Functional divergence of the RNA polymerase II transcription machinery in *Plasmodium falciparum*

By:

Jasmin Knopp

Dissertation presented for the degree of:

Doctor of Philosophy

Molecular and Cell Biology

MCB6002W

University of Cape Town

February 2024

Supervisor

Dr Thomas Oelgeschläger

University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

Acknowledgements

This work is dedicated to all those who gave me the most precious gift of all – time. To those that so freely gave their time in the form of aid and advice, and also to those who loved and supported me unquestionably even when experiments decreased the time that I could give in return. There are so many people without whom this thesis would not have been written. Thank you to my funders the National Research Foundation and Harry Crossley Foundation who provided the means to do these experiments. Thank you to my supervisor Dr Thomas Oelgeschläger for the endless input and support. Thank you doesn't cover the late-night experiments, multi-hour lab discussions and infinite pool of ideas for that one "last experiment"! I am truly honoured to be part of the noodle lab and have taken from this experience far more than the experimental findings detailed here.

Thank you to my fellow lab members (official and unofficial) and friends for acting as the buffers that kept me functional and happy through it all. A particular thanks to those walking down this path with me, Joanna, Robyn, Lize, Keso, Abigail, Leo, Gemma, Luc, Dan, Naadirah, Larnelle and Kristy - this experience would have been much more difficult and far less productive without you.

To my grandparents who instilled in me an appreciation for curiosity and knowledge and who, I believe, would read this thesis cover to cover, even without a degree, were they here today. To my dad who has more excitement about science advances and problems than anyone I have met. Thank you for allowing me to believe in the path I have chosen even on the worst of days and always being there 1000% percent through everything. To my mom for beautifully reminding me also of life outside of science. Thanks for always being there and keeping me balanced. To Arowan, my partner through it all, for believing in me. Always.

A special thanks to Larnelle Garnie without whom none of the *Plasmodium* work would have been possible. Finally, a big thank you to the UCT Confocal & Light Microscope Imaging Facility who made some of the more advanced imaging experiments a reality!

Declaration

I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own. I have used the Harvard convention for referencing. Each contribution to, and quotation in, this thesis from the work(s) of other people has been attributed and has been cited and referenced. This thesis is my own work. I have not allowed and will not allow anyone to copy my work with the intention of passing it off as their own work. I acknowledge that copying someone else's assignment or essay, or part of it, is wrong, and declare that this is my own work.

The PhD research project was conceptualised in collaboration with my supervisor Dr Thomas Oelgeschläger. All experimental work, data analysis, and data interpretation was performed by me. I have acknowledged the use of any reagent or methodology generated and/or developed by others in the appropriate sections of the thesis text, together with the names and affiliations of researchers that have provided access to specific materials.

Signed by candidate

Jasmin Knopp

11 February 2024

Abstract:

This thesis describes the functional characterisation of three *Plasmodium falciparum* general transcription factors - two TBP family proteins designated the TATA-binding protein (TBP) and TBP-like protein (TLP), and the *P. falciparum* orthologue of general transcription factor IIB (TFIIB). Through the biochemical characterisation of protein-DNA interactions, protein-protein interactions and the biomolecular condensation properties of these transcription factors, this study aimed to provide insight into the molecular mechanisms governing promoter recognition and the regulated assembly of the RNA polymerase II transcription initiation complex in *P. falciparum*, which are hitherto not understood.

The work shows that *P. falciparum* expresses two highly divergent TBP family proteins that do not possess sequence-specific TATA box-binding activity. Both *Pf*TBP and *Pf*TLP bind DNA with a general preference for A/T-rich sequences, mediated through interactions with the DNA minor groove, without detectable preference for specific sequence motifs. *Pf*TBP and *Pf*TLP are thus unlikely to contribute to the recognition of specific promoter regions within the A/T-rich context of the *P. falciparum* genome. The study further characterises TBP-independent DNA-binding of *Pf*TFIIB, a *Plasmodium*-specific feature not seen with TFIIB in well-studied eukaryotes. The data presented here show that *Pf*TFIIB binds to DNA in an unspecific manner. Interestingly, *Pf*TFIIB stimulates the DNA-binding activity of *Pf*TBP and *Pf*TLP and forms stable *Pf*TBP-*Pf*TFIIB-DNA and *Pf*TLP-*Pf*TFIIB-DNA nucleoprotein complexes. However, these *Pf*TFIIB interactions do not detectably enhance the sequence-selectivity of *Pf*TBP- or *Pf*TLP-DNA interactions. Thus, recognition of genomic regions at which transcription is preferentially initiated must involve additional factors or may depend to a lesser extent on the recognition of specific core promoter elements by the general transcription machinery.

Transcription condensates play an important role in eukaryotic transcription regulation. In this study, the potential role of *Pf*TFIIB in the formation and regulation of *P. falciparum* transcription condensates is investigated by fluorescence microscopy using a panel of fluorescent protein-tagged *Pf*TFIIB fusion proteins. The work demonstrates that *Pf*TFIIB undergoes condensation at nanomolar concentrations, and partitions with *Pf*TBP and *Pf*TLP into mixed phase separated condensates. Furthermore, assembly and properties of *Pf*TFIIB condensates are shown to be strongly influenced by the presence of DNA and RNA. Interestingly, *Pf*TBP and *Pf*TLP are found to localise in discrete foci in cultured *P. falciparum* blood-stage parasites. Together, these results provide first evidence for the existence of transcription condensates, mediated by transient multivalent interactions between general transcription factors, in the *P. falciparum* malaria parasite.

Table of Contents

List of Abbreviations

A,T,G,C - adenine, thymine, guanine, cytosine

6H - Hexa-histidine tag

aa – amino acid

ACTs - artemisinin-based combination therapies

Ad2ML - Adenovirus type 2 major late

Amp – Ampicillin

ApiAP2 - Apicomplexan AP2

BREd- downstream TFIIB recognition element

BREu- upstream TFIIB recognition element

BSA - Bovine serum albumin

Cam – Chloramphenicol

C-/N- terminus – carboxyl/ amino- terminus

CTD – carboxyl terminal domain

DNA - Deoxyribonucleic acid

DPE - Downstream promoter element

ds – double stranded

DTT – dithiothreitol

EDTA – ethylenediaminetetraacetic acid

EMSA - electrophoretic mobility shift assay

ER - endoplasmic reticulum

FRAP – fluorescence recovery after photobleachin

GBP-130 - glycophorin binding protein 130

GFP – green fluorescent protein

GST - glutathione-S-transferase

HP1 - heterochromatin protein 1

HRP - horseradish peroxidase

IDR – intrinsically disordered region

IMAC- immobilised metal affinity chromatography

KAHRP - Knob associated histidine rich protein

LB – lysogeny broth

LLPS - Liquid-Liquid phase separation

mC - mCherry

Med1 – Mediator subunit 1

mf - mobile fraction

mEGFP – monomeric enhanced green fluorescent protein

MLO – membrane-less organelle

mTdT - mouse terminal deoxynucleotidyl transferase

MW - Molecular weight

NC2 - Negative cofactor 2

NCBI - National Centre for Biotechnology Information

NE – nuclear extract

NP-40 – Nonidet P-40 Substitute

OD – optical density

PCR – polymerase chain reaction

PCV – packed column volume

PEG – polyethylene glycol

***Pf* –** *Plasmodium falciparum*

PIC – pre-initiation complex

PfEMP - *P. falciparum* erythrocyte membrane protein

***Pf*TLP -** *P. falciparum* TBP-like protein

PMSF – phenylmethylsulphonyl fluoride

poly(dG-dC)- poly(deoxyguanylicdeoxycytidylic) acid

poly(dI-dC)-Poly(deoxyinosinic-deoxycytidylic) acid

RNA - Ribonucleic acid

RNA FISH- RNA-fluorescence in situ hybridization

RNA Pol II – RNA polymerase II

RPB1 – RNA PoII II subunit 1

RT – room temperature

Q-sepharose - quaternary-ammonium sepharose

ss – single stranded

SDS - Sodium dodecyl sulfate

SDS-PAGE - Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SELEX - systematic evolution of ligands by exponential enrichment

SP-sepharose - sulphopropyl l-sepharose

TSS – transcription start site

TBP – TATA-binding protein

TF – transcription factor

TFIIB – Transcription Factor IIB

TLP - TBP-like factor

TRF – TBP related factor

tRNAs – transfer RNAs

t½ - half time of recovery

UTR - Untranslated region

YFP – yellow fluorescent protein

WHO - World Health Organisation

Chapter 1: Introduction

1.1. Malaria

1.1.1 The impact of malaria

Plasmodium is a parasitic protozoan belonging to the diverse phylum Apicomplexa. Members of this broad phylum are unicellular obligate parasites with complex life cycles and are characterised by a specialised apical complex that plays a key role in host invasion. Over 200 *Plasmodium* species exist, infecting a broad range of vertebrate hosts, however only six species, *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale curtisi, Plasmodium ovale wallikeri, Plasmodium knowlesi* and *Plasmodium malariae,* infect humans (Oguike *et al.*, 2011; Sato, 2021). Of these it is *Plasmodium falciparum* that has the greatest impact on global mortality and morbidity.

In 2021, there were an estimated 247 million cases of malaria globally and 619 000 malaria related deaths (World Health Organisation, 2022). Most of these deaths occur in Sub-Saharan Africa, where a disproportionally large proportion of malaria related deaths occur in children under the age of five. While a 27% reduction in malaria incidence was observed between 2000 and 2015, the rate of progress towards global malaria eradication has slowed significantly since 2015 (World Health Organization, 2020).

1.1.2 Current malaria treatments

The main methods of malaria prevention include the use of indoor residual spraying, insecticide-treated nets, and chemoprevention in children and pregnant women. Malaria treatment on the other hand makes use of artemisinin-based combination therapies (ACTs) involving the use of a short-lived artemisinin derivative and a long-acting partner drug (Ashley & Phyo, 2018). ACTs were originally developed in the 1990s and have been the recommended treatment globally since 2010, following the rise in resistance to two other widely used antimalarials - chloroquine and pyrimethamine (Thanh *et al.*, 2017; Witkowski *et al.*, 2017; Ashley and Phyo, 2018). In 2007, the first evidence for partial artemisinin resistance was observed in the area of the Tai-Cambodia border (World Health Organization, 2020). This partial resistance to artemisinin has also spurred a rise in resistance to the antimalarial drugs used in combination to artemisinin in ACTs (Thanh *et al.*, 2017). To combat the threat of this resistance spreading the World Health Organisation (WHO) and partners launched the Mekong antimalarial program with the goal of eliminating malaria in the

Mekong region by 2025. This program is currently on track to reach its goal of malaria eradication and no cases of partial resistance have been reported outside of this region. However, the threat of resistance to artemisinin and/or its combination therapies remains high – driving the urgent need to find additional treatments and prevention options.

The development of a safe and efficacious malaria vaccine remains one of the major aims for malaria elimination. However, vaccine development is challenged by the malaria parasites' genetic diversity, complex life cycle and multiple strategies of immune evasion (Bonam *et al.*, 2021). So far, only the RTS,S/AS01 and R21 vaccines have been approved by the WHO, although additional potential vaccines are currently under trial. With RTS,S/AS01 showing low efficacy (Greenwood *et al.*, 2015), and final peer reviewed phase three trial results for the recently approved R21 vaccine still outstanding (Moorthy and Binka, 2021; Nitika, Nema and Bharti, 2023), it remains to be seen whether deployment of either of these vaccines can significantly reduce malaria incidence and fatality. There remains an urgent need for novel anti-malarial treatments and prevention strategies (Ashley & Phyo, 2018).

1.2. *Plasmodium falciparum* biology and gene regulation

1.2.1. *Plasmodium falciparum* life cycle

The *Plasmodium falciparum* life cycle is divided between its human and mosquito hosts. Infection of the human host (reviewed comprehensively in Prudêncio, Rodriguez and Mota, 2006; Menard *et al.*, 2013; Cowman *et al.*, 2016), is initiated by the bite of an infected *Anopheles* mosquito, which deposits saliva containing anti-coagulants, vasodilators and a small but variable number of *P. falciparum* sporozoites (median of 15, range, 0–978) into the dermis (Rosenberg, R., Wirtz, R. A., Schneider, I. & Burge, 1990; Prudêncio, Rodriguez and Mota, 2006). These sporozoites move rapidly via gliding motility, enabling a small proportion of sporozoites to invade the blood vessels, while the rest either remain in the dermis or invade the lymphatic vessels (Amino et al., 2007). Sporozoites that invade the circulatory system play the greatest role in disease progression (Amino *et al.*, 2006). Once transported to the liver sinusoids, these sporozoites interact with heparan sulphate proteoglycans (Vaughan, Noden and Beier, 2002), cross the sinusoidal layer, likely through Kupffer cell invasion, and subsequently invade the hepatocytes.

In the liver, sporozoites traverse across multiple hepatocytes, through disruption of the plasma membrane, before forming a parasitophorous vacuole in the final, destination hepatocyte (Mota *et*

al., 2001). The morphology of the invading sporozoite transforms at this point from the elongated sporozoite to the large circular liver stage of the parasite (Fig. 1), which divides by schizogony to produce a schizont, made up of thousands of merozoites. In some *Plasmodium* species, such as *Plasmodium vivax* and *Plasmodium ovale*, sporozoites may differentiate into hypnozoites (Cogswell, 1992), a dormant form of the parasite responsible for relapse of malaria that can occur between 17 days to a year post infection (Wells, Burrows and Baird, 2010). The production of merozoites in the liver stage takes between 2 – 10 days and culminates in the release of merosomes, membrane bound vesicles of merozoites, that bud off from infected hepatocytes and pass into the circulatory system. Rupture of these merosomes in the microvasculature releases thousands of merozoites into the bloodstream (Fig. 1; Cowman *et al.*, 2016).

Merozoite release into the vasculature marks the beginning of the intraerythrocytic developmental cycle of *P. falciparum* infection. The released merozoites quickly invade host erythrocytes through a complex and dynamic process involving several stages: (i) attachment of the parasite and parasite-driven erythrocyte deformation, (ii) reorientation of the parasite, such that the apical end is adjacent to the erythrocyte membrane, and (iii) erythrocyte invasion, culminating in the sealing off of the parasite within a parasitophorous vacuole within the erythrocyte (Cowman *et al.*, 2016).

Within the erythrocyte, the *P. falciparum* parasite progresses through a ring stage and the highly metabolically active trophozoite stage before finally undergoing schizogony to form a schizont containing, in case of *P. falciparum,* 16-32 merozoites (Fig. 1; Cowman *et al*., 2016). During the intraerythrocytic developmental cycle, the infected erythrocyte is drastically remodelled to facilitate parasite growth and replication (Cowman *et al.*, 2016). A large percentage (60 – 80%) of the erythrocyte haemoglobin is broken down by the parasite to cater for increased energy demands. This process produces heme as a toxic byproduct, which is in turn converted to hemozoin through a biocrystallization process and stored in the parasite food vacuole (Francis, Sullivan and Goldberg, 1997). Additional transport pathways are established to scavenge nutrients and expel waste and the infected erythrocyte is remodelled to increase resistance to mechanical and temperature stress (Cowman *et al.*, 2016).

In parallel to these modifications associated with parasite growth and replication, the surface of infected erythrocytes is drastically remodelled, allowing evasion of the host immune system. Variant antigens, including *P. falciparum* erythrocyte membrane protein 1 (*Pf*EMP1), repetitive interspersed families of polypeptides (RIFINs), and subtelomeric variant open reading frame (stevor) proteins, are displayed on nanoscale protrusions on the surface of infected red blood cells and facilitate adhesion of infected red blood cells in the microvasculature in specific tissues, preventing splenic clearance

and causing severe malaria by blocking blood flow (Wahlgren, Goel and Akhouri, 2017). The best studied of the variant surface antigens are the *Pf*EMP1 proteins (Wahlgren, Goel and Akhouri, 2017). *Pf*EMP1 proteins are encoded by 60 antigenically distinct *var* genes, and represent the major antigen recognized by the host immune system. *Var* gene expression occurs early in the erythrocyte cycle and is tightly regulated, with only one *var* gene being expressed at a given time (Craig and Scherf, 2001; Smith *et al.*, 2001). Switching between the expression of different *var* genes at low frequency causes antigenic variation, which facilitates evasion of the host immune response.

The intraerythrocytic developmental cycle is highly synchronised and lasts 48 hours in *P. falciparum*, culminating in the simultaneous release of merozoites into the blood stream. The synchronous rupture of infected erythrocytes results in a new wave of erythrocyte infections and exposes the host to parasite toxins, initiating a peak in the fever of infected individuals (Oakley *et al.*, 2011).

During the intraerythrocytic developmental cycle, a proportion of asexual parasites differentiate into male and female gametocytes (Fig. 1). Following a lengthy 11-day maturation process in the bone marrow and spleen, these male and female gametocytes are released into the blood stream, primed to infect *Anopholese* mosquitos feeding on the infected vertebrate host (Hawking, Wilson and Kenneth, 1971; Talman *et al.*, 2004; Beri, Balan and Tatu, 2018). In the mosquito midgut lumen, gametocytes quickly transform into mature male and female gametes and fuse to form a zygote (Beier, 1998). The fertilized zygote then differentiates into a mobile ookinete, a process that takes over 24-30 hours in *P. falciparum* (Press, 1992; Jefferson A. Vaughan, Bruce H. Noden, 1994). The ookinete moves through the mosquito midgut epithelium to the basal lamina and forms an oocyst in which thousands of sporozoites develop over a period of at least 10 days (Beier, 1998). Sporozoites released by oocysts travel in the mosquito haemolymph through the body cavity, eventually invading the salivary glands and completing the life cycle.

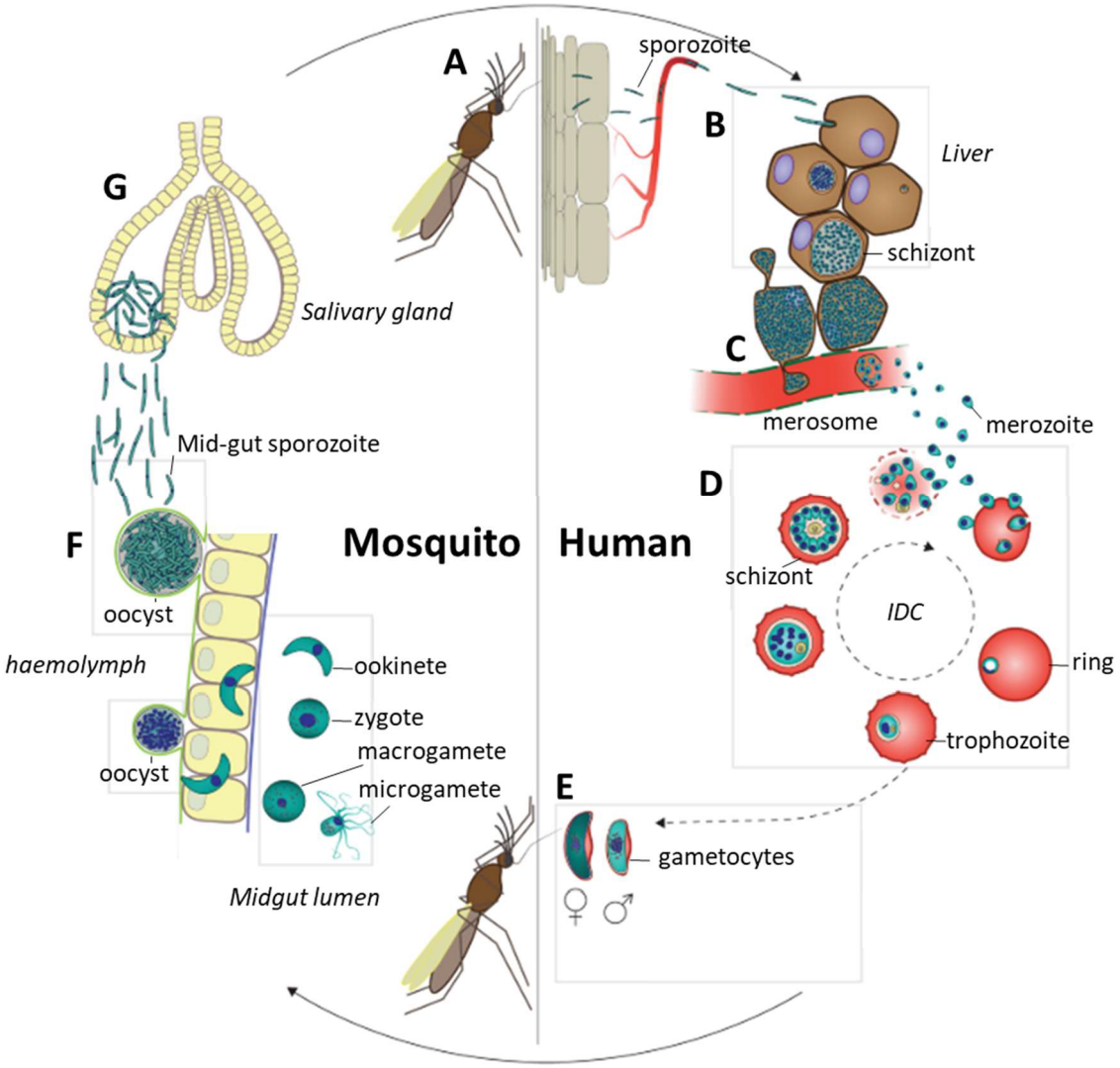


Figure 1. *Plasmodium falciparum* life cycle. (A) The bite of an infected mosquito releases sporozoites, which enter the blood stream. **(B)** Sporozoites infect the liver, form a parasitophorous vacuole in hepatocytes and divide by schizogony to produce a schizont. **(C)** Merosomes bud off from infected hepatocytes and pass into the circulatory system, where they rupture to release thousands of merozoites. **(D)** In the intraerythrocytic developmental cycle (IDC), merozoites invade erythrocytes and progress through ring, trophozoite and schizont stages. Synchronised rupture of infected red blood cells releases merozoites and reinitiates the erythrocyte cycle. **(E)** A proportion of parasites differentiate into male and female gametocytes, which are ingested by mosquitos feeding on the infected human host. **(F)** In the mosquito midgut, male and female gametes fuse to form a zygote, which matures into an ookinete. The ookinete moves through the mosquito midgut epithelium to the basal lamina and forms an oocyst in which thousands of sporozoites develop. **(G)** Sporozoites are released and invade the salivary gland, completing the life cycle. Figure modified from Toenhake & Bártfai, 2019.

1.2.2. The *Plasmodium falciparum* genome

The *Plasmodium falciparum* genome comprises approximately 23 megabase pairs, organised into 14 chromosomes, and encodes approximately 5300 genes (Böhme *et al.*, 2019). The overall gene organisation is similar to that of higher eukaryotes. Genes are monocistronicly transcribed and mRNA production is controlled by a promoter and transactivating factors (Fig. 2A; Horrocks *et al.*, 2009).

The *P. falciparum* genome has an extremely skewed A/T-content of 80.6% overall, compared to 58.9% in *Homo sapiens* (Hamilton et al., 2017), which approaches 90% in the intergenic regions (Gardner *et al.*, 2002). This skewed nucleotide composition has large implications on many aspects of *P. falciparum* transcriptional regulation. At the DNA level, the extremely high A/T-content of the *P. falciparum* genome impacts the diversity and prevalence of sequence motifs available for recognition by DNA-binding proteins (Watzlowik *et al.*, 2021), suggesting evolutionary pressure for the adaptation of transcription factors and chromatin-associated proteins to the A/T-rich genomic context. Furthermore, the extremely high A/T-content has been suggested to decrease local DNA flexibility, impacting chromatin structure and dynamics (Segal and Widom, 2009; Tillo and Hughes, 2009; Ponts *et al.*, 2010; Silberhorn *et al.*, 2016). At the protein level, the A/T-nucleotide bias in the genomic sequence increases the proportion of codons for asparagine, lysine, isoleucine and tyrosine in protein coding sequences and increases the prevalence of low complexity regions, comprised of only a limited repertoire of amino acids, within *P. falciparum* proteins (Watzlowik *et al.*, 2021). These low complexity regions are frequently intrinsically disordered, i.e. they do not conform to a single defined tertiary structure, and may influence *P. falciparum* protein properties and function (Aravind *et al.*, 2003).

1.2.3. Transcription regulation in *Plasmodium falciparum*

The complex life cycle, variant expression of virulence genes such as *Pf*EMP1 (Section 1.2.1) and the ability of *P. falciparum* to adapt to altered host environments, all require tightly regulated gene expression programs. Genome-wide mRNA and proteomic analyses have highlighted the broad existence of distinct programs of temporally expressed genes throughout the *P. falciparum* life cycle (Bozdech *et al.*, 2003; Le Roch *et al.*, 2003; Otto *et al.*, 2010; López-Barragán *et al.*, 2011; Hoo *et al.*, 2016; Toenhake *et al.*, 2018). For many genes there also appears to be a strong correlation between the peak mRNA transcript abundance and the requirement of mRNA protein products as well as a moderately strong correlation between protein abundance and steady state mRNA levels –

supporting that many *P. falciparum* genes are regulated at the level of transcription (Bozdech *et al.*, 2003; Le Roch *et al.*, 2004).

P. falciparum transcription regulation at the level of chromatin structure and chromatin dynamics

A large body of research has focused on delineating the role of chromatin structure, chromatin modifications and chromatin dynamics in *Plasmodium* transcription regulation (Duffy *et al.*, 2014; Toenhake and Bártfai, 2019; Watzlowik *et al.*, 2021; Hollin, Chahine and Le Roch, 2023). In eukaryotes, the genomic DNA is packaged into nucleosomes consisting of DNA wrapped around a histone octamer, comprised of two dimers of histones H2A/H2B and histones H3/H4. Nucleosome composition, positioning, dynamics and modifications play a crucial role in regulating the accessibility of cis-regulatory sequences, thus contributing to transcription regulation. Orthologs of the canonical histone proteins H2A, H2B, H3 and H4, have been identified in *P. falciparum* (Watzlowik *et al.*, 2021). However, both *P. falciparum* histone H2A and H2B are highly divergent from other eukaryotes, with only 64% and 68% amino acid sequence identity relative to their human counterparts (Watzlowik *et al.*, 2021). Furthermore, the linker histone H1, which has important roles in nucleosome spacing and chromatin compaction, is absent in *P. falciparum*. Adding to these differences in histone features, a number of histone variants, some unique to Apicomplexa, have been shown to replace the canonical histones in *P. falciparum* nucleosomes (Hollin and Le Roch, 2020). For example, the histone variant H2A.Z and the Apicomplexa-specific histone variant H2B.Z have been shown to colocalise and form a double variant nucleosome subtype that localises to the A/T-rich intergenic regions of the *P. falciparum* genome (Hoeijmakers *et al.*, 2013). These differences in *P. falciparum* nucleosome composition and histone properties impact the biochemical properties of *P. falciparum* nucleosomes by reducing nucleosome stability, altering sequence-dependent nucleosome positioning and decreasing the nucleosome repeat length (Silberhorn *et al.*, 2016). Likely because of these changes in nucleosome properties, as well as the absence of linker histone H1 and the rigidity of A/T-rich DNA, the *P. falciparum* genome has a generally more open chromatin landscape with reduced nucleosome occupancy relative to other eukaryotes, particularly in the intergenic regions (Ponts *et al.*, 2010).

While the *P. falciparum* genome is overall more accessible, nucleosome positioning still plays a key role in transcriptional regulation. Multiple studies have shown a clear correlation between nucleosome occupancy and mRNA abundance in *P. falciparum* (Ponts *et al.*, 2011; Bunnik *et al.*, 2014; Kensche *et al.*, 2015; Ruiz *et al.*, 2018; Toenhake *et al.*, 2018). As in many other eukaryotes, nucleosome depleted regions (NDRs) are found just upstream of the transcription start site (TSS) of transcriptionally active genes - facilitating transcription factor access to exposed cis regulatory

sequences in these regions (Hollin, Chahine and Le Roch, 2023). In contrast, protein-coding DNA sequences and regulatory regions upstream of the TSS of inactive genes are associated with higher nucleosome occupancy. In addition to local changes in nucleosome positioning, there is also clear evidence for global changes to nucleosome occupancy throughout the *P. falciparum* erythrocyte cycle. Histone occupancy is at its lowest at the trophozoite stages, when transcriptional activity peaks and DNA replication takes place, and at its highest in the schizont phase, when the parasite repackages its genetic material and prepares for the rupture of the infected erythrocyte (Ponts *et al.*, 2010; Bunnik *et al.*, 2014; Hollin, Chahine and Le Roch, 2023).

Histone post-translational modifications (PTMs), including acetylation, phosphorylation, ubiquitylation and methylation of histone tails, also play an important role in *P. falciparum* transcription regulation - modulating nucleosome-DNA interactions and the recruitment of regulatory factors (Hollin, Chahine and Le Roch, 2023). Over 230 different PTMs have been identified in the erythrocyte stages alone (Saraf *et al.*, 2016). The presence and combinations of modifications vary over the erythrocyte cycle, however histone modifications associated with transcription activation, such as histone H3 acetylation at lysine 9 (H3K9ac), histone H4 acetylation at lysine 8 or histone H3 trimethylation at lysine 4 (H3K4me3), are most abundant during the erythrocyte cycle, consistent with the genome-wide abundance of euchromatin during these stages (Saraf *et al.*, 2016).

Multiple studies have shown that chromatin structure and dynamics, histone PTMs and the spatial positioning in the nucleus play a key role in the repression of clonally variant multigene families of virulence genes, such as the *var* gene family of proteins, which encode variants of the *Pf*EMP1 protein (Section 1.2.1). Silencing the 59 inactive *var* genes involves clustering of these genes to repressive heterochromatic regions at the nuclear periphery that are marked by histone H3 trimethylation at lysine 9 (H3K9me3; Freitas-Junior *et al.*, 2000; Salcedo-Amaya *et al.*, 2009; Ay *et al.*, 2014). H3K9me3 marked histones are recognised and bound by heterochromatin protein 1 (HP1), an evolutionary conserved protein with roles in heterochromatin formation (Flueck *et al.*, 2009). Depletion of HP1 disrupts mutually exclusive expression of *Pf*EMP1 proteins in *P. falciparum* parasites, resulting in simultaneous expression of the majority of *var* genes (Brancucci *et al.*, 2014). Trimethylation at lysine 36 of histone H3, mediated by the *P. falciparum* variant-silencing SET gene (*Pf*SET), also plays a key role in repression of inactive *var* genes. Similarly to depletion of HP1, knock-out of *Pf*SET results in simultaneous expression of multiple *var* genes and loss of mutually exclusive *var* gene expression (Jiang *et al.*, 2013).

Considerably less is known about the molecular mechanisms and factors involved in the activation of *var* gene expression. Like repressed *var* genes, the active *var* gene has been observed to localise to the nuclear periphery, occupying a distinct territory to inactive *var* genes. The singular active *var* gene is marked by H3K4me3/2, H3K9ac , an enrichment for histone variants H2A.Z and H2B.Z as well as the chromatin remodeler *Pf*ISWI, which appears to play a direct role in var gene activation (Lopez-Rubio *et al.*, 2007; Petter *et al.*, 2013; Bryant *et al.*, 2020). Acetylation at lysine 8 of histone H4 (H4K8ac) also appears to play an important role in *var* gene regulation, as H4K8ac was observed to be enriched in the intergenic region of the active *var* gene and overexpression of a histone H4 mutant variant in which lysine 8 was substituted by arginine, decreased *var* gene transcript levels (Gupta *et al.*, 2017).

P. falciparum transcription regulation at the level of transcription intitation

In contrast to the significant progress made towards understanding the role of chromatin structure and dynamics in transcriptional regulation, very little progress has been made towards how gene-specific transcription factors and general transcription factors modulate RNA polymerase II (RNA Pol II) transcription machinery assembly and function in *P. falciparum*. (Toenhake and Bártfai, 2019). *In silico* analysis of the *P. falciparum* proteome revealed that *P. falciparum* has a very limited repertoire of gene-specific transcription factors and that many of the general transcription factors are either highly divergent from their homologs in higher eukaryotes or entirely absent (Coulson, Hall and Ouzounis, 2004; Callebaut *et al.*, 2005; Bischoff and Vaquero, 2010). Well-characterised *P. falciparum* gene-specific transcription factors include helix-turn helix transcription factors, C2H2 type zinc finger family transcription factors and a family of apicomplexan-specific transcription factors known as the ApiAp2 family that have been shown to play an essential role in each of the *Plasmodium* life-cycle stages (Jeninga, Quinn and Petter, 2019; Toenhake and Bártfai, 2019).

P. falciparum gene-specific transcription factors are far better characterised than the general transcription factors, most of which remain to be functionally characterised. It is still not clear how the *P. falciparum* general transcription machinery recognises and assembles at the promoter to initiate transcription (Toenhake and Bártfai, 2019).

Two studies have provided preliminary insights into the association between general transcription factor gene occupancy and steady-state mRNA levels. A study which investigated the localisation of the *P. falciparum* TATA box-binding protein (TBP) and transcription factor IIE (TFIIE) found *Pf*TBP and *Pf*TFIIE to be localised at both the 5' and 3' regions of both active and inactive erythrocyte expressed genes during the intraerythrocytic stages (Gopalakrishnan *et al.*, 2009). Similarly, a later study, which investigated RNA Pol II occupancy during the erythrocyte stages using a ChIP-on-chip approach

found that while there was a correlation between RNA Pol II occupancy and mRNA levels for many genes, there were also a large number of genes for which there was a significant delay between peak RNA Pol II occupancy and peak mRNA levels (Rai *et al.*, 2014). These results may point to a potential role of regulatory mechanisms downstream of the assembly of the RNA Pol II machinery, such as transcriptional elongation or mRNA processing and stability, in the regulation of steady-state mRNA levels (Rai *et al.*, 2014). In agreement with this, a study using rapid 4-thiouracil (4-TU) incorporation via pyrimidine salvage to label and quantify mRNA synthesis in relation to steady state mRNA found that mRNA transcription and mRNA stabilisation each make gene-specific contributions to steady-state mRNA levels (Painter *et al.*, 2018).

To date, *Pf*TBP is the only general transcription factor that has been functionally characterised. Based on a series of *in vitro* DNA binding experiments it has been suggested that *Pf*TBP binds to a TGTAA DNA sequence located 81 bp upstream of the GBP-130 TSS and a TGTAA DNA sequence located 184 bp upstream of the KAHRP TSS (Ruvalcaba-Salazar *et al.*, 2005). However, as detailed in Section 1.3.5 of this thesis, some of the results of this study appear inconsistent with the published results of a later study (Gopalakrishnan *et al.*, 2009). In addition, the functional relevance of these observations to promoter recognition within the highly A/T-rich context of *Plasmodium* intergenic regions remains unclear (Toenhake and Bártfai, 2019).

Studies identifying transcription start sites (TSS) during the intraerythrocytic stages have highlighted a number of DNA sequence features which may contribute to TSS selection and promoter recognition/strength (reviewed in Toenhake and Bártfai, 2019). A study, using 5' cap sequencing observed the increased occurrence of a TA dinucleotide at positions -1 and +1 at the TSS (Adjalley *et al.*, 2016). This same study also noted a strong increase in A/T-content around the TSS of the most active TSS cluster as well as an increase in G/C-rich elements approximately 150 and 220 bp downstream of the TSS, which coincides with a localised decrease in nucleosome occupancy (Adjalley *et al.*, 2016). They also observed an increased local G/C-content at the borders of TSS clusters and downstream of clusters of less frequently utilised TSSs. A more recent study employing a directional, amplification-free RNA sequencing (DAFT-seq) approach observed a preference for pyrimidine-purine base pairs at the TSS, with TG being the most frequent dinucleotide (Chappell *et al.*, 2020). This study also highlighted an increase in the prevalence of adenine nucleotides upstream of the TSS and of thymine nucleotides downstream of the TSS which could be observed within 20 bp of the TSS (Chappell *et al.*, 2020). However, it remains unclear to what degree these features contribute to promoter and TSS site selection by the general RNA Pol II machinery. Indeed, many *P. falciparum* genes are transcribed from multiple clusters of TSS sites, which are spread over a large region (Adjalley *et al.*, 2016; Chappell *et al.*, 2020). These results suggest that, in *P. falciparum,* either

the core promoter region driving RNA Pol II transcription initiation, or the distance between core promoters and the site of transcription initiation, may not be well-defined.

Biochemical characterisation of the *P. falciparum* general transcription factor homologs is required to define core promoter identity and the interactions that govern the assembly of the RNA Pol II machinery in *P. falciparum*. One of the goals driving such in depth characterisation of *P. falciparum* general transcription factors is to uncover *Plasmodium*-specific properties, which could be selectively targeted by rational drug design.

1.3. Eukaryotic gene regulation at the level of transcription initiation

Tight regulation of transcription is essential to cellular homeostasis and development. Regulation of transcription is achieved by concerted action of multiple regulatory layers, encompassing the action of gene-specific transcription factors, co-activators and general transcription factors, as well as dynamic changes in chromatin structure that alter the accessibility of cis-regulatory sequences (Fig. 2A).

Transcription is achieved through the action of four types of RNA polymerases and their accessory transcription factors: RNA polymerase I (RNA Pol I), which drives transcription of the 18S and 28S ribosomal RNAs; RNA polymerase II (RNA Pol II), which drives mRNA transcription at all protein-coding genes; RNA polymerase III, which drives transcription of 5S rRNA and tRNAs; and finally RNA polymerase IV, a plant-specific polymerase involved in the production of small interfering RNAs (Thomas and Chiang, 2006). While some transcriptional regulatory mechanisms are shared between all RNA polymerases, this section reviews only the regulation of transcription by RNA Pol II, with specific focus on regulation at the level of transcription initiation, the role of TBP and TBP paralogs and the function of TFIIB in well-studied eukaryotes.

1.3.1. The eukaryotic RNA Pol II core promoter

The RNA Pol II core promoter is defined as the minimal region of DNA required to direct *in vitro* basal levels of transcription (Smale and Kadonaga, 2003). This region typically extends approximately 50 bp upstream and downstream from the transcription start site (TSS) (Haberle and Stark, 2018). Core promoters are highly diverse and can consist of many different combinations of core promoter elements, recognised by general transcription factors (Ngoc *et al.*, 2019), that facilitate the formation of a productive RNA Pol II initiation complex. Fig. 2B summarises some of the well-studied core promoter elements and the specific general transcription factors that facilitate their recognition.

The best studied core promoter element is the TATA box. Typically located between -31 and -24 relative to the TSS, the TATA box has the consensus sequence TATAWAWR and is recognised by the TATA binding protein (TBP), a subunit of transcription factor IID (TFIID; Basehoar, Zanton and Pugh, 2004; Thomas and Chiang, 2006).

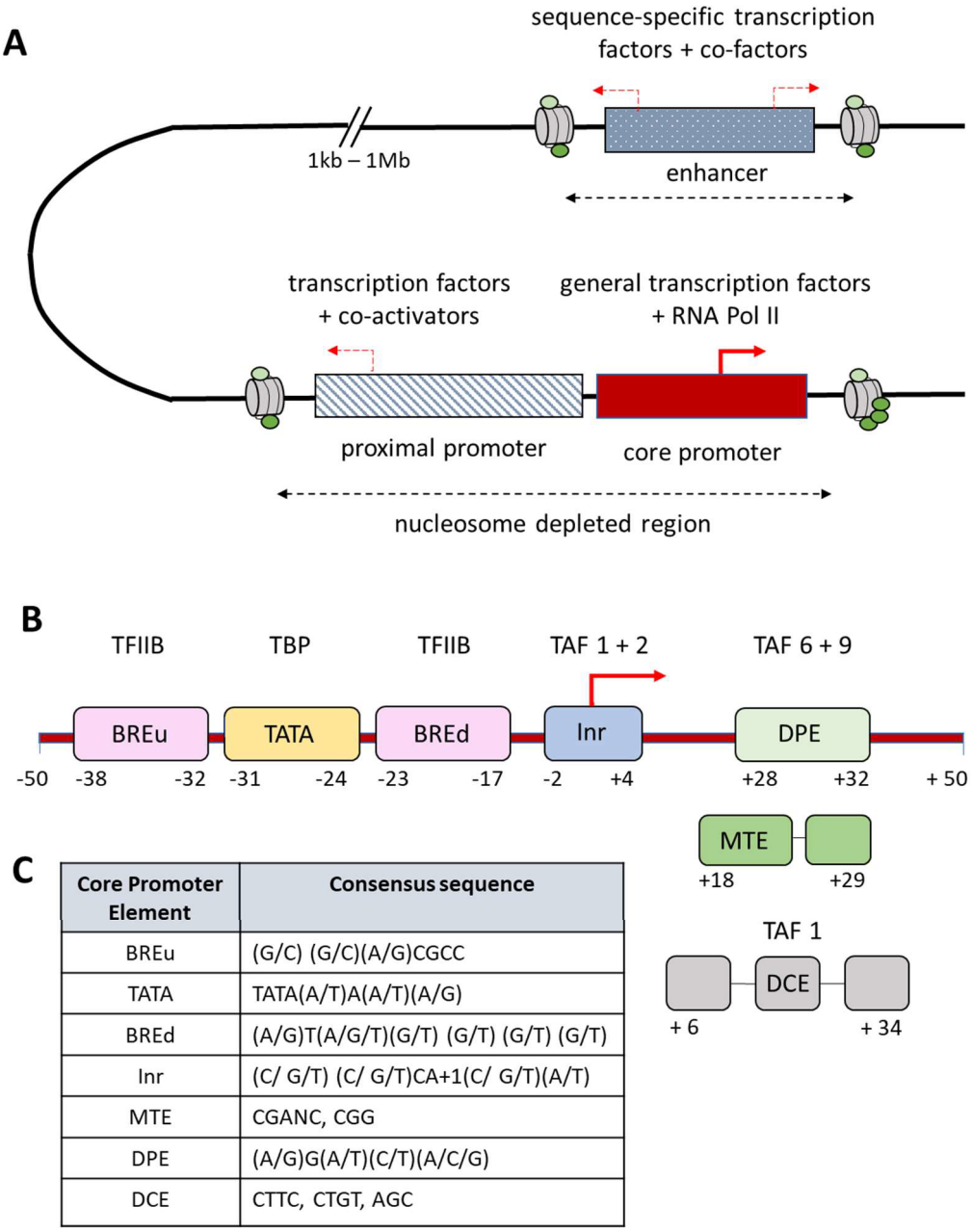
The initiator element is one of the most common core promoter elements (ngoc *et al.*, 2019). In humans, the initiator typically consists of 6 bp spanning the TSS and has the consensus sequence BBCABW (B = C or G or T, W = A or T, ngoc *et al.*, 2017). Sequence motifs matching or closely matching the initiator consensus have been identified at over 50% of focused *H. sapiens* promoters (ngoc *et al.*, 2017). The initiator element can direct basal transcription by itself or act in conjunction with other core promoter elements (Smale and Baltimore, 1989). Multiple factors including TBP-associated factors (TAF) TAF1 and TAF2 (Chalkley and Verrijzer, 1999; Louder *et al.*, 2016), TFIIA and additional cofactors, including the mediator complex (Martinez *et al.*, 1994, 1995, 1998; Xu *et al.*, 2011; Rojas *et al.*, 2022) appear to contribute to initiator-directed transcription, however it is not yet clear which factors directly recognise and bind the initiator element (Rojas *et al.*, 2022). Furthermore, it has been suggested that the factors required for transcription initiation from initiator-containing promoters may vary between species (Rojas *et al.*, 2022). For example, TAFs are required for transcription from initiator-responsive promoters in humans and insects but not in fission yeast (Martinez *et al.*, 1994, 1995; Kaufmann *et al.*, 1998; Chalkley and Verrijzer, 1999; Rojas *et al.*, 2022).

