–3176.

Kirkman, L.A. and Deitsch, K.W., 2012. Antigenic variation and the generation of diversity in malaria parasites. *Current opinion in microbiology*, 15(4), pp.456-462.

Klejman, M.P., Zhao, X., van Schaik, F.M., Herr, W. and Timmers, H.T.M., 2005. Mutational analysis of BTAF1–TBP interaction: BTAF1 can rescue DNA-binding defective TBP mutants. *Nucleic acids research*, 33(17), pp.5426-5436.

Kobayashi, N., Boyer, T.G., and Berk, A.J. 1995. A class of activation domains interacts directly with TFIIA and stimulates TFIIA-TFIID-promoter complex assembly. *Mol Cell Biol,* 15, pp.6465–6473.

Kokubo, T., Swanson, M.J., Nishikawa, J.-I., Hinnebusch, A.G., and Nakatani, Y. 1998. The yeast TAF145 inhibitory domain and TFIIA competitively bind to TATA-binding protein. *Mol Cell Biol,* 18, pp.1003–1012.

Kraemer, S.M., Ranallo, R.T., Ogg, R.C., and Stargell, L.A. 2001. TFIIA interacts with TFIID via association with TATA-binding protein and TAF40. *Mol Cell Biol,* 21, pp.1737–1746.

Krettli, A.U. and Dantas, L.A., 2000. Which routes do Plasmodium sporozoites use for successful infections of vertebrates? *Infection and immunity*, 68(5), pp.3064-3065.

Kravchenko, J.E., Rogozin, I.B., Koonin, E.V. and Chumakov, P.M., 2005. Transcription of mammalian messenger RNAs by a nuclear RNA polymerase of mitochondrial origin. *Nature*, 436(7051), pp.735-739.

Kuddus, R. and Schmidt, M.C., 1993. Effect of the non-conserved N-terminus on the DNA binding activity of the yeast TATA binding protein. *Nucleic acids research*, 21(8), pp.1789-1796.

Kwak, H. and Lis, J.T., 2013. Control of transcriptional elongation. *Annual review of genetics*, 47, pp.483-508.

Kwiatkowski, D., Sambou, I., Twumasi, P., Greenwood, B.M., Hill, A.V.S., Manogue, K.R., Cerami, A., Castracane, J. and Brewster, D.R., 1990. TNF concentration in fatal cerebral, non-fatal cerebral, and uncomplicated Plasmodium falciparum malaria. *The Lancet*, 336(8725), pp.1201-1204.

LaCount, D.J., Vignali, M., Chettier, R., Phansalkar, A., Bell, R., Hesselberth, J.R., Schoenfeld, L.W., Ota, I., Sahasrabudhe, S., Kurschner, C. and Fields, S., 2005. A protein interaction network of the malaria parasite Plasmodium falciparum. *Nature*, 438(7064), pp.103-107.

Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., and Ebright, R.H. 1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev,* 12, pp.34–44.

Lanzer, M., de Bruin, D. and Ravetch, J.V., 1992a. 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), p.1949.

Lanzer, M., de Bruin, D. and Ravetch, J.V., 1992b. 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.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R. and Thompson, J.D., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), pp. 2947-2948.

Lauer, S.A., Rathod, P.K., Ghori, N. and Haldar, K., 1997. A membrane network for nutrient import in red cells infected with the malaria parasite. *Science*, 276(5315), pp.1122-1125.

Lavstsen, T., Salanti, A., Jensen, A.T., Arnot, D.E. and Theander, T.G., 2003. Sub-grouping of Plasmodium falciparum 3D7 var genes based on sequence analysis of coding and non-coding regions. Malaria Journal, 2(1), p.27.
Langelier, M.-F., Forget, D., Rojas, A., Porlier, Y., Burton, Z.F., and Coulombe, B. 2001. Structural and functional interactions of transcription factor (TF) IIA with TFIIE and TFIIF in transcription initiation by RNA polymerase II. *J Biol Chem,* 276, pp.38652–38657.

Lagrange, T., Kim, T.K., Orphanides, G., Ebright, Y.W., Ebright, R.H., and Reinberg, D. 1996. High-resolution mapping of nucleoprotein complexes by site-specific protein-DNA photocrosslinking: organization of the human TBP-TFIIA-TFIIB-DNA quaternary complex. *Proc Natl Acad Sci USA,* 93, pp.10620–10625.

Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D. and Ebright, R.H., 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.

Lee, D.K., DeJong, J., Hashimoto, S., Horikoshi, M. and Roeder, R.G., 1992. TFIIA induces conformational changes in TFIID via interactions with the basic repeat. *Molecular and cellular biology*, 12(11), pp.5189-5196.

Lee, T.I. and Young, R.A. 2000. Transcription of eukaryotic protein-coding genes. *Annu Rev Genet,* 34:77–137.

Lee, D.H., Gershenzon, N., Gupta, M., Ioshikhes, I.P., Reinberg, D. and Lewis, B.A., 2005. Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. *Molecular and cellular biology*, 25(21), pp.9674-9686.

Lenhard, B., Sandelin, A. and Carninci, P., 2012. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature Reviews Genetics*, 13(4), pp.233-245.