Other important core promoter elements include the upstream and downstream TFIIB recognition elements (BREu and BREd), which facilitate promoter recognition by transcription factor IIB (TFIIB) and can have either a stimulatory or inhibitory influence on transcriptional output, depending on promoter context (Lagrange *et al.*, 1998; Evans, Fairley and Roberts, 2001; Deng and Roberts, 2005) ; the downstream promoter element (DPE), which functions in conjunction with the initiator element to enhance RNA Pol II transcription; the motif ten element (MTE), which also functions co-operatively with the initiator element when positioned between +18 to +29 relative to the TSS (Lim *et al.*, 2004); and the downstream core element (DCE) from the human β-globin promoter (ngoc *et al.*, 2019).

Core promoter regions typically contain one to three of these core promoter elements, none of which are universally required for core promoter function. There are also a considerable number of promoters which have no identifiable promoter elements (ngoc et al., 2019).

Based on the pattern of transcription initiation, promoter elements present, nucleotide composition, chromatin environment and the function of the associated gene, core promoters have been broadly

classified into three subtypes (Haberle and Stark, 2018). The first of these subtypes is comprised of the core promoters for adult tissue-specific genes, which display a focused pattern of transcription initiation and often contain initiator or TATA box elements. The second of these subtypes is comprised of the core promoters for housekeeping genes. Housekeeping gene core promoters generally display dispersed patterns of transcription intitation and are demarcated by precisely positioned nucleosomes straddling broad nucleosome-depleted regions. These housekeeping gene core promoters are associated with islands of GC-rich sequences in Mammalia and the occurrence of specific cis-regulatory elements Ohler 1, Ohler 6 and DNA replication element in flies. The third of these subtypes is comprised of the core promoters of developmental transcription factors, which harbour specific nucleosome modifications, thought to enable tissue-specific patterns of expression, and are characterised by islands of GC-rich sequences in Mammalia or the occurrence of a DPE in flies (Haberle and Stark, 2018).



C

Core Promoter Element	Consensus sequence
BREu	(G/C) (G/C)(A/G)CGCC
TATA	TATA(A/T)A(A/T)(A/G)
BREd	(A/G)T(A/G/T)(G/T) (G/T) (G/T) (G/T)
Inr	(C/ G/T) (C/ G/T)CA+1(C/ G/T)(A/T)
MTE	CGANC, CGG
DPE	(A/G)G(A/T)(C/T)(A/C/G)
DCE	CTTC, CTGT, AGC

Figure 2. Overview of cis-regulatory sequences involved in eukaryotic gene regulation. (A) Gene regulation involves the action of multiple transcription factors, activators and co-activators which interact with cis regulatory DNA sequences to drive transcription. The core promoter is recognised by the general transcription factors and acts as a landing platform for the RNA Pol II complex to facilitate basal transcription (red arrow). Core promoter activity can be modulated by gene-specific transcription factors and co-activators, which interact with the proximal promoter and enhancer elements, as well as epigenetic regulators, which alter the chromatin landscape and the accessibility of cis-regulatory sequences. Nucleosomes are represented as grey cylinders, with modifications represented as green circles. Note that enhancers are transcribed, producing so-called enhancer RNAs (eRNAs) and that many core promoters are transcribed bi-directionaly, producing short, unstable anti-sense RNAs. eRNAs and anti-sense RNAs are represented as red dashed arrows. **(B)** Schematic of core promoter structure with key general transcription factors involved in core promoter element recognition elements indicated. Positions of core promoter elements relative to the TSS indicated. **(C)** Consensus sequences of core promoter elements. Panels (B) and (C) are adapted from (ngoc *et al.,* 2019) and (Thomas and Chiang, 2006).

1.3.2. Assembly of the RNA Pol II machinery and transcription initiation at focused promoters containing a TATA box

The general RNA Pol II transcription machinery consists minimally of RNA Pol II and six general transcription factors (GTFs; TFIIA, -B, -D, -E, -F, -H), which are required for the recognition of core promoter elements and the assembly and positioning of RNA Pol II. According to the classical stepwise assembly model (Fig. 3), assembly of the RNA polymerase II complex at TATA-containing promoters is initiated by binding of the TATA-binding protein (TBP), a subunit of TFIID, to the TATA box (Thomas and Chiang, 2006). Binding of TBP to the minor groove of the TATA box sequence results in dramatic distortion of the DNA helix, which is facilitated by the insertion of two pairs of phenylalanine residues between the first and last base pairs of the TATA box DNA sequence (Nikolov *et al.*, 1996). The TBP-DNA complex is then recognised and bound by TFIIA and TFIIB, which make additional DNA contacts and stabilise the TBP-DNA complex (Barberis *et al*., 1993; Geiger *et al.*, 1996a; Hisatake *et al.*, 1993; Imbalanzo *et al.*, 1993; Lagrange *et al.*, 1996; Oelgeschläger *et al*., 1996; Thomas & Chiang, 2006). Binding of TFIIB facilitates the recruitment of RNA Pol II and TFIIF (Ha *et al.*, 1993; Fang and Burton, 1996), which in turn facilitates the subsequent recruitment of TFIIE and TFIIH. In the presence of ATP, a 10-15 bp stretch of DNA is then separated by TFIIE and TFIIH, forming a 'transcription bubble', and the single-stranded DNA template is positioned in the RNA Pol II catalytic cleft to initiate transcription (Thomas and Chiang, 2006; Sainsbury, Bernecky and Cramer, 2015).

Transcription of the full-length nascent RNA transcript does not necessarily occur immediately following transcription initiation. Instead, RNA Pol II can go through many rounds of abortive transcription, generating very short transcripts less than ten nucleotides in length until clashes between the growing RNA and TFIIB trigger conformational changes within RNA Pol II and the release of TFIIB (Saunders, Core and Lis, 2006). Following this period of abortive transcription, RNA Pol II frequently transcribes a 20-50 nucleotide transcript and then pauses due to the action of regulatory factors such as 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole sensitivity-inducing factor (DSIF) and negative elongation factor NELF (Lee and Young, 2013). Release from this paused state is facilitated by positive transcription elongation factor b (P-TEFb) kinase which phosphorylates the repressive DSIF-NELF complex as well as the RNA Pol II C-terminal domain (CTD). Phosphorylation of the DSIF-NELF complex and the RNA Pol II CTD triggers disassociation of NELF, RNA Pol II release from its paused state, and the association of the RNA Pol II machinery with RNA processing factors and chromatin modifiers (Adelman and Lis, 2012). This regulation of the transition to transcriptional elongation constitutes another important level of transcriptional control (Adelman and Lis, 2012).

While the above model for transcriptional initiation holds true for many TATA-containing promoters, it is important to highlight that only a small subset of eukaryotic promoters (5-7% of eukaryotic promoters) contain a TATA box (Roy and Singer, 2015). How the transcription machinery recognises and assembles at TATA-less promoters is still not understood and may deviate considerably from the canonical stepwise assembly pathway (Donczew and Hahn, 2018; Chen and Xu, 2022; Rojas *et al.*, 2022). In addition, while general transcription factors involved in the assembly of the RNA Pol II transcription intitation complex were initially regarded to be universally required for RNA Pol II transcription initiation, it is important to emphasise that the constituents of the RNA Pol II machinery may vary between promoters, cell types and across species (Akhtar and Veenstra, 2011; White-cooper and Davidson, 2011; Duttke, 2015). Gene duplication events have led to an expanded repertoire of general transcription factors. These paralogs often display tissue- or stage-specific patterns of expression, distinct DNA-binding properties and protein interactions and can form alternative preinitiation complexes to drive the expression of cell type-specific genes (Akhtar and Veenstra, 2011).

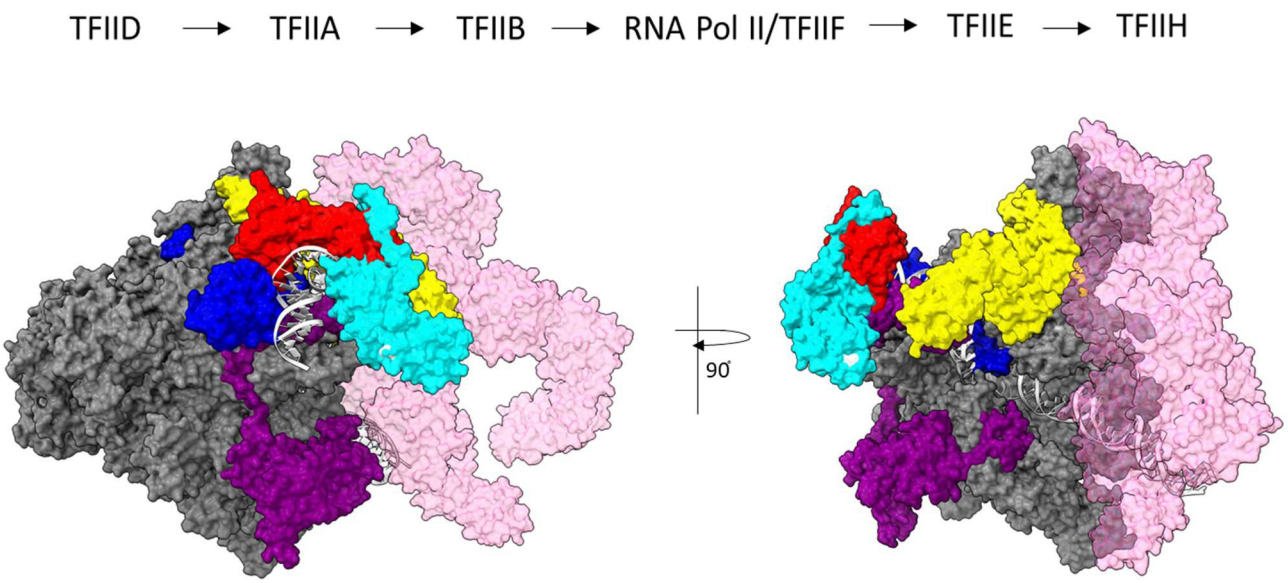


Figure 3. Model for the stepwise assembly of the RNA Pol II transcription initiation complex at TATA-containing promoters. Schematic of the stepwise assembly pathway with cartoon representation of the structure of the RNA Pol II transcription initiation complex following promoter opening, modelled from the structure coordinates determined by electron microscopy for the human pre-initiation complex in the open state (5IY7; He *et al*., 2016). TBP is shown in red, TFIIA in light blue, TFIIB in dark blue, TFIIE in yellow, TFIIF in purple, RNA Pol II in grey, TFIIH in light pink, promoter DNA in white.

1.3.3. The TATA-binding protein in well-studied eukaryotes

TBP is one of a number of general transcription factors which have undergone gene duplication events throughout metazoan evolution. Interestingly, TBP is the only general transcription factor universally required for transcription by RNA polymerases I, II and III. TBP is an integral component of specific multi-protein complexes that direct the assembly of RNA polymerase I, II and III at the transcription start site, including selectivity factor I (SL1), required for transcription initiation at ribosomal RNA (rRNA) genes, TFIID, required for transcription initiation at protein-coding genes, and TFIIIB, which functions in RNA Pol III transcription (Akhtar and Veenstra, 2011).

TBP has a bipartite structure comprised of an intrinsically disordered N-terminal region of variable length and amino acid sequence, and an evolutionary highly conserved C-terminal DNA-binding domain. The TBP C-terminal DNA-binding domain has a saddle-like ternary structure consisting of two quasi-symmetric repeats each made up of two α-helices and five β-sheets (Fig. 4). The underside of the 'saddle' is made up of β-strands that form a concave DNA interaction interface, while the two alpha helices form seat and stirrup on the convex surface (Kim, Nikolov and Burley, 1993; Nikolov *et al.*, 1995, 1996; Geiger *et al.*, 1996a), which engage in direct interactions with TFIIB, TFIIA and other TBP interaction partners (Geiger *et al.*, 1996b; Nikolov *et al.*, 1996).

TBP interacts with DNA through an induced fit mechanism in which TBP interaction with the minor groove of the 'TATAWAWR' recognition sequence (W = A or T; R = G or A) results in severe distortion of the DNA helix (Bucher, 1990; Thomas and Chiang, 2006). Key to stabilising this distortion is intercalation of two phenylalanine residues (F288 and F197 in *Homo sapiens* TBP) between the first and second and between the seventh and eighth base pairs of the TATA box sequence, respectively (Nikolov *et al.*, 1996). Two additional phenylalanine residues (F214 and F305 in *Hs*TBP; Fig. 4) buttress the inserting phenylalanine residues and make additional stabilising DNA contacts. Intercalation of F288 of *Hs*TBP between the first and second base pair of the TATA box alters the DNA trajectory and partially unwinds the B-DNA helix, resulting in a widened major groove and facilitating further TBP-DNA interactions. This partial unwinding is reversed by the action of the second pair of phenylalanine residues, F197 and F214 of *Hs*TBP, which re-establish the B-DNA conformation. The two DNA kinks caused by phenylalanine intercalation result in an overall bending of the DNA helix by approximately 80° towards the major groove and a local unwinding of the DNA helix by approximately 120° (Nikolov *et al.*, 1996). Sequence-selective TBP-DNA interactions occur predominantly through the indirect readout of sequence-dependent DNA structure and deformability rather than through direct readout which involves hydrogen bond formation between

the amino acid side chains and DNA (Bareket-Samish, Cohen and Haran, 2000; Faiger, Ivanchenko and Haran, 2007).

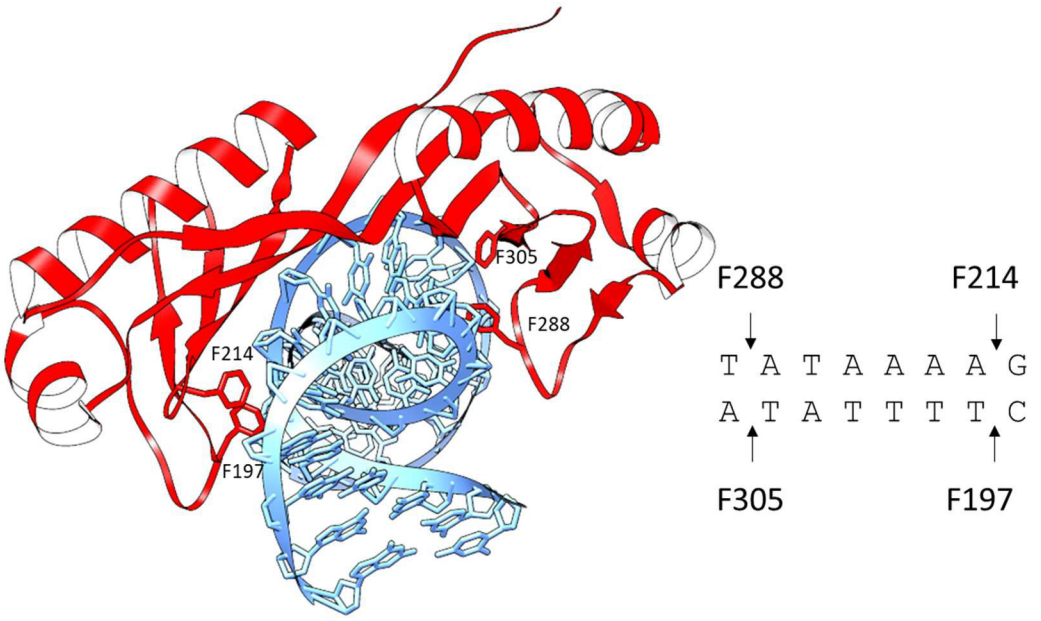


Figure 4. Structure of *Hs*TBP bound to the adenovirus 2 major late (Ad2ML) promoter TATA box.
Left: Cartoon representation of the *Hs*TBP DNA-binding domain bound to the Ad2ML promoter TATA box sequence modelled from the co-ordinates for the TBP-TFIIB-DNA complex (1C9B) determined by X-ray crystallography (Tsai and Sigler, 2000). TBP is shown in red, DNA in blue. *Right:* schematic of the Ad2ML sequence with positions of intercalating phenylalanine residues indicated.

The length and sequence composition of the N-terminal intrinsically disordered region (IDR) of TBP varies greatly between species. While its precise role remains enigmatic, the TBP N-terminal IDR has been shown to destabilise yeast TBP-DNA interactions and alter TBP DNA bending *in vitro* (Kuddus and Schmidt, 1993). In addition, a monoclonal antibody directed against the N-terminus of *Hs*TBP was shown to significantly reduce *in vitro* transcription from TATA-containing promoters without impacting TBP-TATA or TBP-TFIIB interactions, suggesting a role for the TBP IDR subsequent to initiation complex assembly (Lescure et al., 1994). However, deletion of a large portion of the N-terminal IDR in mouse primary fibroblast cells did not significantly impact transcription by RNA Pol I or II from a range of promoters (Schmidt *et al.*, 2003). Instead, the N-terminal IDR appears to be required for suppression of the maternal immune response towards post transplantation mouse embryos through unknown mechanisms (Hobbs *et al.*, 2006). The N-terminal IDR of *Hs*TBP has also been shown to drive *in vitro* liquid-liquid phase separation (LLPS) and has been proposed to play a key role in TBP incorporation into transcriptional condensates (Basu *et al.*, 2020a, Karamanof, 2022, Section 1.4). Misregulation of TBP LLPS and TBP incorporation into transcriptional condensates, caused by changes in TBP IDR amino acid sequence, has been proposed as a possible driver of

spinocerebellar ataxia type 17, a progressive neurodegenerative disease (Nakamura, 2001; Basu *et al.*, 2020).

1.3.4. TBP-related factors

Gene duplication events have resulted in the evolution of multiple TBP paralogs over the course of evolutionary history (Akhtar & Veenstra, 2011). To date, five TBP family proteins have been identified in metazoans: the insect-specific TBP paralog, TBP-related factor 1 (TRF1), the Bilateria-specific TBP-related factor 2 (TRF2), also known as TBP-like protein 1 (TBPL1), the vertebrate-specific TRF3 and the Drosophila-specific TRF4 and TRF5. These TBP family proteins often display divergent protein-DNA and protein-protein interactions and have been shown to play important roles in development and differentiation (Akhtar and Veenstra, 2011; Duttke *et al.*, 2014). The emergence of these TBP-family proteins has been proposed to have facilitated the evolution of alternative pre-initiation complexes driving key species- and cell-type specific gene expression programs in metazoans (Jacobi *et al.*, 2007; Akhtar and Veenstra, 2011; Duttke, 2015). To illustrate the divergence in TBP protein family interactions and functions, each of the TBP paralogs are discussed below.

TRF1 - an insect-specific TBP paralog with a key role in RNA Pol III transcription

The insect-specific TBP paralog TRF1 is highly expressed in the nervous tissue and gonads from stage 13 in developing *Drosophila* embryos (Crowley *et al.*, 1993; Hansen *et al.*, 1997). Knockout of TRF1 results in male sterility and nervous system defects, highlighting a central role of this TBP paralog in *Drosophila* neuronal functionality and germ cell development (Crowley *et al.*, 1993). TRF1 shares 60% amino acid sequence identity and 80% amino acid sequence similarity with the *Drosophila* TBP (*Dm*TBP) DNA-binding domain, with a high degree of identity of amino acid residues involved in DNA binding. Similar to TBP, TRF1 has been shown to bind the Adenovirus 2 major late (Ad2ML) TATA sequence in DNA footprinting experiments and in electrophoretic mobility shift assays (EMSAs). In this context, TRF1 TATA box-binding is strongly stimulated by the presence of TFIIB and TFIIA (Crowley *et al.*, 1993; Hansen *et al.*, 1997). TRF1 was also found to be sufficient to direct RNA Pol II *in vitro* transcription from the adenovirus E4 promoter, with TRF1-mediated transcription being diminished by omission of TFIIA and abolished by omission of TFIIB (Hansen *et al.*, 1997).

While TRF1 and *Dm*TBP share similar DNA-binding properties and interactions with TFIIA and TFIIB, they differ in their interactions with many key transcription factors and with respect to gene-specific promotor occupancy. TRF1 does not complex with TBP-associated factors (TAFs) that form, together with TBP, the TFIID complex (Hansen *et al.*, 1997) and instead is found strongly associated with BRF1,

a RNA Pol III transcription factor (Takada *et al.*, 2000). Possibly as a consequence of these differences in protein-protein interactions, differences in the DNA-binding interface, and alternative interactions with activator proteins, TRF1 and TBP were found to localise to different chromosomal sites on *Drosophila* polytene chromosomes (Hansen *et al.*, 1997; Holmes and Tjian, 2000). Antibody staining of polytene chromosomes suggested that TBP is ubiquitously involved in transcription, whereas TRF1 localised to a limited subset of sites, many of which corresponded to tRNA genes, suggesting a selective role of TRF1 in RNA Pol III transcription (Hansen *et al.*, 1997). Indeed, TRF1-depleted *Drosophila* cell extracts were shown to be unable to support *in vitro* transcription from an array of RNA Pol III promotors, unless rescued by the addition of recombinant TRF1-BRF complex (Takada *et al.*, 2000). In line with this, a genome wide Chip-on-chip study found both TRF1 and BRF1 to be localised at the majority of RNA Pol III promotors and only at a subset of RNA Pol II promotors (Vaughan, Noden and Beier, 2007). However, more recent experiments revealed an essential role of TBP, and not TRF1, in RNA Pol III transcription of U6 spliceosomal RNA and suggest that, in insects, different classes of RNA Pol III promotors assemble different preinitiation complexes containing either TBP or TRF1 (Verma *et al.*, 2013).

TRF2 (TBPL1) - a metazoan TBP paralog lacking TATA-box binding activity with a species-specific role in development

TRF2, also known as TBP-like protein (TLP) or TBP-like protein 1 (TBPL1) is present in all metazoans and is approximately 40 % identical and 60 % similar in amino acid sequence to the TBP DNA-binding domain (Dantonel *et al.*, 1999). Multiple sequence alignment of TRF2 proteins has highlighted a large degree of evolutionary divergence within the TRF2 family relative to TBP, suggesting that this TBP paralog may have species-specific functions (Dantonel *et al.*, 1999). Strikingly, while the saddle-shaped structure of the TBP DNA-binding domain appears conserved, three of the four phenylalanine residues involved in intercalation into the TATA box DNA sequence are mutated in TRF2 proteins (Dantonel *et al.*, 1999). Given the mutation of these phenylalanine residues strongly diminishes TBP TATA box binding, it is perhaps unsurprising that although both *Drosophila* and human TRF2 can bind to DNA and associate with TFIIA and TFIIB, they do not bind to a prototypical TATA box DNA in a sequence-specific manner (Moore *et al.*, 1999; Rabenstein *et al.*, 1999). In line with this, mouse TRF2 was shown to be unable to stimulate RNA Pol II transcription *in vitro* from the adenovirus E4 or Ad2ML promotors (Ohbayashi, Makino and T. A. Tamura, 1999). Furthermore, human TRF2 inhibited transcription from Ad2ML and HSV TK TATA box-containing promotors in human nuclear extract, possibly through sequestration of TFIIA (Moore *et al.*, 1999).

While TRF2 appears to play no role in the transcription of TATA-containing promotors (Moore *et al.*, 1999; Wang *et al.*, 2014), TRF2-containing complexes have been shown to play a key role in stimulating transcription from an array of alternative core promotor elements. Elegant experiments have shown that TRF2 plays an important stimulatory role in driving transcription from TCT core promoter element-dependent ribosomal protein gene promotors in humans and *Drosophila* (Wang *et al.*, 2014). Binding to a subset of these ribosomal protein promotors has recently been shown to involve recruitment by transcription factor M1BP, although it is unlikely that this is the only TRF2 interaction-partner involved (Baumann and Gilmour, 2017). TRF2 has also been shown to interact with the DNA replication-related element-binding factor (DREF) and to upregulate gene expression at many key promotors containing a DNA replication-related element (DRE; Hochheimer *et al.*, 2002; Baumann and Gilmour, 2017). In addition, *Drosophila* TRF2 plays a key role in transcription from DPE-dependent promotors, one of the most prevalent core promotor elements in Drosophila, and positively regulates transcription from the histone H1 promotor, which lacks TATA, TCT or DPE elements (Isogai *et al.*, 2007; Kedmi *et al.*, 2014).

TRF2 ablation has highlighted crucial species-specific roles of this TBP paralog at the organismal level. In worms, frogs and fish knock down of TRF2 protein levels results in early embryonic arrest (Akhtar and Veenstra, 2011), while in mice TRF2 knockdown specifically affects spermatogenesis in males. Several models have been proposed as to how TRF2 regulates gene expression in different organisms, not all of which involve alternative TRF2-containing preinitiation complexes. The association of TRF2 with components of the nucleosome remodelling factor complex (NURF) and DREF in *Drosophila* embryos (Hochheimer *et al.*, 2002), as well as with the association of murine TRF2 with TIPT, a testis-specific protein with strong interactions with heterochromatin proteins HP1α and HP1γ (Brancorsini *et al.*, 2008), has led to the proposal that TRF2 plays a direct regulatory role in chromatin remodelling during spermatogenesis (White-cooper and Davidson, 2011). In addition, recent work by Martianov *et al.* suggests that murine TRF2 is recruited together with TBP to haploid cell promotors and may function in tandem with TBP as a subunit of TFIIA or the testis-specific TFIIA paralog ALF (TFIIA-like factor), thus challenging earlier models proposing that TRF2 replaces TBP in driving RNA Pol II transcription initiation complex assembly (Martianov *et al.*, 2016). While a great deal of work remains to establish the exact mechanisms by which TRF2 imparts its regulatory functions, the importance of this TBP paralog in gene regulation and development of metazoans has been firmly established.

TRF3 – a vertebrate-specific TBP paralog that is closely related to TBP

TRF3 is a vertebrate-specific TBP paralog with 95 % amino acid sequence identity to the structured TBP DNA-binding domain (Jallow *et al.*, 2004). Similarly to TBP, TRF3 interacts with both TFIIB and TFIIA, displays prototypical TATA-box binding and supports basal RNA Pol II transcription *in vitro* (Balduf *et al.*, 2004; Jallow *et al.*, 2004; Akhtar and Veenstra, 2009). TBP and TRF3 have been shown to be functionally redundant across a variety of promotor contexts both *in vitro* and *in vivo* (Jallow *et al.*, 2004; Akhtar and Veenstra, 2009). However, TBP and TRF3 are not completely interchangeable with regard to their function as some promotors specific for either TBP or TRF3 have been identified (Hart *et al.*, 2007; Jacobi *et al.*, 2007).

TRF3 is highly expressed in *Xenopus* and mouse oocytes, where TBP protein expression is undetectable. A switch from TRF3 to TBP occurs during embryonic development, facilitated by proteolytic cleavage of TRF3 in the late stages of oocyte meiotic maturation and the translation of stored maternal TBP mRNA before the mid-blastula stage (Akhtar and Veenstra, 2009). In line with these observations, TRF3 expression is required for embryonic development in fish and frogs (Balduf *et al.*, 2004; Jallow *et al.*, 2004). In zebrafish, TRF3 interacts with TAF3 and binds to the promoter of the *mespa* transcription factor to increase *mespa* protein levels and commit the mesoderm to the haematopoietic lineage (Hart *et al.*, 2007, 2009). In *Xenopus,* TRF3 plays a key role in the expression of vertebrate specific embryonic genes and dorso-ventral patterning (Jacobi *et al.*, 2007). More recently, TRF3 has been found to be required for mesendoderm specification in human embryonic stem cells, where TRF3 was observed to localise to the promoters of mesendoderm signature genes and promote their transcription (Liang *et al.*, 2020). However, this role in development is not conserved in mice. TRF3-null mice are viable but females born infertile due to defects in folliculogenesis (Altobelli, Poch and Gazdag, 2009).

TRF4 and TRF5 – Drosophila specific TBP paralogs

Gene duplication has led to the existence of two additional TBP family proteins in *Drosophila* designated TRF4 and TRF5. These proteins display a greater degree of divergence to TBP than TRF1 and TRF2, and have both lost their nuclear localisation sequence (Kurshakova *et al.*, 2019). TRF4 was found to be more highly expressed than TRF5, with peak expression in the testes (Kurshakova *et al.*, 2019). TRF4 was shown to have a cytoplasmic localisation and appears to associate with the endoplasmic reticulum (ER). In agreement with this, TRF4 copurifies with ER-associated proteins. These observations suggest neofunctionalization of TRF4 and the adoption of ER-associated

functions. The localisation of TRF5 and its association with interaction partners have not yet been assessed.

TBP family proteins in basal eukaryotes

Initially, the emergence of TBP paralogs was thought to be a feature of metazoans and theorised to have aided the evolution of complex body plans (Jacobi *et al.*, 2007; Akhtar and Veenstra, 2011; Duttke, 2015). However, the discovery of divergent TBP family proteins in basal eukaryotes such as *Entomoeba histolytica* (Castañon-Sanchez *et al.*, 2010), *Trypanosoma brucei* (Ruan *et al.*, 2004), *Cryphecodinium cohini* (Guillebault *et al.*, 2002) and *Trichomonas vaginalis* (Parra-Marín *et al.*, 2019) suggests that TBP gene duplication events occurred either earlier in evolution, or multiple times independently. TBP family proteins from basal eukaryotes often display altered DNA interactions, with protozoan TBPs either binding alternative DNA sequence motifs or having lost sequence-selective DNA binding altogether (Guillebault *et al.*, 2002; Parra-Marín *et al.*, 2019). The altered biochemical properties of these TBP family proteins highlights potentially divergent roles for TBP superfamily proteins within protozoa.

1.3.5. TBP family proteins in *Plasmodium falciparum*

Plasmodium falciparum has two TBP family proteins, here referred to as *P. falciparum* TATA box-binding protein (*Pf*TBP; PlasmoDB ID PF3D7_0506200) and *P. falciparum* TBP-like protein (*Pf*TLP; PlasmoDB ID PF3D7_1428800). Of these, only *Pf*TBP has been studied functionally *in vitro* (McAndrew *et al.*, 1993; Ruvalcaba-Salazar *et al.*, 2005; Gopalakrishnan *et al.*, 2009). As noted in Section 1.2.3., these studies concluded that *Pf*TBP binds to TGTAA and TATAA sequences in a sequence-specific manner (Ruvalcaba-Salazar *et al.*, 2005) and associates with DNA regions located 5' and 3' of genes expressed during the parasite erythrocyte stages, irrespective of gene activity (Gopalakrishnan *et al.*, 2009). However, data shown in these publications appear inconsistent. Specifically, immunoblot analyses using a polyclonal antibody rased against *Pf*TBP in the study by Ruvalcaba-Salazar *et al.* identified *Pf*TBP as a 26 kDa protein in *P. falciparum* extracts (Ruvalcaba-Salazar *et al.*, 2005), whereas immunoblot analysis by Gopalakrishnan *et al.* using a different polyclonal antibody identified a band with an apparent molecular mass of 42 kDa (Gopalakrishnan *et al.*, 2009), broadly consistent with the 38 kDa molecular mass of *Pf*TBP predicted by *P. falciparum* genome sequence analysis (The Plasmodium Genome Consortium, 2001). Interestigly, both studies also make use of the same anti-human TBP antibody to support the validity of their immunoblot results, again reporting contrasting results despite using the identical antibody.

Furthermore, and with these potential inconsistancies in mind, the results of DNA binding experiments published so far, supported in part by antibody supershift experiments performed with the polyclonal antibody produced by Ruvalcaba-Salazar *et al*., appear not to allow clear conclusions regarding *Pf*TBP DNA sequence specificity. It is also unclear how binding of *Pf*TBP to TGTAA and TATAA sequences would contribute to promoter recognition within the A/T-rich context of the *Plasmodium* genome. So far, systematic evolution of ligands by exponential enrichment (SELEX) experiments carried out in our group showed some preference for DNA sequences enriched in A and T nucleotides but failed to to identify specific DNA sequence motifs preferentially recognised by *Pf*TBP (Milton, 2017). Thus the molecular properties of *Pf*TBP and DNA motifs preferentially bound by *Pf*TBP remain to be clearly defined.

There are currently no published studies on *Pf*TLP properties or function. However, initital characterisation of *Pf*TLP by our group has demonstrated *Pf*TLP DNA-binding activity that remains to be further characterised (Milton, 2017 ; Bing, 2015).

1.3.6. Transcription factor IIB (TFIIB)

TFIIB is a multifunctional RNA Pol II general transcription factor, which plays key roles in transcription initiation, transcription start site (TSS) selection and transcription termination (O'Brien and Ansari, 2022).

TFIIB is composed of two regions, a highly structured C-terminal region, which engages in interactions with TBP and DNA and a largely structurally disordered N-terminal region, which is required for recruitment of RNA Pol II to the preinitiation complex (Fig. 5; Deng and Roberts, 2007).

The structured C-terminal region is composed of two imperfect cyclin-like repeats, each consisting of five α-helices that form a compact globular domain (Nikolov *et al.*, 1995). The two cyclin-like domains are connected by a short linker that bridges the cleft between the two domains. Limited hydrophobic contacts between the two cyclin-like domains allow for a degree of flexibility in interdomain conformation (Nikolov *et al.*, 1995). TFIIB binds to the TBP-TATA box nucleoprotein complex through interactions between amino acids located in the basic TFIIB cleft and the acidic stirrup of TBP. Binding of TFIIB to the TBP-TATA box complex facilitates extensive TFIIB interactions with DNA upstream and downstream of the TATA box (Nikolov et al., 1995; Tsai & Sigler, 2000). These contacts stabilise the TBP-TATA box complex, contribute to promoter selection and have been proposed to facilitate the correct orientation of the preinitiation complex on the promoter (Hisatake, Roeder and Horikoshi, 1993; Nikolov *et al.*, 1995; Tsai and Sigler, 2000).

TFIIB has been shown to recognise so-called B-recognition elements located upstream (BREu) or downstream (BREd) of the TATA box core promoter element (Lagrange *et al.*, 1998; Deng and Roberts, 2005). TFIIB binds the BREu core promoter element (5'-G/C-G/C-G/A-C-G-C-C-3') through contacts between a helix-turn-helix motif, located in the second cyclin-like domain of TFIIB, and the DNA major groove (Lagrange *et al.*, 1998). Binding of TFIIB to the BREu can occur in the absence of TBP, suggesting a possible role for TFIIB-BREu interactions in promotor recognition at TATA-less promoters (Lagrange *et al.*, 1998). Relevant to the characterisation of *Plasmodium* TFIIB presented in this thesis, this helix-turn-helix motif appears not to be evolutionary conserved and to be absent in yeast, plants and *Plasmodium* TFIIB (Lagrange et al., 1998; Bing, 2015). Subsequent investigations of TFIIB-DNA interactions identified a downstream BRE (BREd consensus sequence: G/A-T/G/A-T/G-G/T-T/G-T/G; Deng and Roberts, 2005). BREd is recognised by the loop located between H2 and H3 of the TFIIB first direct repeat, which appears to be broadly conserved across eukaryotes, including *Plasmodium* (Deng & Roberts, 2005; Bing, 2015).

Interestingly, TFIIB contacts with the BREd appear to only take place in the context of the ternary TBP-TFIIB-TATA box complex. Furthermore, BREd was shown to modulate promotor activity either positively or negatively in a context-dependent manner, influenced by the presence or absence of the BREu motif (Deng and Roberts, 2005).

The largely structurally disordered TFIIB N-terminal portion can be subdivided into several regions based on function, including the B-ribbon, the B-reader (also known as the B-finger), and the B-linker (Fig. 5; Kostrewa *et al.*, 2009). TFIIB is vital for RNA Pol II recruitment to the TBP-TATA box complex. This recruitment involves the B-ribbon which interacts with the RNA Pol II dock domain and which is correctly positioned by additional contacts between the structured TFIIB C-terminal domain and the RNA Pol II wall (Buratowski and Zhou, 1993; Hisatake, Roeder and Horikoshi, 1993; Pardee, Bangur and Ponticelli, 1998; Chen and Hahn, 2003; Kostrewa *et al.*, 2009). Following recruitment of RNA Pol II, the B-linker domain of TFIIB has been found to play a key role in promoter opening, which occurs 20 bp downstream of the TATA-box (Kostrewa *et al.*, 2009). The transcription bubble is then stabilised by the B-reader region, comprised of an α-helix and a mobile loop, which also functions to scan the template strand and aid in recognition of the TSS (Pinto *et al.*, 1994; Pardee, Bangur and Ponticelli, 1998; Cho and Buratowski, 1999; Faitar, Brodie and Ponticelli, 2001; Zhang, Carson and Ma, 2002; Kostrewa *et al.*, 2009).

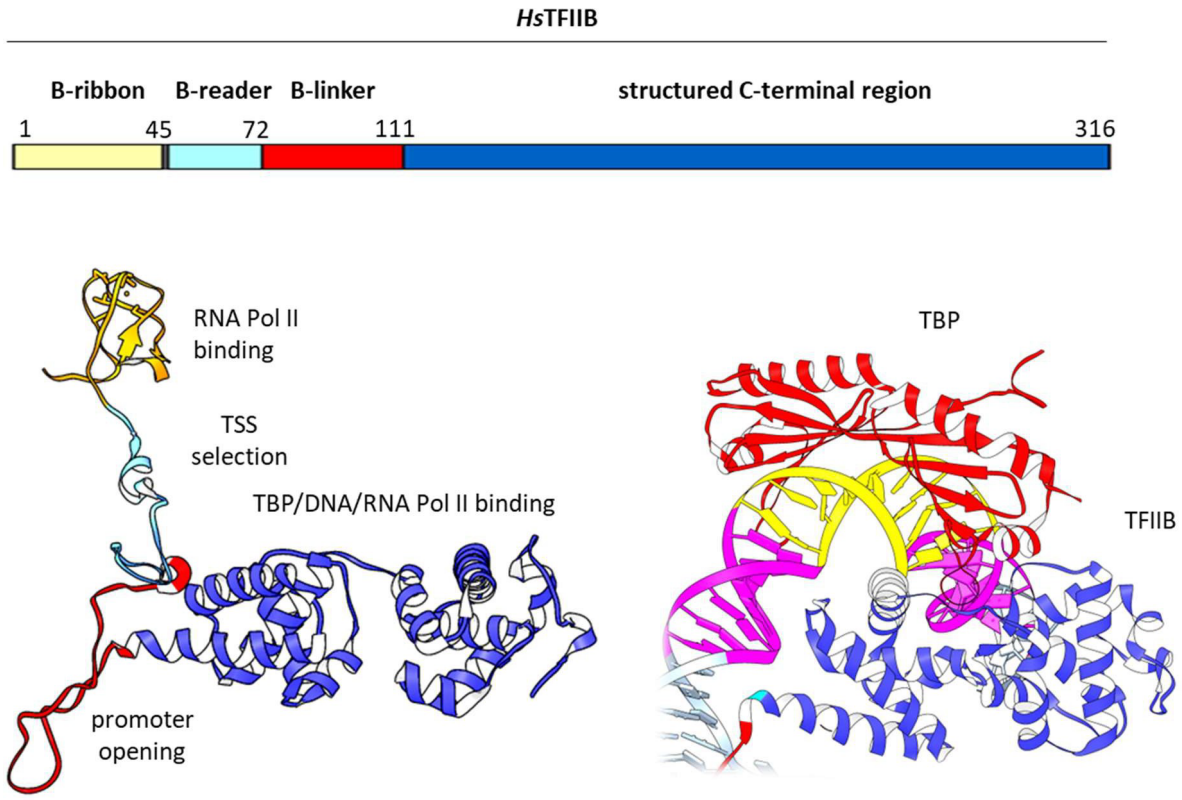


Figure 5. Structure of *Hs*TFIIB. *Top:* Schematic of human TFIIB (*Hs*TFIIB) structure with B-ribbon, B-reader, B-linker and structured C-terminal region indicated. *Bottom left*: Cartoon representation of the *Hs*TFIIB structure. TFIIB regions are coloured according to the schematic structure shown at the top with associated functions indicated. *Bottom right*: Cartoon representation of the ternary TBP-TFIIB-TATA box complex structure, TBP is shown in red, the structured C-terminal domain of TFIIB is shown in blue. The DNA TATA box sequence is coloured yellow, upstream (BREu) and downstream (BREd) TFIIB recognition elements are coloured pink. Structures of isolated *Hs*TFIIB and of the ternary TBP-TFIIB-TATA box complex were modelled using the atomic structure coordinates for the human preinitiation complex (PIC) in the open state (5IY7 and 5IYB respectively), determined by electron microscopy (He *et al*., 2016).