Le Roch, K.G., Zhou, Y., Blair, P.L., Grainger, M., Moch, J.K., Haynes, J.D., De la Vega, P., Holder, A.A., Batalov, S., Carucci, D.J. and Winzeler, E.A., 2003. Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science*, 301(5639), pp.1503-1508.

Le Roch, K.G., Johnson, J.R., Florens, L., Zhou, Y., Santrosyan, A., Grainger, M., Yan, S.F., Williamson, K.C., Holder, A.A., Carucci, D.J. and Yates, J.R., 2004. Global analysis of transcript and protein levels across the Plasmodium falciparum life cycle. *Genome Research*, 14(11), pp.2308-2318.

Leung, D., Jung, I., Rajagopal, N., Schmitt, A., Selvaraj, S., Lee, A.Y., Yen, C.A., Lin, S., Lin, Y., Qiu, Y. and Xie, W., 2015. Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature*, 518(7539), pp.350-354.

Lewis, B.A., Kim, T.K. and Orkin, S.H., 2000. A downstream element in the human β-globin promoter: evidence of extended sequence-specific transcription factor IID contacts. Proceedings of the *National Academy of Sciences*, 97(13), pp.7172-7177.

Li, X.-Y., Virbasius, A., Zhu, X., and Green, M.R. 1999. Enhancement of TBP binding by activators and general transcription factors. *Nature,* 399, pp. 605–609.

Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J. and Sim, H.S., 2012. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell*, 148(1), pp.84-98.

Lieberman, P.M., Schmidt, M.C., Kao, C.C. and Berk, A.J., 1991. Two distinct domains in the yeast transcription factor IID and evidence for a TATA box-induced conformational change. *Molecular and Cellular Biology*, 11(1), pp.63-74.

Lim, C.Y., Santoso, B., Boulay, T., Dong, E., Ohler, U. and Kadonaga, J.T., 2004. The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes & Development*, 18(13), pp.1606-1617.

Lin, C.S., Uboldi, A.D., Epp, C., Bujard, H., Tsuboi, T., Czabotar, P.E. and Cowman, A.F., 2016. Multiple Plasmodium falciparum merozoite surface protein 1 complexes mediate merozoite binding to human erythrocytes. *Journal of Biological Chemistry*, 291(14), pp.7703-7715.

Liu, D., Ishima, R., Tong, K.I., Bagby, S., Kokubo, T., Muhandiram, D.R.,Kay, L.E., Nakatani, Y., and Ikura, M. 1998. Solution structure of aTBP-TAFII230 complex: protein mimicry of the minor groove surface of the TATA box unwound by TBP. *Cell,* 94, pp. 573–583.

Llanos-Cuentas, A., Lacerda, M.V., Rueangweerayut, R., Krudsood, S., Gupta, S.K., Kochar, S.K., Arthur, P., Chuenchom, N., Möhrle, J.J., Duparc, S. and Ugwuegbulam, C., 2014. Tafenoquine plus chloroquine for the treatment and relapse prevention of Plasmodium vivax malaria (DETECTIVE): a multicentre, double-blind, randomised, phase 2b dose-selection study. *The Lancet*, 383(9922), pp.1049-1058.

Luse, D.S. and Jacob, G.A., 1987. Abortive initiation by RNA polymerase II in vitro at the adenovirus 2 major late promoter. *Journal of Biological Chemistry*, 262(31), pp.14990-14997.

Maldonado, E., 1999. Transcriptional functions of a new mammalian TATA-binding protein-related factor. *Journal of Biological Chemistry*, 274(19), pp.12963-12966.

Mandal, S., Moudgil, M.N. and Mandal, S.K., 2009. Rational drug design. *European Journal of Pharmacology*, 625(1), pp.90-100.

Martianov, I., Fimia, G.M., Dierich, A., Parvinen, M., Sassone-Corsi, P. and Davidson, I., 2001. Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene. *Molecular Cell,* 7(3), pp.509-515.

Masters, K.M., Parkhurst, K.M., Daugherty, M.A. and Parkhurst, L.J., 2003. Native human TATA-binding protein simultaneously binds and bends promoter DNA without a slow isomerization step or TFIIB requirement. *Journal of Biological Chemistry*, 278(34), pp.31685-31690.

Malecová, B., Gross, P., Boyer-Guittaut, M., Yavuz, S. and Oelgeschläger, T., 2007. The initiator core promoter element antagonizes repression of TATA-directed transcription by negative cofactor NC2. *Journal of Biological Chemistry*, 282(34), pp.24767-24776.l

Marbach-Bar, N., Bahat, A., Ashkenazi, S., Golan-Mashiach, M., Haimov, O., Wu, S.Y., Chiang, C.M., Puzio-Kuter, A., Hirshfield, K.M., Levine, A.J. and Dikstein, R., 2016. DTIE, a novel core promoter element that directs start site selection in TATA-less genes. *Nucleic Acids Research*, 44(3), pp.1080-1094.

Martianov, I., Velt, A., Davidson, G., Choukrallah, M.A. and Davidson, I., 2016. TRF2 is recruited to the pre-initiation complex as a testis-specific subunit of TFIIA/ALF to promote haploid cell gene expression. *Scientific Reports*, 6.