In addition to its key role in transcription initiation, TFIIB has also been proposed to play an important role in the regulation of transcription by transcription activators. TFIIB has been demonstrated to interact with the transactivation domains of a number of transcription activators including VP16, CTF1 and p53 (Deng and Roberts, 2007). Studies suggests that these interactions facilitate transcription activation through enhanced recruitment of TFIIB to the core promoter, by modulating TFIIB interaction with BRE elements, or by altering TFIIB conformation (Deng and Roberts, 2007). Several studies have demonstrated intramolecular interactions between the N-terminal B-reader and C-terminal second cyclin-like repeat of TFIIB, promoting a transcriptionally inactive 'closed' TFIIB conformation (Roberts and Green, 1994; Hayashi *et al.*, 1998; Bangur *et al.*, 1999; Wu and Hampsey, 1999; Hawkes, Evans and Roberts, 2000; Zhang *et al.*, 2000; Glossop, Dafforn and Roberts, 2004). TFIIB interactions with transcriptional activators disrupt these TFIIB

intramolecular interactions and thus favour the 'open', transcriptionally active form of TFIIB. In support of the importance of the interplay of intramolecular TFIIB and intermolecular TFIIB-activator interactions for transcriptional activation, TFIIB mutants with enhanced intramolecular interactions displayed reduced activator-TFIIB interactions, reduced activator-dependent recruitment of TFIIB to promotor-bound TBP, and reduced activated transcription levels (Hawkes, Evans and Roberts, 2000; Glossop, Dafforn and Roberts, 2004).

Following RNA Pol II transcription initiation complex formation, promoter opening and the initiation of transcription, TFIIB has additional important regulatory functions in RNA Pol II promotor escape and in the transition between transcription initiation and productive elongation. Steric clashes between the growing RNA transcript and the B-reader in the RNA Pol II active site prevent initial transition of the RNA Pol II complex from initiation to elongation resulting instead in consecutive rounds of abortive transcription, producing RNAs less than seven nucleotides in length. Escape towards productive elongation requires the release of TFIIB which is thought to occur upon the abrupt re-annealing of the upstream DNA strands after RNA Pol II promoter clearance (Pal, Ponticelli and Luse, 2005; Kostrewa *et al.*, 2009).

TFIIB has also been shown to play a role in RNA Pol II transcription termination (Henriques *et al.*, 2012; Allepuz-Fuster *et al.*, 2019; O'Brien and Ansari, 2022). TFIIB was found to localise to both the 5' and 3' regions of many transcribed genes (O'Brien and Ansari, 2022). In addition, TFIIB has been shown to interact with a number of termination factors, including Ssu72, a 3' end processing-termination factor and subunit of the CPF 3' end processing-termination multiprotein complex in budding yeast (Wu *et al.*, 1999; Dichtl *et al.*, 2002; Bratkowski *et al.*, 2018), and Rna15, a subunit of the CF1 3' end processing-termination multiprotein complex (El Kaderi *et al.*, 2009). These results have been further corroborated by a recent quantitative proteomic study that showed that TFIIB associates with all three multiprotein termination complexes, CF1, CPF and Rat1 in yeast (O'Brien and Ansari, 2024). In agreement with this, depletion of TFIIB resulted in genome-wide transcription termination defects in the mammalian HAP-1 cell line (Santana *et al.*, 2022). Different models have been proposed to explain a role of TFIIB in transcription termination. TFIIB may be involved indirectly in transcriptional termination and facilitate recruitment of termination factors to the terminator region or, alternatively, TFIIB could act directly in termination, by insertion of the B-reader into the RNA exit channel of RNA Pol II, thereby displacing the native RNA strand (O'Brien and Ansari, 2022).

Finally, multiple studies have demonstrated a key role for TFIIB in gene loop formation, involving the juxtaposition of the promoter and terminator regions observed at many yeast and mammalian genes

(Singh and Hampsey, 2007; Wang, Fairley and Roberts, 2010; Medler *et al.*, 2011; Allepuz-Fuster *et al.*, 2019). Gene looping has been proposed to play a role in transcriptional regulation by facilitating the recycling of transcribing RNA Pol II from the terminator region back to the core promoter (El Kaderi *et al.*, 2009), by acting as a barrier to transcriptional interference between closely spaced genes (Henriques *et al.*, 2012), and by contributing to promoter directionality by recruiting termination factors to the upstream promoter region to mediate the termination of antisense RNA transcription (Tan-Wong *et al.*, 2012; Al Husini, Kudla and Ansari, 2013).

1.3.7. TFIIB in *Plasmodium falciparum*

P. falciparum TFIIB (PlasmoDB ID PF3D7_0110800) was one of the first general transcription factors to be identified by *in silico* analyses (Coulson, Hall and Ouzounis, 2004). Expression of *Pf*TFIIB has been detected by mass-spectrometry in sporozoites from mosquito salivary glands (Lindner *et al.*, 2013), during the intraerythrocytic stages (Treeck *et al.*, 2011), and in gametocytes, with greater expression in male gametocytes compared to female gametocytes (Lasonder *et al.*, 2016). AlphaFold structure prediction suggests that the tertiary structure of *Pf*TFIIB is conserved in that its C-terminal region folds into two cyclin-like repeats as seen in higher eukaryotes (see Results, Fig. 36). Furthermore, multiple amino acid sequence alignment suggests that, within *Pf*TFIIB, the B-reader amino acid residues, which play a key role in TSS selection, and the charged cluster region of the B-reader, that engages in intramolecular interactions with the C-terminal domain, are highly conserved. Similarly, the B-linker residues involved in promotor opening are also conserved in *Pf*TFIIB (Bing, 2015).

In contrast to key interacting amino acids within the largely disordered N-terminal region of *Pf*TFIIB that show a large degree of conservation, there is considerable divergence at a number of key interaction interfaces within the structured C-terminal cyclin-like *Pf*TFIIB domains, including the interface predicted to interact with *Pf*TBP or *Pf*TLP (Buendía-Orozco *et al.*, 2005; Bing, 2015). In addition, the helix-turn-helix comprised of H4' and H5' within the TFIIB second cyclin-like repeat, which mediates TBP-independent contacts of *Hs*TFIIB with the BREu, is also not conserved in *Plasmodium*. However, there are a number of positively charged residues in the region which could contribute to DNA interactions (Bing, 2015). Notably, the loop between H2 and H3 of the first direct repeat, which mediates *Hs*TFIIB interactions with the BREd also harbours extreme substitutions in *Pf*TFIIB (Bing, 2015). Finally, *Pf*TFIIB possesses additional basic amino acid sequences at the N-terminus (aa1-37) and the C-terminus (aa350-367) that are absent in TFIIB in previously characterised model eukaryotes and which may contribute to divergent *Pf*TFIIB function (Talvik, 2016). Initial characterisation of *Pf*TFIIB in our laboratory revealed that *Pf*TFIIB is able to engage in TBP-independent DNA interactions and forms a ternary *Pf*TBP-*Pf*TFIIB-DNA complex (Talvik, 2016).

1.4. Emerging complexities of transcriptional regulation - the importance of fuzzy interactions

Up until recently, models for transcriptional regulation in eukaryotes had largely focused on the role of specific protein-protein and protein-DNA interactions, generally emphasising the role and interactions of protein domains with a defined three-dimensional structure. However, less specific 'fuzzy interactions' have recently been recognised as a key regulatory component of many cellular processes – including transcription. The next section reviews the current understanding of the role of biomolecular condensation, a process largely driven by "fuzzy" biomolecular interactions, in transcription regulation, with focus on the existing models for the formation and regulation of transcriptional condensates.

1.4.1. Biomolecular condensation as an important regulator of cellular function

Spatiotemporal separation of biomolecular processes is key to cellular function. One mechanism whereby cells achieve compartmentalisation is through the utilisation of phospholipid bilayers forming membrane-enclosed organelles such as the nucleus, mitochondria and the endoplasmic reticulum - another is through biomolecular condensation.

Biomolecular condensation refers to the formation of membrane-less non-stoichiometric assemblies enriched in certain biomolecules (Choi, Holehouse and Pappu, 2020; Mittag and Pappu, 2022). These assemblies have been found to play a key role the regulation of biochemical reactions within living cells by selectively partitioning and concentrating specific biomolecules. In the cell, biomolecular condensates are referred to as membrane-less organelles (MLO) to distinguish them from organelles surrounded by a lipid bilayer. Examples of MLOs include the nucleolus, P-bodies, stress granules, DNA damage foci, and, relevant to gene regulation, RNA transport granules and super-enhancer condensates (Fig. 6A; Sabari, Dall'Agnese and Young, 2020; Mitrea *et al.*, 2022).

Importantly, biomolecular condensation is an umbrella term referring to the formation of concentrated membrane-less assemblies through a variety of different, but not mutually exclusive, mechanisms. Biomolecular condensates can form through the physicochemical process of liquid-liquid phase separation (LLPS), the condition-dependent de-mixing of one phase of biomolecules into two distinct liquid phases, one more concentrated in certain biomolecules than the other (Section 1.4.2; Shin and Brangwynne, 2017). Alternatively, biomolecular condensates may form through percolation, referred to as the local non-stochiometric clustering of interacting biomolecules into 'pre-percolation clusters' through the establishment of networks of weak

molecular interactions, which become system-spanning above a specific threshold concentration (Fig. 6B; Deniz, 2022; Mittag and Pappu, 2022; Das and Deniz, 2023). Active processes, such as the targeted deposition of biomolecules by molecular motors, have also been shown to play a key role in the formation of biomolecular condensates. For example, dynein has been shown to transport abortive translation initiation complexes to stress granules and P-bodies (Fig. 6B; Loschi *et al.*, 2009; Berry, Brangwynne and Haataja, 2018).

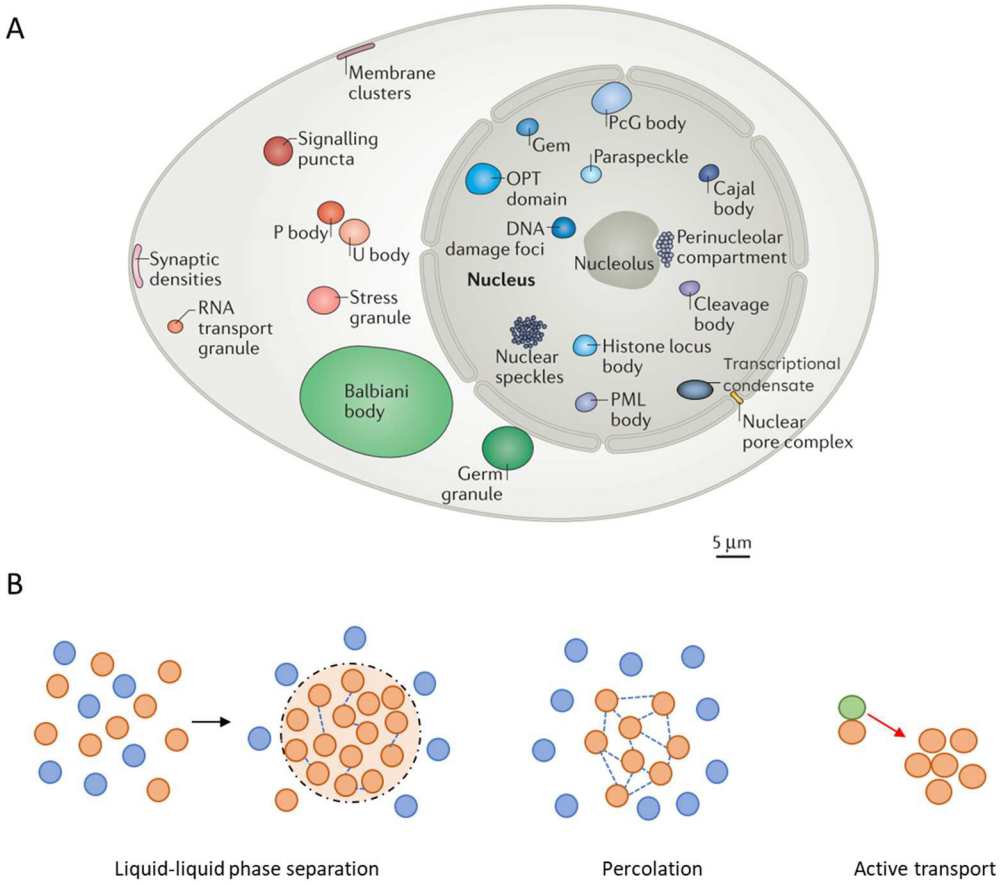


Figure 6. **Biomolecular condensation as a regulator of cellular function**. **(A)** Schematic of known cellular membrane-less organelles formed through biomolecular condensation (adapted from Banani *et al*. 2017). **(B)** Illustration of driving forces leading to the formation of biomolecular condensates. Note that these processes are not mutually exclusive. Orange circles represent biomolecules enriched in condensates, blue circles represent molecules excluded from phase separated condensates or pre-percolation clusters, black dashed line represents the phase boundary, dashed blue lines represent dynamic interactions of multivalent biomolecules, a black arrow represents a change in environmental conditions that triggers liquid-liquid phase separation, green circles represent molecular motors.

Understanding the molecular mechanisms governing the assembly and regulation of biomolecular condensation is key for our understanding of cell physiology and, also, to our understanding of disease. A plethora of recent studies have linked cellular dysregulation of biomolecular condensation to the occurrence of many diseases, including neurodegenerative diseases and cancer (Shin and

Brangwynne, 2017; Alberti and Dormann, 2019). In addition, given that biomolecular condensates are likely to play a key role in pathogen cellular biology, the selective targeting of pathogen biomolecular condensates has emerged as a promising avenue for research into pathogen therapeutics.

1.4.2. Liquid-liquid phase separation as a driver of biomolecular condensation

LLPS refers to a spontaneous de-mixing of a single phase of solutes into two distinct liquid phases, one more concentrated in a given solute than the other (Fig. 7A). This de-mixing can be observed by the formation of liquid droplets within which the phase-separating molecule is concentrated (the 'concentrated phase'), which are separated from the surrounding 'dilute phase' by a phase boundary (Fig. 7A). LLPS is driven by intermolecular interactions between biomolecules, including proteins and nucleic acids, as well as molecule-solvent interactions (Fig. 7B). Key to the phenomenon of biomolecular phase separation is the concept of multivalency - which refers to the propensity of a given molecule to dynamically engage simultaneously in multiple weak non-covalent interactions, either with different molecules (heterotypic interactions) or with molecules of the same type (homotypic interactions) (Banani *et al.*, 2017). Multivalent interactions allow for the formation of complex and dynamic molecular interaction networks in three dimensional space that can drive phase separation under conditions when the combined strength of intermolecular interactions between phase-separating molecules and intermolecular interactions between surrounding solvent molecules exceeds the combined strength of interactions between phase-separating molecules and solvent molecules and the entropic cost of de-mixing (Fig. 7A; Banani *et al.*, 2017).

Within these networks of interactors, constituents of condensates formed through LLPS have been grouped into two categories, 'scaffolds' and 'clients' (Banani *et al.*, 2016). Biomolecules that are classified as scaffolds are typically highly multivalent and are strong drivers of condensate assembly whilst clients are molecules that selectively partition into condensates formed by scaffold molecules (Banani *et al.*, 2016). The high valency of scaffold proteins is generally conferred by repeated modular interaction domains, and/or intrinsically disordered regions (IDRs), which do not adopt a defined tertiary structure but dynamically sample an array of possible conformations (Banani *et al.*, 2017). IDRs are often amino acid sequences of low complexity, comprised of only a limited subset of amino acids, most commonly, serine, glycine, glutamine, asparagine, phenylalanine and tyrosine (Banani *et al.*, 2017; Alberti, Gladfelter and Mittag, 2019). Some IDRs also contain a high fraction of charged amino acid residues (Banani *et al.*, 2017; Lyons *et al.*, 2023). These amino acid residues can be arranged in blocks or repeated patterns within IDRs and mediate multiple weak transient interactions, such as pi-pi, cation-pi, electrostatic interactions and weak hydrophobic interactions

(Fig. 7B; Alberti, Gladfelter and Mittag, 2019; Dignon, Best and Mittal, 2020). The ability of IDRs to engage in transient multivalent interactions combined with their intrinsic conformational flexibility makes IDRs common and strong drivers of LLPS. Nucleic acids are also by nature multivalent polymers capable of engaging in multiple simultaneous homotypic and heterotypic interactions (Jain and Vale, 2017; Larson *et al.*, 2017; Shin and Brangwynne, 2017; Claeys Bouuaert *et al.*, 2021; Alemasova and Lavrik, 2022). RNA in particular has been shown to be a strong driver of LLPS and to play a key role in modulating biomolecular condensation in the cellular context.

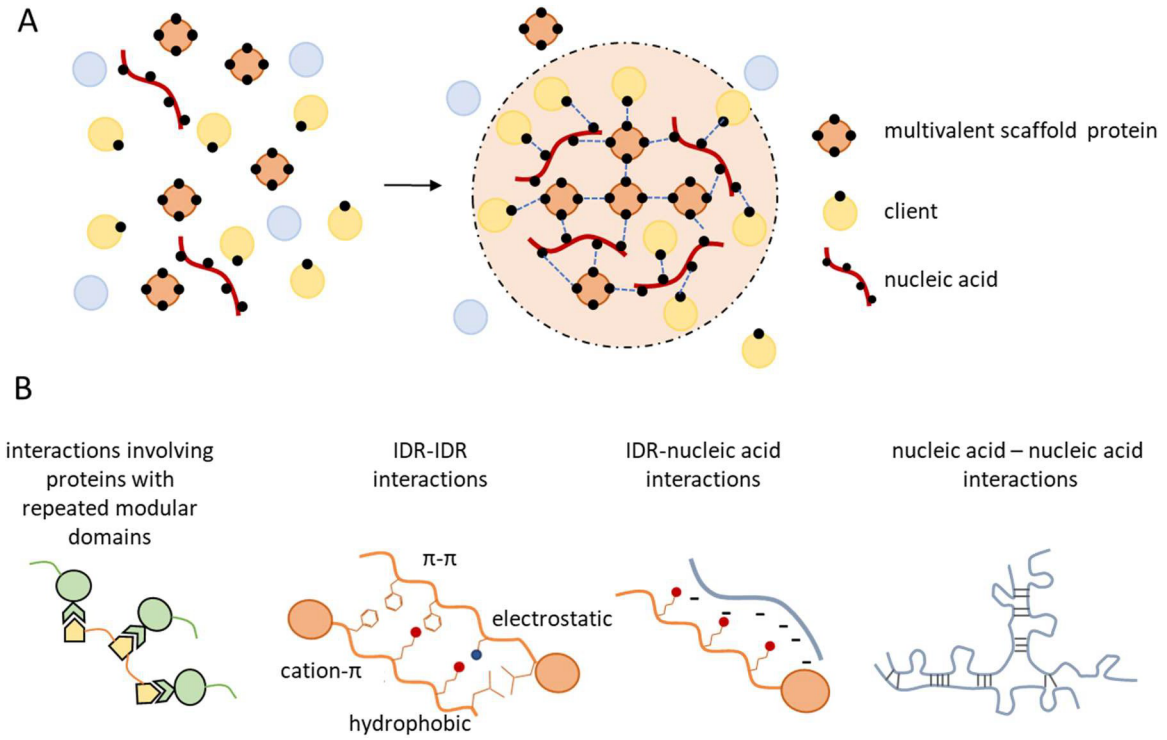


Figure 7. Multivalency drives liquid-liquid phase separation.
A) Schematic depicting phase separation in response to an environmental change such as alterations in biomolecule concentration, temperature or pH. Phase separation occurs as a consequence of multivalent interactions that allow for the formation of interaction networks, and compensate for the entropic costs of de-mixing in the phase separating system. (B) Examples of non-covalent interactions that contribute to the valency of biomolecules and thus increase the propensity of a biomolecule to undergo phase separation.

Given that intermolecular interactions drive phase separation, LLPS is highly sensitive to changes in conditions that affect homotypic and heterotypic interactions between biomolecules, including changes in the concentration of biomolecules and environmental conditions such as temperature, pH, and ionic strength (Dignon, Best and Mittal, 2020). Condensates formed through LLPS are reversable, and rapidly assemble and disassemble in response to changes in these environmental conditions. Importantly, condensates formed through LLPS are defined by properties associated with liquids. At a macroscopic level, liquid biomolecular condensates are characterised by a spherical

morphology, resulting from interfacial tension, and the ability to fuse and relax into larger spherical condensates upon contact (Alberti, Gladfelter and Mittag, 2019). At the molecular level, phase-separated molecules within liquid biomolecular condensates are mobile both within the condensate and across the phase boundary, a property which can be experimentally assessed using fluorescence recovery after photobleaching (FRAP) experiments (Alberti, Gladfelter and Mittag, 2019). Finally, simple binary LLPS systems consisting of a singles species of phase-separating biomolecules and a solvent are characterised by a system-specific threshold concentration of the phase separating molecule, referred to as the saturation concentration (C_{sat}), below which the phase separating molecules and solvent molecules are homogenously distributed in a single phase, and above which the system de-mixes into coexisting phases, a dense phase, containing the phase-separating molecule at high concentration, surrounded by a dilute phase, containing the phase-separating molecule at a lower concentration. Increasing the concentration of the phase separating biomolecule above C_{sat} results in a corresponding increase in the volume fraction of the condensed phase without affecting the concentration of the phase separating molecule in both phases. The successful experimental determination of C_{sat} and the observation that increasing the concentration of the phase separating molecule above C_{sat} increases the volume fraction of the condensed phase without increasing the concentration of the phase separating molecule in either phase, are generally taken as strong evidence that biomolecular condensation results from LLPS. (Alberti, Gladfelter and Mittag, 2019). However, percolation and phase separation are often coupled processes, and in some systems, non-stochiometric pre-percolation mesoscale clusters of phase-separating molecules can be detected below the saturation concentration as a result of interaction network formation in the absence of phase separation (Mittag and Pappu, 2022).

Together, a defined C_{sat}, an increase in the volume of the condensed phase with increasing concentration of the phase separating molecule, spherical condensate morphology, the observation of fusion events between condensates, liquid-like diffusion kinetics of molecules within condensates and condensate reversibility are considered defining characteristics of liquid phase-separated droplets. It is important to note, however, that condensates that were initially formed by LLPS may possess properties that deviate from these hallmarks of LLPS at the time point of observation. This is because condensates formed through LLPS can display ‘metastability’, a propensity to transition from the liquid state to a gel-like or solid state and to form gel-, glass- or fibril-like assemblies over time. This transition has also been referred to as hardening or maturation and is proposed to occur as a consequence of increased or novel intermolecular interactions, for example β-strand interactions, resulting from the high concentration of molecules in the condensed phase of liquid condensates, vitrification resulting from interactions with a high rate constant for association (k_{on})

and a low low rate constant for dissociation (k_{off}) and entanglement of polypeptide chains within the condensed phase (Fig. 8;Banani *et al.*, 2017). Maturation of biomolecular condensates within cells is proposed to be regulated by a variety of mechanisms including the action of chaperones and disaggregases as well as buffering condensate components such RNA (Banani *et al.*, 2017; Maharana *et al.*, 2018). Mis-regulation of maturation has been linked to the formation of pathological aggregates involved in neurodegenerative diseases.

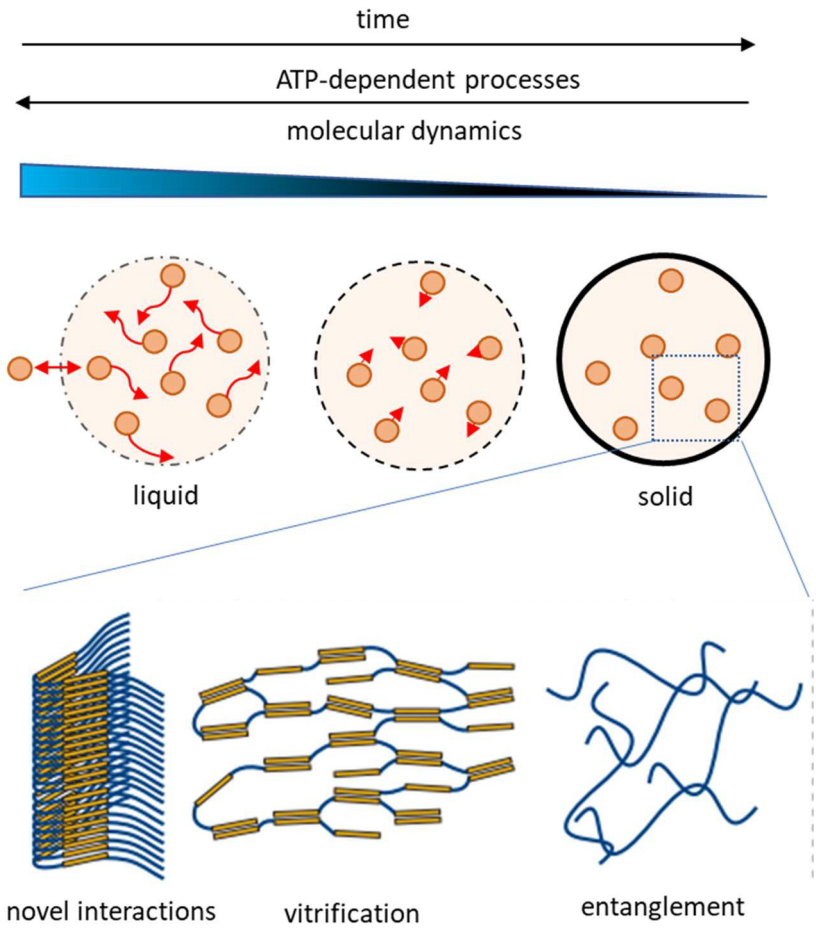


Figure 8. Condensates formed through LLPS can undergo maturation or hardening. Schematic depicting the process of maturation whereby initially liquid condensates transition into gel-like or solid assemblies in which molecules have reduced mobility. Inset depicts processes proposed to result in maturation including the establishment of novel interactions which can result in the formation of amyloid-like fibrils, vitrification and the entanglement of polypeptide chains within the condensed phase. These processes are time dependent, can be regulated by the presence of buffering components such as RNA and can be reversed by ATP-dependent processes such as the action of disaggregases and chaperones. Figure adapted from Banani *et al* 2017.

1.4.3. Transcription condensates

Clusters of RNA Pol II in the nucleus were first observed in the early 90s, (Jackson *et al.*, 1993; Wansink *et al.*, 1993; Iborra *et al.*, 1996) leading to the proposal of a transcription factory model in which a small number of transcription factories or clusters of RNA Pol II and other transcription factors were responsible for driving transcription of specific genes in eukaryotic cells. These early observations were supported by later studies that showed that distal transcribed genes colocalise to RNA Pol II foci (Osborne *et al.*, 2004; Mitchell and Fraser, 2008; Miranda *et al.*, 2010) and demonstrated the dynamic recruitment of positive regulators of transcription such as CDK9 into active transcription factories (Ghamari *et al.*, 2013).

The transcription factory model is useful in that it provides a simple mechanism whereby coregulation of distant genes can be achieved and helps to explain the discontinuous nature of transcription activity at active gene promoters, commonly termed transcriptional bursting. However, while the concept of transcription factories is now widely accepted, the nature of these factories has become a topic of great debate (Chong *et al.*, 2018; Palacio and Taatjes, 2022). Two main models for transcription factories in eukaryotic gene regulation are currently under debate, one that proposes that transcription factories are liquid phase separated condensates and another that proposes the formation of localised transcription hubs that are not separated by a phase boundary from the surrounding solution (Chong *et al.*, 2018; Palacio and Taatjes, 2022).

1.4.4. The phase separation model for the formation and regulation of transcription condensates

The phase separation model for transcriptional control was first proposed by Hnisz *et al.* upon reflecting on the properties of transcription factories, specifically the properties of super enhancers, which consist of clusters of enhancers and associated transcription factors (Hnisz *et al.*, 2017). Super enhancers are characterised by high concentrations of transcription factors, are formed in a highly co-operative manner, and can dissipate spontaneously upon perturbation of a single cofactor. Furthermore, transcription factors enriched at super enhancers have a high prevalence of low complexity domains, which are often intrinsically disordered and capable of driving *in vitro* phase transitions (Kato *et al.*, 2012; Kwon *et al.*, 2013; Hnisz *et al.*, 2017). The model proposes that binding of multiple gene-specific DNA-binding activators and their coactivators to multiple enhancer regions promotes de-mixing of these factors into concentrated transcription condensates, capable of partitioning RNA Pol II and positive-acting regulators of RNA Pol II transcription, resulting in robust transcription from multiple promotors at the same time (Fig. 9). Critically, in this model, the formation of phase separated transcription condensates is proposed to be tuned by modifications to the system that alter the valency or affinity of the phase-separating biomolecules or by the absence or presence of additional regulatory factors, allowing the rapid formation and dissolution of transcription condensates in response to environmental changes or regulatory cues (Hnisz *et al.*, 2017).

Shortly after its original proposal, a flurry of research results provided evidence in support of the phase separated model for transcriptional control. This included the demonstration of *in vitro* phase separation of Med1 (the largest mediator subunit), of the fully assembled mediator complex, and of key coactivator BRD4 (Boija *et al.*, 2018; Sabari *et al.*, 2018) and the demonstration that Med1 condensates formed *in vitro* selectively partition transcriptional activators, key signalling proteins from the WNT, TGF-β and JAK/STAT pathways and RNA Pol II (Boija *et al.*, 2018; Sabari *et al.*, 2018;

Zamudio *et al.*, 2019). These transcription factors were also found to localise to super enhancer-associated puncta in cultured cells, which displayed hallmarks of phase separated condensates (Sabari *et al.*, 2018). Importantly, amino acid residues found to be key to the *in vitro* phase separation of a number of these transcription factors were also shown to be required for mediator recruitment and transcription output *in vivo* (Boija *et al.*, 2018; Lyons *et al.*, 2023). Elegant experiments have demonstrated that selective partitioning of positive regulators of transcription into mediator condensates is driven by blocks of charged amino acids within the IDRs of Med1 and partitioned transcription regulators (Lyons *et al.*, 2023). Scrambling or deleting charged blocks within the IDRs of Med1 or partitioned transcription regulators, was shown to diminish partitioning into Med1 condensates and decrease transcription, while addition of charged blocks into the IDRs of non-partitioning proteins increased incorporation of these factors into Med1IDR condensates with ramifications on transcription (Lyons *et al.*, 2023). These results provide first insights into the molecular grammar driving selective partitioning of transcriptional regulators into transcription condensates.

The phase separated model for transcription regulation is further supported by experiments demonstrating that the CTD of RPB1, the largest subunit of RNA Pol II (Boehning *et al.*, 2018), is a potent driver of LLPS. Phase separation of the RPB1 CTD was found to be regulated by phosphorylation, with phosphorylation diminishing incorporation of RNA Pol II into transcriptional condensates and, instead, increasing partitioning of phosphorylated RNA PoI II into splicing condensates (Guo *et al.*, 2019). Together, these results highlight the important role of post-translational protein modifications in regulating the composition of transcription condensates formed through LLPS. More specifically, these results reveal a key role for the phosphorylation status of RNA Pol II in the shuttling of transcribing RNA Pol II between specialised nuclear condensates promoting transcription initiation or co-transcriptional RNA processing.

RNA is now also understood to play a key role in the regulation of transcription condensate assembly and disassembly. It turns out that low RNA-to-protein ratios, resulting from the production of short RNAs during transcription initiation, promote the formation of transcription condensates while high RNA-to-protein ratios, produced during productive transcription, lead to condensate dissolution due to unfavourable charge ratios (Henninger *et al.*, 2021; Sharp *et al.*, 2022). This phenomenon explains transcriptional bursting, the discontinuous nature of transcript production, and has been proposed to function as a feedback control mechanism to regulate the transcriptional output of phase separated transcription condensates (Henninger *et al.*, 2021).

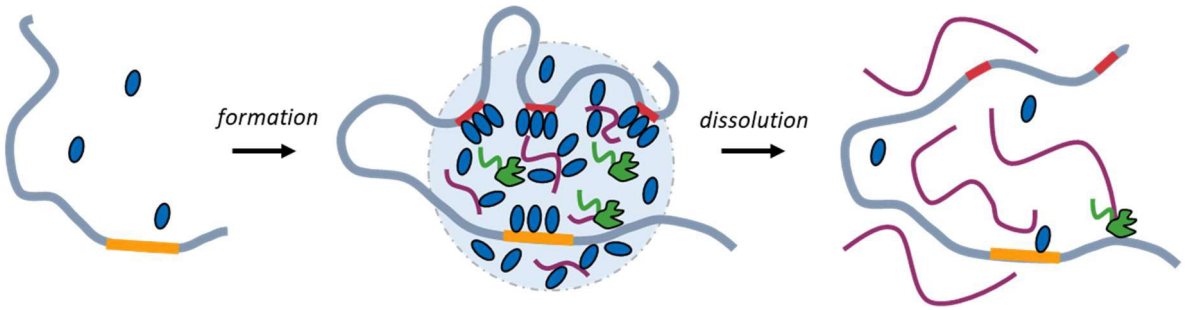


Figure 9. A phase separation model for the formation and dissolution of transcription condensates. High concentrations of transcription activators and co-factors (represented by blue ovals) form a local interaction network with RNA Pol II (green) and positive-acting regulators of RNA Pol II transcription which results in the formation of a phase separated condensate. Condensate formation is promoted by low concentrations of RNA (purple) and increases transcriptional output. Condensate dissolution is triggered by including high RNA concentrations, or other environmental changes or regulatory cues. Binding regions for transcription factors within promoters and enhancers are represented by orange and red squares.

1.4.5. The hub model for the formation and regulation of transcription condensates

In many ways the hub model for transcription regulation is highly similar to the phase separation model. Like the phase separation model, it proposes that RNA Pol II, trans-activators, mediator and other co-factors as well as chromatin re-modellers cluster in a relatively small number of nuclear locations. These clusters form as a result of interaction networks between multivalent proteins and contain high concentrations of proteins facilitating mRNA transcription. Similar to the LLPS model for transcriptional control, multiple promotors and enhancers can form part of this network and interactions can be modulated by modifications or additional factors such as RNA (Palacio and Taatjes, 2022).

Where this model differs from the LLPS model is on the question of whether or not transcription condensates represent a distinct phase. This distinction may at first seem to be semantic in nature, but whether or not transcription condensates represent a distinct phase has important implications with regard to the way molecules interact within the condensates. Important properties of phase-separated condensates that differ from the properties of local concentrations of transcription factors that form in the absence of phase separation include water content, which can be reduced relative to the surrounding dilute phase, and the dielectric constant of the system (Palacio and Taatjes, 2022). Changes in these parameters can have significant impact on biomolecular interactions and can, for example, even result in the partial unwinding of DNA (Nott, Craggs and Baldwin, 2016). Furthermore, the presence of a phase boundary may significantly impact the exchange of molecules between transcription condensate and surrounding solution. Finally, the

formation of phase separated transcription condensates implies a concentration threshold for biomolecules driving the phase transition, which would not be required for the formation of transcriptional hubs resulting from networking transitions in the absence of phase separation.

Most of the debate around whether or not phase separated condensates represents a distinct phase stems from the difficulty in demonstrating whether transcription condensates represent a distinct liquid phase *in vivo* and whether phase separation actually correlates with transcriptional activity (Palacio and Taatjes, 2022). Many of the condensate properties, such as spherical morphology, the dynamics of molecules within condensates and the exchange of molecules between condensates and surrounding environment, are shared between LLPS and alternative mechanisms of condensate formation. One of the first examples for a role for condensate formation by means other than LLPS came from work on viral replication compartments which, despite displaying many of the hallmarks of LLPS, were found not to represent phase separated compartments (McSwiggen *et al.*, 2019). Instead these viral replication compartments were demonstrated to form as a consequence of non-specific low-affinity interactions between DNA-binding proteins and naked viral DNA which increase the residence times of DNA-binding proteins within the viral replication compartment and result in a local accumulation of DNA-binding proteins in the absence of a phase boundary (McSwiggen *et al.*, 2019). Importantly to the debate surrounding the nature of transcription factories, recent work investigating the role of phase separation in the activity of a number of transcriptional activators, found that, although there was a correlation between the propensity of these activators to undergo LLPS and transcription activation levels, the actual occurrence of phase separation had a neutral or even negative impact on transcriptional activity (Trojanowski *et al.*, 2022). In support of these findings, expression of endogenous genes regulated by the oncogenic transcription factor EWS:FLI1, were reported to be decreased when LLPS is artificially induced by increasing EWS::FLI1 homotypic interactions beyond a certain threshold (Chong *et al.*, 2022). These findings are further consistent with earlier work showing that local hubs of transcription activators, with the capacity to stabilise DNA binding, recruit RNA Pol II and increase transcription of target genes, can be detected in the absence of phase separation using advanced imaging techniques (Chong *et al.*, 2018). Together these observations suggest that there may be different 'sweet spots' for different gene regulatory systems, for optimal clustering of transcription factors through multivalent interactions, and that further concentration of factors resulting in phase separation can have a neutral or even repressive effect.

Given that these studies have only investigated the role of a limited number of transcription factors in regulating specific target genes or reporter gene arrays, further research is required to disentangle the role of phase separation in transcription from the role of non-phase separated

hubs of transcription factors, across multiple gene regulatory contexts. At this time, both the condensate and hub models for transcription condensates should be considered.

1.4.6. Biomolecular condensation in *Plasmodium falciparum* biology

Early *in silico* analysis of the *P. falciparum* proteome highlighted that *P. falciparum* proteins have an extraordinarily high prevalence of low complexity regions with high potential to be structurally disordered (Aravind *et al.*, 2003). These regions of *P. falciparum* proteins are comprised of only a limited set of amino acids and are often highly enriched for asparagine (Aravind *et al.*, 2003). Interestingly, 87% of *P. falciparum* proteins contain at least one low complexity region compared to 60% to 78% in other eukaryotes (DePristo, Zilversmit and Hartl, 2006). The high prevalence of low complexity regions in *P. falciparum* can be explained by the high genomic A/T-content as well as high levels of genomic recombination in *P. falciparum*, which have been proposed to cause increased replication slippage and unequal crossover recombination (DePristo, Zilversmit and Hartl, 2006). However, this does not exclude that these low complexity regions may play important functional roles in *P. falciparum* biology. A number of possible functional advantages for the high prevalence of disordered proteins in *P. falciparum* have been proposed. These include a possible role for disordered regions in protein-DNA and protein-protein interactions (Karlin *et al.*, 2002) and a role in generating antigenic diversity, for example by acting as a hotspot for recombination between members of the *var* gene family. (DePristo, Zilversmit and Hartl, 2006). In light of the recent wealth of papers demonstrating the importance of low complexity sequences and intrinsically disordered regions in driving biomolecular condensation and phase separation (Boija *et al.*, 2018; Alberti, Gladfelter and Mittag, 2019), the high prevalence of these regions in *P. falciparum* may suggest a key role of biomolecular condensation through multivalent interactions of structurally disordered regions in the cellular compartmentalisation of this parasite.

To our knowledge, only one study has so far directly investigated the phase separation properties of specific *P. falciparum* proteins (Voß *et al.*, 2023). This study demonstrates Ca^{2+}-induced phase separation of centrins to be dependent on a disordered N- terminal region, and further provides evidence for liquid-like properties of centrin condensates *in vivo*.

1.4.7. Transcription condensates in *Plasmodium falciparum*

Experiments monitoring transcription in *P. falciparum* using 5'-bromouridine 5'-triphosphate modified nucleotides to label nascent RNA demonstrated that transcription occurs in

discrete nuclear foci during the *P. falciparum* blood stages (Moraes *et al.*, 2013). These foci appear to be developmentally regulated and to be localised to sites of reduced chromatin density (Moraes *et al.*, 2013). Furthermore, activation of the virulence-associated var genes has been proposed to occur by translocation of *var* genes to a specific compartment located at the nuclear periphery where the *var* gene promoter is incorporated into a transcription factory (Duraisingh *et al.*, 2005; Ralph, Scheidig-Benatar and Scherf, 2005). These observations support the proposal that transcription regulation in *P. falciparum* involves transcription condensates. However, *P. falciparum* transcription factors shown to play a key role in transcription condensate formation in higher eukaryotes are either highly divergent in structure or altogether absent, suggesting that *P. falciparum* transcription condensates may be distinct with respect to their composition and physicochemical properties.

There are currently no published studies reporting on the condensation properties of *Plasmodium* transcription factors. Work by past and current group members aimed to address this knowledge gap and to investigate the potential role of biomolecular condensation in *P. falciparum* transcription regulation. This research has so far demonstrated that a number of *Plasmodium* general transcription factors, including *Pf*TBP, *Pf*TLP, the IDR of the putative *Pf*TFIIA-αβ subunit and the C-terminal domain of the largest *P. falciparum* RNA Pol II subunit, *Pf*RBP1-CTD, have a high propensity to undergo LLPS (Karamanof, 2022; Pegram 2022; Michowicz, Joanna, in preparation). Phase separation of these *P. falciparum* transcription factors was shown to occur at close to physiological concentrations, with C_{sat} values in the low micromolar range, suggesting that these factors may be drivers in the formation of transcription condensates. Consistent with their highly divergent structural properties, the condensation properties of these *P. falciparum* transcription factors differed significantly from their human counterparts (Karamanof, 2022; Pegram 2022). Given that different small therapeutic molecules selectively partition into specific cellular condensates in cultured human cells (Klein *et al.*, 2020), *Plasmodium*-specific transcription condensate properties provide an attractive new opportunity for targeted drug design and the production of potent novel antimalarials.