Matsui, T., Segall, J., Weil, P.A. and Roeder, R.G., 1980. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *Journal of Biological Chemistry*, 255(24), pp.11992-11996.

McAndrew, M.B., Read, M., Sims, P.F. and Hyde, J.E., 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.

McLean, A.R.D., Ataide, R., Simpson, J.A., Beeson, J.G. and Fowkes, F.J.I., 2015. Malaria and immunity during pregnancy and postpartum: a tale of two species. *Parasitology*, 142(08), pp.999-1015.

McNamara, C.W., Lee, M.C., Lim, C.S., Lim, S.H., Roland, J., Nagle, A., Simon, O., Yeung, B.K., Chatterjee, A.K., McCormack, S.L. and Manary, M.J., 2013. Targeting Plasmodium PI (4) K to eliminate malaria. *Nature*, 504(7479), pp.248-253.

Mehlin, C., Boni, E., Buckner, F.S., Engel, L., Feist, T., Gelb, M.H., Haji, L., Kim, D., Liu, C., Mueller, N. and Myler, P.J., 2006. Heterologous expression of proteins from Plasmodium falciparum: results from 1000 genes. *Molecular and Biochemical Parasitology*, 148(2), pp.144-160.

Meister, S., Plouffe, D.M., Kuhen, K.L., Bonamy, G.M., Wu, T., Barnes, S.W., Bopp, S.E., Borboa, R., Bright, A.T., Che, J. and Cohen, S., 2011. Imaging of Plasmodium liver stages to drive next-generation antimalarial drug discovery. *Science*, 334(6061), pp.1372-1377.

Meisterernst, M. and Roeder, R.G., 1991a. Family of proteins that interact with TFIID and regulate promoter activity. *Cell*, 67(3), pp.557-567.

Miao, J., Fan, Q., Cui, L., Li, J., Li, J. and Cui, L., 2006. The malaria parasite Plasmodium falciparum histones: organization, expression, and acetylation. *Gene*, 369, pp.53-65.

Milton, R., 2012. Transcription regulation in Plasmodium falciparum (Pf): Heterologous expression of recombinant PfTBP and PfTFIIA. (Honours dissertation, University of Cape Town).

Min, I.M., Waterfall, J.J., Core, L.J., Munroe, R.J., Schimenti, J. and Lis, J.T., 2011. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes & Development*, 25(7), pp.742-754.

Miotto, O., Amato, R., Ashley, E.A., MacInnis, B., Almagro-Garcia, J., Amaratunga, C., Lim, P., Mead, D., Oyola, S.O., Dhorda, M. and Imwong, M., 2015. Genetic architecture of artemisinin-resistant Plasmodium falciparum. *Nature Genetics,* 47(3), pp.226-234.

Mnzava, A.P., Knox, T.B., Temu, E.A., Trett, A., Fornadel, C., Hemingway, J. and Renshaw, M., 2015. Implementation of the global plan for insecticide resistance management in malaria vectors: progress, challenges and the way forward. *Malaria Journal*, 14(1), p.173.

Mok, S., Ashley, E.A., Ferreira, P.E., Zhu, L., Lin, Z., Yeo, T., Chotivanich, K., Imwong, M., Pukrittayakamee, S., Dhorda, M. and Nguon, C., 2015. Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. *Science*, 347(6220), pp.431-435.

Moore, P.A., Ozer, J., Salunek, M., Jan, G., Zerby, D., Campbell, S. and Lieberman, P.M., 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.

Mudeppa DG, Pang CK, Tsuboi T, Endo Y, Buckner FS, Varani G, Rathod PK. Cell-free production of functional Plasmodium falciparum dihydrofolate reductase-thymidylate synthase. *Molecular and biochemical parasitology*. 2007 Feb 28;151(2), pp.216-9.

Murakami, K., Elmlund, H., Kalisman, N., Bushnell, D.A., Adams, C.M., Azubel, M., Elmlund, D., Levi-Kalisman, Y., Liu, X., Gibbons, B.J. and Levitt, M., 2013. Architecture of an RNA polymerase II transcription pre-initiation complex. *Science*, 342(6159), p.1238724.

Murphy, F.V. and Churchill, M.E., 2000. Nonsequence-specific DNA recognition: a structural perspective. *Structure*, 8(4), pp.R83-R89.

Nakadai, T., Shimada, M., Shima, D., Handa, H. and Tamura, T.A., 2004. Specific interaction with transcription factor IIA and localization of the mammalian TATA-binding protein-like protein (TLP/TRF2/TLF). J*ournal of Biological Chemistry*, 279(9), pp.7447-7455.
Vancouver

Nikolov, D.B., Hu, S.-H., Lin, J., Gasch, A., Hoffman, A., Horikoshi, M., Chua, N.-H., Roeder, R.G., and Burley, S.K. 1992. Crystal structure of TFIID TATA-box binding protein. *Nature,* 360, pp.40–46.

Nikolov, D.B., Chen, H., Halay, E.D. and Usheva, A.A., 1995. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature*, 377(6545), p.119.

Nikolov, D.B., Chen, H., Halay, E.D., Hoffman, A., Roeder, R.G. and Burley, S.K., 1996. Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc. of the Nati. Acad. of Sci*., 93(10), pp.4862-4867.

Ogawa, N. and Biggin, M.D., 2012. High-throughput SELEX determination of DNA sequences bound by transcription factors in vitro. Gene Regulatory Networks: *Methods and Protocols*, pp.51-63.

Ohbayashi, T., Shimada, M., Nakadai, T., Wada, T., Handa, H. and Tamura, T., 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.

Ohler, U., Liao, G.C., Niemann, H. and Rubin, G.M., 2002. Computational analysis of core promoters in the Drosophila genome. *Genome Biology*, 3(12), pp.research0087-1.

Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O. and Pikaard, C.S., 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell*, 120(5), pp.613-622.

Orphanides, G., Lagrange, T. and Reinberg, D., 1996. The general transcription factors of RNA polymerase II. *Genes & Development*, 10(21), pp.2657-2683.

O'neill, P.M., 2004. Medicinal chemistry: a worthy adversary for malaria. *Nature*, 430(7002), pp.838-839.

Ozer, J., Moore, P.A., Bolden, A.H., Lee, A., Rosen, C.A. and Lieberman, P.M., 1994. Molecular cloning of the small (gamma) subunit of human TFIIA reveals functions critical for activated transcription. *Genes & Development*, 8(19), pp.2324-2335.

Ozer, J., Lezina, L.E., Ewing, J., Audi, S., and Lieberman, P.M. 1998. Association of transcription factor IIA with TBP is required for transcriptional activation of a subset of promoters and cell cycle progression in S. cerevisiae. *Mol Cell Biol* , 18, pp.2559–2570.

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

Papai, G., Weil, P.A. and Schultz, P., 2011. New insights into the function of transcription factor TFIID from recent structural studies. *Current opinion in genetics & development*, 21(2), pp.219-224.

Pei, X., An, X., Guo, X., Tarnawski, M., Coppel, R. and Mohandas, N., 2005. Structural and functional studies of interaction between Plasmodium falciparum knob-associated histidine-rich protein (KAHRP) and erythrocyte spectrin. *Journal of Biological Chemistry*, 280(35), pp.31166-31171.

Pelechano, V. and Steinmetz, L.M., 2013. Gene regulation by antisense transcription. *Nature reviews: Genetics,* 14(12), p.880.

Penny, M.A., Verity, R., Bever, C.A., Sauboin, C., Galactionova, K., Flasche, S., White, M.T., Wenger, E.A., Van de Velde, N., Pemberton-Ross, P. and Griffin, J.T., 2016. Public health impact and cost-effectiveness of the RTS, S/AS01 malaria vaccine: a systematic comparison of predictions from four mathematical models. *The Lancet*, 387(10016), pp.367-375.

Pereira, L.A., van der Knaap, J.A., van den Boom, V., van den Heuvel, F.A.J., and Timmers, H.T.M. 2001. TAFII170 interacts with the concave surface of TATA-binding protein to inhibit its DNA binding activity. *Mol Cell Bio*l, 21, pp.7523–7534.

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.

Petri, V., Hsieh, M. and Brenowitz, M., 1995. Thermodynamic and kinetic characterization of the binding of the TATA binding protein to the adenovirus E4 promoter. *Biochemistry*, 34(31), pp.9977-9984.

Phillips, M.A., Lotharius, J., Marsh, K., White, J., Dayan, A., White, K.L., Njoroge, J.W., El Mazouni, F., Lao, Y., Kokkonda, S. and Tomchick, D.R., 2015. A long-duration dihydroorotate dehydrogenase inhibitor (DSM265) for prevention and treatment of malaria. *Science Translational Medicine*, 7(296), pp.296ra111-296ra111.

Pikaard, C.S., Haag, J.R., Ream, T. and Wierzbicki, A.T., 2008. Roles of RNA polymerase IV in gene silencing. *Trends in plant science*, 13(7), pp.390-397.

Pizzi, E. and Frontali, C., 2001. Low-complexity regions in Plasmodium falciparum proteins. *Genome Research*, 11(2), pp.218-229.

Plattner, F. and Soldati-Favre, D., 2008. Hijacking of host cellular functions by the Apicomplexa. Annu. *Rev. Microbiol.*, 62, pp.471-487.

Protein Data Bank (PDB) (http://www.rcsb.org/pdb/static.do?p=general_information/about_pdb/index.html)
Peterson, M.G., Tanese, N., Pugh, B.F., and Tjian, R. 1990. Functional domains and upstream activation properties of cloned human TATA binding protein. *Science* 248, pp.1625–1630.

Prudêncio, M., Rodriguez, A. and Mota, M.M., 2006. The silent path to thousands of merozoites: the Plasmodium liver stage. *Nature Reviews Microbiology*, 4(11), pp.849-856.

Pugh, B.F. and Tjian, R. 1991. Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes & Development,* 5, pp.1935–1945.

QIAexpressionist, A., 2002. A handbook for high-level expression and purification of 6xhis-tagged proteins. Qiagen. pp. 1-125.