1.5. Aims of this study

This study aims to address the current dearth of knowledge pertaining to the molecular mechanisms governing gene regulation at the level of transcription initiation in *P. falciparum*. The study focuses on the functional characterisation of three key proteins involved in promoter recognition and assembly of the RNA Pol II transcription initiation complex - the general transcription factors *Pf*TBP, *Pf*TLP and *Pf*TFIIB. The results of this work are presented in three parts, as outlined below.

PART 1: Characterisation of *Pf*TBP and *Pf*TLP DNA-binding properties

TBP binding to the core promoter is considered to be the first step in the assembly of the RNA Pol II transcription initiation complex during gene transcription. *In silico* analyses have identified two TBP-family proteins in *Plasmodium falciparum*, here referred to as *Pf*TBP and *Pf*TLP. A key aim of this study was to determine if *Pf*TBP or *Pf*TLP display prototypical TATA box-binding activity, with sequence specificity for the TATAWAWR TATA box consensus sequence.

PART 2: Characterisation of *Pf*TFIIB

In well-studied eukaryotes, TFIIB binds to the TBP-TATA nucleoprotein complex, recruits RNA Pol II during transcription initiation complex assembly, and plays important roles in transcription start site selection and the initiation of transcription. In contrast to previously characterised TFIIB in well-studied model organisms, which is unable to stably bind to DNA on its own, previous work by our group demonstrated that *Pf*TFIIB possesses strong TBP-independent DNA-binding activity. This observation suggested that in *P. falciparum* TFIIB may contribute, in addition to *Pf*TBP or *Pf*TLP, to promotor selection and the initiation of RNA Pol II transcription initiation complex assembly. The focus of this work was to further characterise TBP-independent *Pf*TFIIB-DNA interactions, and to investigate functional interactions of *Pf*TFIIB with *Pf*TBP and *Pf*TLP.

PART 3: Biomolecular condensation of *Pf*TFIIB and the potential role of *Pf*TFIIB in transcription condensate assembly

Previous work in the laboratory had demonstrated that *Pf*TBP and *Pf*TLP can undergo liquid-liquid phase separation at physiological concentrations and has identified IDRs in *Pf*TFIIA as well as the CTD of the largest *Pf*RNA Pol II subunit, RPB1, as potent drivers of phase separation. The work presented here provides a first characterisation of the condensation properties of *Pf*TFIIB in isolation, in the presence of nucleic acids, and in combination with other *P. falciparum* general transcription factor components.

Chapter 2: Materials and Methods

2.1 Characterisation of *Pf*TBP and *Pf*TLP

2.1.1 Multiple sequence alignment and phylogenetic analysis

Sequences of TBP family proteins from select higher eukaryotes, archaea and previously characterised protozoa were retrieved from the National Centre for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov/). Additional apicomplexan TBP-family protein sequences were retrieved by Blast-P search (Altschul *et al.*, 1990; McGinnis and Madden, 2004) using *Hs*TBP, *Pf*TBP and *Pf*TLP protein sequences as queries. The species and accession numbers for sequences used in the phylogenetic analysis are listed in Supplementary Table 1.

Sequence alignments were performed using the Expresso mode of T-coffee (Armougom *et al.*, 2006), a consistency-based alignment program, with default settings on the T-coffee online server. The Expresso mode of T-coffee was chosen as this mode also integrates structural information, where available, which can increase alignment accuracy, particularly the accuracy of alignments of highly divergent sequences (Armougom *et al.*, 2006). Minor manual edits were made to further optimise the alignment. These edits were made to regions of the alignment that border structured regions including the region directly preceding S1 corresponding to G160-P163 in *Hs*TBP, the amino acid position corresponding to L165 in *Hs*TBP, the region between H2 and S1' corresponding to L244 – P247 in *Hs*TBP, the amino acid position corresponding to K337 in *Hs*TBP. The ClustalW consensus grouping system was used to assess conservation (Thompson, Higgins and Gibson, 1994).

Amino acid sequences corresponding to the conserved structured C-terminal DNA-binding domain of *Hs*TBP (G160-K337) were used to generate a maximum likelihood phylogram in Mega X (Kumar *et al.*, 2018). There were a total of 174 aa positions in the final dataset used for the construction of the phylogenetic tree. The evolutionary history was inferred by using the Maximum Likelihood method and Le Gascuel substitution model (Le and Gascuel, 2008). Initial trees for the heuristic search were obtained automatically by applying Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances, which were estimated using the JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories; +G, parameter = 1,4646). Tree node confidence was determined by bootstrap (1000 bootstrap replicates).

2.1.2. Tertiary structure Prediction

Tertiary structure predictions were retrieved from the AlphaFold protein structure database (Jumper *et al.*, 2021; Varadi *et al.*, 2022), available at https://alphafold.ebi.ac.uk/, using the uniport identifiers Q8I440 (*Pf*TBP, 38 kDa), P32086 (*Pf*TBP, 26 kDa) and A0A144A3A5 (*Pf*TLP). The quality of these models was independently verified using PROCHECK (Laskowski *et al.*, 1993) and Verify-3D (Luthy, Bowie and Eisenberg, 1992) using the SAVES online server (https://saves.mbi.ucla.edu/). The values for the experimentally determined TBP crystal structures as determined by Buendía-Orozco *et al*. were included for comparison (Table S2, Buendía-Orozco, Guerrero and Pastor, 2005). The predicted models were visualised using UCSF Chimera (Pettersen *et al.*, 2004), available at https://www.cgl.ucsf.edu/chimera/.

2.1.3. Bacterial expression vectors for 6His:*Pf*TBP and 6His:*Pf*TLP

Unless otherwise stated, all bacterial expression vectors used in this study are derivatives of pET11d (Novagen) expressing N-terminal six histidine-tagged recombinant proteins (Hoffmann and Roeder, 1991) under control of the T7 promoter. For protein expression, pET11d derivatives were transformed into *Escherichia coli* BL21(DE3), which express T7 polymerase under control of the lacUV5 promoter, which can be induced by addition of the allolactose mimic isopropyl β-d-1-thiogalactopyranoside (IPTG) to the culture medium.

pET11d expression vectors for the expression of 26 kDa 6His:*Pf*TBP and 38 kDa full-length 6His:*Pf*TBP, 6His:*Pf*TLP and for 6His:*Pf*TLP variants with F-to-A amino acid substitutions F60A, F283A and the double substitution F60A/F283A were previously generated in the laboratory (Milton, 2017; Bing; 2015; Adebolajo, 2016).

2.1.4. Bacterial expression and purification of 6His:*Pf*TBP and 6His:*Pf*TLP

E.coli BL21(DE3) CodonPlus RIL (*Agilent technologies*) were used for recombinant expression of all recombinant *P. falciparum* proteins generated in this study. *E.coli* BL21 CodonPlus (DE3)-RIL are specifically designed to increase expression of heterogenous proteins from organisms with A/T-rich genomes and carry additional copies of the argU, ileY, and leuW tRNA genes, which encode tRNAs for the arginine codons AGA and AGG, the isoleucine codon AUA, and the leucine codon CUA, respectively. Transformation of *E.coli* BL21(DE3) CodonPlus RIL cells was carried out using the Inoue protocol (Inoue, Nojima and Okayama, 1990) before plating onto LB agar plates supplemented with 34 µg/ml chloramphenicol (Cam), 100 µg/ml ampicillin (Amp), and 55 mM glucose.

For protein expression, *E.coli* cells from a single freshly transformed colony were used to inoculate 50 ml starter cultures in LB medium supplemented with 55 mM glucose, 100 µg/ml Amp and 34 µg/ml Cam and incubated for 16 h at 37°C with shaking.

6His:*Pf*TBP 26 kDa and 38 kDa variants were produced by leaky expression in the absence of inducer. For this purpose, starter cultures were pelleted (4000 x g, 10 min, RT) and resuspended in fresh LB containing 100 µg/ml Amp to remove β-lactamase activity. 500 ml expression cultures (LB containing 100µg/ml Amp) were inoculated with 1/100 volume resuspended overnight starter culture and grown at 37°C to an optical density at 600nm of 0.5 (OD_{600} 0.5), rapidly cooled on ice to 30°C and incubated for further 18 h at 30°C. Cells were harvested by centrifugation (4500 x g, 20 min, 10°C).

Prior to sonication, *E. coli* cell pellets were subjected to osmotic shock following a published protocol designed to increase the yield of low abundance recombinant proteins purified by immobilised metal affinity chromatography (IMAC) from *E.coli* cells (Magnusdottir *et al.*, 2009a). In brief, pelleted *E.coli* cells were resuspended in 5 ml, per gram cell pellet, sucrose buffer (50 mM HEPES pH 7.9, 20% sucrose, 1 mM EDTA) and centrifugated again (7 000 x g, 30 min, 4°C). After removal of the supernatant, the cell pellet was resuspended in 5 ml 5 mM $MgSO_4$ per gram cell pellet and incubated on ice for 10 min. Finally, *E.coli* cells were again pelleted by centrifugation (4500 x g, 20 min, 4°C) and the supernatant, containing periplasmic material, which increases metal ion leakage from immobilized metal affinity chromatography (IMAC) resins (Magnusdottir *et al.*, 2009), was discarded.

Following osmotic shock, pelleted cells were resuspended in 5 ml sonication buffer (50 mM Tris-Cl pH 7.9 at 4°C, 500 mM NaCl, 5 mM imidazole pH 8, 0.1 % NP-40, 8 mM β-mercaptoethanol) supplemented with 50 µl of protease inhibitor cocktail (P8340, Sigma-Aldrich) and 0.2 µl benzonase (E1014, Sigma-Aldrich) per gram cell pellet. Cells were lysed by sonication on ice, using six 10 s pulses followed by a 30 s cooling period at power level 4 (MSE Soniprep 150). The sonicated lysate was incubated on ice for 1 h to allow for benzonase digestion of nucleic acids and cleared twice by centrifugation (20000 x g, 35 min, 4°C).

6His:*Pf*TBP proteins were purified from the cleared lysate using TALON metal affinity resin (Clontech Laboratories) in a Poly-Prep® chromatography column (Bio-Rad) at 4°C according to the manufacturers' instructions. Prior to protein elution, the resin was washed with 25 packed column volumes (PCV) BC-buffer (20 mM TrisHCl pH7.9 at 4°C, 20% v/v glycerol, 0.2 mM EDTA, 0.1% v/v NP-40, 10 mM β-mercaptoethanol) containing 300 mM KCl and 20 mM imidazole. Proteins were eluted with 10 PCV BC-buffer containing 300 mM KCl and 250 mM imidazole. Elution fractions (300 µl) were kept on ice and analysed by SDS-PAGE before fractions containing 6His:*Pf*TBP protein were pooled, flash frozen in liquid N_2 and stored at -80°C.

Bacterial expression vectors for wildtype 6His:*Pf*TLP and mutant variants used in this study were generated previously in the laboratory (van der Linden, 2019) and expression and purification of 6His:*Pf*TLP proteins was carried out previously as described (van der Linden, 2019).

2.1.5. Affinity purification of anti-*Pf*TBP, anti-*Pf*TLP and anti-*Pf*TFIIB antibodies

Custom polyclonal rabbit antibodies for *Pf*TBP, *Pf*TLP and *Pf*TFIIB were generated by *BioGenes* (Berlin, Germany). *Pf*TBP antibodies were raised against the peptide SEYDNNEKEKSDDL, corresponding to aa 125 - 138 in the disordered N-terminal region of the predicted 38 kDa *Pf*TBP protein (PlasmoDB ID PF3D7_0506200). *Pf*TLP antibodies were raised against the peptide NDENKSNDNKEQND, corresponding to aa 294-307, located in IDR1 between S3' and S4' (Section 3.1.2). *Pf*TFIIB antibodies were raised against the peptide sequence LKQKYLSEDKRKKN, corresponding to aa 354 – 367 located at the extreme C-terminus. Antisera were further purified by antibody affinity chromatography following a standard laboratory protocol (Harlow and Lane, 1988). Peptide-coupled affinity resins used for antibody purification were previously generated in the laboratory (Talvik, 2016) using the peptides used for rabbit immunisation (Biogenes) and SulfoLink™ coupling resin (Thermo Scientific) according to the manufacturer's instructions. Antisera were diluted with one volume 20 mM Tris pH 7.5 and centrifuged (1400 x g, 10 min, 4°C) to remove particulates and passed six times over the appropriate peptide-coupled resin at 4°C. Unbound material was removed by washing with 20 PCV 20 mM Tris pH 7.5 followed by 20 PCV of 20 mM Tris pH 7.5, 500 mM NaCl. Antibodies were first eluted with 5 PCV 100 mM Glycine pH 2.5. Elution fractions were collected and neutralised by drop-wise addition of equal volume 1 M Tris pH 8.8. The column was then equilibrated with 20 PCV 20 mM Tris pH 8.8 and an additional elution was performed with 5 PCV 100 mM Triethylamine pH 11.5, 15 ml of elution was collected and neutralised by dropwise addition of with 780 µl neutralisation solution, made up by mixing 600 µl 1M Tris pH 7.5 with 180 µl concentrated HCl. Elution fractions were pooled and dialysed overnight against PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4), pH 7.4. After dialysis, antibodies were precipitated by dropwise addition of saturated $(NH_4)_2SO_4$ to a final concentration of 50%. Precipitated antibodies were pelleted by centrifugation (12900 x g, 30 min, 4°C), and resuspended in PBS. Finally, a buffer exchange into PBS was carried out using PD Miditrap G-10 columns (Merck) to remove any excess salt from the antibody preparations. The concentrations of antibody preparations were determined by BioRad Protein Assay according to the manufacturer's instructions and the purity and integrity of the antibodies were verified by SDS-PAGE.

2.1.6. *Plasmodium falciparum* culture

P. falciparum DD2 (Wellems *et al.*, 1990) trophozoites from sorbitol synchronised cultures (Trager and Jansen, 1976) were kindly provided by Larnelle Garnie from the University of Cape Town Bioinorganic Research Group (Garnie, in preparation). Briefly, parasites were cultured in pre-washed type O^+ human erythrocytes (2% haematocrit) in RPMI 1640 media supplemented with AlbuMax (1.04% (w/v) RPMI-1640, 0.4% (w/v) D-glucose, 0.6% (w/v) HEPES, 0.0088 (w/v) hypoxanthine, 0.5% (w/v) AlbuMax and 0.005% (w/v) gentamycin, 0.21% sodium bicarbonate). Cultures were maintained at 37˚C and gassed with mixed air (5% CO_2, 5% O_2, 90% N_2). Parasites were synchronised by treatment with pre-warmed (37°C) 5% sorbitol for 10 min. Parasitaemia and the stage of cultures was determined by microscopy of sample smears stained with Giemsa.

Parasites were harvested at the trophozoite stage at parasitaemia between 5% and 15% by saponin lysis (0.15% saponin), pelleted by centrifugation (750 x g, 4 min, RT) and washed twice with PBS before resuspension in PBS.

2.1.7. Indirect immunofluorescence microscopy of *P. falciparum* blood-stage parasites

Indirect immunofluorescence microscopy was performed according to a published protocol for the immunolocalization of nuclear antigens in *P. falciparum* (Moll *et al.*, 2013). Red blood cells (RBCs) containing trophozoite stage parasites were pelleted (750 x g, 4 min, RT) to produce 100 µl of packed RBC. The RBC were resuspended in 2 ml of RPMI growth media and saponin was added to a final concentration of 0.07 % to induce red blood cell lysis and release of the parasites. The parasites were pelleted by centrifugation (750 x g, 4 min, RT) and washed twice in RPMI growth media. For antibody staining and immunofluorescence analysis, parasites were fixed in solution by resuspension in ice cold 4 % formaldehyde for 10 min, pelleted again and washed once with ice cold PBS, before being resuspended in ice cold PBS and stored at 4°C. Stored parasites were used within one week of original fixation as suggested (Moll *et al.*, 2013). For immunostaining, parasites were incubated with affinity purified custom polyclonal rabbit primary antibodies for *Pf*TBP or *Pf*TLP (Section 2.1.5, 1:100 dilution) in PBS containing 1% (w/v) BSA for 30 min at 37°C. Parasites were then pelleted (3400 x g, 1 min, RT) and washed twice with PBS before addition of donkey anti-Rabbit IgG Alexa-488 conjugated antibody (Amersham 711-546-152). After incubation for 30 min at 37°C, the parasites were washed twice with PBS and incubated with 1 µg/ml 4',6-diamidino-2-phenylindole (DAPI) for 5 min. The parasites were washed again three times with PBS, deposited on a coverslip and air dried. Samples were mounted with Miowol Mounting Media with n-propylgallate antifade (Sigma-Aldrich) and stored protected from light before visualisation.

Image acquisition was performed with a Nikon Ti-E inverted widefield fluorescence microscope equipped with a Nikon Intensilight C-HGFIE HG Pre-centered mercury-fiber illuminator. DAPI (320 - 380 nm excitation (Ex), 435 - 485 nm emission (Em), Nikon) or FITC (460 – 500 nm Ex, 505 - 560 nm Em, Nikon) filter sets were used for visualisation. The microscope was fitted with a CFI Apochromat-TIRF 100 x objective lens (Nikon) with 1.49 NA and a temperature-change correction ring, with a DS-Fi2 camera (Nikon) for acquisition. NIS-Elements AR software (Nikon) was used for image acquisition and processing.

2.1.8. Generation of *Plasmodium falciparum* extracts

Whole parasite extracts were prepared by resuspension of 5 x 10^5 to 5 x 10^6 *P. falciparum* blood-stage parasites prepared by saponin lysis of infected RBCs in SDS-PAGE loading buffer (250 mM Tris-Cl pH 6.8, 0.02% (w/v) Bromophenol blue, 30% (v/v) glycerol, 10% (w/v) SDS) and heating at 90 ^{o}C for 5 min. Samples were briefly centrifuged (13 000 x g, 30 s) to remove hemozoin before SDS-PAGE.

Nuclear and cytoplasmic extracts were prepared following a published protocol with minor modifications (Voss *et al.*, 2002). Early and late stage trophozoites were isolated from infected RBCs by saponin lysis (Section 2.1.6) and stored at -20°C. 1-6 x 10^9 parasites were resuspended in 800 µl lysis buffer (20 mM HEPES, pH 7.8, 10 mM KCl, 1 mM EDTA, 1 mM DTT, 1 mM PMSF, 0.65% NP-40) and incubated on ice for 5 min. Nuclei were pelleted by centrifugation (2500 x g, 5 min, 4°C) and the cytoplasmic extract was carefully removed and stored at -80°C. The nuclei pellet was washed twice with lysis buffer and resuspended in extraction buffer (20 mM HEPES pH 7.8, 800 mM KCl, 1 mM EDTA, 1 mM DTT, 1 mM PMSF). The resuspended nuclei were shaken vigorously for 30 min at 4°C before the extract was cleared by centrifugation (13000 x g, 30 min, 4°C) and the supernatant was diluted 1:1 in dilution buffer (20 mM HEPES pH 7.8, 1 mM EDTA, 1 mM DTT, 30% glycerol). The protein concentration of all fractions was determined using the BioRad Protein Assay according to the manufacturer's instructions, and fractions were aliquoted, flash-frozen in liquid N_2 and stored at -80°C.

2.1.9 Immunoblot analysis of *P. falciparum* proteins

Proteins were resolved by SDS-PAGE and transferred to Immobilon PVDF membrane (Millipore) in transfer buffer (25 mM Tris, 250 mM glycine, 10% methanol) using a Mini Trans-Blot® Electrophoretic Transfer Cell (BioRad) for 1 h 45 min at 100 V at 4°C. After transfer, the membrane was stained with 0.1% (w/v) Ponceau S (Sigma-Aldrich) in 5% (v/v) glacial acetic acid for 10 min, to visualise proteins. The stained membrane was then washed four times for 5 min with TBST (20 mM Tris pH 8.0, 150 mM NaCl, 0.1% (v/v) Tween-20) and blocked for 1 h with 5% (w/v) skim milk powder in TBST. The blocked membrane was incubated overnight at 4°C with primary antibody, diluted in 5%

skim milk powder in TBST. Commercial mouse anti-His antibody (MCA1396, BioRad) was used at a dilution of 1:5000, and immunoaffinity-purified anti-*Pf*TBP antibody at 1:5000, anti-*Pf*TLP antibody at 1:2000, and anti-*Pf*TFIIB at 1:2000 dilution. After over-night incubation, the primary antibody solution was removed and the membrane washed three times for 5 min in TBST before incubation with secondary goat anti-mouse IgG-horseradish peroxidase conjugate (A4416, Sigma-Aldrich, 1:3000 in TBST) for 1 h. After removal of the secondary antibody solution, the membrane was washed again three times for 5 min in TBST. Antibody signals were detected using the WesternBright Sirius (Advansta) chemiluminescence detection kit, following manufacturer's instructions, and visualised on Super RX X-ray film (Fuji).

2.1.10. Analysis of *Pf*TBP and *Pf*TLP DNA binding using electrophoretic mobility shift assays (EMSAs)

DNA-binding assays were carried out with 5'-biotinylated dsDNA probes to allow for chemiluminescent detection using streptavidin-conjugated horseradish peroxidase. The sequences of DNA probes used to test for specific binding of *PfTBP* or *PfTLP* to a prototypical TATA box are listed in Table 1 below. These included (i) a 40 bp Ad2ML-TATA probe, containing the adenovirus 2 major late promoter TATA-box region, a standard promoter probe used to study the interactions of TBPs from well-studied eukaryotes with a consensus TATA-box (Sawadogo and Roeder, 1985; Kim, Nikolov and Burley, 1993; Kuddus and Schmidt, 1993; Nikolov *et al.*, 1996; Zhao and Herr, 2002), (ii) a mutant variant of the 40 bp Ad2ML-TATA DNA probe, Ad2ML-GAGA, harbouring T-G substitutions at the first and third positions of the TATA-sequence, shown to abolish *Hs*TBP-TATA interactions (Gilfillan *et al.*, 2005), (iii) a 40 bp sequence encompassing the transcription start site of the mouse terminal deoxynucleotidyl transferase (mTdT) promoter, a natural G/C-rich TATA-less promoter; and (iv) a mutant variant of the mTdT promoter containing the consensus TATA sequence TATAAAA (Martinez *et al.*, 1998; Malecová *et al.*, 2007).

To generate short dsDNA probes (Tables 1 and 3), complementary single-stranded 5'-biotinylated oligonucleotides (inqaba Biotech) were combined in TE buffer (10 mM Tris-HCl pH 8.0, 1mM EDTA) containing 50 mM NaCl, heated for 5 min at 95°C and slowly cooled to RT over >1 h. After hybridisation, DNA probes were resolved by 5% PAGE in 1x TBE buffer, detected by staining with ethidium bromide, excised from the gel, recovered by passive diffusion and purified using the QIAEX II Gel Extraction Kit (Qiagen) according to the manufacturers protocol for the extraction of DNA fragments from polyacrylamide gels. The integrity of short dsDNA probes was verified by analytical 5% PAGE in 1x TBE and ethidium bromide staining.

Long EMSA probes (Table 2) were generated by PCR using biotinylated primers. Putative *P. falciparum* promoter regions used in this study include the promoter for the 130 kDa glycophorin-binding protein (GBP-130; Lanzer, de Bruin and Ravetch, 1992b) and the promoter for the knob-associated histidine-rich protein (KAHRP; Lanzer, De Bruin and Ravetch, 1992a).

pGEM-7Zf (Promega) derivative plasmids, pTHG5GBP-130, containing the DNA region from -179 to + 57 relative to the GBP-130 TSS (Lanzer, de Bruin and Ravetch, 1992b) and pTHG5KAHRP, containing the DNA region from -171 to +65 relative to the KAHRP TSS (Lanzer, De Bruin and Ravetch, 1992a) were generated previously in the laboratory (Hessler, 2014). Biotinylated dsDNA probes encompassing putative promoter regions and containing additional plasmid sequences used for PCR as indicated in Table 2, were generated by PCR using primers pTOTdTPCR-FWD and pTOTdTPCR-REV (Table S2). Probes GA, encompassing GBP-130 region from -179 to -58, and GB, encompassing the GBP-130 promoter region from -81 to +57 relative to the GBP-130 TSS, were generated by PCR using the primer pairs pTOTdTPCR-FWD and GBP-130-REV3 (GA) and GBP-130-FWD3 and pTOTdTPCR-REV (GB, Table S2). Probes KA and KB encompassing KAHRP promoter regions from -171 to -55 and -67 to + 65 relative to the KAHRP TSS were generated using the primer pairs pTOTdTPCR-FWD and pTOTdTPCR-REV (KA) and KAHRP-FWD3 and pTOTdTPCR-REV (KB, Table S2). PCR amplification was carried out using Kapa Taq polymerase (Merck) and pTHG5GBP-130 or pTHG5KAHRP as templates with an elongation temperature of 60°C, shown to enhance amplification of A/T-rich sequences (Su *et al.*, 1996). All PCR products used in EMSAs were gel-purified using the GeneJET gel extraction kit (Thermo Fisher Scientific) according to the manufacturer's instructions.

Plasmids pTOG5TdT(-41+33) and pTOG5TdT(-41/TATA/+33) (Malecov *et al.*, 2007) were used to PCR amplify the 266 bp mTdT and 260 bp mTdT-TATA probes using 5'-biotinylated primers pTOPCR-US-FWD and pTOTdTPCR-REV (Table S2).

DNA-binding reactions (10 µl) were initiated by addition of 4 µl template mix, containing 5 fmol 5'-biotinylated DNA probe, 50 mM HEPES pH 8.0, 12.5 mM magnesium acetate, 12.5 mM DTT, 1.25% NP-40, 250 ng/µl BSA, to 6 µl protein mix, containing the protein of interest in BC-buffer (Section 2.1.5.) containing 100 mM KCl. For EMSA competition experiments, unlabelled competitor DNA was added to the template mix. Binding reactions were incubated for 45 min on ice unless otherwise indicated. Following incubation, 2 µl loading buffer containing 15% (w/v) Ficoll 400 (F2637, Sigma-Aldrich) was added to the DNA-binding reactions and free DNA and protein-DNA complexes were resolved by either agarose gel electrophorese or by native PAGE.

EMSA agarose gel electrophoresis was performed using 1.4% agarose gels (TopVision Agarose, Thermo Scientific), in 0.5 x TBE (45 mM Tris-Cl pH 8.3, 45 mM boric acid, 1 mM EDTA) supplemented with 5 mM magnesium acetate where indicated. The electrophoresis buffer consisted of 0.5 x TBE

buffer supplemented with 2 mM magnesium acetate where indicated. Gel electrophoresis was carried out at 4°C and 100V for 1 h to resolve protein complexes formed with DNA probes 40 – 60 bp in length or for 2 h 45 min when DNA probes >200 bp were used. Native PAGE was also carried out at 4°C, using 5% 37:1 polyacrylamide gels in 0.5 x TBE and 0.5 mM DTT, and 0.5 x TBE running buffer. Gels were pre-run for 1 h 45 min before samples were loaded and protein-DNA complexes and free DNA resolved at 100 V for 1 h 45 min.

After electrophoresis, free DNA and protein-DNA complexes were transferred onto Immobilon nylon membrane (Millipore). For agarose gel EMSAs, this was achieved by vacuum transfer using a Slab Gel Dryer (Hoefer Scientific). Free DNA and protein complexes resolved in PAGE were transferred in 1 x TBE at 20 V for 30 min using a semi-dry MAXI 20 transfer apparatus (Cleaver Scientific). Wet membranes were crosslinked using a Hoefer UVC 500 crosslinker with four 1200 mJ/cm^2 pulses. Free DNA and DNA-protein crosslinker were visualised using a chemiluminescent Nucleic Acid Detection Module (Thermo Scientific) according to the manufacturer's instructions.

Table 1. DNA sequences of 40 bp EMSA dsDNA probes. dsDNA probes were generated by hybridisation of 5'-biotinylated complementary strands and gel purified as described in Section 2.1.10.

Name	Sequence
Ad2ML-TATA (+)	GTTCCTGAAGGGGGGCTATAAAAGGGGGTGGGGGCGCGTT
Ad2ML-TATA (-)	AACGCG CCCCCACCCCCTTTTATAGCCCCCCTTCAGGAAC
Ad2ML-GAGA (+)	GTTCCTGAAGGGGGGCGAGAAAAGGGGGTGGGGGCGCGTT
Ad2ML-GAGA (-)	AACGCGCCCCCACCCCCTTTTCTCGCCCCCCTTCAGGAAC
mTdT-40 (+)	CCAGGGTGGTACCTATGGGTCTGCTGGTGAGAGGACATCA
mTdT-40 (-)	TGATGTCCTCTCACCAGCAGACCCATAGGTACCACCCTGG
mTdT-TATA (+)	CCAGGGTGGTACCTATGTATAAAATGGTGAGAGGACATCA
mTdT-TATA (-)	TGATGTCCTCTCACCATTTTATACATAGGTACCACCCTGG

Table 2. EMSA DNA probes generated by PCR using 5'-biotinylated primers. Nucleotides labelled in blue denote sequences used for PCR amplification. The TSSs identified for the GBP-130 and KAHRP promoters are highlighted in red (Lanzer, de Bruin and Ravetch, 1992a; Lanzer, De Bruin and Ravetch, 1992b).

Promoter	Sequence	Length	A/T-content
GBP-130	AACGCGTTGGGAGCTCTCCGGATCCTTTGAAGTACACTCAAAATAAGTTA TATACCATATGTTTTTAACATATATTATATATATATATATATATAATATA ATATAATATTATAATTATTTTTTAATATTATTAAATTGAACATAATTATTTTA TATCTTACTATTATTTTTAGAAAATTTATTATATATACATGCAATCATAAAT AATGTTTTCCCTGAACCTTTTTTCAATGAAATAAGTTAACACACCATTCCTT TTGAATTCTAGAGTCTCCGCTCG	282 bp	78 %
GA	GAACATAATTATTTTATATCTTACTATTATTTTTAGAAAATTTATTATATATA CATGCAATCATAAATAATGTTTTCCCTGAACCTTTTTTCAATGAAATAAGTT AACACACCATTCCTTTTGAATTCTAGAGTCTCCGCTCG	143 bp	74.8 %
GB	AACGCGTTGGGAGCTCTCCGGATCCTTTGAAGTACACTCAAAATAAGTTA TATACCATATGTTTTTAACATATATTATATATATATATATATAATATA ATATAATATTATAATTATTTTTTAATATTATTAAATTGAACATAATTATTTTA TATCTTAC	163 bp	81.6 %
KAHRP	AACGCGTTGGGAGCTCTCCGGATCCTGATGATATGTTTTATATTTCTTGTT TTAAATTATTAGAATAAAAAGAAATAATTATTATTTCATGGAAATAATATA ATATTATTATTTTTTTTTTTAATTATTTAGTAGTTATGTTTTGTCGTTTTTTCT CATTTATTTATTATAATTTACCTATAGTATATACTATGATGTATATTTACTCT AGTATGAAGAATAAAGTTAATGTAAAATATTACTACACTACATGCAGTTTT AGAATTCTAGAGTCTCCGCTCG	282 bp	77 %
KA	GTCGTTTTTTCTCATTTATTTATTATAATTTACCTATAGTATATACTATGATG TATATTTACTCTAGTATGAAGAATAAAGTTAATGTAAAATATTACTACACT ACATGCAGTTTTAGAATTCTAGAGTCTCCGCTCG	138 bp	73.9 %
KB	AACGCGTTGGGAGCTCTCCGGATCCTGATGATATGTTTTATATTTCTTGTT TTAAATTATTAGAATAAAAAGAAATAATTATTATTTCATGGAAATAATATA ATATTATTATTTTTTTTTTTAATTATTTAGTAGTTATGTTTTGTCGTTTTTTCT C	157 bp	78.3 %
mTdT-WT	AATTGGGCCCGACGTCGCATGCTCCTCTAGAGCTTGCATGCCTGCAGGTC GGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCC TCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAG ACTCTAGAATTCGAGCTCGGTACCTATGGGTCTGCTGGTGAGAGGACATC AGAGCCCTCATTCTGGAGACACCACCTGATGGCACAGACAGAGGATCCG GAGAGCTCCCAACGCGTT	266 bp	40.6 %

mTdT_ TATA	AATTGGGCCCGACGTCGCATGCTCCTCTAGAGCTTGCATGCCTGCAGGTC GGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCC TCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAG ACTCTAGAATTCGGTACCTATGTATAAAATGGTGAGAGGACATCAGAGCC CTCATTCTGGAGACACCACCTGATGGCACAGACAGAGGATCCGGAGAGC TCCCAACGCGTT	260 bp	42.7 %

Table 3. DNA sequences of 50 bp GBP-130 EMSA dsDNA probes dividing GB with 10 bp overlap. dsDNA probes were generated by hybridisation of 5'-biotinylated complementary strands and gel purified as described in Section 2.1.10. The TSS is highlighted in red.

Name	Sequence	A/T-content
GB1 +	ATATAATATAATATTATAATTATTTTTTAATATTATTAAATTGAACATAA	96 %
GB1 -	TTATGTTCAATTTAATAATATTAAAAAATAATTATAATATTATATTATAT	96 %
GB2 +	TATGTTTTTAACATATATTATATATATATATATATATAATATAATATA	96 %
GB2 -	TATATTATATTATATATATATATATATATATAATATATGTTAAAAACATA	96 %
GB3 +	CCGGATCCTTTGAAGTACACTCAAAATAAGTTATATACCATATGTTTTTA	70 %
GB3 -	TAAAAACATATGGTATATAACTTATTTTGAGTGTACTTCAAAGGATCCGG	70 %

2.2. Characterisation of *Pf*TFIIB

2.2.1. Bacterial expression vectors for wild type 6His:*Pf*TFIIB and mutant variants

pET11D expression vectors for the expression of 6His:*Pf*TFIIB, 6His:*Pf*TFIIBΔ37, 6His:*Pf*TFIIBΔ140 and 6His:*Pf*TFIIBΔ140Δ350 were previously generated (Talvik, 2016; Eelu, 2016). The expression vector for 6His:*Pf*TFIIBΔ350, lacking aa 350-367, was generated by insertion of the PCR-amplified open reading frame for *Pf*TFIIB from aa 1-350 into pET11d using NdeI and BamHI sites. The sequence of pET11d-6His:*Pf*TFIIBΔ350 (Appendix A) was verified by sequencing using the Central Analytical Facilities DNA Sequencing Unit at the University of Stellenbosch.

2.2.2. Bacterial expression and purification of 6His:*Pf*TFIIB proteins

Pilot expression experiments of 6His:*Pf*TFIIB Δ140Δ350 were performed to determine expression conditions for *Pf*TFIIB proteins. For this purpose, 200 ml expression cultures in LB containing

100 µg/ml Amp and 55 mM glucose were inoculated with 1/20 volume of starter culture as described in Section 2.1.4. and protein expression was carried out as described in the Results Section 4.1.1. In order to compare the amounts of soluble 6His:*Pf*TFIIB141-350 expressed under different conditions, cells were pelleted by centrifugation (9000 x g, 10 min, 4°C), and cleared lysate was generated as described in Section 2.1.4. 1 ml of cleared lysate obtained for each expression protocol was used to purify soluble 6His:*Pf*TFIIB141-350 using TALON metal affinity resin (Clontech) in a batch procedure. In brief, 50 µl metal affinity resin was equilibrated with sonication buffer and pelleted by centrifugation at 700 x g for 2 min at 4°C in a benchtop microcentrifuge. Cleared lysate (1 ml) was added to the pre-equilibrated resin, mixed, and incubated with agitation for 20 min at room temperature. Unbound material was removed by centrifugation (700 x g, 2 min, 4°C) and the resin was washed twice with 1 ml sonication buffer. Bound material was released by mixing the resin with of 50 µl BC-buffer (Section 2.1.4.) containing 100 mM KCl and 250 mM imidazole, followed by centrifugation (700 x g, 1 min, 4°C) and collection of supernatant. This elution step was repeated twice. To compare the levels of soluble 6His:*Pf*TFIIB141-35, the two eluates obtained for the various expression conditions were compared by SDS-PAGE and Coomassie staining.

For large scale expression of 6His:*Pf*TFIIB proteins, overnight starter cultures were generated as described above (Section 2.1.4). 500 ml LB containing 100 µg/ml Amp and 55 mM glucose were inoculated with 1/20 volume of starter culture and incubated at 37˚C. Upon reaching an OD_{600} of 1 the culture was placed on ice to rapidly reduce the temperature to 25°C. IPTG was added to a final concentration of 1 mM to induce protein expression, cultures were further incubated for 3 h at 30°C and cells harvested by centrifugation (4500 x g, 20 min, 10°C). Typically, a total of six 500 ml cultures (total culture volume 3 l) were used for expression of wildtype 6His:*Pf*TFIIB or 6His:*Pf*TFIIB mutant variants. Following protein expression, cell pellets were treated by osmotic shock, cells where lysed by sonication, and 6His:tagged recombinant proteins purified by metal-affinity chromatography using TALON resin (Clontech) as described in Section 2.1.4. IMAC purified preparations of 6His:*Pf*TFIIB and 6His:*Pf*TFIIB mutants were buffer exchanged into BC-buffer containing 100 mM KCl (BC-100) using PD SpinTrap G-25 Microspin columns (Merck) and passed over Q-Sepharose ion exchange resin (Merck) equilibrated in BC-100 buffer. The flow-through was collected, protein-containing fractions identified by SDS-PAGE, aliquoted, snap-frozen in liquid N_2 and stored at -80°C. Initially, preparations passed through Q-Sepharose were further purified by SP-Sepharose ion exchange chromatography (Talvik, 2017). However, this step did not significantly enhance the purity of preparations (Fig. 33) and reduced overall yield of the recombinant protein of interest.

2.2.3. Analysis of *Pf*TFIIB DNA binding using electrophoretic mobility shift assays (EMSAs)

DNA probes utilised in the characterisation of *Pf*TFIIB-DNA interactions included (i) a 60 bp region of the putative KAHRP promoter (Table 4), spanning positions -15 to -74 relative to the TSS (Lanzer, De Bruin, et al., 1992a), previously shown to be bound by 6His:*Pf*TFIIBΔ140 in EMSA assays (Talvik, 2016), (ii) a 60 bp region of the putative *Pf*EMP1 (PF3D7_1240600) promoter, spanning positions -1181 to -1247 relative to the *Pf*EMP1 gene ATG start codon (Table 4), also shown to bind to 6His:*Pf*TFIIBΔ140 in EMSA assays (Talvik, 2016), (iii) a 60 bp probe containing the Ad2ML promoter (Table 4; see Section 2.1.10) and (iv) a 282 bp GBP-130 promoter probe generated by PCR as described above (Table 2; Section 2.1.10).

EMSAs were performed as described in Section 2.1.10. For experiments involving two protein components or antibodies, proteins were incubated for 30 min on ice before addition of template mix.

Quantification of DNA binding shown in Fig. 37 was performed by density analysis of bands corresponding to free probe in ImageJ (Schneider, Rasband and Eliceiri, 2010). Data plotting and non-linear regression was performed using Graphpad Prism 9.5.1 for Windows (www.graphpad.com).

Table 4. DNA sequences of 60 bp EMSA dsDNA probes. dsDNA probes were generated by hybridisation of complementary 5'-biotinylated oligonucleotides as described in Section 2.1.10.

Promoter	Sequence
Ad2ML-TATA (+)	ACTCTAGGGGTGTTCCTGAAGGGGGGCTATAAAAGGGGGTGGGGGCGCGTTCGTCCTCAC
Ad2ML-TATA (-)	GTGAGGACGAACGCGCCCCCACCCCCTTTTATAGCCCCCCTTCAGGAACACCCCTAGAGT
*Pf*EMP1 (+)	ATATATGTATTTATTTTCATAGAAATGTGGTAGATAATATAGATAGAAAGGTAATATTTC
*Pf*EMP1 (-)	GAAATATTACCTTTCTATCTATATTATCTACCACATTTCTATGAAAATAAATACATATAT
KAHRP (+)	AATAAATGAGAAAAAACGACAAAACATAACTACTAAATAATTAAAAAAAAAAAATAATAAT
KAHRP (-)	ATTATTATTTTTTTTTTTAATTATTTAGTAGTTATGTTTTGTCGTTTTTTCTCATTTATT

2.2.4. Analysis of *Pf*TFIIB interactions using GST pull-down assays.

5 µl glutathione magnetic beads (Pierce) were equilibrated twice with 300 µl immobilisation buffer (125 mM Tris-Cl pH 8.0, 150 mM NaCl, 100 µg/ml BSA, 5 mM DTT, 0.1 % (v/v) NP-40). Between washing steps, beads were separated from buffer using a magnetic stand. 450 ng 6His:GST-*Pf*TFIIB (Talvik, 2016) or molar-equivalent GST control (Talvik, 2016) was incubated with the pre-equilibrated glutathione magnetic beads in 125 µl immobilisation buffer with constant agitation for 2 h at 20 °C. Following immobilisation, 6His:GST beads and 6His:GST-*Pf*TFIIB beads were washed once with 300 µl immobilisation buffer and equilibrated twice with 300 µl binding buffer, comprised of 4 parts binding mix (50 mM HEPES pH 8.0, 12.5 mM magnesium chloride, 12.5 mM DTT, 0.25% NP-40, 250 ng/µl BSA) and 6 parts BC buffer containing 100 mM KCl. 50 nM 6His:*Pf*TLP, 6His:*Pf*TBP or 6His:*Hs*TBP or 25 nM 6His:*Pf*TFIIB were added to 250 nM 6His:GST or 6His:GST*Pf*TFIIB in binding buffer and incubated at 20°C for 4 h. Unbound proteins were removed and immobilised proteins washed twice with 300 µl binding buffer. Following removal of the wash buffer, immobilised GST or GST-fusion protein along with interacting proteins were eluted with 20 µl elution buffer (20 mM Tris-Cl pH 8.0, 0.2 mM EDTA, 60 mM KCl, 10 mM β-mercaptoethanol, 0.1 % NP-40, 50 mM glutathione) for 30 min at 20°C.

2.3. Biomolecular condensation of *Pf*TFIIB

2.3.1. Bioinformatic analysis of *Pf*TFIIB

Disorder prediction was conducted using the meta prediction servers Dismeta (Huang, Acton and Montelione, 2014) available at https://montelionelab.chem.rpi.edu/dismeta/ and MobiDB (Piovesan *et al.*, 2021) available at https://mobidb.bio.unipd.it/.

Net charge per residue was calculated using CIDER (Holehouse et al., 2015), available at http://157.245.85.131:8000/CIDER/

The tertiary structure prediction for *Pf*TFIIB was retrieved from the Alphafold protein structure database (Jumper *et al.*, 2021; Varadi *et al.*, 2022), available at https://alphafold.ebi.ac.uk/, and visualised using UCSF Chimera (Pettersen *et al.*, 2004), available at https://www.cgl.ucsf.edu/chimera/. Figures were generated using the IBS web server (Liu *et al.*, 2015) available at http://ibs.biocuckoo.org/online.php.

2.3.2. Generation of expression vectors for the expression of fluorescent protein fusion constructs

pET11d expression vectors for the expression of 6His:mCherry-*Pf*TFIIB, 6His:*Pf*TFIIB1-140-mCherry, 6His:*Pf*TFIIB38-140-mCherry, 6His:mCherry-*Pf*TFIIBΔ140, 6His:YFP and 6His:mCherry were generated by overlap extension PCR (Heckman and Pease, 2007) using the primer sequences listed in Table S3 and vectors pET11d-6His:*Pf*TFIIB (Section 2.2.1) and pCD3-975-mCherry-vector for amplification of *Pf*TFIIB and mCherry ORFs. pCD3-975-mCherry-vector was kindly provided by Llewelyn van de Pas (University of Cape Town, Department of Molecular and Cell Biology). Open reading frames for the fusion proteins of interest were inserted into pET11d using NdeI and BamHI sites, in frame with the coding sequence for the N-terminal 6His-tag.

The pET11d vector for expression of 6His:YFP-*Pf*TFIIB expression vector was constructed previously in the laboratory (Davies, 2022).

For expression of 6His:mCherry-*Pf*TFIIBΔ140, the pETEC expression vector was used (Sabari *et al.*, 2018), which was kindly provided by Richard Young (Whitehead Institute, MIT, Boston, USA). Similar to pET11d, the pETEC vector allows the expression of N-terminal 6His:tagged recombinant proteins under control of the T7 promoter in *E.coli*. pETEC-6His:mCherry-*Pf*TFIIBΔ140, was generated in collaboration with Abigail Russel (Russel, in preparation). The open reading frame for *Pf*TFIIBΔ140 was generated by PCR and digested with BglII and EcoRI. The digested fragment was then inserted into the pETEC multiple cloning site using BamHI and EcoRI sites.

pET11d expression vectors for expression of 6His:mCherry and 6His:YFP proteins used as control in biomolecular condensation assays were generated by insertion of PCR amplified ORFs for mCherry or YFP into pET11d using NdeI and BamHI sites. All plasmid constructs were verified by sequencing using the Central Analytical Facilities DNA Sequencing Unit at the University of Stellenbosch, South Africa. Detailed plasmid maps can be found in Appendix A.

2.3.3. Expression and purification of fluorescent protein-tagged fusion proteins

Purified 6His:mCherry-*Pf*TBP (Pegram, 2022), 6His:YFP-*Pf*TBP-IDR, 6His:mEGFP-*Pf*RPB1-CTD and 6His:YFP (Karamanof, 2022), and 6His:YFP-*Pf*TLP and 6His:YFP-*Pf*TLP-IDR1 (Michowicz, in preparation) were previously generated in the laboratory.

6His:mCherry-*Pf*TFIIB, 6His:YFP-*Pf*TFIIB and 6His:mCherry-*Pf*TFIIBΔ140 were expressed in *E.coli* BL21(DE3) CodonPlus RIL (Aligent technologies) as described for 6His:*Pf*TFIIB (Section 2.2.2) with minor modifications. LB culture media for expression of 6His:mCherry-*Pf*TFIIBΔ140 was supplemented

with 50 µg/ml kanamycin and 55 mM glucose to select for the pETEC-6His:mCherry-PfTFIIBΔ140 plasmid providing kanamycin resistance. Cells were harvested, subjected to osmotic shock prior to cell lysis and proteins purified by IMAC as described in Section 2.1.4.

6His:*Pf*TFIIB1-140-mCherry, 6His:*Pf*TFIIB38-140-mCherry and 6His:mCherry proteins were expressed at high levels, allowing for a simplified expression protocol. Briefly, expression cultures were inoculated with 1/100 volume of overnight starter culture (Section 2.1.4) and incubated at 37 ^{o}C with shaking. Protein expression was induced at an OD_{600} of 0.5 by the addition of 1 mM IPTG and the cultures were incubated for a further 3 h at 37°C with shaking.

IMAC purified 6His:mCherry-*Pf*TFIIBΔ140 was further concentrated by ultrafiltration, using a Nanosep® Centrifugal Device (Pall) with a 10 kDa cutoff, according to the manufacturer's instructions. The concentrated 6His:mCherry-*Pf*TFIIBΔ140 preparation was centrifugated at 13000 x g for 30 min at 4°C and the supernatant transferred into a fresh Eppendorf tube to remove any protein aggregates before being flash frozen in liquid N_2 and stored at -80°C.

2.3.4. Analysis of biomolecular condensation *in vitro* using widefield epifluorescence microscopy

Purified proteins were thawed on ice, diluted in BC-buffer (Section 2.1.4) supplemented with the desired concentration of KCl and centrifugated at 13000 x g for 10 min at 4°C to pellet any aggregates that may have accumulated. Proteins were mixed by gentle pipetting and combined with an equal volume of the relevant nucleic acid component in TE buffer (Section 2.1.10) or TE buffer control and incubated for 5 min at RT. Condensation was induced by adding an equal volume of condensation buffer (20 mM Tris-Cl pH 7.9 at 4°C, 20% (v/v) PEG 20000 (Sigma-Aldrich), 10 mM magnesium acetate, 1 mM DTT), containing 150 mM KCl unless otherwise indicated, to the protein mix. Total yeast RNA was denatured by heating for 3 min at 90°C, followed by rapid cooling on ice, before use in *in vitro* condensation assays (van Treeck *et al.*, 2018). In reactions containing DNA, a final concentration of 50 µM DAPI was included in the reaction to allow for DNA visualisation by fluorescence microscopy.

For microscopic visualisation of biomolecular condensates, condensation reactions were applied to a custom-made imaging chamber (Alberti *et al.*, 2018), assembled as follows. Glass slides and coverslips were salinised with Sigmacote (Sigma-Aldrich) and thoroughly washed, once with distilled water and once with coverslip wash buffer (20 mM Tris-Cl pH 7.9, 150 mm KCl, 1 mM DTT). Two parallel strips of double-sided adhesive tape were then mounted lengthwise onto the glass slide, roughly at the widths of the coverslip. 3 µl of the condensation reaction were pipetted between the strips of adhesive tapes and covered by a coverslip. With the coverslip fixed

by the adhesive tape, the assembly was inverted and the reaction incubated for 1 h to allow condensates to settle onto the coverslip.

Image acquisition was performed with a Nikon Ti-E inverted widefield fluorescence microscope equipped with a Nikon Intensilight C-HGFIE HG Pre-centered mercury-fiber Illuminator. DAPI (320-380 nm excitation (Ex), 435 - 485 nm emission (Em), Nikon), GFP (460-500nm Ex, 505-560 nm Em, Nikon), YFP (490-510 nm Ex, 520-550 nm, Nikon) or mCherry (540-580 nm Ex, 590-670 nm 24 Em, Nikon) filter sets were used to visualise respective fluorescent proteins. The microscope was fitted with a CFI Apochromat-TIRF 100 X objective lens (Nikon) with 1.49 NA and a temperature-change correction ring, and with a DS-Fi2 camera (Nikon) for acquisition. Image acquisition was achieved using NIS-Elements AR software (Nikon). For each condensation reaction, images of at least ten different field of view were captured using an automated script that moved the microscope stage a fixed distance in one direction. For any specific condensation reaction, shutter speed and exposure time were kept at a constant value during imaging of condensates formed at different experimental conditions unless explicitly stated otherwise.

2.3.5. Quantitative analysis of fluorescent microscope images

Image analysis was performed using a Python custom script developed in the laboratory (Karamanof, 2022) based on Python scripts shared by Richard Young, Whitehead Institute, MIT, Boston, USA (Henninger *et al.*, 2021). The script makes use of Scikit-image (Van Der Walt *et al.*, 2014), SciPy (Virtanen *et al.*, 2020), NumPy (Harris *et al.*, 2020) and Matplotlib (Hunter, 2007) packages for image processing, mathematical operations and generation of graphs and is freely available at https://github.com/leetheflee/droplet-analysis.

To calculate the condensed fraction, defined as the fraction of each field of view that is occupied by condensates, the script first calculates the mean brightness of the field of view. It then identifies areas occupied by condensate droplets and produces a 'droplet mask' to calculate the total area occupied by condensates in a field of view. Condensate droplets are identified as areas larger than 20 pixels, with a pixel intensity greater than 4 standard deviations above the mean brightness of the field of view. Original images and their calculated droplet masks, calculated at different threshold values, were manually compared to determine the threshold values for standard deviation and droplet size that produced a droplet mask closely matching the original image.

These settings were applied for the analysis of all condensate images generated in this study, with the exception of *Pf*TFIIB-RNA condensates shown in Fig. 56. Here, a threshold of pixel intensity greater than 3 standard deviations above the mean brightness of the field of view had to

be used to produce a droplet mask that accurately matched the boundaries of *Pf*TFIIB-RNA condensates. The values of the condensed fraction, calculated as ratio of the total droplet area over the total area in a field of view is displayed as a function of the independent variable for condensate formation. In data figures displaying condensate size as a function of a variable, the area of each condensate identified in the field of view was plotted.

Merged images and plots of pixel intensity across assemblies were generated using ImageJ (Schneider, Rasband and Eliceiri, 2010).

2.3.6. Fluorescence Recovery After Photobleaching (FRAP)

FRAP was performed on entire droplets settled on the coverslip, following 1-2 h incubation of condensation reactions at room temperature, unless otherwise indicated. Image acquisition and bleaching were performed using a Zeiss LSM 880 confocal microscope and a 63 x/1.4 NA oil immersion objective with detector gain set to 700. Zenn software (Zeiss) was used for image acquisition. Bleaching of mCherry fluorescence was carried out at a wavelength of 561 nm and bleaching of YFP at 488 nm, using 100% laser power, until less than 10% of the initial fluorescence signal remained. Fluorescence recovery was recorded immediately after bleaching at a frame rate of 1.002 s / frame. FRAP measurements were typically taken over 200 frames with 10 - 50 frames of pre-bleach measurements taken unless otherwise stated. The imaging area was set to 26.99 μm^2, equivalent to a pixel area of 1000 x 1000, with the exception of Fig. 52, where the imaging area was doubled. ImageJ was used to measure average fluorescence intensities of bleached droplet areas, unbleached droplets and background at specific time points in the same field of view. Recovery curves were calculated using easyFRAP (Koulouras *et al.*, 2018; https://easyfrap.vmnet.upatras.gr/) and double normalisation (Phair, Gorski and Misteli, 2003). Estimated half time of recovery ($t_{1/2}$) and mobile fractions were estimated using the double exponential function of easyFRAP. Average FRAP recovery curves were generated from the data obtained from the analysis of at least three different condensates per condition.

Chapter 3: Results PART 1

Characterisation of *Pf*TBP and *Pf*TLP

Two TBP family proteins have been identified in *Plasmodium falciparum,* TATA box-binding protein (*Pf*TBP) and TBP-like protein (*Pf*TLP; Coulson *et al*., 2004). A few studies have begun to examine the functional properties and role of *Pf*TBP (McAndrew *et al.*, 1993; Ruvalcaba-Salazar *et al.*, 2005; Gopalakrishnan *et al.*, 2009), however these studies offer only a very preliminary functional characterisation of *Pf*TBP and data provided by these studies appears to fall short of providing convincing evidence for some of the conclusions posited (see Discussion Section 6.1). Initial studies performed by our research group investigated some aspects of *Pf*TLP function (Bing, 2015; Milton 2017; van der Linden 2019; Michowicz, in preparation), but more systematic studies to thoroughly compare the biochemical properties of *Pf*TBP and *Pf*TLP have so far not been carried out. This section describes the functional characterisation of *Pf*TBP and *Pf*TLP with focus on their relationship to TBP family proteins, their conservation at the level of primary, secondary and tertiary structure, their expression in *P. falciparum* trophozoites and their DNA-binding characteristics.

3.1. *In silico* analyses of *Pf*TBP and *Pf*TLP

As a point of entry into understanding the functional divergence of these two TBP family proteins, *in silico* analyses were performed to: (i) understand the relationship between *Pf*TBP, *Pf*TLP and other eukaryotic and protozoan TBP proteins, (ii) reveal the conservation/divergence at amino acid residues which in higher eukaryotic TBPs have been established to be involved in DNA and TFIIB interactions, and (iii) determine the predicted structure of *Pf*TLP and *Pf*TBP.

3.1.1. Phylogenetic analysis of *Pf*TBP and *Pf*TLP

To establish the relationship between *Pf*TBP and *Pf*TLP and other TBP family proteins from protozoa, higher eukaryotes and archaea, a panel of TBP family proteins were retrieved from the NCBI database and by BlastP searches (Altschul *et al.*, 1990; McGinnis and Madden, 2004) using *Pf*TBP, *Pf*TBP and *Hs*TBP protein sequences as queries. Hits from all major apicomplexan species and previously characterised TBP proteins from non-apicomplexan protozoa, archaea and other eukaryotes were chosen for analysis. Selected sequences were aligned using the Expresso suite of T-coffee (Armougom *et al.*, 2006) and a maximum likelihood phylogenetic tree was constructed in Mega X (Kumar *et al.*, 2018b, Fig. 10; Materials and Methods, Section 2.1.1.). As shown in Fig. 10, phylogenetic analysis of the 57 TBP family proteins chosen for analysis groups *Pf*TBP and *Pf*TLP into two separate well-supported clades containing putative *Pf*TBP apicomplexan homologs (ApiTBP) or

*Pf*TLP apicomplexan homologs (ApiTLP), respectively. Homologs of *Pf*TBP and *Pf*TLP were identified across all apicomplexans included in this study, with the exception of *Eimeria tenella*, for which only a ApiTBP family protein could be identified.

Neither the ApiTBP nor the ApiTLP clades group robustly with the well-characterised TBPs of higher eukaryotes, or with the previously identified TRF4 TBP-family proteins found in Trypanosomatidae. Interestingly *Eh*TBP, one of the few protozoan TBPs to display sequence-specific binding to the TATA box DNA sequence, groups with TBP proteins of higher eukaryotes. In contrast, the more divergent protozoan Api-TBP, Api-TLP, TRF2, archaeal TBP and TRF4 clades, many of which contain members with divergent DNA-binding properties, all group separately from higher eukaryotic TBPs. Both the Api-TLP clade and the Api-TBP clade encompass highly divergent members, as demonstrated by the long branch lengths, suggesting that the functional properties of these proteins may be species-specific. It was not possible to draw meaningful conclusions regarding deeper evolutionary relationships between the strongly supported Api-TLP, TRF2, ApiTBP, TRF4 or archaeal TBP clades due to the low bootstrap support for deeper branches. Even if these relationships had high bootstrap support, interpretations of phylogenies with mixtures of short and long branches should be approached with caution as these phylogenies are prone to bias by long branch attraction - a phenomenon in which distantly related branches are grouped together based on attraction between long branch lengths (Kapli, Yang and Telford, 2020). It is thus important to emphasise that the designation of Api-TLP and TRF2 groups as sister clades does not imply a shared evolutionary history. Indeed, there is convincing evidence from previous bioinformatics studies for a model in which the gene duplication event that led to the evolution of TRF2 only occurred after the split between bilaterian and non-bilaterian animals (Duttke *et al.*, 2014), suggesting that ApiTLP and TRF2 indeed arose from separate gene duplication events.

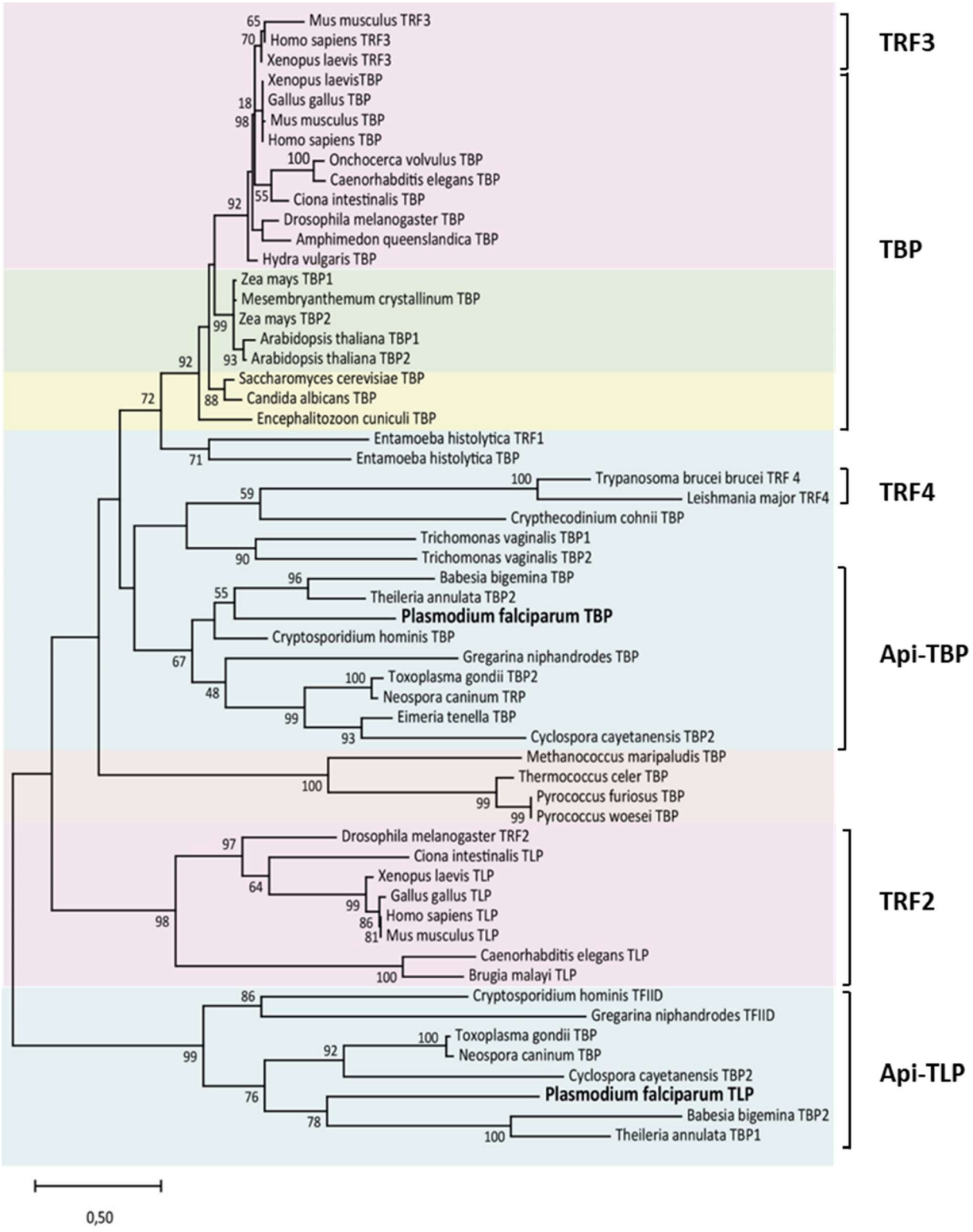


Figure 10. Unrooted phylogram of TBP-type proteins with clades classified as prototypical TBP,-TBP-related factors (TRFs), and TBP-like proteins (TLP). Evolutionary history of various TBP-type proteins was inferred from an alignment of the C-terminal structured core domain of 57 TBP sequences (Supplementary Table 1) using a maximum likelihood method. The tree with the highest log likelihood (-10735,24) is shown. Percentage of bootstrap replicates (out of 1000) that supported each node is indicated. Only bootstrap values above 50 are displayed. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. TBP, TATA-binding protein; TRF, TBP-related factor; TLP, TBP-like protein.

3.1.2. Multiple sequence alignment of *Pf*TBP and *Pf*TLP

In order to gain more detailed insight into the functional divergence of *Pf*TBP and *Pf*TLP from well-characterised TBPs of higher eukaryotes, the Expresso multiple sequence alignment generated in Section 3.1.1. was used to compare *Pf*TBP and *Pf*TLP to prototypical TBP-family proteins of higher eukaryotes with particular emphasis on conservation of amino acid residues involved in TATA-box recognition and TFIIB interactions. The more divergent TRF2 family, shown to have divergent DNA-binding properties (Dantonel *et al*., 1999; Ohbayashi *et al.,* 1999; Rabenstein *et al.,* 1999), was also included in this analysis as a point of reference for TBP family proteins that display atypical DNA interactions (Fig. 11, Section 1.3.4).

TBP family proteins consist of a variable unstructured N-terminal domain and a structured saddle-like C-terminal domain consisting of two quasi-symmetric repeats. While the overall tertiary structure of the C-terminal domain appears conserved across even the more divergent members of the TBP family, the length and amino acid composition of the N-terminal domain varies considerably between species (Nikolov *et al.*, 1996; Lee and Struhl, 2001; Basu *et al.*, 2020). For this reason, analysis of only the conserved structured C-terminal region is shown (Fig. 11).

Importantly, there is a still a degree of uncertainty in the current literature with regard to the start of the *Pf*TBP open reading frame, with some studies supporting the existence of a 26 kDa TBP protein (McAndrew *et al.*, 1993; Ruvalcaba-Salazar *et al.*, 2005), while others support expression of a larger 38 kDa protein product (Gopalakrishnan *et al.*, 2009; Treeck *et al.*, 2011), with a much longer intrinsically disordered N-terminal domain. The 38 kDa *Pf*TBP protein was chosen as reference for sequence alignment and phylogenetic analysis as this is the variant supported by the most recent genome annotation and protein mass spectrometry evidence (Treeck *et al.*, 2011). *Pf*TLP has a much shorter unstructured N-terminal region, composed of only 23 amino acid residues. Instead, *Pf*TLP harbours two insertions, comprising intrinsically disordered regions (IDRs), in the structured C-terminal domain, a longer IDR1 (131 aa residues) inserted between β-strands S3 and S4 and a shorter IDR2 (33 amino acid residues) inserted between the corresponding β-strands S3’ and S4’ in the second repeat (Fig. 11, Bing, 2015, van der Linden, 2019).

*Pf*TBP displays divergence at key amino acids involved in sequence-specific DNA binding while most TFIIB-interacting residues are conserved.

*Pf*TBP displays a high degree of divergence in primary amino acid sequence relative to *Hs*TBP, with the core structured domain showing only approximately 38% amino acid sequence identity to *Hs*TBP (70% similarity as defined by scoring higher than 0.5 on the Gonnet PAM 250 matrix). In contrast, TBP orthologs from higher eukaryotes tend to display a much greater degree of conservation (≥ 70% sequence identity; McAndrew *et al.,* 1993). This high degree of conservation amongst prototypical TATA box proteins from higher eukaryotes can be clearly observed in the conserved blocks of amino acids that span important DNA, TFIIA and TFIIB interaction interfaces (Fig. 11). In contrast, very few of these conserved blocks are maintained in *Pf*TBP and *Pf*TLP or in the more divergent TLP/TRF2 family proteins. Strikingly, many of the amino acids conserved in *Pf*TBP and *Pf*TLP are also conserved in functionally divergent members of the TLP/TRF2 family which display divergent DNA-binding properties.

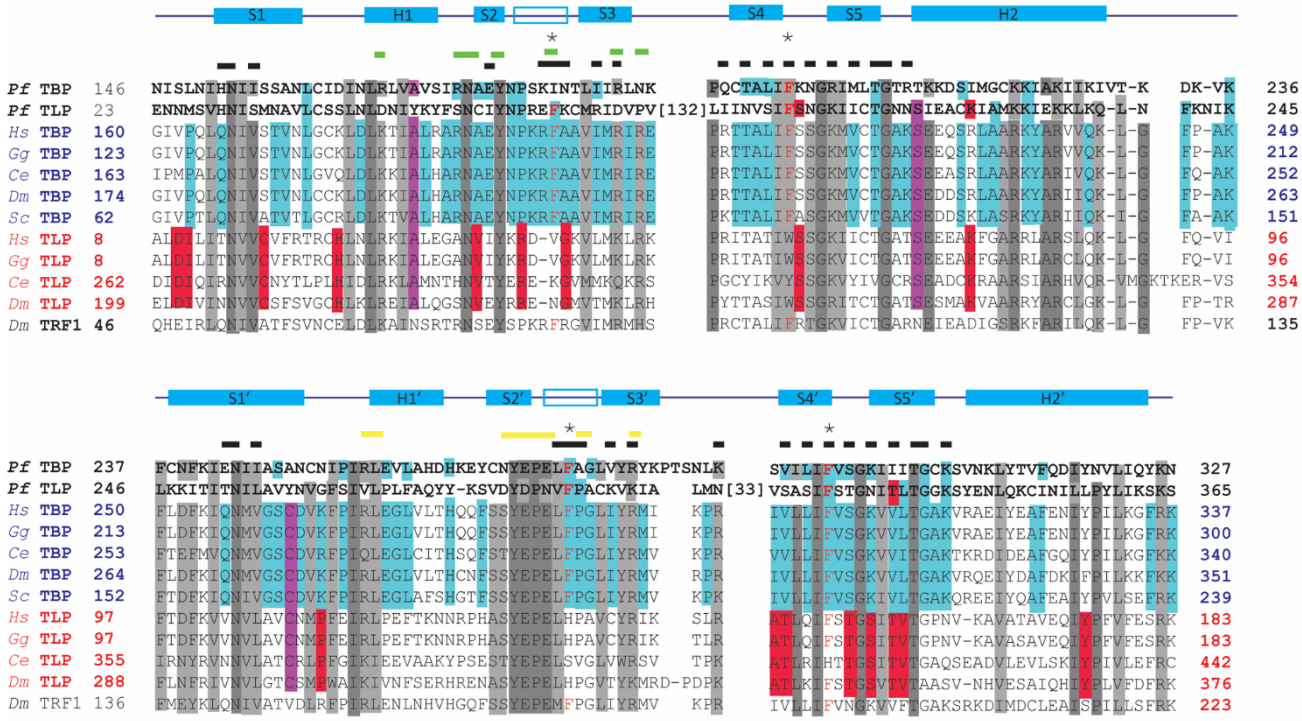


Figure 11. Multiple sequence alignment of *Plasmodium falciparum* TBP family proteins and TBP and TRF2/TLP homologs from well-studied eukaryotes. Amino acid residues highlighted in grey are identical or conserved between all groups (clustalW); residues highlighted in blue are identical across TBPs, residues highlighted in red are identical across TRF2/TLP proteins. Black lines indicate *Hs*TBP residues involved in DNA binding, green lines indicate *Hs*TBP TFIIA-interacting residues and yellow lines indicate *Hs*TBP IIB-interacting residues. Stars indicate the location of the four phenylalanine residues involved in TATA-box recognition. Sequence alignment generated with T-coffee from the following protein sequences: *Plasmodium falciparum (Pf)* TBP (P32086); *Plasmodium falciparum* (*Pf*) TLP (AAA80641.1); *Homo sapiens (Hs)* TBP (AAA36731.1); Gallus gallus (*Gg*) TBP (BAA20298.1); *Caenorhabditis elegans* (*Ce*) TBP (AAA03582.1); *Drosophila melanogaster* (*Dm*) TBP (AAA68629.1); *Saccharomyces cerevisiae* (*Sc*) TBP (NP 011075.3); *Homo sapiens (Hs)* TLP (AAD28785.1); *Gallus gallus* (*Gg*) TLP (BAA75886.1); *Caenorhabditis elegans* (Ce) TLP (CAB03082.2); *Drosophila melanogaster* (Dm) TRF2 (NP 996377.1); *Drosophila melanogaster* (Dm) TRF1 (Q27896.1).

Despite this high divergence in primary structure, some of the residues involved in TATA-box recognition in higher eukaryotes do appear to be conserved in *Pf*TBP. Of the thirty-seven residues shown to interact with the TATA-box in *Hs*TBP (Nikolov *et al.*, 1996), twenty-one residues, including all of those involved in hydrogen bond interactions with nucleotide bases, are identical between *Hs*TBP and *Pf*TBP (Fig. 11 and Table 5).

Perhaps the most notable non-identical substitution observed in *Pf*TBP relative to TBPs from well-studied eukaryotes is the phenylalanine-to-isoleucine substitution observed at one of the four highly conserved phenylalanine residues involved in TATA box recognition and DNA bending (F197 in *Hs*TBP; Fig. 11 and Table 5). In prototypical TBP proteins, this phenylalanine residue has been shown to insert between the seventh and eighth base pair of the TATA box DNA sequence, resulting in severe distortion of the DNA helix (Nikolov *et al.*, 1996). While no studies to date have assessed the impact of substitution at F197 in *Hs*TBP, mutation of any of the other three phenylalanine residues involved in TATA box recognition (F214, F288, F305) has been shown to abrogate binding of *Hs*TBP to the prototypical TATA box sequence (Reddy and Hahn, 1991; Klejman *et al.*, 2005). This includes conservative substitution of phenylalanine by leucine, shown to inhibit *Sc*TBP TATA box binding (Reddy and Hahn, 1991). These observations, as well as the fact that F197 in *Hs*TBP is strictly conserved in TBPs from higher eukaryotes (Fig. 11), but not in TRF2 family proteins and TBP-family proteins from protozoa, which do not display selective binding to a TATA sequence (Dantonel *et al.,* 1999; Ohbayashi, Makino and Tamura, 1999; Rabenstein *et al*., 1999; Guillebault *et al*., 2002; Parra-Marín *et al*., 2019), strongly suggest that the phenylalanine-to-isoleucine substitution at position 183 in *Pf*TBP will significantly alter *Pf*TBP-DNA interactions.

Additional substitutions with clear potential to influence *Pf*TBP-DNA interactions include: (i) substitution of alanine A198 in *Hs*TBP, which neighbours the intercalating phenylalanine and engages in interactions with nearby ribose, with asparagine N184 in *Pf*TBP, (ii) the substitution of serine S216 in *Hs*TBP, involved in H-bonding with the phosphate backbone (Nikolov *et al.*, 1996), with asparagine N202 in *Pf*TBP, (iii) and the substitution of proline P289 in *Hs*TBP, which interacts with the adenine nitrogenous base (1') and ribose backbone (Nikolov *et al.*, 1996), with alanine A276 in *Pf*TBP (Fig. 11 and Table 5). Aside from these non-conservative mutations of key DNA-interacting amino acid residues, *Pf*TBP also harbours an insertion of three amino acids in the "RXXXXXXVLLIF" motif to generate "RXXXXXXXXXVLLIF" (McAndrew *et al.*, 1993). This change in the relative position

of important DNA-interacting residues may also significantly alter DNA interactions and thus sequence specificity of *Pf*TBP relative to prototypical TBP proteins (McAndrew *et al.*, 1993).

Unlike the TBP DNA-interacting amino acid residues, TBP amino acid residues involved in interaction with TFIIB (Nikolov *et al.*, 1995) appear largely conserved (Fig. 11 and Table 6). All three amino acid residues shown by mutational analysis to be vital to TBP-TFIIB complex formation (*Hs*TBP E284, *Hs*TBP E286 and *Hs*TBP L287; T. K. Kim *et al.*, 1994; Tang *et al.*, 1996) and to engage in important van der Waals and H-bond interactions with TFIIB (Nikolov *et al.*, 1995) are conserved in *Pf*TBP. In addition to key interacting amino acid residues identified in TBP mutation studies, *Pf*TBP also displays significant conservation of other residues shown by crystallographic studies to contribute to interactions with TFIIB (Nikolov *et al.*, 1995, Tang *et al.*, 1996).

*Hs*TBP	DNA Interaction	Deleterious	*Pf*TBP	*Pf*TLP
Q^{166}	R		H^{152}	H^{29}
N^{167}	R,B		N^{153}	N^{30}
V^{169}	B	!	I^{155}	S^{32}
E^{191}	W		E^{177}	I^{54}
R^{196}	P,R		K^{182}	E^{59}
F^{197}	R, i		I^{183}	F^{60}
A^{198}	R		N^{184}	K^{61}
I^{201}	R		I^{187}	R^{64}
R^{203}	W, P	!	R^{189}	D^{66}
R^{208}	W	!	Q^{194}	I^{203}
T^{210}	R, P		T^{196}	N^{205}
L^{212}	B	!	L^{198}	S^{207}
F^{214}	B, R	!	F^{200}	F^{209}
S^{216}	R, P		N^{202}	N^{211}
K^{218}	R, W	!	R^{204}	K^{213}
V^{220}	R, B		M^{206}	I^{215}
T^{222}	R, B		T^{208}	T^{217}
G^{223}	R, W		G^{209}	G^{218}
K^{225}	R		R^{211}	N^{220}
Q^{256}	R		E^{243}	T^{252}
N^{257}	R,B		N^{244}	N^{253}
V^{259}	B	!	I^{246}	L^{255}
L^{287}	R	!	L^{274}	V^{282}
F^{288}	R, i	!	F^{275}	F^{283}
P^{289}	R, B		A^{276}	P^{284}
I^{292}	R	!	V^{279}	K^{287}
R^{294}	P	!	R^{281}	K^{289}
R^{299}	W	!	K^{289}	N^{294}
V^{301}	R	!	V^{291}	S^{329}
L^{303}	R, B	!	L^{293}	S^{331}
F^{305}	R, B	!	F^{295}	F^{333}
S^{307}	P		S^{297}	T^{335}
K^{309}	W, R	!	K^{299}	N^{337}
V^{311}	R, B		I^{301}	T^{339}
T^{313}	R, B		T^{303}	T^{341}
G^{314}	W		G^{304}	G^{342}
K^{316}	P	!	K^{306}	K^{344}

Table 5. Conservation of key DNA-binding residues in *Pf*TBP and *Pf*TLP with reference to *Hs*TBP. DNA Interactions by *Hs*TBP indicated as follows: R = interacts with ribose; B = interacts with nucleotide base; P = interacts with phosphate; W = water mediated contact; i = intercalating; ! denotes residues which decrease the affinity of TBP for DNA when mutated (Y. Kim *et al.*, 1993; Nikolov *et al.*, 1996; Patikoglou *et al.*, 1999). Dark blue indicates identity, light blue indicates conservation between groups of strongly similar properties as determined by ClustalW, yellow indicates conservation between groups of weakly similar properties as determined by ClustalW, orange indicates extreme substitution.

Table 6. Conservation of TFIIB-interacting amino acid residues with reference to *Hs*TBP. TFIIB interactions by *Hs*TBP are indicated as follows: VdW= van der Waals interaction; H = hydrogen-bond; sb = salt bridge; ! denotes residues shown to be required for TBP-TFIIB interaction (T. K. Kim *et al*., 1994, Nikolov *et al.*, 1995, Tang *et al*., 1996). Dark blue indicates identity, light blue indicates conservation between groups of strongly similar properties as determined by ClustalW, yellow indicates conservation between groups of weakly similar properties as determined by ClustalW, orange indicates extreme substitution.

***Hs*TBP**	TFIIB Interaction	Deleterious	*Pf*TBP	*Pf*TLP
E^{271}	VdW		E^{258}	P^{267}
Y^{283}	VdW		Y^{270}	Y^{278}
E^{284}	H, VdW	!	E^{271}	D^{279}
P^{285}	VdW		P^{272}	P^{280}
E^{286}	H, VdW	!	E^{273}	N^{281}
L^{287}	VdW	!	L^{274}	V^{282}
K^{337}	sb		N^{327}	S^{365}

*Pf*TLP displays a high degree of divergence at key residues involved in sequence-specific DNA interactions and moderate conservation of amino acid residues involved in interaction with TFIIB

*Pf*TLP displays a greater degree of sequence divergence from prototypical TBPs compared to *Pf*TBP, with only 25% of amino acid residues of the *Pf*TLP structured core domain (excluding the two disordered insertions) identical to *Hs*TBP (48% similarity as defined by scoring higher than 0.5 on the Gonnet PAM 250 matrix). In line with this low degree of overall conservation, *Pf*TLP displays divergence at many amino acid positions that are key to sequence-specific interactions of TBP with the TATA box (Fig. 11 and Table 5). Only 13 of the 37 amino acids shown to engage in direct interactions with the TATA box DNA sequence (Nikolov *et al.*, 1996) are identical between *Pf*TLP and *Hs*TBP, eight residues have been substituted by amino acids with strongly similar properties, five have been substituted with amino acids with vaguely similar properties (< 0.5 and > 0 in the Gonnet PAM 250 matrix) and eleven represent extreme substitutions (scoring <0 on the Gonnet PAM 250 matrix; Table 5).

Extreme substitutions observed in *Pf*TLP include the substitution of V169, L212 and L303 in *Hs*TBP by serine in *Pf*TLP (S32, S207, S331; Fig. 11, Table 5). In *Hs*TBP, these residues have been shown to interact through van der Waals interactions with the nucleotide bases of the TATA box and mutation any one of these residues has been shown to significantly diminish TBP TATA box-binding activity

(Kim *et al.*, 1993). Thus, the substitution of these hydrophobic residues for serine, a polar amino acid, are expected to significantly alter *Pf*TLP-DNA interactions.

*Pf*TLP also displays considerable divergence at amino acid residues which, in *Hs*TBP, are involved in interactions with the ribose groups of the DNA backbone. These include substitution of alanine A198 in *Hs*TBP with lysine K61 in *Pf*TLP, substitution of isoleucine I201 in *Hs*TBP with arginine R64 in *Pf*TLP, substitution of glutamine Q256 in *Hs*TBP with threonine T252 in *Pf*TLP, substitution of isoleucine I292 in *Hs*TBP with lysine K287 in *Pf*TLP, substitution of valine V301 in *Hs*TBP with serine S329 in *Pf*TLP, and substitution of leucine L303 in *Hs*TBP with serine S331 in *Pf*TLP. Additional noteworthy extreme substitutions shown to decrease *Hs*TBP sequence-specific DNA-binding activity include substitution of arginine R203 in *Hs*TBP, which engages in H-bond interactions with the phosphate backbone, with aspartic acid D66 in *Pf*TLP and mutation of multiple residues involved in water-mediated interactions with the DNA backbone.

Finally, *Pf*TLP contains two intrinsically disordered regions (IDRs) inserted between β-strands S3 and S4, and S3' and S4' (Fig. 11), which may affect DNA interactions of *Pf*TLP by altering the relative position of DNA-interacting amino acids in the "RXXXXXXTALIF" and "RXXXXXXVLLIF" motifs (as hypothesised for *Pf*TBP; McAndrew *et al*., 1993), by interacting with DNA and stabilising *Pf*TLP-DNA interactions and/or assisting *Pf*TLP facilitated diffusion on DNA, by interfering with *Pf*TLP-DNA interactions through steric hinderance, or by engaging in inhibitory protein-protein interactions. Interestingly, despite the overall high degree of divergence at key DNA-interacting amino acid residues and the presence of IDR insertions, all four phenylalanine residues that facilitate the extreme distortion of the TATA-box upon TBP DNA binding are conserved in *Pf*TLP. These phenylalanine residues are generally mutated in divergent TBP family proteins such as members of the TRF2/TLP clade or TRF4 TBP proteins, which have lost the ability to recognise the TATA box DNA sequence motif (Fig. 11, Dantonel *et al.*, 1999; Ohbayashi, Makino and T. Tamura, 1999; Rabenstein *et al.*, 1999; Bischoff and Vaquero, 2010; Parra-Marín *et al.*, 2019).

Similar to *Pf*TBP, the *Pf*TLP-TFIIB interaction interface is slightly more conserved with 4 identical, 3 conserved and 2 non-conserved residues relative to *Hs*TBP (Fig. 11, Table 6). Importantly, all three amino acid residues shown to be required for TFIIB interaction (Kim *et al.*, 1994; Tang *et al.*, 1996; see above) have been replaced with residues with similar properties (*Hs*TBP E284 - *Pf*TLP D279, *Hs*TBP E286 – *Pf*TLP N281, *Hs*TBP L287 – *Pf*TLP V282)

3.1.3. Tertiary structure prediction of *Pf*TBP and *Pf*TLP

Given the high degree of divergence of both *Pf*TBP and *Pf*TLP from prototypical TBPs at the primary structure level, it was important to assess whether these proteins adopt the saddle-like tertiary structure characteristic for the DNA-binding domain of members of the TBP family. To this end, tertiary structure predictions for *Pf*TLP and *Pf*TBP were obtained from the AlphaFold protein structure database (Jumper *et al.*, 2021; Varadi *et al.*, 2022). The quality of these models was independently verified using PROCHECK (Laskowski *et al.*, 1993) and Verify-3D (Luthy, Bowie and Eisenberg, 1992) (Table S4). As shown in Fig. 12, *Pf*TLP and *Pf*TBP are predicted to adopt the characteristic saddle-shaped structure of the C-terminal TATA-binding domain of prototypical TATA-binding proteins (Thomas and Chiang, 2006).