Rabenstein, M.D., Zhou, S., Lis, J.T. and Tjian, R., 1999. TATA box-binding protein (TBP)-related factor 2 (TRF2), a third member of the TBP family. *Proc.of the Nat. Acad. of Sci.*, 96(9), pp.4791-4796.

Ralph, S.A., Scheidig-Benatar, C. and Scherf, A., 2005. Antigenic variation in Plasmodium falciparum is associated with movement of var loci between subnuclear locations. *Proc.of the Nat. Acad. of Sci.*, 102(15), pp.5414-5419.

Ralph, S.A. and Scherf, A., 2005. The epigenetic control of antigenic variation in Plasmodium falciparum. *Current opinion in microbiology*, 8(4), pp.434-440.

Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J.K., Hagen, G., Guilfoyle, T.J., Paša-Tolić, L. and Pikaard, C.S., 2009. Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Molecular cell*, 33(2), pp.192-203.

Reddy, P. and Hahn, S., 1991. Dominant negative mutations in yeast TFIID define a bipartite DNA-binding region. *Cell*, 65(2), pp.349-357.

Reinberg, D., Horikoshi, M., and Roeder, R.G. 1987. Factors involved in specific transcription in mammalian RNA polymerase II. Functional analysis of initiation factors IIA and IID and identification of a new factor operating at sequences downstream of the initiation site. *J Biol Chem*, 262, pp. 3322–3330.

Riley, E.M. and Stewart, V.A., 2013. Immune mechanisms in malaria: new insights in vaccine development. *Nature medicine*, 19(2), pp.168-178.

Robinson, M.M., Yatherajam, G., Ranallo, R.T., Bric, A., Paule, M.R., and Stargell, L.A. 2005. Mapping and functional characterization of the TAF11 interaction with TFIIA. *Mol Cell Biol,* 25, pp.945–957.

Roeder, R.G. and Rutter, W.J. 1969. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature,* 224, pp.234–237.

Roeder, R.G. and Rutter, W.J. 1970. Specific nucleolar and nucleoplasmic RNA polymerases. *Proc Natl Acad Sci USA*, 65, pp.675–682.

Roy, A.L. and Singer, D.S., 2015. Core promoters in transcription: old problem, new insights. *Trends in Biochemical Sciences*, 40(3), pp.165-171.

Ruvalcaba-Salazar, O.K., del Carmen Ramírez-Estudillo, M., Montiel-Condado, D., Recillas-Targa, F., Vargas, M. and Hernández-Rivas, R., 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.183-196.

Ruvalcaba-Salazar, O.K., Romero-Ramírez, H., Santos-Argumedo, L., Vargas, M. and Hernández-Rivas, R., 2006. Preparation and characterization of a monoclonal antibody specific to Plasmodium falciparum TATA binding protein. *Hybridoma and Hybridomics*, 25(6), pp.367-371.

Sachs, J. and Malaney, P., 2002. The economic and social burden of malaria. *Nature*, 415(6872), pp.680-685.

Salmon, B.L., Oksman, A. and Goldberg, D.E., 2001. Malaria parasite exit from the host erythrocyte: a two-step process requiring extraerythrocytic proteolysis. *Proc. of the Nat. Acad. of Sci.*, 98(1), pp.271-276.

Sambrook, J. and Russel, D. W. 2001. Molecular Cloning: A Laboratory Manual. Coldspring harbor laboratory press.

Saunders, A., Core, L.J. and Lis, J.T., 2006. Breaking barriers to transcription elongation. *Nature Reviews Molecular Cell Biology*, 7(8), pp.557-567.

Sawadogo, M. and Roeder, R.G., 1985. Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative in vitro assay. *Proc. of the Nat. Acad. of Sci.*, 82(13), pp.4394-4398.

Scherf, A., Lopez-Rubio, J.J. and Riviere, L., 2008. Antigenic variation in Plasmodium falciparum. *Annu. Rev. Microbiol*., 62, pp.445-470.

Scherf, A., Hernandez-Rivas, R., Buffet, P., Bottius, E., Benatar, C., Pouvelle, B., Gysin, J. and Lanzer, M., 1998. Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in Plasmodium falciparum. *The EMBO journal*, 17(18), pp.5418-5426.

Seder, R.A., Chang, L.J., Enama, M.E., Zephir, K.L., Sarwar, U.N., Gordon, I.J., Holman, L.A., James, E.R., Billingsley, P.F., Gunasekera, A. and Richman, A., 2013. Protection against malaria by intravenous immunization with a nonreplicating sporozoite vaccine. *Science*, 341(6152), pp.1359-1365.

Shao, H., Revach, M., Moshonov, S., Tzuman, Y., Gazit, K., Albeck, S., Unger, T. and Dikstein, R., 2005. Core promoter binding by histone-like TAF complexes. *Molecular and Cellular Biology*, 25(1), pp.206-219.

Shimizu, Y., Kuruma, Y., Ying, B.W., Umekage, S. and Ueda, T., 2006. Cell-free translation systems for protein engineering. *FEBS Journal*, 273(18), pp.4133-4140.

Shin, S.C.J., Vanderberg, J.P. and Terzakis, J.A., 1982. Direct infection of hepatocytes by sporozoites of Plasmodium berghei. *The Journal of protozoology*, 29(3), pp.448-454.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J. and Thompson, J.D., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(1), p.539.
Vancouver

Slattery, M., Zhou, T., Yang, L., Machado, A.C.D., Gordân, R. and Rohs, R., 2014. Absence of a simple code: how transcription factors read the genome. *Trends in Biochemical Sciences*, 39(9), pp.381-399.

Smale, S.T. and Baltimore, D. (1989) The ‘‘Initiator’’ as a transcription control element. *Cell,* 57, pp.103–113

Smale, S.T. and Kadonaga, J.T. 2003. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* 72 pp.449–479.

Smith, J.D., Rowe, J.A., Higgins, M.K. and Lavstsen, T., 2013. Malaria's deadly grip: cytoadhesion of Plasmodium falciparum–infected erythrocytes. *Cellular Microbiology*, 15(12), pp.1976-1983.

Solow, S., Salunek, M., Ryan, R., and Lieberman, P.M. 2001. TAFII250 phosphorylates human transcription factor IIA on serine residues important for TBP binding and transcription activity. *J Biol Chem,* 276, pp.15886–15892.

Spangenberg, T., Burrows, J.N., Kowalczyk, P., McDonald, S., Wells, T.N. and Willis, P., 2013. The open access malaria box: a drug discovery catalyst for neglected diseases. *PloS one*, 8(6), pp.e62906.

Spring, M., Murphy, J., Nielsen, R., Dowler, M., Bennett, J.W., Zarling, S., Williams, J., de la Vega, P., Ware, L., Komisar, J. and Polhemus, M., 2013. First-in-human evaluation of genetically attenuated Plasmodium falciparum sporozoites administered by bite of Anopheles mosquitoes to adult volunteers. *Vaccine*, 31(43), pp.4975-4983.

Straimer, J., Gnädig, N.F., Witkowski, B., Amaratunga, C., Duru, V., Ramadani, A.P., Dacheux, M., Khim, N., Zhang, L., Lam, S. and Gregory, P.D., 2015. K13-propeller mutations confer artemisinin resistance in Plasmodium falciparum clinical isolates. *Science*, 347(6220), pp.428-431.

Sturm, A., Amino, R., Van de Sand, C., Regen, T., Retzlaff, S., Rennenberg, A., Krueger, A., Pollok, J.M., Menard, R. and Heussler, V.T., 2006. Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. *Science*, 313(5791), pp.1287-1290.

Su, X.Z., Wu, Y., Sifri, C.D. and Wellems, T.E., 1996. Reduced extension temperatures required for PCR amplification of extremely A+ T-rich DNA. *Nucleic Acids Research*, 24(8), pp.1574-1575.

Sun, X., Ma, D., Sheldon, M., Yeung, K. and Reinberg, D., 1994. Reconstitution of human TFIIA activity from recombinant polypeptides: a role in TFIID-mediated transcription. *Genes & Development*, 8(19), pp.2336-2348.

Suzuki, Y., Tsunoda, T., Sese, J., Taira, H., Mizushima-Sugano, J., Hata, H.,Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., Suyama, A., Sakaki, Y.,Morishita, S., Okubo, K., and Sugano, S. 2001. Identification and characterization of the potential promoter regions of 1031 kinds of human genes. *Genome Res.* 11, pp. 677–684.

Talvik, G., 2016. Transcription regulation in Plasmodium falciparum: functional characterisation of general transcription factor IIB (Doctoral dissertation, University of Cape Town).

Tan, S., Hunziker, Y., Sargent, D.F. and Richmond, T.J., 1996. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature*, 381(6578), pp.127.

Tavares, J., Formaglio, P., Thiberge, S., Mordelet, E., Van Rooijen, N., Medvinsky, A., Ménard, R. and Amino, R., 2013. Role of host cell traversal by the malaria sporozoite during liver infection. *Journal of Experimental Medicine*, 210(5), pp. 905-915.

Taylor, H.M., Kyes, S.A. and Newbold, C.I., 2000. Var gene diversity in Plasmodium falciparum is generated by frequent recombination events. *Molecular and Biochemical Parasitology*, 110(2), pp.391-397.

Teichmann, M., Wang, Z., Martinez, E., Tjernberg, A., Zhang, D., Vollmer, F., Chait, B.T. and Roeder, R.G., 1999. Human TATA-binding protein-related factor-2 (hTRF2) stably associates with hTFIIA in HeLa cells. *Proc. of the Nat. Acad. of Sci*, 96(24), pp.13720-13725.

Theisen, M., Roeffen, W., Singh, S.K., Andersen, G., Amoah, L., van de Vegte-Bolmer, M., Arens, T., Tiendrebeogo, R.W., Jones, S., Bousema, T. and Adu, B., 2014. A multi-stage malaria vaccine candidate targeting both transmission and asexual parasite life-cycle stages. *Vaccine*, 32(22), pp.2623-2630.

Thomas, M.C. and Chiang, C.M., 2006. The general transcription machinery and general cofactors. *Critical Reviews in Biochemistry and Molecular Biology*, 41(3), pp.105-178.