Importantly, the intrinsically disordered N-terminal regions of *Pf*TBP and *Pf*TLP as well as the intrinsically disordered regions, inserted between beta strands S3 and S4, and S3' and S4' of *Pf*TLP, could not be modelled with high confidence (Fig. S1). This is due to the inherent difficulties in predicting the conformations of disordered regions which have relatively flat free energy landscapes (Turoverov, Kuznetsova and Uversky, 2010; Fisher and Stultz, 2011; Uversky, 2013). Instead of conforming to a single secondary to tertiary structure, these disordered regions may be best described as highly dynamic regions that can adopt an ensemble of distinct structures (Fisher and Stultz, 2011).

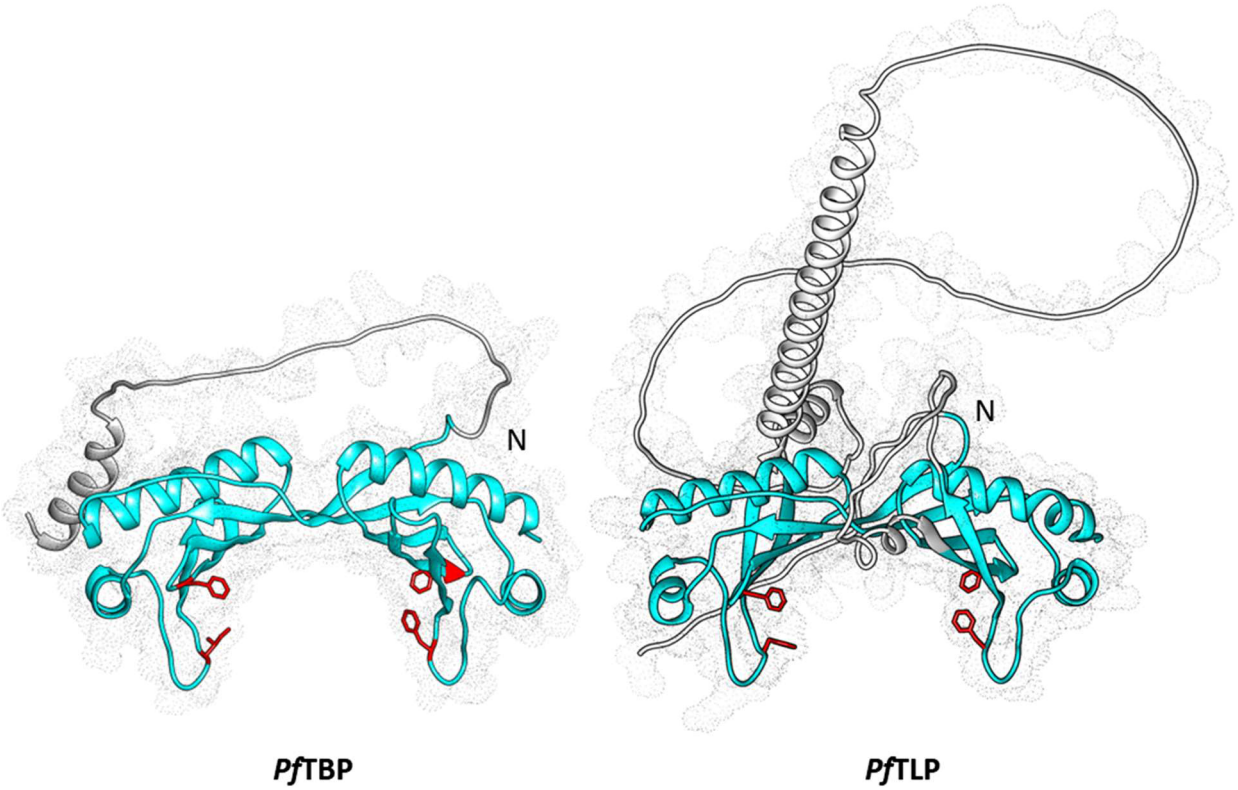


Figure 12. AlphaFold predicted structures of *Pf*TBP and *Pf*TLP. Cartoon representations of the predicted *Pf*TBP and *Pf*TLP structures are shown embedded within surface rendering. The position of phenylalanine residues, or substituted residues, corresponding to those important for TATA-specific DNA binding of prototypical TBP family proteins are highlighted in red, the predicted C-terminal TBP DNA-binding domain is shown in cyan, and regions predicted to be intrinsically disordered are shown in grey. Note that the predicted structure of 26 kDa *Pf*TBP (see Section 3.1.2) is shown, lacking most of the intrinsically disordered N-terminal region, in order to visualise the predicted DNA-binding domain structure unobscured.

3.2. Expression and purification of recombinant 6His:*Pf*TBP

As discussed in section 3.1.2, there have been conflicting reports regarding the start of the *Pf*TBP open reading frame in the existing literature, with some studies supporting the expression of a 26 kDa *Pf*TBP protein (McAndrew *et al.*, 1993; Ruvalcaba-Salazar *et al.*, 2005), while others support the expression of a 38 kDa protein product with a much longer intrinsically disordered N-terminal region (Gopalakrishnan *et al.*, 2009a; Treeck *et al.*, 2011). For this reason, both *Pf*TBP variants were expressed and purified.

Conditions for the bacterial expression of 6His:*Pf*TBP proteins by leaky overnight expression were previously established in the laboratory (Milton, 2017, Materials and Methods, Section 2.1.4.). For purification of 6His:*Pf*TBP proteins from *E.coli* cells, an osmotic shock protocol was adapted, which had been shown to increase the yield of recombinant proteins expressed at low levels in *E.coli* purified by immobilised metal affinity chromatography (IMAC) up to ten fold (Magnusdottir *et al.*, 2009). The osmotic shock method allows the removal of problematic periplasmic material through sequential incubation in sucrose and low salt buffer before lysis of *E.coli* cells and protein purification, (Materials and Methods Section 2.1.4.). Indeed, incorporation of this protocol increased the yield of 6His:*Pf*TBP from about 5 µg/l to about 50 µg/l culture.

Following osmotic shock, the cell pellet was sonicated, cell debris removed by centrifugation and the resulting cleared lysate purified by IMAC (Materials & Methods, Section 2.1.4., Fig. 13A). SDS-PAGE analysis of IMAC purified 6His:*Pf*TBP revealed dominant bands for the 6His-tagged *Pf*TBP variants with a calculated molecular mass of 26 kDa ("6His:*Pf*TBPcore", Fig. 13) and 38 kDa ("6His:*Pf*TBP FL", Fig. 13) migrating at an apparent molecular mass of 26 kDa and 50 kDa respectively. An approximately 10 kDa larger apparent molecular mass based on SDS-PAGE mobility (~ 62 kDa) compared to calculated molecular mass (47 kDa) was also observed for recombinant 6His:YFP:*Pf*TBP-IDR fusion protein, previously generated by fellow research group member Leo Karamanof, which contains only the unstructured N-terminal region of the predicted full-length 38 kDa *Pf*TBP (Fig. S2). The additional N-terminal region present in the 38 kDa protein is highly acidic, with a calculated pI of 4.37. Thus, the reduced mobility of the 38 kDa 6His:*Pf*TBP-FL protein in SDS-PAGE can be explained by reduced binding of SDS to the highly acidic N-terminal region that is present in the 38 kDa 6His:*Pf*TBP-FL protein but absent in the shorter 26 kDa variant .

SDS-PAGE analysis of the 6His:*Pf*TBP preparations further revealed the presence of additional high molecular weight bands that migrate at a position corresponding to approximately twice the mass of the 6His:*Pf*TBP bands (Fig 13 B, lane 1 and 2). However, protein bands of twice the expected

apparent molecular mass were also seen in the control lane containing purified BSA (Fig. 13B, lane 3). Thus these appear to be gel artifacts rather than high molecular mass contaminant *E.coli* proteins.

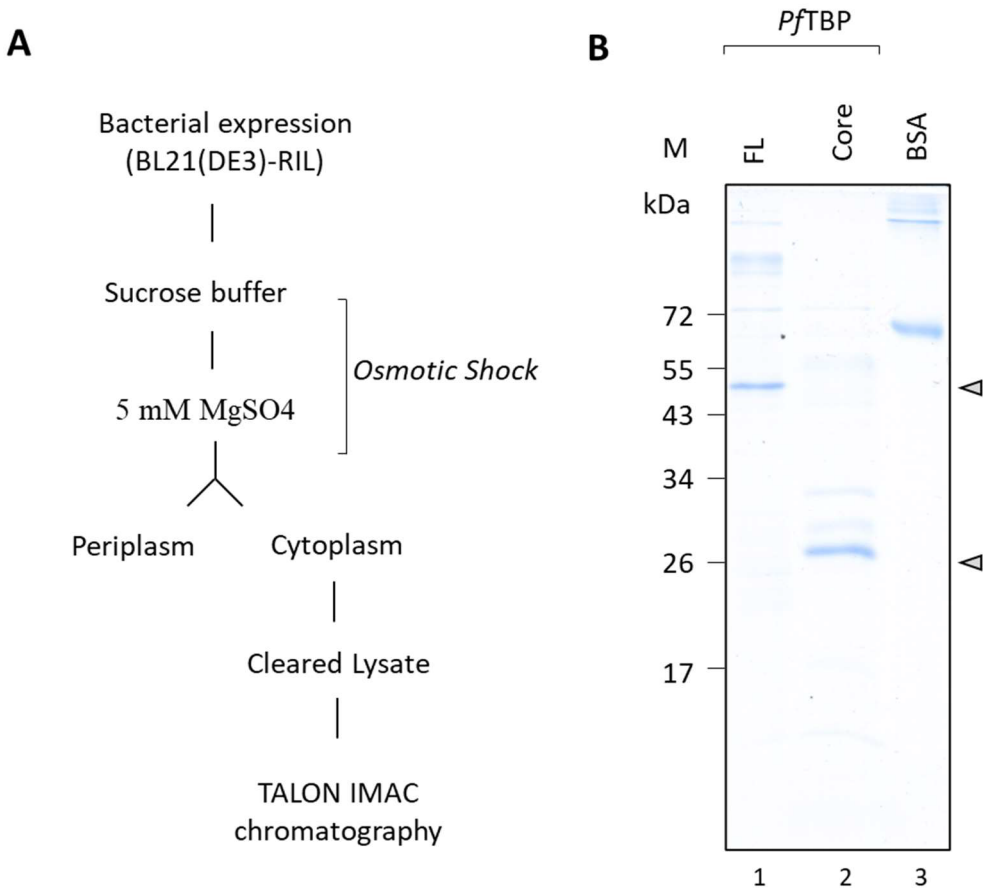


Figure 13. Purification of 6His:*Pf*TBP. (A) Schematic illustration of 6His:*Pf*TBP purification. **(B**) Analysis of purified 6His:*Pf*TBP-FL (38 kDa) and 6His:*Pf*TBP-Core (26 kDa) proteins by SDS-PAGE. 400 ng total protein were resolved by 12% SDS-PAGE and visualised by Coomassie staining.

3.3. Purification and characterisation of rabbit polyclonal antibodies for *Pf*TBP and *Pf*TLP

Polyclonal peptide antibodies for *Pf*TBP and *Pf*TLP were generated by *BioGenes* (Berlin, Germany) and further purified by antigen affinity chromatography and ammonium sulphate precipitation (Materials and Methods, Section 2.1.5).

*Pf*TBP antibodies were raised against the peptide sequence SEYDNNEKEKSDDL located in the N-terminal disordered region of *Pf*TBP and present in both the predicted 38 kDa and 26 kDa *Pf*TBP proteins. Affinity purified anti-*Pf*TBP antibody detected recombinantly expressed 6His:*Pf*TBP-Core (26 kDa) and 6His:*Pf*TBP-FL (38 kDa) in immunoblot assays (Fig. 14A, lanes 3-4). In contrast, no reactivity was observed for the pre-immune control (Fig. 14A, lanes 1-2).

To test whether the anti-*Pf*TBP antibody detects expression of *Pf*TBP in *P. falciparum* cell extracts, immunoblot analysis of whole trophozoite lysates obtained from cultured blood-stage parasites (Materials and Methods, Section 2.1.8) was carried out. A prominent 50 kDa band was detected, consistent with the apparent molecular mass of recombinant 38 kDa 6His:*Pf*TBP seen in SDS-PAGE (Fig. 14B, lane 4-6). In addition, the antibody detected a less prominent band with an apparent molecular mass of 56 kDa. This lower mobility form may represent a phosphorylated or otherwise modified form of *Pf*TBP. A phosphorylated form of *Hs*TBP with similar reduced mobility can be detected in nuclear extracts of cultured human cells (Biggs, Ahn and Kraft, 1998). Neither of these bands were detected when using the pre-immune serum as a control (Fig. 14B, lane 1-3). A 26 kDa *Pf*TBP variant could not be detected in late trophozoite whole cell extracts, suggesting that this potential isoform was either not expressed in the cultured parasites or expressed at undetectable levels.

To further investigate the subcellular localisation of *Pf*TBP, trophozoites were fractionated (Materials and Methods, Section 2.1.8) and the resulting nuclear and cytoplasmic extracts were analysed by immunoblotting. A prominent 50 kDa band was detected only in the nuclear extract (Fig. 14C, compare lane 3 to lane 4), suggesting nuclear localisation of *Pf*TBP. In addition, a much less prominent band corresponding to an apparent molecular mass of 45 kDa was also detected by the anti-*Pf*TBP antibody. This band, also enriched in the nuclear extract, may indicate some breakdown of *Pf*TBP that occurred during extract preparation.

*Pf*TLP antibodies were raised against the peptide sequence NDENKSNDNKEQND located in the intrinsically disordered region IDR2, located between S3' and S4' of the *Pf*TLP core domain (aa 294-307). Affinity purified anti-*Pf*TLP antibody, but not the pre-immune control, detected purified

recombinant 6His:*Pf*TLP protein (Fig. 15A, compare lanes 3 and 4 with lanes 1 and 2). Furthermore, purified anti-*Pf*TLP antibody, but not the pre-immune control, showed strong reactivity against a protein with an apparent molecular mass of approximately 44 kDa in *P. falciparum* whole trophozoite extracts, consistent with the calculated molecular mass of *Pf*TLP of 42 kDa (Fig. 15B, lanes 3 and 4). These results validate the anti-*Pf*TLP antibody and provide strong evidence for the expression of *Pf*TLP in *P. falciparum* trophozoites. Unfortunately, immunoblot analysis of nuclear and cytoplasmic trophozoite extracts with the *Pf*TLP antibody resulted in very high background reactivity, which complicate interpretation of the results (Fig. 15C). However, the data suggest the presence of *Pf*TLP in both nuclear and cytoplasmic extracts.

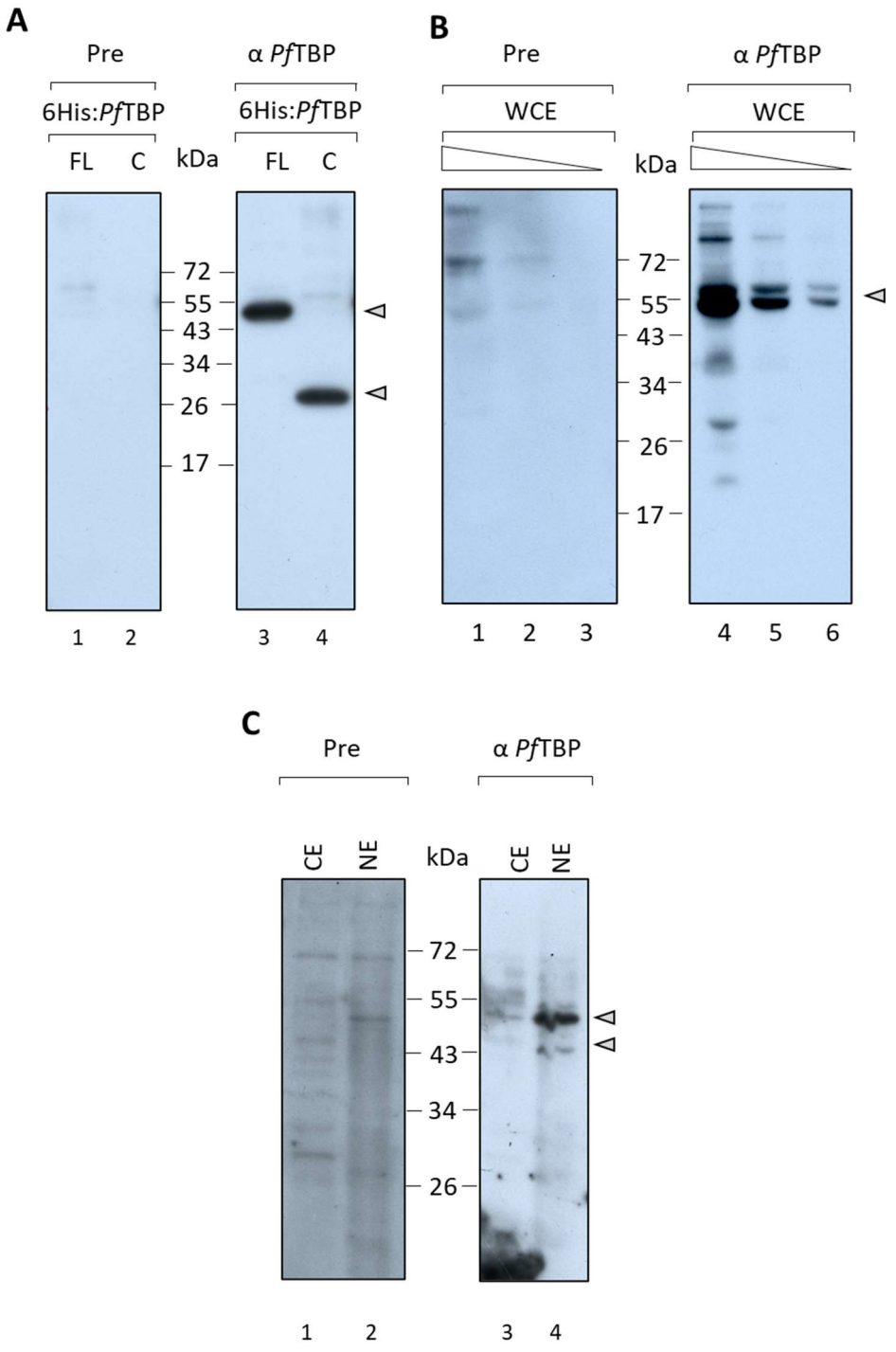


Figure 14. Immunoblot analysis of purified 6His:*Pf*TBP and *P. falciparum* extracts using an anti-*Pf*TBP antibody. (A) 500 fmol purified recombinant 26 kDa 6His:*Pf*TBP (C) or 38 kDa 6His:*Pf*TBP (FL), **(B)** *P. falciparum* trophozoite whole cell extract (WCE) from 1 x 10^6, 5 x 10^5 or 2.5 x 10^5 *P. falciparum* trophozoites, or **(C)** 8 µg total protein *P. falciparum* trophozoite nuclear (NE) or cytoplasmic (CE) extract were separated by 12% SDS-PAGE, transferred to PVDF membrane, probed with affinity purified anti-*Pf*TBP antibody or pre-immune serum (Pre) as indicated, and visualised by chemiluminescence on X-ray film. Immunoreactive bands corresponding to the protein of interest are indicated.

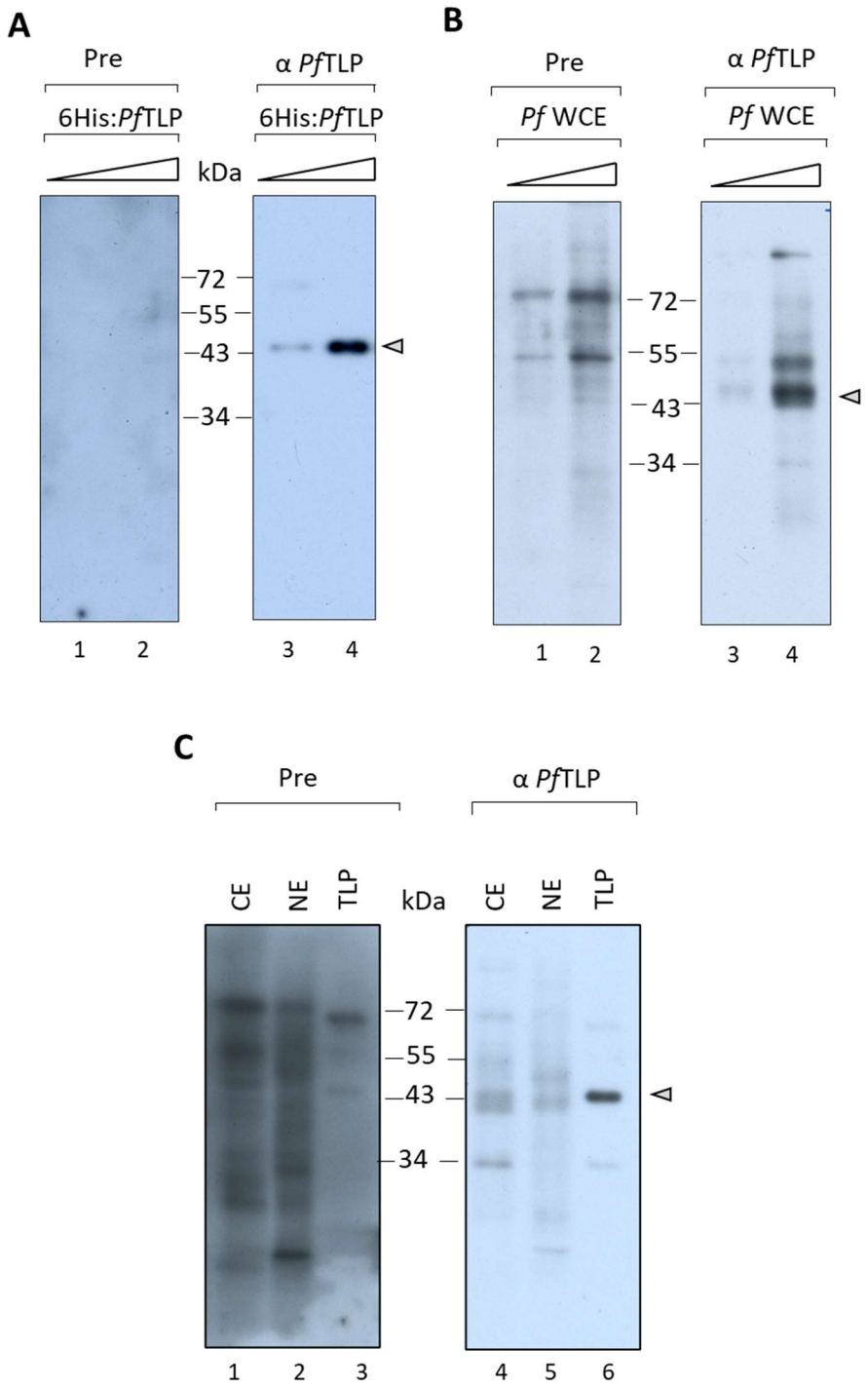


Figure 15. Immunoblot analysis of purified 6His:*Pf*TLP and *P. falciparum* extracts using an anti-*Pf*TLP antibody. (A) 125-250 fmol purified recombinant 6His:*Pf*TLP, **(B)** *P. falciparum* trophozoite whole cell extract (WCE) from 1 x 10^6 or 4 x 10^6 *P. falciparum* trophozoites, or **(C)** 8 µg total protein *P. falciparum* trophozoite nuclear (NE) or cytoplasmic (CE) extract or 110 fmol 6His:*Pf*TLP were separated by 12% SDS-PAGE, transferred to PVDF membrane, probed with affinity purified anti-*Pf*TLP antibody or pre-immune serum (Pre) as indicated, and visualised by chemiluminescence on X-ray film. Immunoreactive bands corresponding to the protein of interest are indicated.

3.4. *Pf*TBP and *Pf*TLP are localised in discrete puncta in *P. falciparum* trophozoites and schizonts

The intracellular localisation of *Pf*TBP and *Pf*TLP in the *P. falciparum* trophozoite and schizont stages was further investigated by indirect immunofluorescence assays (IFA) using the affinity purified anti-*Pf*TBP and anti-*Pf*TLP antibodies and DAPI staining as a marker for the nucleus (Materials and Methods, Section 2.1.7.). In these assays, fixation of samples and antibody staining of the samples was performed in solution rather than on the microscope slide in order to better preserve the nuclear architecture (Moll *et al.*, 2013).

Both the anti-*Pf*TBP and the anti-*Pf*TLP antibodies produced strong signals in trophozoites and schizonts not seen with the AlexaFlour 488 anti-rabbit secondary antibody-only control (Figs. 16 and 17), supporting the conclusion from the immunoblot analysis of parasite extracts (Figs. 14 and 15) that *Pf*TBP and *Pf*TLP are expressed during the parasite blood stages. The antibody staining of both proteins was not uniform but localised to discrete puncta, suggesting that both *Pf*TBP and *Pf*TLP are concentrated in specific cellular compartments during the *P. falciparum* blood stages. A similar punctate localisation of TBP has been reported in cultured murine cells and has been suggested to reflect the partitioning of TBP into transcription condensates (Basu *et al.*, 2020; Shao *et al.*, 2022).

*Pf*TLP puncta did not colocalise exclusively with the DAPI stain but did appear to be excluded from the food vacuole, visible under bright field as a dark pigmented region (Fig. 16). Assessment of nuclear versus cytoplasmic localisation by microscopy is complicated by the complex nuclear architecture of *P. falciparum* parasites, which consists of a central nuclear region stained by DAPI, surrounded by a perinuclear zone that is not stained by DAPI (Volz *et al.*, 2010). The perinuclear zone appears to play an important role in the regulation of some genes, in particular members of the *var* gene family, which localise either to repressive centres or to a small active zone in the nuclear periphery (Volz *et al.*, 2010). Due to lack of an appropriate marker for the perinuclear zone, it was not possible to conclusively determine if the *Pf*TLP foci, which did not colocalise with the DAPI stain, localise to the perinuclear zone or to the cytoplasm during these stages in the parasite life cycle.

The *Pf*TBP puncta also did not overlap strictly with the DAPI nuclear stain, suggesting that *Pf*TBP may also localise in part to the nuclear periphery or to the cytoplasm (Fig. 17). In addition, the anti-*Pf*TBP antibody recognised a component of a spherical structure surrounding some trophozoites (Fig. 17). This likely reflects a reactivity of the anti-*Pf*TBP antibody to a component of the ruptured infected red blood cell membrane, which can remain attached to a proportion of parasites following saponin lysis (Quadt *et al.*, 2020).

Finally, it should be noted that in the fluorescence microscopy assays reported here, only a fraction of parasites examined was immunostained with the anti-*Pf*TBP and anti-*Pf*TLP antibodies. This was likely due to variable permeabilization of isolated parasites prior to antibody treatment, potentially as a result of the absence of a separate permeabilization step but may alternatively reflect heterogeneity in the expression of these proteins amongst the parasite population. Further experiments that make use of different fixation and permeabilization protocols are needed to clarify this issue.

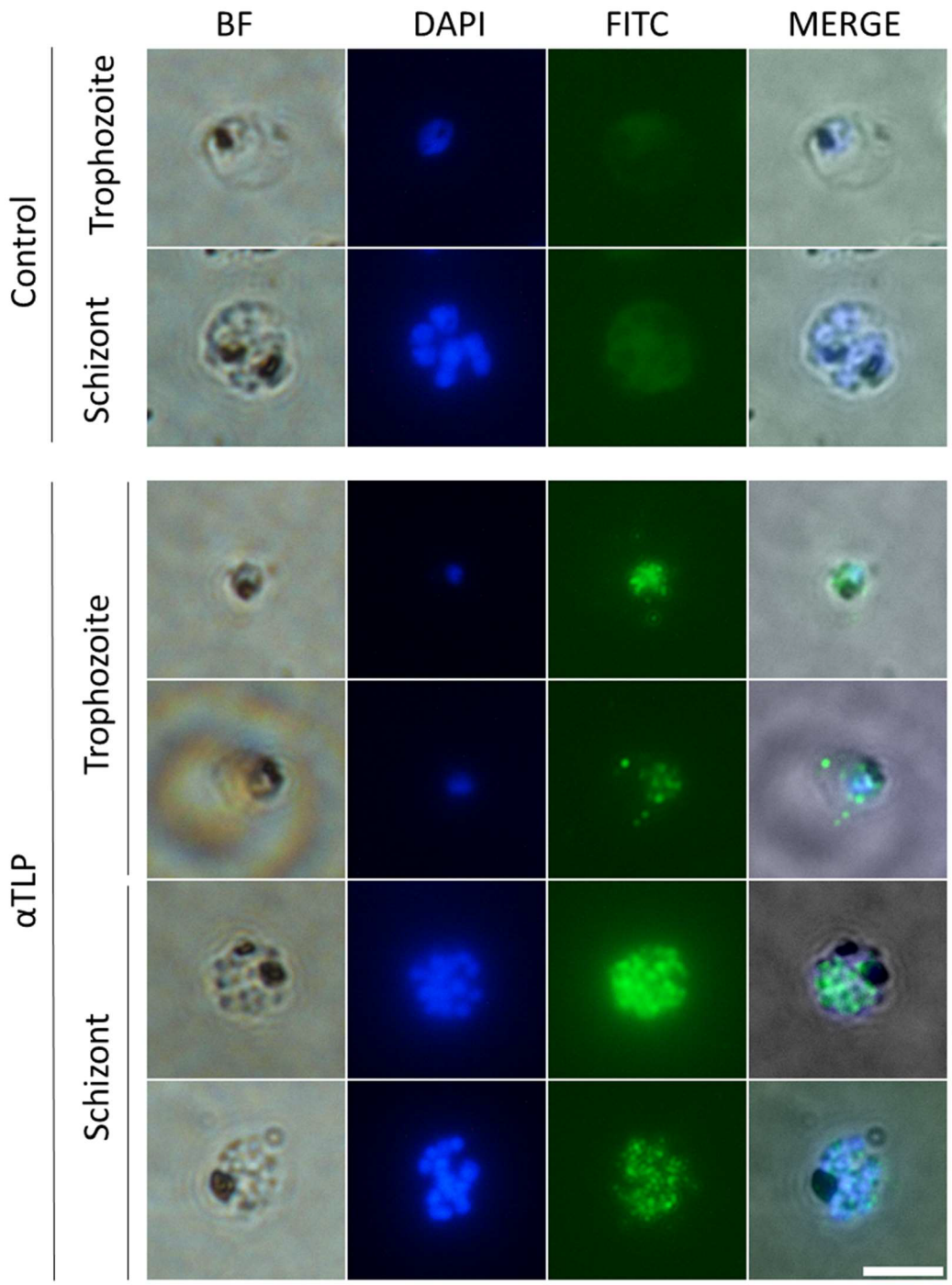


Figure 16. *Pf*TLP antibodies recognise discrete puncta in *Plasmodium falciparum* trophozoites and schizonts. Indirect immunofluorescence assay using affinity-purified rabbit anti-*Pf*TLP primary antibody. Infected red blood cells from a synchronised culture were saponin lysed and fixed in 4% formaldehyde before incubation with anti-*Pf*TLP antibody. Following two wash steps, parasites were incubated with anti-rabbit AlexaFlour 488 secondary antibody and counterstained with 4,6-diamidino-2-phenylindole (DAPI). As a control, parasites were stained with DAPI and secondary antibody only. Parasites were visualised by brightfield (BF) and the relevant channels for detection of DAPI and AlexaFlour 488 fluorescence. Scale bar: 5µM.

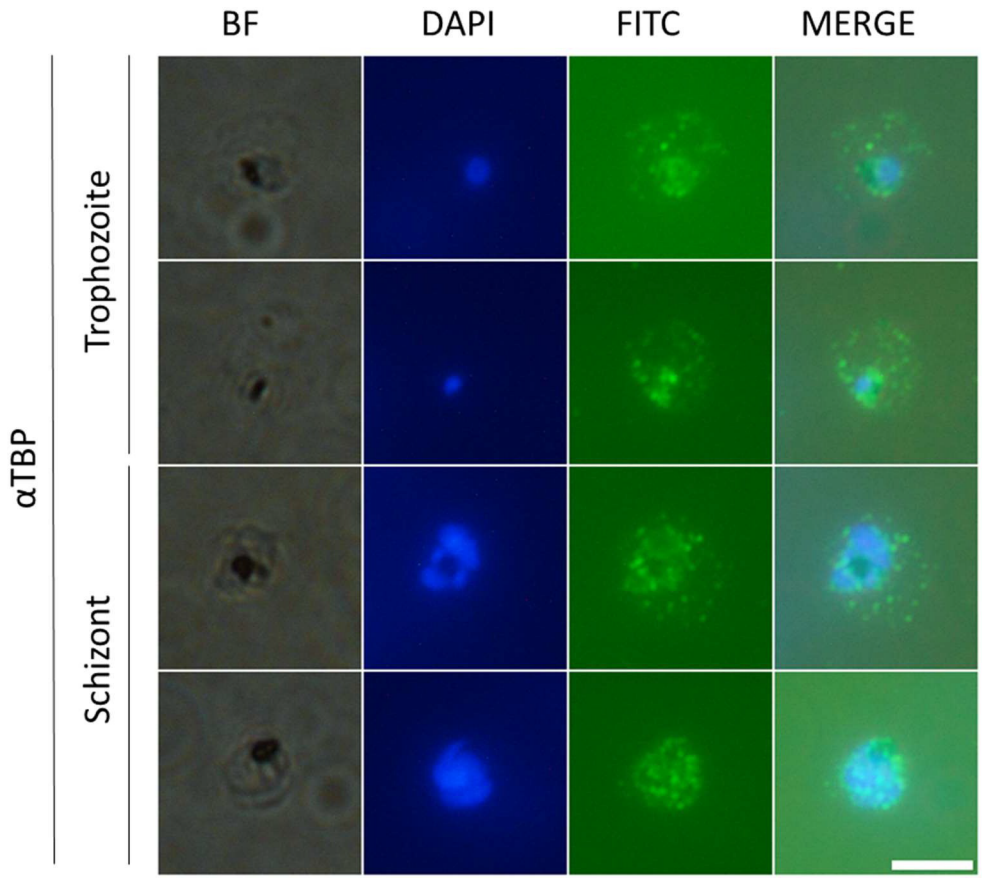


Figure 17. *PfTB*P antibodies recognise discrete puncta in *Plasmodium falciparum* trophozoites and schizonts. Indirect immunofluorescence assay using rabbit anti-*Pf*TBP primary and anti-rabbit AlexaFlour 488 conjugated secondary antibody. Samples were processed as described in the legend to Fig. 16 above. Scale bar: 5µM.

3.5. Characterisation of *Pf*TBP and *Pf*TLP DNA-binding activity

3.5.1. The intrinsically disordered *Pf*TBP N-terminus inhibits *Pf*TBP-DNA interactions

While the role of the conserved C-terminal saddle-like DNA-binding domain of TBPs has been well characterised, the function of the variable N-terminal intrinsically disordered region (IDR) has remained somewhat enigmatic. In *Saccharomyces cerevisiae*, the N-terminal IDR of *Sc*TBP has been shown to increase dissociation of *Sc*TBP-TATA box complexes during electrophoresis and reduce the association rate of *Sc*TBP-DNA complex formation (Lieberman *et al.*, 1991; Kuddus and Schmidt, 1993). These observations were further supported by later footprinting studies and solution studies utilising fluorescent dyes, which showed that the N-terminal region of *Sc*TBP interacts with the TBP core domain DNA-binding interface and impacts TBP DNA-binding affinity, DNA-binding kinetics and DNA bending (Gupta *et al.*, 2007; Khrapunov and Brenowitz, 2007; Delgadillo *et al.*, 2009).

To determine if the extended N-terminal region present in 38 kDa *Pf*TBP-FL affects *Pf*TBP-DNA interactions, purified recombinantly expressed 26 kDa 6His:*Pf*TBP-C, with a shorter 44 amino acid residue N-terminal IDR, and 38 kDa 6His:*Pf*TBP-FL, with the full-length N-terminal IDR of 143 amino acid residues, were normalised by immunoblot analysis using an anti-His antibody (Fig. 18A) and their relative DNA-binding activity compared in electrophoretic mobility shift assays (EMSAs; Fig. 18B). To this end, increasing concentrations of 26 kDa 6His:*Pf*TBP-C and 38 kDa 6His:*Pf*TBP-FL were incubated with a 282 bp DNA probe containing the *P. falciparum* GBP-130 promoter region from -179 to + 57 relative to the predicted TSS (Lanzer, De Bruin and Ravetch, 1992), to which binding of *Pf*TBP had previously been demonstrated in the laboratory (Milton, 2017). Previous studies had shown that inhibition of *Sc*TBP DNA binding by its N-terminal IDR was strongly increased when binding reactions were carried out on ice, compared to binding reactions incubated at 30˚C (Lieberman *et al.*, 1991). For this reason, DNA-binding reactions with *Pf*TBP-C and *Pf*TBP-FL were carried out on ice before being resolved by agarose gel electrophoresis in the presence of magnesium.

Consistent with previous results, 6His:*Pf*TBP-C bound strongly to the GBP-130 probe, with a prominent 6His:*Pf*TBP-C-GBP-130 complex detected at a concentration of 62 nM (Fig. 18B, lane 2). Further increasing 6His:*Pf*TBP-C concentrations resulted in a corresponding increase of bound DNA probe and the formation of smeary *Pf*TBP-DNA complexes with increasingly reduced mobility, suggesting multiple *Pf*TBP binding events (Fig. 18B, lanes 2-4). In contrast, DNA complexes formed with the 38 kDa 6His:*Pf*TBP-FL were barely detectable even at the highest concentration (250 nM) of

6His:*Pf*TBP-FL (Fig. 18B, lane 8). This result suggests that the longer N-terminal IDR of *Pf*TBP-FL may indeed inhibit *Pf*TBP-DNA interactions.

Given that the variable N-terminal disordered region of TBP is not required for sequence-specific DNA binding by TBPs in well-characterised eukaryotes, 6His:*Pf*TBP-C was used for further characterisation of *Pf*TBP DNA-binding activity.

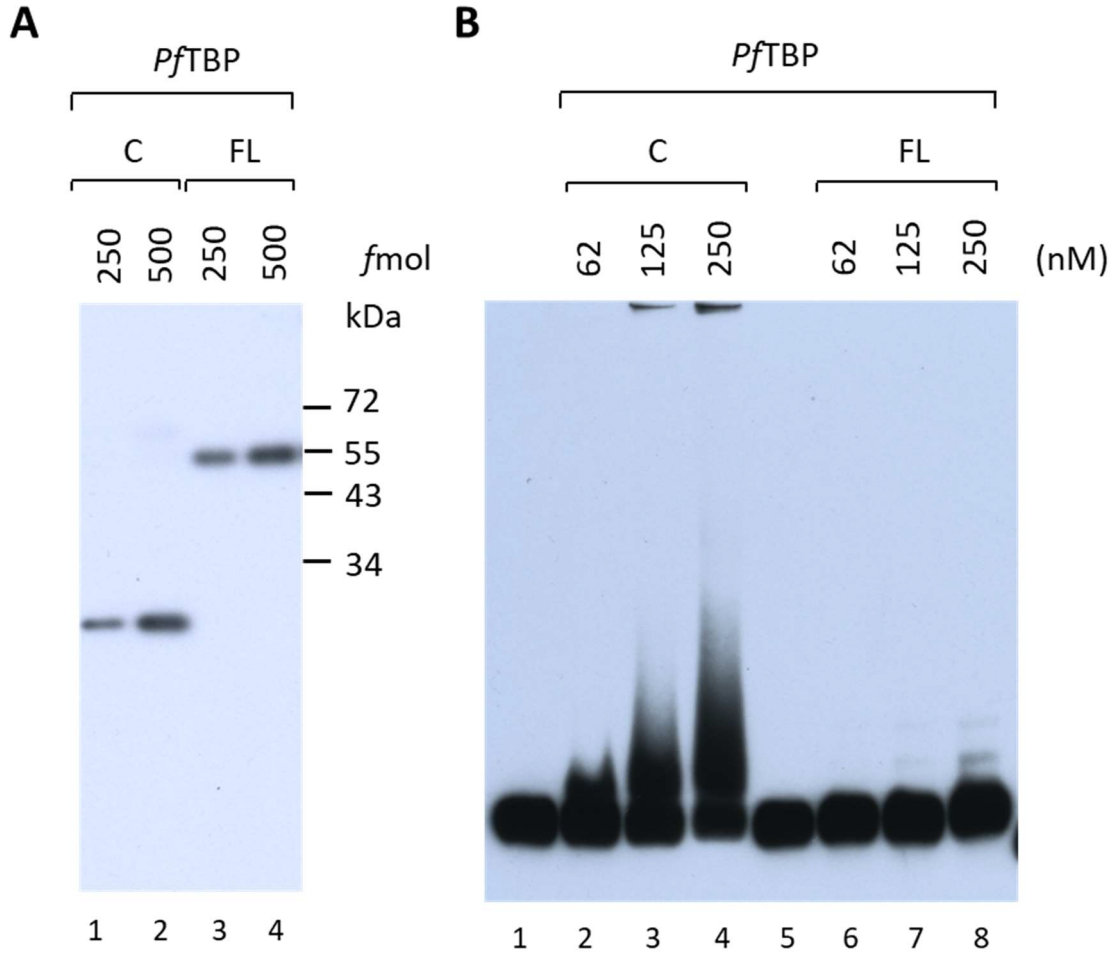


Figure 18. The intrinsically disordered *Pf*TBP N-terminus inhibits *Pf*TBP DNA-binding activity.
(A) Normalisation of purified recombinant 26 kDa *Pf*TBP-C (C) and full-length 38 kDa *Pf*TBP-FL (FL) preparations by immunoblot analysis. Samples were resolved by 12% SDS-PAGE, transferred to a PVDF membrane and probed with mouse anti-His antibody. The relative position of MW standards is indicated. **(B)** DNA-binding reactions (10 µl) contained 5 fmol 282 bp biotinylated GBP-130 DNA probe, 30 ng poly(dG-dC) and indicated concentrations of *Pf*TBP-C or *Pf*TBP-FL. DNA-binding reactions were resolved by 1.4% agarose gel electrophoresis in the presence of Mg^{2+}.