Tilley, L., Sougrat, R., Lithgow, T. and Hanssen, E., 2008. The Twists and Turns of Maurer's Cleft Trafficking in P. falciparum–Infected Erythrocytes. *Traffic*, 9(2), pp.187-197.

Timmers, H.T.M. and Sharp, P.A. 1991. The mammalian TFIID protein is present in two functionally distinct complexes. *Genes & Development*, 5, pp.946–1956.

Timmers, H.T.M., Meyers, R.E., and Sharp, P.A. 1992. Composition of transcription factor B-TFIID. *Proc Natl Acad Sci ,* 89, pp 8140–8144.

Tolia, N.H. and Joshua-Tor, L., 2006. Strategies for protein coexpression in Escherichia coli. *Nature methods*, 3(1), pp. 55-64.

Triglia, T. and Kemp, D.J., 1991. Large fragments of Plasmodium falciparum DNA can be stable when cloned in yeast artificial chromosomes. *Molecular and Biochemical Parasitology*, 44(2), pp.207-211.

Tsuboi, T., Takeo, S., Iriko, H., Jin, L., Tsuchimochi, M., Matsuda, S., Han, E.T., Otsuki, H., Kaneko, O., Sattabongkot, J. and Udomsangpetch, R., 2008. Wheat germ cell-free system-based production of malaria proteins for discovery of novel vaccine candidates. *Infection and immunity*, 76(4), pp.1702-1708.

Urban, B.C., Ferguson, D.J., Pain, A., Willcox, N., Plebanski, M., Austyn, J.M. and Roberts, D.J., 1999. Plasmodium falciparum-infected erythrocytes modulate the maturation of dendritic cells. *Nature*, 400(6739), pp.73-77.

Van Dyke, M.W., Roeder, R.G. and Sawadogo, M., 1988. Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. *Science*, 241(4871), pp.1335.

Vedadi, M., Lew, J., Artz, J., Amani, M., Zhao, Y., Dong, A., Wasney, G.A., Gao, M., Hills, T., Brokx, S. and Qiu, W., 2007. Genome-scale protein expression and structural biology of Plasmodium falciparum and related Apicomplexan organisms. *Molecular and biochemical parasitology*, 151(1), pp.100-110.

Veenstra, G.J.C., Weeks, D.L. and Wolffe, A.P., 2000. Distinct roles for TBP and TBP-like factor in early embryonic gene transcription in Xenopus. *Science*, 290(5500), pp.2312-2315.

Vennerstrom, J.L., Arbe-Barnes, S., Brun, R., Charman, S.A., Chiu, F.C., Chollet, J., Dong, Y., Dorn, A., Hunziker, D., Matile, H. and McIntosh, K., 2004. Identification of an antimalarial synthetic trioxolane drug development candidate. *Nature*, 430(7002), pp.900-904.

Voss, T.S., Healer, J., Marty, A.J., Duffy, M.F., Thompson, J.K., Beeson, J.G., Reeder, J.C., Crabb, B.S. and Cowman, A.F., 2006. A var gene promoter controls allelic exclusion of virulence genes in Plasmodium falciparum malaria. *Nature*, 439(7079), pp.1004-1008.

Voss, T.S., Tonkin, C.J., Marty, A.J., Thompson, J.K., Healer, J., Crabb, B.S. and Cowman, A.F., 2007. Alterations in local chromatin environment are involved in silencing and activation of subtelomeric var genes in Plasmodium falciparum. *Molecular microbiology*, 66(1), pp.139-150.

Vugt, M.V., Wilairatana, P., Gemperli, B., Gathmann, I., Phaipun, L., Brockman, A., Luxemburger, C., White, N.J., Nosten, F. and Looareesuwan, S., 1999. Efficacy of six doses of artemether-lumefantrine (benflumetol) in multidrug-resistant Plasmodium falciparum malaria. *The American journal of tropical medicine and hygiene*, 60(6), pp.936-942.

Wang, Y.L., Duttke, S.H., Chen, K., Johnston, J., Kassavetis, G.A., Zeitlinger, J. and Kadonaga, J.T., 2014. TRF2, but not TBP, mediates the transcription of ribosomal protein genes. *Genes & development*, 28(14), pp.1550-1555.

Wang, J., Zhao, S., He, W., Wei, Y., Zhang, Y., Pegg, H., Shore, P., Roberts, S.G. and Deng, W., 2017. A Transcription Factor IIA Binding Site Differentially Regulates RNA Polymerase II-Mediated Transcription in a Promoter Context-dependent Manner. *Journal of Biological Chemistry*, pp. jbc-M116.

Warfield, L., Ranish, J.A., and Hahn, S. 2004. Positive and negative functions of the SAGA complex mediated through interaction of Spt8 with TBP and the N-terminal domain of TFIIA. *Genes & Development*, 18, pp. 1022–1034.

Wassmer, S.C., Taylor, T.E., Rathod, P.K., Mishra, S.K., Mohanty, S., Arevalo-Herrera, M., Duraisingh, M.T. and Smith, J.D., 2015. Investigating the pathogenesis of severe malaria: a multidisciplinary and cross-geographical approach. The *American journal of tropical medicine and hygiene*, 93(3 Suppl), pp.42-56.

Weinmann, R. and Roeder, R.G. 1974. Role of DNA-dependent RNA polymerase III in the transcription of the tRNA and 5S RNA genes. *Proc Natl Acad Sci USA,* 71, pp.1790–1794.

Weil, P.A., Luse, D.S., Segall, J. and Roeder, R.G., 1979. Selective and accurate initiation of transcription at the Ad2 major late promotor in a soluble system dependent on purified RNA polymerase II and DNA. *Cell*, 18(2), pp.469-484.

Weiss, G.E., Gilson, P.R., Taechalertpaisarn, T., Tham, W.H., de Jong, N.W., Harvey, K.L., Fowkes, F.J., Barlow, P.N., Rayner, J.C., Wright, G.J. and Cowman, A.F., 2015. Revealing the sequence and resulting cellular morphology of receptor-ligand interactions during Plasmodium falciparum invasion of erythrocytes. *PLoS Pathog*, 11(2), pp. e1004670.

Weng, H., Guo, X., Papoin, J., Wang, J., Coppel, R., Mohandas, N. and An, X., 2014. Interaction of Plasmodium falciparum knob-associated histidine-rich protein (KAHRP) with erythrocyte ankyrin R is required for its attachment to the erythrocyte membrane. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1838(1), pp.185-192.

Wang, J., Zhao, S., He, W., Wei, Y., Zhang, Y., Pegg, H., Shore, P., Roberts, S.G. and Deng, W., 2017. A Transcription Factor IIA Binding Site Differentially Regulates RNA Polymerase II-Mediated Transcription in a Promoter Context-dependent Manner. *Journal of Biological Chemistry*, pp. jbc-M116.

Willy, P.J., Kobayashi, R., and Kadonaga, J.T. 2000. A basal transcription factor that activates or represses transcription. *Science,* 290, pp.982–985.

Winzeler, E.A. and Manary, M.J., 2014. Drug resistance genomics of the antimalarial drug artemisinin. *Genome biology*, 15(11), p.544.

World Health Organization, 2016. *World malaria report 2016*. Geneva: WHO.

Wu, S.-Y. and Chiang, C.-M. 1998. Properties of PC4 and an RNA polymerase II complex in directing activated and basal transcription in vitro. *J Biol Chem*, 273 pp.12492–12498.

Wu, S.-Y., Kershnar, E., and Chiang, C.-M. 1998. TAFII-independent activation mediated by human TBP in the presence of the positive cofactor PC4. *EMBO J,* 17, pp.4478–4490.

Wu, Y., Sinden, R.E., Churcher, T.S., Tsuboi, T. and Yusibov, V., 2015. Chapter Three-Development of Malaria Transmission-Blocking Vaccines: From Concept to Product. *Advances in parasitology*, 89, pp.109-152.

Yokomori, K., Admon, A., Goodrich, J.A., Chen, J.-L., and Tjian, R. 1993. Drosophila TFIIA-L is processed into two subunits that are associated with the TBP/TAF complex. *Genes & Development,* 7, pp.2235–2245.

Yokomori, K., Zeidler, M.P., Chen, J.L., Verrijzer, C.P., Mlodzik, M. and Tjian, R., 1994. Drosophila TFIIA directs cooperative DNA binding with TBP and mediates transcriptional activation. *Genes & Development*, 8(19), pp.2313-2323.

Young, R.A. 1991. RNA polymerase II. *Annu Rev Biochem,* 60, pp.689–715.

Yudkovsky, N., Ranish, J.A. and Hahn, S., 2000. A transcription reinitiation intermediate that is stabilized by activator. *Nature*, 408(6809), pp.225-229.

Zanella, F., Lorens, J.B. and Link, W., 2010. High content screening: seeing is believing. *Trends in biotechnology*, 28(5), pp.237-245.

Zeeman, A.M., van Amsterdam, S.M., McNamara, C.W., Voorberg-van der Wel, A., Klooster, E.J., van den Berg, A., Remarque, E.J., Plouffe, D.M., van Gemert, G.J., Luty, A. and Sauerwein, R., 2014. KAI407, a potent non-8-aminoquinoline compound that kills Plasmodium cynomolgi early dormant liver stage parasites in vitro. *Antimicrobial agents and chemotherapy*, 58(3), pp.1586-1595.

Zehring, W.A. and Greenleaf, A.L. 1990. The carboxyl-terminal repeat domain of RNA polymerase II is not required for transcription factor Sp1 to function in vitro. *J Biol Chem* 265 pp. 8351–8353.

Zhang, D., Penttila, T.L., Morris, P.L., Teichmann, M. and Roeder, R.G., 2001. Spermiogenesis deficiency in mice lacking the Trf2 gene. *Science*, 292(5519), pp.1153-1155.

Zhao, X. and 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(5), pp.615-627.

Zhou, Q., Boyer, T.G., and Berk, A.J. 1993. Factors (TAFs) required for activated transcription interact with TATA box-binding protein conserved core domain. *Genes Dev*, 7, pp.180–187.