3.5.2. Establishing EMSA conditions for the resolution of *Pf*TBP and *Pf*TLP interactions with short DNA probes

The electrophoretic mobility shift assay (EMSA) is a sensitive and efficient method for the characterisation of protein-nucleic interactions (Hellman and Fried, 2007) and has been instrumental in classical studies of TBP-TATA interactions (Horikoshi *et al.*, 1989; Kuddus and Schmidt, 1993; Coleman and Pugh, 1995; Starr, Hoopes and Hawley, 1995; Bareket-Samish, Cohen and Haran, 2000; Zhao and Herr, 2002; Gilfillan *et al.*, 2005). This technique involves the incubation of a protein of interest with a labelled nucleic acid probe and subsequent separation of protein-nucleic acid complexes from unbound nucleic acid by electrophoresis. The gel conditions used for the electrophoresis step have a significant impact on the stability and migration of the protein-nucleic acid complex during electrophoresis. Poor choice of gel conditions may decrease the detection of protein-nucleic acid complexes, may result in complex dissociation, or decrease the separation of protein-DNA complexes from free DNA. For these reasons, optimising gel conditions is a prerequisite for studying protein-nucleic acid interaction by EMSA. While some protein-nucleic acid complexes are well resolved in native PAGE gels, others are better resolved in agarose gels with larger mesh sizes. Furthermore, additives such as glycerol, ethylene glycol or magnesium may be required for complex stability during electrophoresis (Kuddus and Schmidt, 1993; Hellman and Fried, 2007).

Given the high degree of divergence at the DNA-interaction interface of *Pf*TLP and *Pf*TBP (Section 3.1.2), this study aimed to determine whether *Pf*TBP and *Pf*TLP bind a prototypical TATA box DNA sequence in a sequence-specific manner. To this end, the adenovirus 2 major late promoter (Ad2ML), widely used in seminal studies to determine TATA-specific DNA binding of eukaryotic TBPs *in vitro*, (Sawadogo and Roeder, 1985; Hahn *et al.*, 1989; Horikoshi *et al.*, 1989; Kim, Nikolov and Burley, 1993; Kuddus and Schmidt, 1993; Nikolov *et al.*, 1996; Weideman *et al.*, 1997; Zhao and Herr, 2002) was considered an ideal DNA probe to examine interactions of *Pf*TLP and *Pf*TBP with a prototypical TATA box DNA sequence.

As an initial means to assess *Pf*TBP and *Pf*TLP interactions with the Ad2ML TATA box, DNA-binding reactions were resolved by PAGE gel. To this end increasing concentrations of either 6His-*Pf*TBP-C, 6His-*Pf*TLP or 6His-*Hs*TBP were incubated with a 40 bp Ad2ML probe (Ad2ML-TATA) in standard transcription buffer before free DNA and protein-DNA complexes were resolved by native 5% PAGE. These initial conditions, previously established for the detection of *Hs*TBP-Ad2ML complexes (Horikoshi *et al.*, 1989; Weideman *et al.*, 1997; Gilfillan *et al.*, 2005), proved inadequate to resolve either *Pf*TBP or *Pf*TLP complexes with the Ad2ML-TATA probe (Fig. 19). Consistent with previously

published results (Gilfillan *et al.*, 2005), *Hs*TBP formed two well-resolved complexes with different mobility upon binding to the Ad2ML-TATA probe (Fig. 19 lanes 5-7, panel A and B). In contrast, increasing concentrations of *Pf*TBP with the Ad2ML-TATA probe resulted in the formation of a number of less distinct complex bands (Fig. 19A, lanes 2-4). In addition, a significant proportion of *Pf*TBP-Ad2ML complexes were unable to enter the gel and were instead retained in the wells. *Pf*TLP complexes with the Ad2ML-TATA probe could not be resolved at all, and increasing *Pf*TLP concentrations resulted in increasing retention of the DNA probe in the wells without detectable complexes in the gel (Fig. 19B, lanes 2-4). These observations suggest multiple binding of *Pf*TBP and *Pf*TLP to the 40 bp Ad2ML-TATA probe and/or aggregation of protein-DNA complexes in the binding reactions.

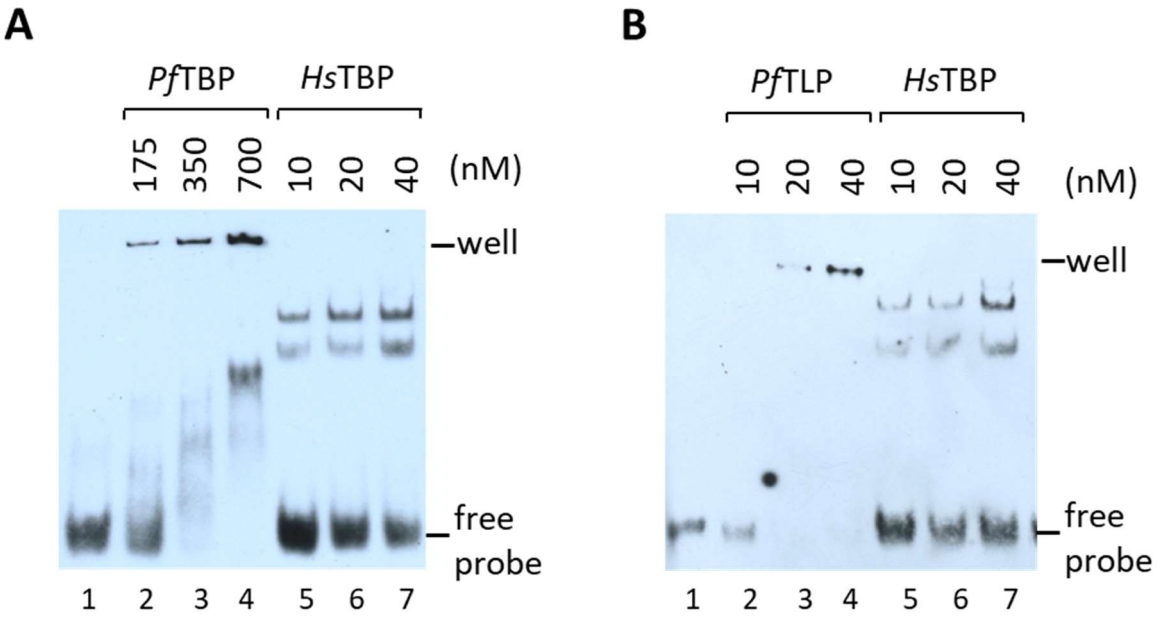


Figure 19. Analysis of *Pf*TBP and *Pf*TLP binding to a 40 bp Ad2ML DNA probe by PAGE EMSA.
Binding reactions (10 µl) contained 5 *f*mol 40 bp biotinylated Ad2ML-TATA DNA probe and indicated concentrations of *Pf*TBP, *Pf*TLP or *Hs*TBP. Reactions were incubated at 27°C for 45 min and resolved by 5% (37:1) PAGE in the absence of Mg^{2+}. The positions of free DNA probe and the gel wells are indicated.

Agarose EMSAs have also been used to study TBP-DNA interactions (Lieberman and Berk, 1994) and have been shown to offer better resolution of large multiprotein complexes such as the TFIID complex (Zerby and Lieberman, 1997). More recently, agarose EMSAs have also been successfully used to study interactions between *Pf*TBP and *Pf*TLP and long 200 bp DNA probes (Talvik, 2016; Milton, 2017; van der Linden, 2019). To test whether *Pf*TBP and *Pf*TLP DNA interactions could be resolved by altering the gel type or by the presence of magnesium ions, binding reactions were repeated and resolved by 1.4% agarose gel electrophoresis in either the presence or absence of magnesium ions, which have been shown to stabilise *Hs*TBP-TATA box interactions during gel electrophoresis (Horikoshi *et al.*, 1989; Reddy and Hahn, 1991; Kuddus and Schmidt, 1993).

As expected, *Hs*TBP-Ad2ML complexes appeared more stable in gels run in the presence of 2 mM Mg^{2+} in the gel and running buffer (Fig. 20, lanes 13 - 17 compare upper and lower panels). In contrast, the presence of Mg^{2+} decreased detection of *Pf*TBP-Ad2ML and *Pf*TLP-Ad2ML complexes. Gel electrophoresis in the absence of Mg^{2+} resulted in the detection of multiple *Pf*TBP-Ad2ML complex bands (upper panel lanes 2-6), suggesting the occurrence of multiple DNA binding events or alternatively the existence of different protein-DNA complex conformations, for example resulting from differential DNA-bending. However, multiple complexes are not resolved in EMSA gels in the absence of Mg^{2+} where only a single complex band was observed at high *Pf*TBP concentrations (Fig. 20, compare lanes 2-6 upper and lower panels). Similarly, EMSA electrophoresis in the absence of Mg^{2+} allowed detection of a single low mobility complex formed in the presence of increasing *Pf*TLP concentrations, which could not be resolved from the free probe when electrophoreses was carried out in the presence of Mg^{2+} (Fig. 20, lane 11, compare upper and lower panels).

Taken together, the results presented above show that, in contrast to human TBP-TATA interactions, which are stabilised by the presence of Mg^{2+} during electrophoresis, DNA binding of *Pf*TBP or *Pf*TLP is best studied in agarose EMSAs, without magnesium present in the agarose gel or electrophoresis buffer.

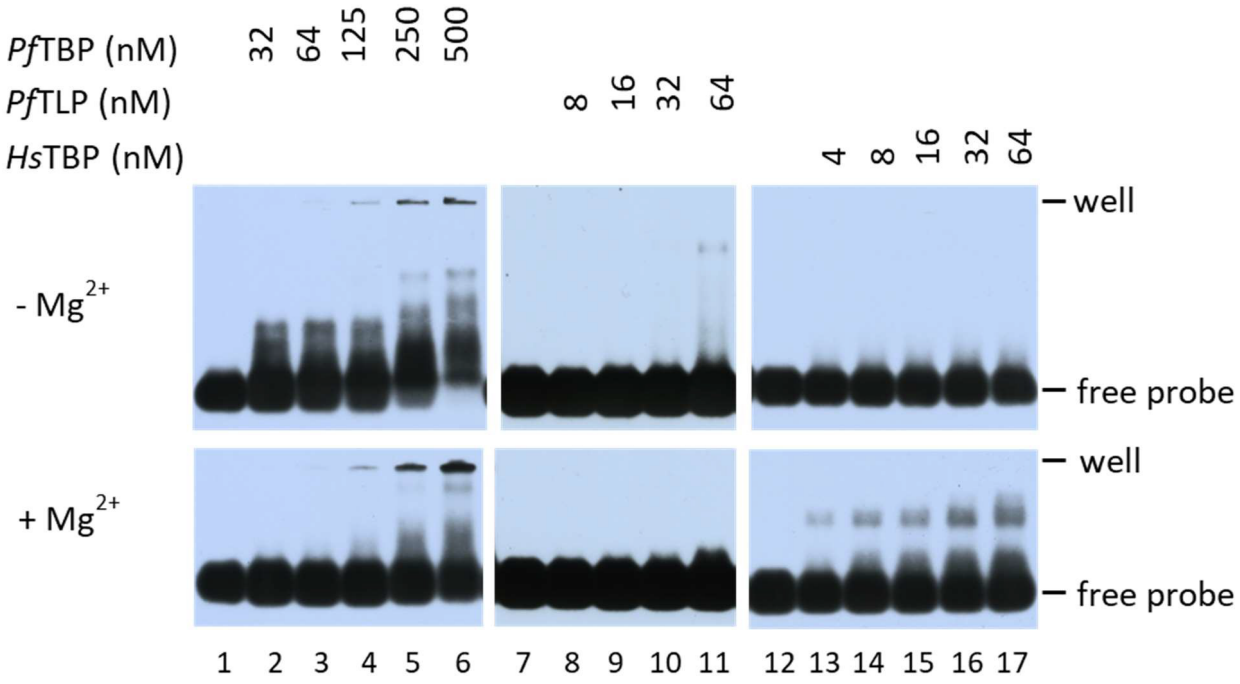


Figure 20. Detection of binding of *Pf*TBP or *Pf*TLP to Ad2ML promoter DNA in agarose EMSAs is negatively affected by the presence of Mg^{2+} in the gel and electrophoresis buffer. Binding reactions (10 µl) contained 5 fmol 40bp biotinylated Ad2ML DNA probe and indicated concentrations of *Pf*TBP, *Pf*TLP or *Hs*TBP. After incubation for 45 min at 27 ° C, one half of the binding reaction was resolved by 1.4% agarose gel electrophoresis in the presence of 5 mM magnesium acetate and the other half by 1.4% agarose electrophoresis in the absence of Mg^{2+}. Data panels shown for the presence or absence of Mg^{2+}originate from the same gel, respectively.

3.5.3. *Pf*TBP and *Pf*TLP do not possess sequence-specific TATA box-binding activity

Prototypical TATA box-binding proteins in well-studied eukaryotic organisms are characterised by sequence-specific binding to the 'TATAWAWN' TATA box consensus sequence (Thomas & Chiang, 2008). Substituting thymine nucleotides at the first and third positions of the TATA box sequence with guanine has been shown to drastically reduce *Hs*TBP DNA binding as detected by EMSAs (Gilfillan *et al.*, 2005) and, at the same time, to drastically diminish promoter activity in human and yeast cell extracts (Wobbe and Struhl, 1990). To investigate whether *Pf*TBP and *Pf*TLP interactions with the Ad2ML promoter (Fig. 20) are sensitive to substitution of thymine by guanine at the first and third positions of the 'TATAWAWN' consensus sequence, the affinity of *Pf*TBP and *Pf*TLP for Ad2ML-TATA and Ad2ML-GAGA DNA probes was compared by protein titration EMSA and EMSA competition experiments (Materials and Methods, Section 2.1.10).

For protein titration EMSAs, 40 bp biotinylated Ad2ML-TATA or Ad2ML-GAGA DNA probes were incubated with increasing concentrations of 6His:*Hs*TBP, 6His:*Pf*TBP-C or 6His:*Pf*TLP. Bound and free DNA were separated by agarose gel electrophoresis either in the presence (*Hs*TBP) or in the absence (*Pf*TBP, *Pf*TLP) of Mg^{2+} (see Section 3.5.2). A representative result is shown in Fig. 21. Consistent with the established specificity of prototypical eukaryotic TBPs (Gilfillan *et al.*, 2005; Thomas and Chiang, 2006), *Hs*TBP binding to Ad2ML-GAGA was drastically reduced relative to Ad2ML-TATA (Fig. 21A, compare lanes 2 and 3 to 5 and 6). In contrast, both *Pf*TBP and *Pf*TLP bound the AD2ML-TATA and GAGA probes with comparable affinity (Fig. 21B, compare lanes 2-4 to 6-8, Fig. 21C, compare lanes 2-5 to 7-10; also see Fig. S3).

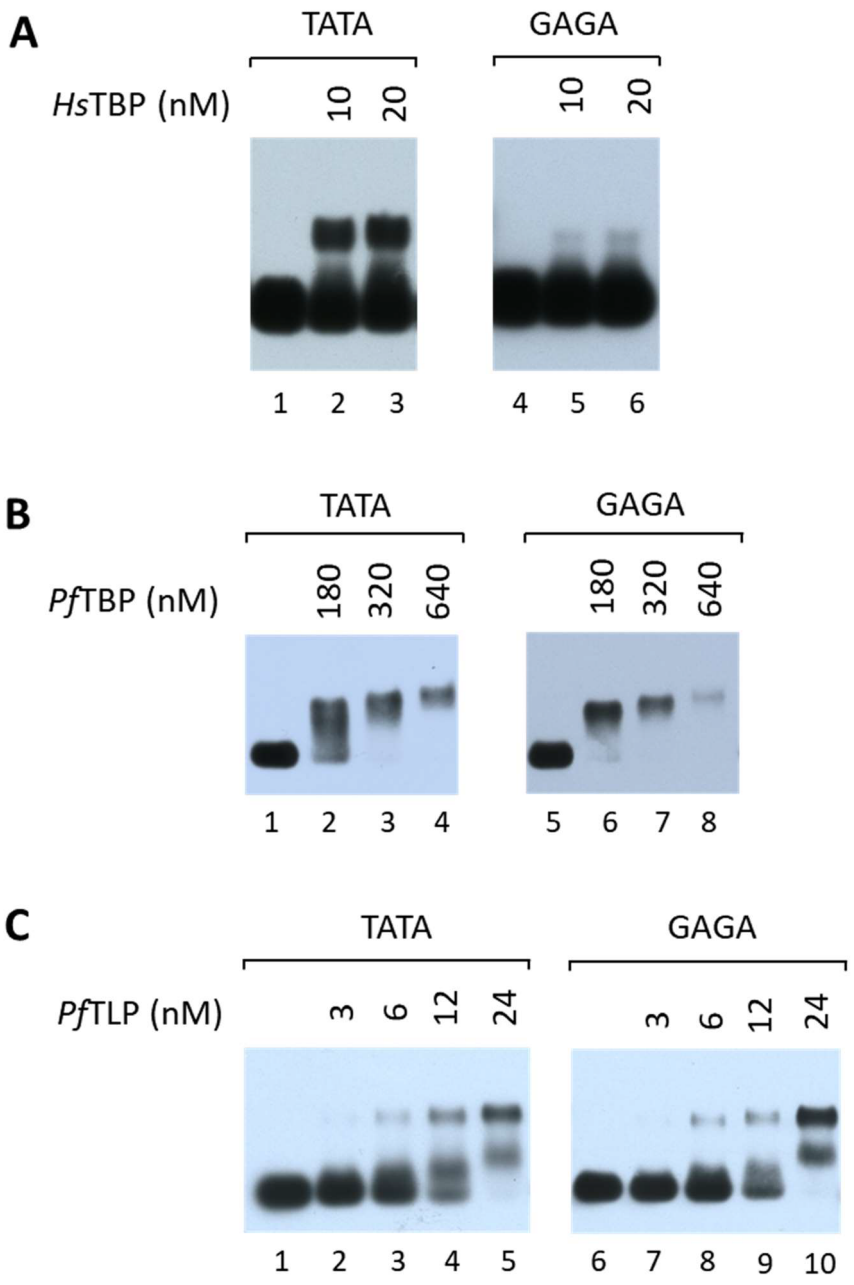


Figure 21. *Pf*TLP and *Pf*TBP show no preference for a *bona fide* TATA box DNA sequence in EMSA protein titration assays. DNA-binding reactions (10 µl) contained 10 fmol **(A)** or 5 fmol **(B)** of 40 bp Ad2ML-TATA or Ad2ML-GAGA DNA probe and indicated concentrations of *Hs*TBP, *Pf*TBP or *Pf*TLP. Reactions were incubated at 27°C for 45 min and resolved by 1.4% agarose gel electrophoresis in the presence **(A)** or absence **(B and C)** of 5 mM Mg^{2+}.

*Pf*TBP and *Pf*TLP TATA-binding specificity was further investigated by EMSA competition experiments. EMSA competition assays involve the incubation of the protein of interest with a labelled DNA probe in the presence of increasing concentrations of unlabelled DNA probes. The unlabelled DNA probe competes with the labelled DNA probe for DNA-protein interaction, decreasing the formation of labelled protein-DNA complex. By comparing the degree of competition exerted by different unlabelled DNA sequences, the relative affinity of the protein of interest for these unlabelled probes can be reliably determined (Carey, Peterson and Smale, 1998). This approach offers two main advantages over protein titration EMSAs. First, binding affinity to different unlabelled competitor DNA sequences is not influenced by the relative stability of protein DNA-complexes during electrophoresis. Second, the same labelled probe is used in all reactions, eliminating potential differences in labelling and overall integrity of the DNA probe used to form detectable protein-DNA complexes.

For the EMSA competition experiments, a total of five different competitor DNAs were selected. These included a 40 bp Ad2ML promoter probe with the wildtype TATA box sequence TATAAAA (TATA), a TATA box mutant GAGAAAA (GAGA), a 40 bp probe containing the G/C-rich and TATA-less mouse terminal deoxynucleotidyl transferase (mTdT-WT) promoter (Smale and Baltimore, 1989; Martinez *et al.*, 1998), a mTdT promoter mutant variant containing the high affinity TATA sequence TATAAAA (mTdT-TATA) (Martinez *et al.*, 1998; Malecová *et al.*, 2007)), as well as poly(dG-dC) DNA, consisting entirely of alternating G-C base pairs.

Comparison of the competition exerted by these different competitors on the formation of the Ad2ML-TATA-*Hs*TBP complex revealed that *Hs*TBP binds to the Ad2ML-TATA and mTdT-TATA sequences with much higher affinity than to the Ad2ML-GAGA and mTdT-WT sequences (Fig. 22). Notably, equivalent nucleotide amounts of poly(dG-dC) competitor had no significant effect on *Hs*TBP-TATA box interactions (Fig. 22, top panel, lanes 18-21).

In line with expectations, these results clearly confirm the specificity of *Hs*TBP-TATA box interactions under the experimental conditions, and further confirm the validity of this assay as a test for *bona fide* TBP-TATA interactions.

In contrast to the results seen with *Hs*TBP, *Pf*TBP-Ad2ML and *Pf*TLP-Ad2ML complex formation was equally outcompeted by unlabelled Ad2ML-TATA, Ad2ML-GAGA, mTdT-WT and mTdT-TATA competitors (Fig. 22). These results clearly show that neither *Pf*TBP nor *Pf*TLP exhibit sequence-specific binding to the prototypical TATA-box DNA sequence. In addition, and perhaps most significantly, poly(dG-dC) outcompeted both the *Pf*TBP-Ad2ML-TATA and *Pf*TLP-Ad2ML-TATA

complexes as well as all other competitors, strongly suggesting that both *Pf*TBP and *Pf*TLP interact non-specifically with the Ad2ML-TATA DNA.

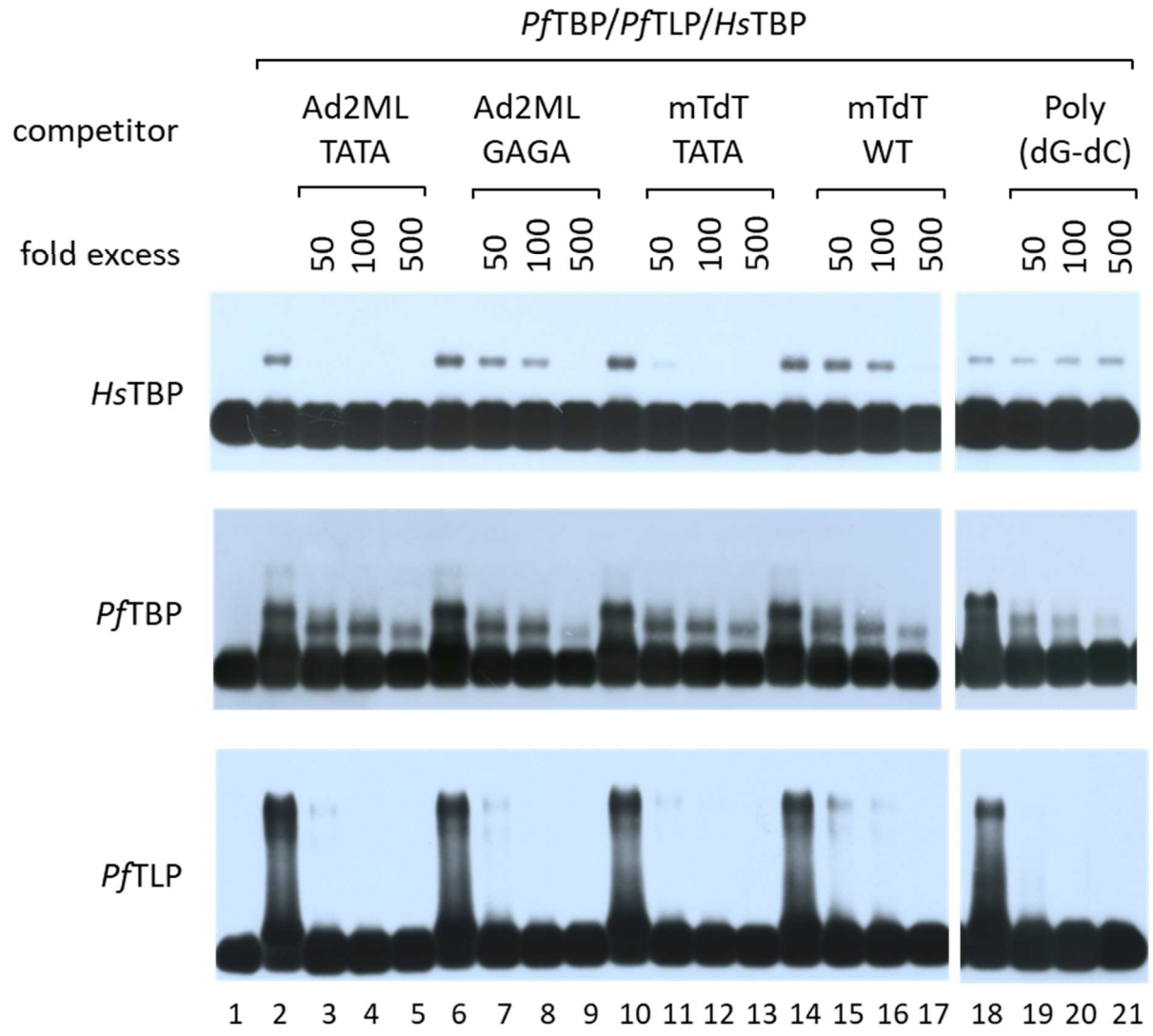


Figure 22. *Pf*TLP and *Pf*TBP bind DNA containing a *bona fide* TATA box in a non-specific manner.
DNA-binding reactions (10 µl) contained 5 fmol Ad2ML-TATA, 0.7 nM *Hs*TBP, 12 nM *Pf*TLP or 63 nM *Pf*TBP and the indicated fold excess of unlabelled competitor or poly(dG-dC) competitor. Fold excess poly(dG-dC) was calculated in molar nucleotides. Binding reactions were incubated at 27°C for 45 min. Complexes were resolved by 1.4% agarose gel in the presence (*Hs*TBP) or absence (*Pf*TBP and *Pf*TLP) of Mg^{2+}.

3.5.4. Analysis of *Pf*TBP interactions with *P. falciparum* GBP-130 and KAHRP putative promoter regions

3.5.4.1. PfTBP interacts selectively with 282 bp KAHRP and GBP-130 promoter DNA probes

Prior experiments by our group have shown *Pf*TBP to exhibit sequence-selective binding to a 282 bp DNA probe containing the *P. falciparum* glycophorin-binding protein 130 (GBP-130) putative promoter region (spanning bp -179 to +57 (Lanzer, de Bruin and Ravetch, 1992) compared to a 260 bp DNA probe containing the G/C-rich mouse TdT promoter (Milton, 2017). However, the DNA region(s) conferring sequence-selective binding had not been identified. Given that *Pf*TBP did not display sequence-specific binding to the prototypical TATA box DNA sequence, this provided opportunity to investigate the sequence determinants of selective *Pf*TBP-DNA interactions. As a first step, *Pf*TBP binding to four DNA probes was compared. These included a 282 bp GBP-130 probe, a 282 bp probe containing the *P. falciparum* knob-associated histidine-rich protein (KAHRP) putative promoter region from -171 to +65 (Lanzer, De Bruin and Ravetch, 1992), a 266 bp DNA probe containing the G/C-rich wildtype mouse TdT promoter (mTdT-WT), and a mouse TdT promoter variant (mTdT-TATA) containing the consensus TATA box sequence TATAAAA 30 bp upstream of the transcription initiation site (Martinez *et al.*, 1998; Malecová *et al.*, 2007).

As shown in Fig. 23, *Pf*TBP binds to the A/T-rich GBP-130 and KAHRP DNA probes with comparable affinity (panel A, compare lanes 2-4 with lanes 6-8; panel B, compare lanes 2-5 with lanes 7-10). *Pf*TBP binding to these putative *P. falciparum* promoter regions is sequence-selective, as *Pf*TBP only interacts with the G/C-rich mTdT DNA probes at significantly higher *Pf*TBP concentrations (panel A, compare lanes 2-4 and 6-8 with lanes 10-12 and 14-16, panel B, compare lanes 2-5 and 7-10 with lanes 12-15 and 17-20). In line with the results of Section 3.5.3 (Fig. 21 and 22), the presence of the consensus TATA box sequence 'TATAAAA' did not affect *Pf*TBP interactions with the mTdT promoter probes (Fig. 23 panel B, compare lanes 12-15 with lanes 17-20). Thus, the TATAAAA sequence was not sufficient to mediate *Pf*TBP nucleoprotein complex formation at concentrations at which complexes can be observed with the KAHRP and GBP-130 probes (Fig. 23A). Together, these results suggest that *Pf*TBP interactions with the GBP-130 and KAHRP DNA probes are distinct from established TBP interactions with a prototypical TATA-box DNA sequence. Although *Pf*TBP EMSA results were broadly consistent in gels run in the presence or absence of magnesium ions (Fig. 23, compare panels A and B), *Pf*TBP complexes formed with GBP-130 and KAHRP sequences appeared more stable during electrophoresis in the presence of magnesium. Electrophoresis in the absence of magnesium ions resulted in smearing of the free probe rather than formation of a distinct complex band, suggesting dissociation of *Pf*TBP-DNA complexes during electrophoresis (Fig. 23, compare

panels A and B). For this reason, Mg^{2+} agarose gels were used for further analysis of *Pf*TBP binding to these A/T-rich *P. falciparum* putative promoter regions.

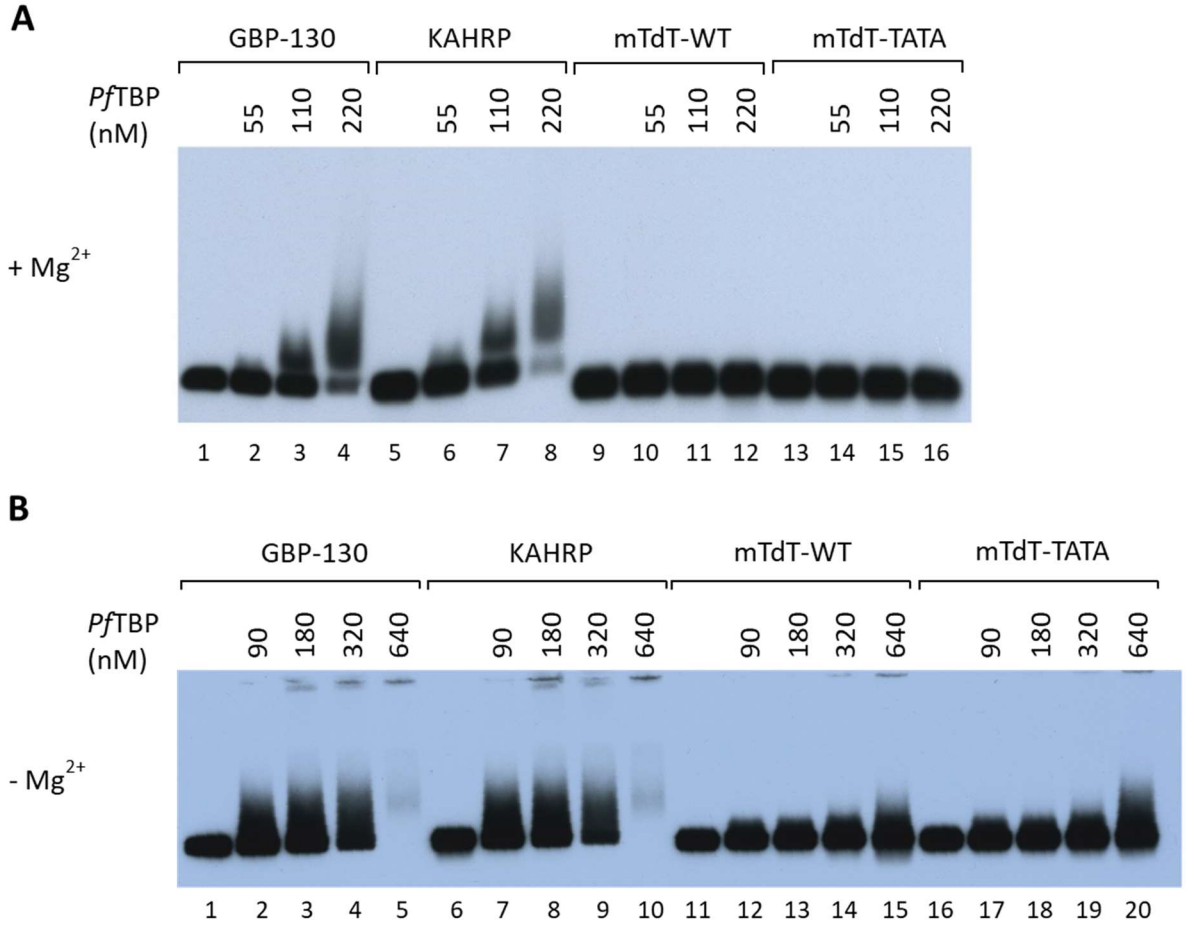


Figure 23. Selective binding of *Pf*TBP to GBP-130 and KAHRP putative *P. falciparum* promoter regions. DNA-binding reactions (10 µl) contained 5 fmol of GBP-130 (282 bp), KAHRP (282 bp), mTdT-WT (260 bp) or mTdT-TATA (266 bp) DNA probes, 30 ng **(A)** or 100 ng **(B)** poly(dG-dC) and indicated concentrations of *Pf*TBP. Reactions were incubated on ice for 45 min before being resolved by 1.4% agarose gel electrophoresis in the presence **(A)** or absence **(B)** of Mg^{2+}.

3.5.4.2. PfTBP displays moderate sequence selectivity towards specific regions within GBP-130 and KAHRP promoter DNA probes.

To identify regions within the GBP-130 and KAHRP probes preferentially bound by *Pf*TBP, two overlapping DNA probes dividing the GBP-130 and KAHRP promoter regions, designated GA and GB (GBP-130), and KA and KB (KHARP), were generated by PCR (Materials and Methods, Section 2.1.10, Fig. 24A). Protein titration experiments using the GA, GB, KA and KB DNA probes, revealed that *Pf*TBP bound to both subregions of the GBP-130 and KAHRP probes (Fig. 24B). However, *Pf*TBP-DNA complex formation could be detected with an approximately two-fold lower *Pf*TBP concentration with the GB DNA probe compared to the GA DNA probe (Fig. 24B, compare lanes 2 and 3 with lanes 7 and 8), suggesting that *Pf*TBP has a slightly higher affinity for the GB DNA probe. Similarly, a 2-4-fold lower concentration was required to observe *Pf*TBP-DNA complex formation with the KB DNA probe compared to the KA DNA probe (Fig. 24B, compare lanes 12 and 13 with 17 and 18).

In order to ascertain that the differences in binding of *Pf*TBP to the different GBP-130 or KHARP DNA fragments reflects an actual difference in binding affinity rather than relative stability of *Pf*TBP-DNA complexes during electrophoresis, EMSA competition experiments were performed. In these experiments, *Pf*TBP DNA complex formation with labelled GB probe was challenged with excess unlabelled GA and GB DNA fragments. In agreement with the results of the protein titration experiments (Fig. 24B), the EMSA competition results show that approximately four-fold higher concentrations of the GA DNA probe were required compared to the GB DNA probe to outcompete the *Pf*TBP-GB complex (Fig. 24C, compare lane 5 with lane 13). In addition, while *Pf*TBP interactions with the biotinylated GB DNA probe resulted in an upwards smear of free probe, characteristic of less stable non-specific interactions, the inclusion of GA DNA competitor in the binding reaction promoted the visualisation of a well-resolved complex (Fig. 24C, lanes 9-13). This observation may suggest that GA competitor DNA outcompeted relatively unstable, non-specific interactions, but not a population of stably bound *Pf*TBP-GB complexes. In contrast, no discrete complex was formed in the presence of GB competitor DNA, suggesting equivalent competition of both stable and relatively unstable *Pf*TBP-DNA complexes. Taken together, the results of the EMSA competition experiments (Fig. 24C), confirm the results of the protein titration EMSAs (Fig. 24B) and indicate that *Pf*TBP interactions with the GBP-130 DNA probe are, at least to some extent, sequence-selective.

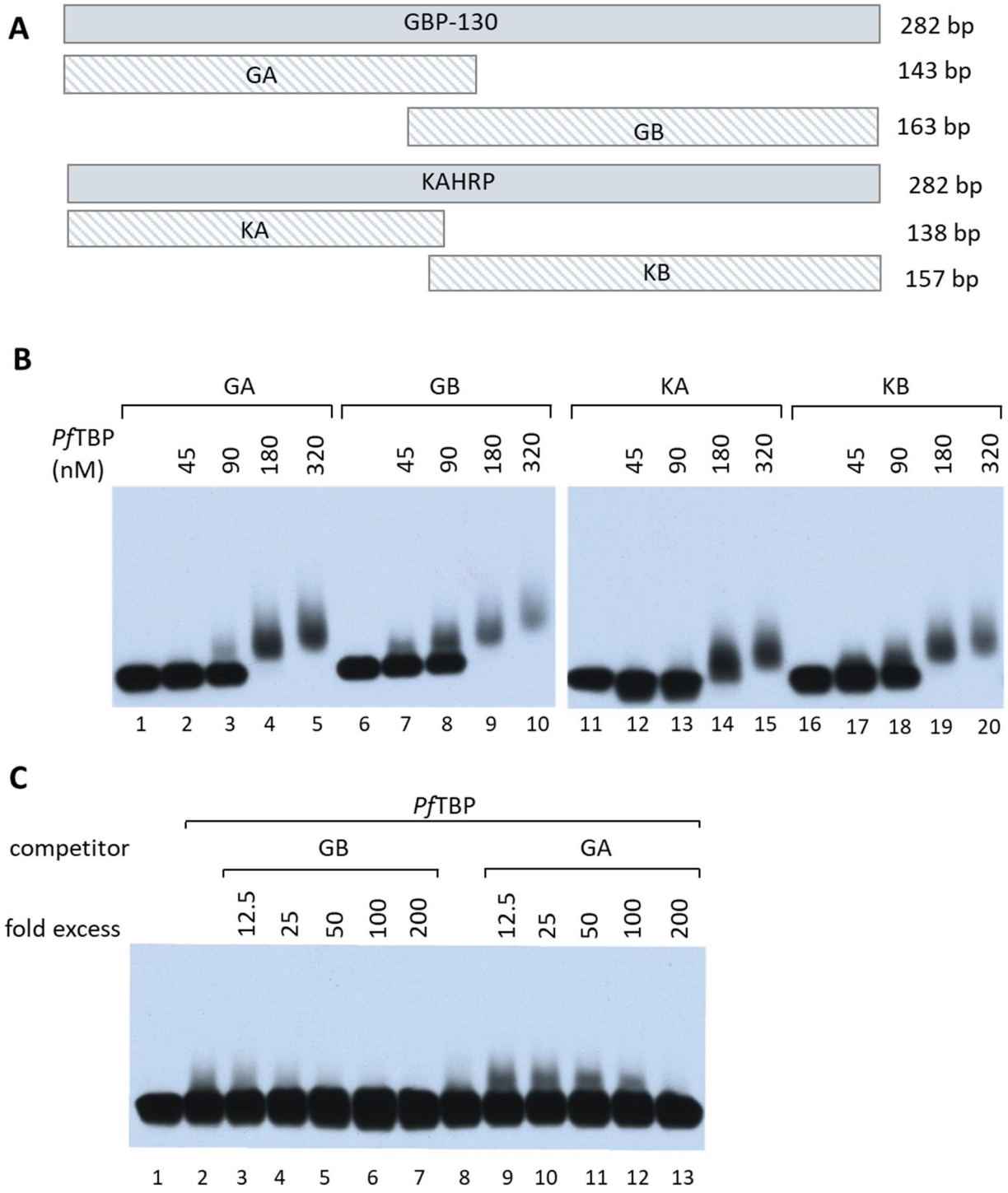


Figure 24. *Pf*TBP displays moderate sequence selectivity towards specific GBP-130 and KAHRP promoter regions. (A) Schematic representation of DNA probes containing specific regions within GBP-130 and KAHRP promoters. **(B)** DNA-binding reactions (10 µl) contained 5 fmol biotinylated GA (143 bp), GB (163 bp), KA (138 bp) or KB (157 bp) DNA probes, 30 ng poly(dG-dC) and indicated concentrations of *Pf*TBP. **(C)** DNA-binding reactions (10 µl) contained 5 fmol biotinylated GB, 30 ng poly(dG-dC), 16 nM *Pf*TBP and the indicated fold excess unlabelled competitor. All reactions (B and C) were incubated on ice for 45 min before being resolved by 1.4% agarose gel in the presence of Mg^{2+}.

3.5.4.3. PfTBP binds multiple sites within the GB DNA probe

Considering the relatively small differences in affinity for *Pf*TBP binding to GA and GB and KA and KB DNA probes and the very low degree of DNA sequence complexity within the highly A/T-rich *P. falciparum* GBP-130 and KHARP promoter regions, it seemed unlikely that *Pf*TBP specifically recognises a single well-defined DNA sequence motif within the KHARP or GBP-130 probes. However, a previous study reported that *Pf*TBP interacted specifically with 'TGTAA' and 'TATAA' motifs within the *P. falciparum* GBP-130 and KAHRP promoter regions (Ruvalcaba-Salazar *et al.*, 2005). Therefore, *Pf*TBP binding to the GB probe was further investigated in order to find out if *Pf*TBP binding to a defined interaction site can be demonstrated. To delineate regions within the GB fragment of the GBP-130 promoter region selectively bound by *Pf*TBP, 50 bp DNA probes spanning GB with a 10 bp overlap were generated (Fig. 25A). Because *Pf*TBP showed higher affinity for the GB over the GA fragment (Fig. 24), the region of the GB fragment overlapping with the GA fragment was excluded from the analysis. Also excluded were DNA sequences used for PCR amplification, present in all promoter DNA probes, including the mTdT probes to which *Pf*TBP does not bind (Fig. 23; see also Materials and Methods, Section 2.1.10).

As shown in Fig. 25, *Pf*TBP displayed roughly equivalent binding activity to each of the three 50 bp fragments, with the GB3 probe bound with slightly lower affinity (Fig. 25B, compare lanes 4 and 9 to lane 14). Furthermore, *Pf*TBP bound the three 50 bp fragments with lower affinity compared to the complete 163 bp GB probe (Fig. 25B, compare lanes 2-4, 7-9 and 12-14 with lanes 17-19). These results are not consistent with a model in which *Pf*TBP binds specifically to a single site and suggest instead that *Pf*TBP interacts with multiple binding sites within the GB probe.

The results presented above can be explained by a model in which *Pf*TBP possesses broad selectivity for binding A/T-rich DNA regions and forms a stable complex with a wide variety of A/T-rich sequence motifs with only minor differences in binding affinity. The slightly lower *Pf*TBP affinity for probe GB3 compared to probes GB1 and GB2 may be explained by an overall lower A/T-content of the GB3 probe (70%) compared to the to the GB1 and GB2 probes (97%; Table 3), which would impact overall *Pf*TBP affinity if *Pf*TBP binds preferentially to A/T-rich regions (Fig. 25B). Similarly, the slightly higher affinity of *Pf*TBP for the GB DNA probe compared to the GA DNA probe and the slightly higher affinity for the KB DNA probe compared to the KA probe, noted previously (Fig. 24), can be explained by higher A/T-content of GB (81.6%A/T) and KB (78.3% A/T) DNA probes compared to GA (74.8% A/T) and KA (73.9% A/T) DNA probes. Furthermore, GB and KB DNA probes were slightly longer (163 bp and 157 bp respectively) than GA and KA (143 bp and 148 bp respectively) DNA probes (Table. 2), and thus may contain a larger number of potential *Pf*TBP DNA-binding sites.

Considering the lack of evidence for well-defined *Pf*TBP recognition sites within the context of A/T-rich *P. falciparum* promoter regions, no further attempts were made to determine a defined *Pf*TBP DNA recognition sequence.

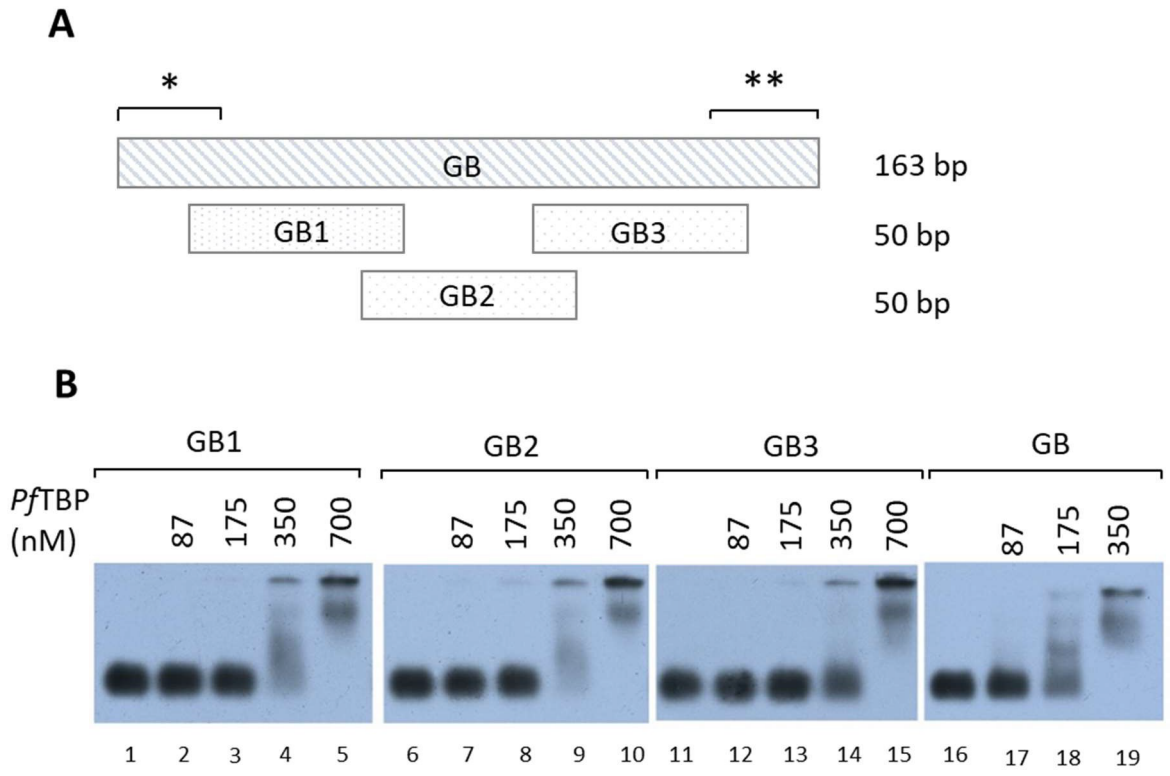


Figure 25. *Pf*TBP binding to the GB probe cannot be attributed to interactions with specific DNA regions or DNA sequence motifs. (A) Schematic representation of GB probes GB1-3. The 5' region of GB marked *, overlapping with probe GA, and the 3' region of GB marked **, containing plasmid-derived DNA sequences common to all mTdT and GBP-130 probes, were excluded from further analysis. **(B)** DNA-binding reactions (10 µl) contained 5 fmol biotinylated GB (163 bp), or mass equivalent (16 fmol) GB probes 1-3, 10 ng poly(dG-dC) and indicated *Pf*TBP concentrations. Binding reactions were incubated on ice before being resolved by 1.4% agarose gel electrophoresis in the presence of Mg^{2+}.

3.5.5. *Pf*TBP binds preferentially to the minor groove of A/T-rich DNA

TBPs from higher eukaryotes interact predominantly through hydrophobic interactions and hydrogen bonding with the DNA minor groove of the TATA-box (Starr and Hawley, 1991; Kim, Nikolov and Burley, 1993; Patikoglou *et al.*, 1999). To determine if DNA binding of *Pf*TBP to the A/T-rich GBP-130 and KHARP DNA probes are mediated by similar interactions with the DNA minor groove of A/T-rich DNA, the affinity of *Pf*TBP for GBP-130, poly(dI-dC) and poly(dG-dC) DNA was compared in EMSA competition experiments.

Poly(dI-dC) is a synthetic polymer that has been previously used to test AT-hook containing proteins for interaction with the minor groove of A/T-rich DNA (Banks, Mohr and Reeves, 1999; Metcalf and Wassarman, 2006). Substitution of the TATA box with dI-dC nucleotide base pairs was also used in early studies to demonstrate that sequence specific binding of prototypical TBP to the TATA box occurs through the DNA minor groove (Starr and Hawley, 1991; Lee and Hahn, 1995). The usefulness of poly(dI-dC) in differentiating between major and minor groove interactions in the context of DNA sequences composed of A and T nucleotides stems from the chimeric nature of poly(dI-dC) with regard to chemical signatures present in the major and minor grooves. The major groove of poly(dI-dC) contains electron donors and acceptors found in poly(dG-dC), while the minor groove contains the chemical signatures corresponding to poly(dA-dT) (Fig. 26A). As shown in Fig. 26B, *Pf*TBP bound to poly(dI-dC) and A/T-rich GBP-130 competitor DNA with comparable affinity (compare lanes 3-6 with lanes 8-11). In contrast, *Pf*TBP bound to poly(dG-dC) with significantly lower affinity than poly(dI-dC) and A/T-rich GBP-130 DNA competitor (compare lanes 13-17 with lanes 3-6 and 8-11). These results show that minor groove interactions are indeed sufficient to facilitate selective binding of *Pf*TBP to A/T-rich DNA.

Furthermore, equivalent competition with poly(dI-dC) and GBP-130 DNA also demonstrates that *Pf*TBP binds with comparable affinity to DNA comprised entirely of alternating A and T DNA nucleotides (poly(dI-dC)) and DNA containing a large variety of potential A/T-rich sequence motifs (GBP-130). These observations further confirm that *Pf*TBP binds a wide variety of A/T-rich sequences with comparable affinity.

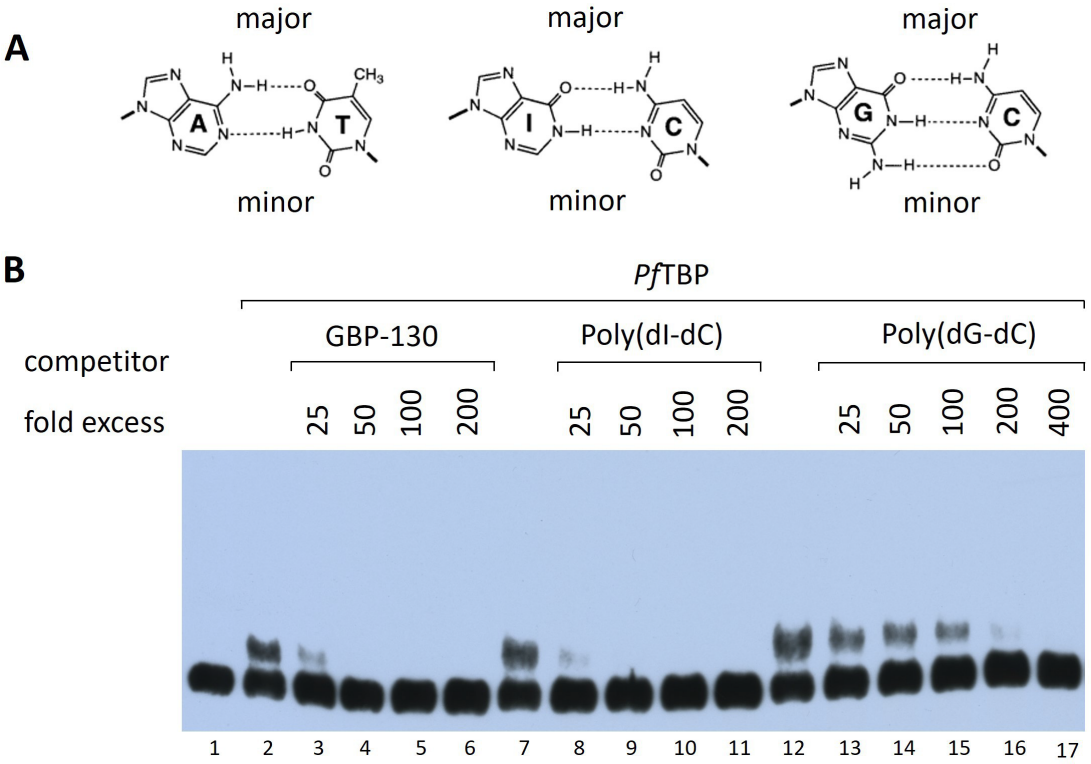


Figure 26. *Pf*TBP binds preferentially to the minor groove of A/T-rich DNA.
(A) Schematic showing the chemical signatures of A-T, I-C and G-C purine-pyrimidine base pairs presented in the major (top) and minor (bottom) grooves of B-DNA (adapted from Bailly *et al*., 1996). **(B)** DNA-binding reactions (10 µl) contained 5 fmol biotinylated 282bp GBP-130 DNA probe, 30 ng poly(dG-dC), 110 nM *Pf*TBP and added fold excess unlabelled competitor DNA as indicated. Fold excess poly(dG-dC) and poly(dI-dC) were calculated in molar nucleotides. DNA-binding reactions were incubated on ice for 45 min before being resolved by 1.4% agarose gel electrophoresis in the presence of Mg^{2+}**.**

3.5.6. *Pf*TLP binds selectively to the GBP-130 and KAHRP putative promoter regions

Like *Pf*TBP, *Pf*TLP has also been shown to interact with GBP-130 and KAHRP 282 bp DNA probes in agarose EMSA experiments (Bing, 2015; Milton, 2017; van der Linden 2019). Furthermore, preliminary experiments indicated that *Pf*TLP also binds to the GBP-130 promoter DNA probe with higher affinity than G/C-rich DNA probes (Milton, 2017). To characterise *Pf*TLP-DNA interactions in greater detail, EMSA titration assays were perfomed with GBP-130 and KARHP promoter probes. *Pf*TLP-DNA complexes were resolved by agarose electrophoresis in the absence of magnesium ions, as the presence of magnesium ions greatly reduced the difference in electrophorectic mobility between free probe and *Pf*TLP-DNA complexes (Fig. S4). Similar to the results seen with *Pf*TBP, increasing concentrations of *Pf*TLP resulted in the formation of *Pf*TLP-DNA complexes with both the GBP-130 and KAHRP putative *P. falciparum* promoter sequences (Fig. 27). No interaction was observed when equivalent concentrations of *Pf*TLP were incubated with the G/C-rich mTdT 260 bp DNA probe, showing that *Pf*TLP interactions with the GBP-130 and KAHRP probes are sequence-selective (Fig. 27, compare lanes 2-5 and lanes 7-8 with lanes 12-15). However, the *Pf*TLP-DNA complexes detection of defined *Pf*TLP-DNA complexes was highly variable between different electrophoresis runs, with longer durations of electrophoresis resulting in increased complex dissociation.

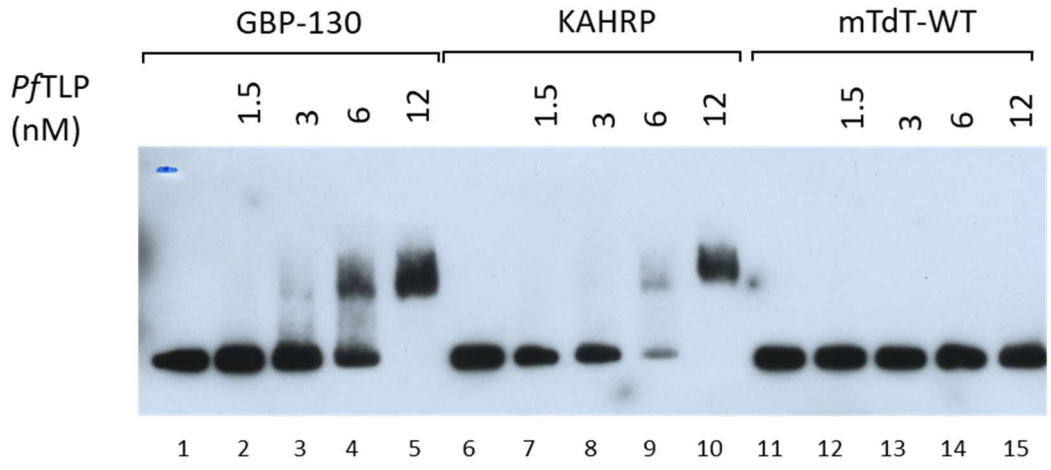


Figure 27. Selective binding of *Pf*TLP to GBP-130 and KAHRP putative *P. falciparum* promoter regions. DNA binding reactions (10 µl) contained 5 fmol of either GBP-130 (282 bp), KAHRP (282 bp) or mTdT-WT (260 bp) biotinylated DNA probes, 10 ng poly(dG-dC) and indicated concentrations of *Pf*TLP. Reactions were incubated on ice for 45 min before being resolved by 1.4% agarose gel electrophoresis in the absence of Mg^{2+}.

To further delineate the minimal DNA regions that may be responsible for *Pf*TLP sequence-selective binding, *Pf*TLP interaction with shorter overlapping regions within the the GBP-130 (probes GA and GB) or KAHRP (probes KA and KB, Fig. 24A) was examined by protein titration EMSAs as described for *Pf*TBP. Similarily to *Pf*TBP (Fig. 24), *Pf*TLP bound to all of the short GBP-130 and KAHRP probes, with only slightly greater affinity for the slightly longer GB and KB probes compared to GA and KA probes (Fig. 28A, compare lanes 2-4 with lanes 6-8 and lanes 10-12 with lanes 14-16). None of the observed *Pf*TLP interactions resulted in the formation of a distinct well-resolved complex (Fig. 28A), suggesting dissociation of *Pf*TLP-DNA complexes during electrophoresis.

Additional protein titration experiments revealed similar affinity of *Pf*TLP for all three 50 bp GB probe fragments (Figs. 25A and 28B). Further, and consistent with similar results seen with *Pf*TBP (Fig. 25B), the affinity of *Pf*TLP for 50 bp probe fragments appeared reduced compared to the parental 163 bp GB probe (Fig. 28B). These results suggest that *Pf*TLP interacts in similar manner with multiple regions of the GB DNA probe and suggest that *Pf*TLP, like *Pf*TBP, binds DNA with a broad preference for A/T-rich sequences, rather than defined sequence motifs within the A/T-rich promoter DNA context.

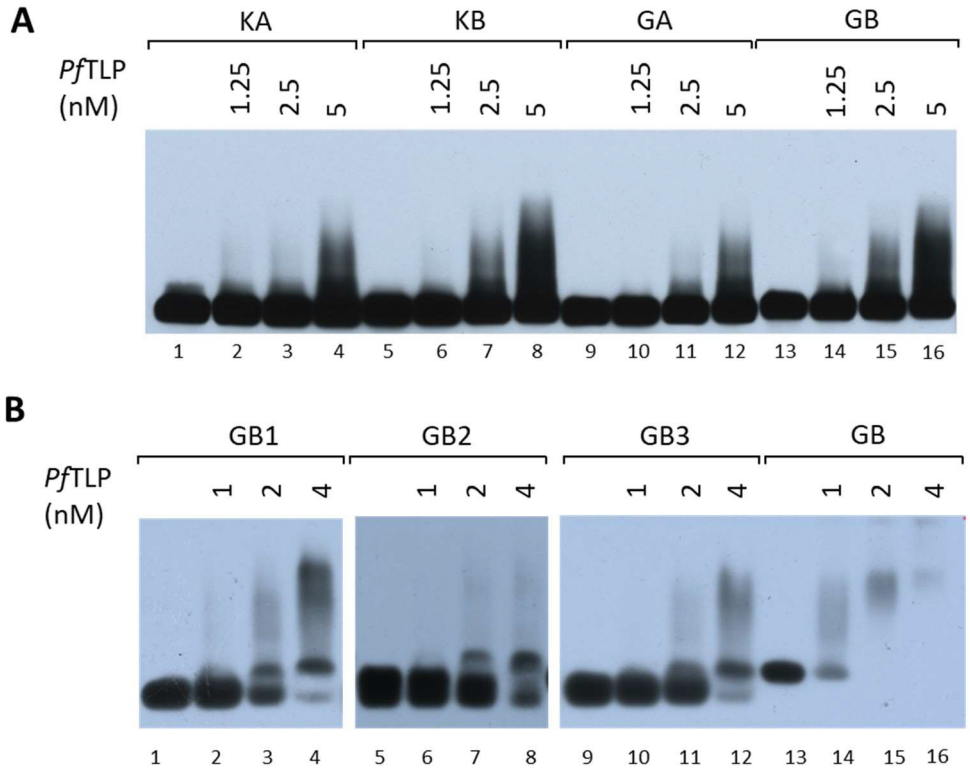


Figure 28. *Pf*TLP interacts with multiple regions within GBP-130 promoter DNA. DNA binding reactions (10 µl) contained 5 fmol biotinylated GA (143 bp), GB (163 bp), KA (138 bp) or KB (157 bp) DNA probes **(A)** or ng equivalent GB1, GB2 or GB3 DNA probes (50 bp) **(B)**, 10 ng poly(dG-dC) and indicated concentrations of *Pf*TLP. DNA-binding reactions were incubated on ice for 45 min before being resolved by 1.4% agarose gel in the absence of Mg^{2+}. Note for (B), exposure of second panel was adjusted to compensate for reduced biotinylation of GB2.

3.5.7. *Pf*TLP binds preferentially to the minor groove of A/T-rich DNA

To determine if *Pf*TLP sequence-selective DNA binding to A/T-rich DNA is also mediated by minor groove interactions, an EMSA competition assay was performed to compare the affinity of *Pf*TLP for GBP-130, poly(dI-dC) and poly(dG-dC) unlabelled competitors. As described in Section 3.5.5., this assay allows determination of the contributions of minor and major grooves for protein-DNA interactions in the context of A/T-nucleotide sequences.

As seen in Fig. 29, *Pf*TLP bound poly(dI-dC) with much higher affinity than poly(dG-dC) (compare lanes 8-11 with 13-16) and with roughly equivalent affinity to GBP-130 promoter DNA (compare lanes 3-6 with 8-11). These results clearly demonstrate that *Pf*TLP interactions with the DNA minor groove drive sequence-selective binding of *Pf*TLP to A/T-rich DNA, similar to *Pf*TBP-DNA interactions (Fig. 26).

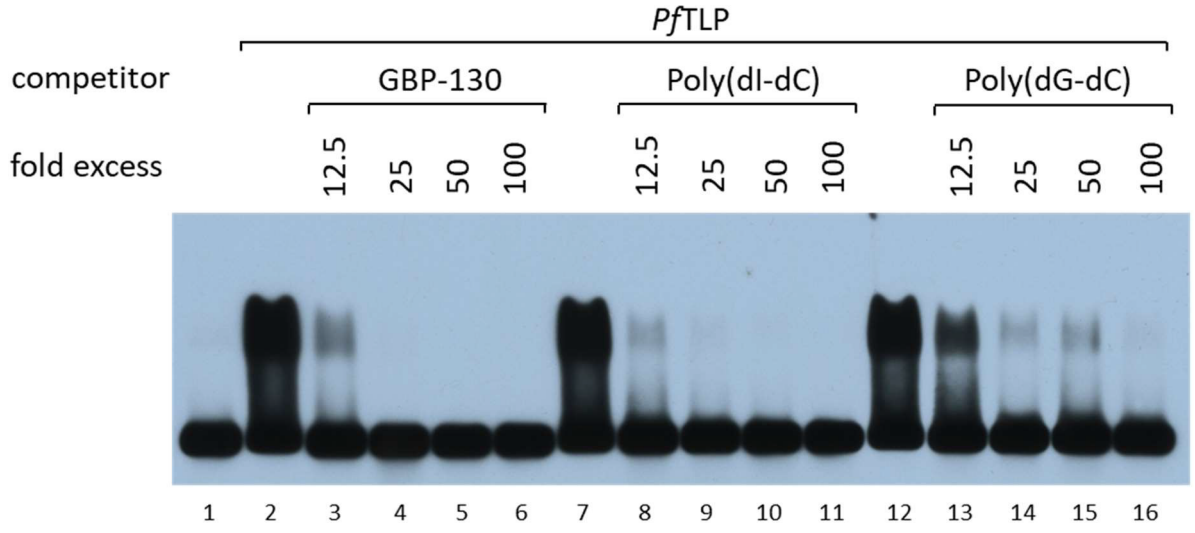


Figure 29. *Pf*TLP binds preferentially to the minor groove of A/T-rich DNA. DNA-binding reactions (10 µl) contained 5 fmol biotinylated 282 bp GBP-130 DNA probe, 3 nM *Pf*TLP and increasing fold excess unlabelled competitor as indicated. Fold excess poly(dG-dC) and poly(dI-dC) were calculated in molar nucleotides. DNA-binding reactions were incubated on ice for 45 min before being resolved by 1.4% agarose gel in the absence of Mg^{2+}.

3.5.8. Evolutionarily conserved phenylalanine residues are not required for *Pf*TLP-DNA interactions but impact *Pf*TLP-DNA complex mobility during electrophoresis

As described in Section 3.1.2., the four phenylalanine residues, involved in the severe distortion of the TATA box DNA, are strictly conserved amongst prototypical TATA box-binding proteins. In higher eukaryotic TBP family proteins, mutations of phenylalanine residues corresponding to F214, F288 or F305 in *Hs*TBP decrease TBP-DNA interactions in DNA-binding assays to undetectable levels (Reddy and Hahn, 1991; Klejman *et al.*, 2005). To investigate the role of the corresponding phenylalanine residues in *Pf*TLP, a panel of mutants, F60A, F283A and F60AF283A were previously generated in our laboratory (Adebolajo, 2016; van der Linden, 2019). F60 and F283 of *Pf*TLP correspond to the two phenylalanine residues which in prototypical TBP-TATA interactions intercalate between the first and second base pairs (F283) and second last and last base pairs (F60) of the TATA box sequence (Nikolov *et al.*, 1996; Patikoglou *et al.*, 1999).

EMSA experiments were carried out to compare DNA binding of wildtype *Pf*TLP and F-to-A mutants to the GBP-130 promoter probe to which sequence-selective binding has been established (Section 3.5.6.). As shown in Fig. 30A, addition of increasing concentrations of *Pf*TLP F60, F283 and the F60F283 double mutant resulted in concentration-dependent disappearance of free probe and concomitant formation of a protein-DNA complex. Interestingly, the proportion of bound DNA was consistent between *Pf*TLP WT and mutants at comparable *Pf*TLP concentrations, indicating that F60 and F283 are not required for *Pf*TLP-DNA binding. However, the electrophoretic mobility of *Pf*TLP F-A mutants was markedly increased relative to the wild type *Pf*TLP-DNA complex. This increase in mobility could be attributed to a reduced number of *Pf*TLP mutant proteins binding to the DNA or to differences in the conformation of nucleoprotein complexes, for example reduced DNA bending by *Pf*TLP F-A mutant proteins.

To further investigate differences in the mobility of nucleoprotein complexes formed by wild type *Pf*TLP and *Pf*TLP F-to-A mutants, and to limit the number of potential *Pf*TLP interaction sites, experiments with a shorter DNA probe, consisting of 60 bp of putative KHARP promoter region (KHARP 60 bp), were carried out. As seen in Fig. 30B, *Pf*TLP WT binds this KAHRP DNA probe in a concentration-dependent manner without formation of well-defined low mobility complexes, suggesting that *Pf*TLP-DNA interactions with the 60 bp KHARP DNA probe are unstable during electrophoresis. This instability is consistent with the binding observed on other A/T-rich short probes (Fig.28B). Comparison of *Pf*TLP WT and F60A, F283A and F60AF283A binding activity showed that all proteins bound the 60 bp KAHRP probe with comparable affinity, indicated by the equivalent disappearance of the free probe, once again suggesting that F60 and F283

are not required for *Pf*TLP DNA binding. However, unlike with the *Pf*TLP wildtype protein, DNA-protein complexes formed by all three of the F-A mutants did not dissociate significantly during electrophoresis and formed a prominent band of significantly greater mobility compared to the unstable complexes formed by wildtype *Pf*TLP. Given the short length of the 60 bp KAHRP probe it appears less likely (although not impossible) that this difference in mobility between wildtype *Pf*TLP-DNA complexes and DNA complexes formed with F-A *Pf*TLP mutants is caused by differences in the number of molecules bound to the DNA probe. Instead, these results support a model in which the *Pf*TLP F60 and F283 residues are required to induce a conformational change in the *Pf*TLP nucleoprotein complex that in turn causes a lower mobility and, at the same time, a greater rate of complex dissociation during electrophoresis.

Temperature has been shown to impact the degree of DNA bending of the TATA box by human and yeast TBP (Delgadillo *et al.*, 2009), as well as the stability of the bent TBP-TATA complex formed by archaeal TBPs (Gietl *et al.,* 2014). In order to assess whether the temperature of DNA-binding reactions affects the electrophoretic mobility of *Pf*TLP nucleoprotein complexes, DNA-binding reactions were incubated at 27°C (Fig. 31) instead on ice (Fig. 30) prior to agarose gel electrophoreses. As shown in Fig. 31, incubation of DNA-binding reactions at 27°C had little effect on overall DNA-binding activity. However, the mobility of *Pf*TLP-DNA complexes relative to free DNA was significantly altered compared to binding reactions incubated on ice (Fig. 30B). When incubated at 27°C, *Pf*TLP F283-DNA complexes showed similar mobility to *Pf*TLP WT (Fig. 31, compare lanes 2-4 with 10-12). Additionally, after incubation at 27°C, the F60A and F60AF283A *Pf*TLP double mutants appeared to form multiple distinct protein-DNA complexes with distinct electrophoretic mobility (Fig. 31, compare lanes 6-8 and 14-16 with lanes 2-4). These results suggest that higher temperatures during DNA binding may at least partially compensate for the impact of phenylalanine mutations on *Pf*TLP-DNA nucleoprotein complex structure and, consequently, mobility. This compensatory effect may be due to increased flexibility of DNA at higher temperatures, which in turn may decrease the energy required for DNA deformation and thus the formation of a bent *Pf*TLP-DNA complex.

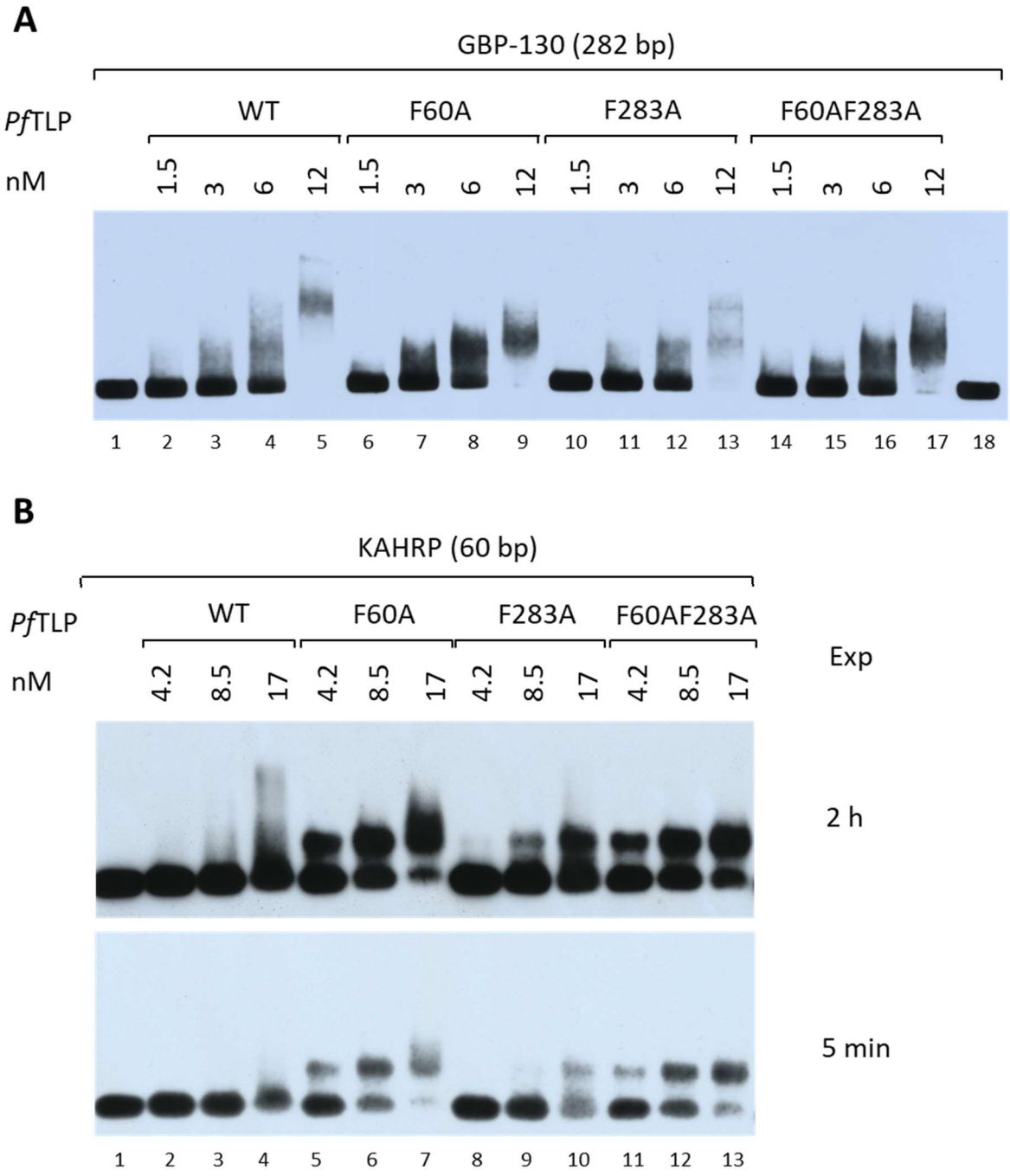


Figure 30. Mutation of conserved phenylalanine residues does not impact *Pf*TLP affinity but alters *Pf*TLP-GBP-130 complex mobility. DNA-binding reactions (10 µl) contained 5 fmol biotinylated 282 bp GBP-130 **(A)** or 60 bp KAHRP **(B)** in the presence of 10 ng poly(dG-dC) **(A)** or 5 ng poly(dG-dC) background competitor **(B),** 120 mM KCl **(A)** or 180 mM KCl **(B)** and increasing concentrations of *Pf*TLP WT, F60A, F283A or F60AF283A as indicated. DNA binding reactions were incubated on ice for 45 min before being resolved by 1.4% agarose gel in the absence of Mg^{2+}. Exposure times indicated on the left.

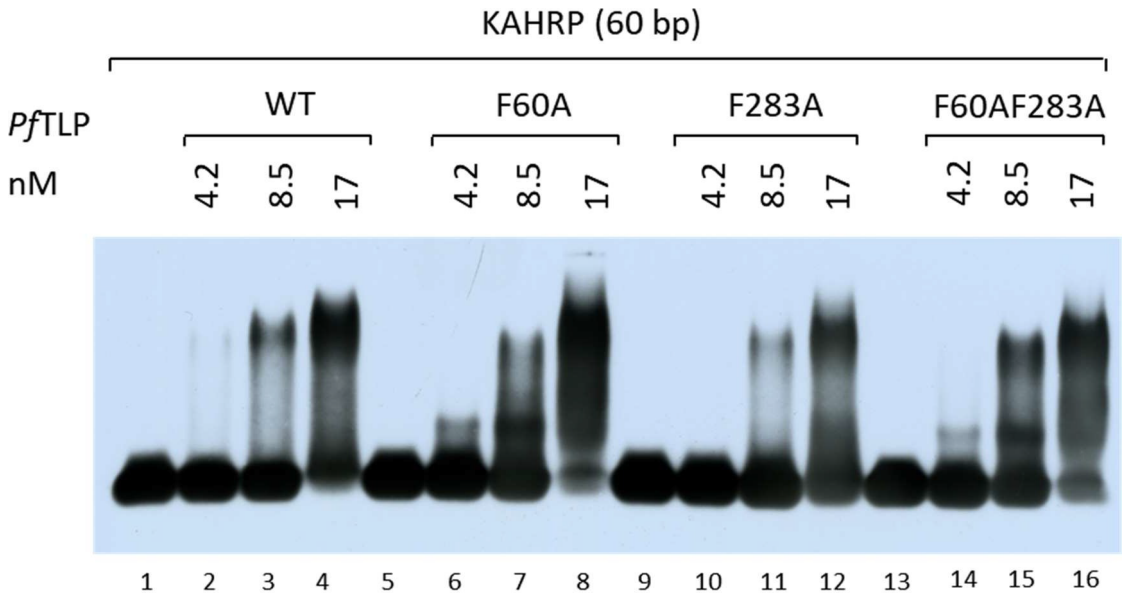


Figure 31. Incubation at 27°C partially compensates for the impact of F60A and F283A mutations on *Pf*TLP-DNA complex mobility. DNA-binding reactions (10 µl) contained 5 fmol biotinylated 60 bp KAHRP, indicated concentrations of *Pf*TLP WT, F60A, F283A or F60AF283A and 180 mM KCl, except for lanes 2, 6, 10 and 14, which contained 120 mM KCl. DNA-binding reactions were incubated at 27°C for 45 min before being resolved by 1.4% agarose gel in the absence of Mg^{2+}.

Chapter 4: Results Part 2

Functional characterisation of *P. falciparum* TFIIB

Previous work by our group had shown that *Pf*TFIIB possesses TBP-independent DNA-binding activity (Bing, 2015; Talvik, 2016). Since neither *Pf*TBP nor *Pf*TLP display selective DNA-interactions beyond a broad preference for A/T-rich DNA (Section 3.5), other general transcription factors such as *Pf*TFIIB may play a more prominent role in promoter selection in *P. falciparum*. This section of work aimed to further characterise *Pf*TFIIB TBP-independent DNA interactions and to investigate interactions between *Pf*TFIIB and *Pf*TBP and *Pf*TFIIB and *Pf*TLP and the potential role of these interactions in the assembly of the RNA Pol II transcription initiation complex.

4.1. Expression of recombinant *Pf*TFIIB

4.1.1. Initial optimisation of *Pf*TFIIB expression

pET-11d derivative expression vectors for the recombinant expression of 6His-tagged *Pf*TFIIB and a panel of *Pf*TFIIB deletion mutants have been previously generated by members of our laboratory (Materials and Methods Section 2.2.1; Bing, 2015; Talvik, 2016; Eelu, 2016). Pilot expression of one of these deletion mutants, 6His:*Pf*TFIIBΔ140Δ350, consisting only of the *Pf*TFIIB structured core domain (Fig. 36), was carried out in *E.coli* BL21-DE3-RIL to compare soluble expression of 6His:*Pf*TFIIBΔ140Δ350 at a variety of conditions.

Previous work within our research group found that heterologous expression of *Pf*TFIIB inhibits bacterial growth and demonstrated that higher levels of *Pf*TFIIB expression could be observed from uninduced cultures compared to induced bacterial cultures (Talvik, 2016). Therefore, expression of the *Pf*TFIIB core domain was first examined in the absence of the inducer IPTG. Such 'leaky' expression of *Pf*TFIIB can occur due to basal expression of T7 polymerase under control of the LacUV5 promoter (Talvik, 2016). However, no detectable expression of *Pf*TFIIBΔ140Δ350 was observed in cultures grown in the absence of inducer (Fig. 32, condition i).

Next, *Pf*TFIIB expression was induced by addition of IPTG at late log phase to maximise cell numbers before induction and to compensate for the detrimental effect of *Pf*TFIIB expression on *E.coli* growth. In addition to increasing cell number, induction of protein expression at late log phase has also been shown to increase soluble expression of some proteins, presumably by slowing down the rate of protein production (Galloway, Sowden and Smith, 2003). In this protocol, the bacterial culture was subjected to cold shock by rapid cooling in an ice bath at an OD_{600} of 1, to induce the

expression of cold shock proteins (Phadtare, Alsina and Inouye, 1999), before addition of IPTG inducer and incubation at 30°C for 3 h. Under these conditions soluble *Pf*TFIIBΔ140Δ350 could be purified from cell lysates (Fig. 32, condition ii).

Further pilot experiments were carried out to test additional conditions. These included induction at OD_{600} 1.7 with 1 mM IPTG followed by 1 h expression at 37°C (Fig. 32, condition iii) and induction at OD_{600} 1.7 with 1 mM IPTG followed by cold shock and overnight expression at 17°C (Fig. 32, condition iv). Lower temperatures during protein expression in bacteria have been shown to increase protein folding (Rosano and Ceccarelli, 2014). It was therefore hypothesised that overnight incubation at 17°C may result in higher yields of soluble *Pf*TFIIBΔ140Δ350 by decreasing protein misfolding. While expression using conditions iii and iv produced detectable levels of *Pf*TFIIBΔ140Δ350 (Fig. 32, lanes 3-6), maximum soluble expression was achieved using condition ii (Fig. 32), which included cold shock at 1 OD_{600} prior to addition of 1 mM IPTG inducer, followed by 3 h protein expression at 30 ̊C. This expression protocol was used for further large scale expression of 6His:*Pf*TFIIB and 6His:*Pf*TFIIB deletion mutants.

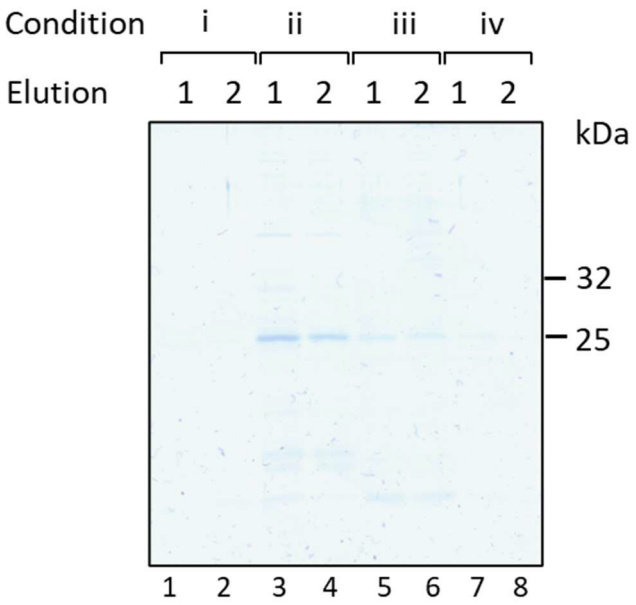


Fig. 32. Expression of 6His:*Pf*TFIIBΔ140Δ350 using different conditions.
200 ml of LB were inoculated with 10 ml saturated overnight culture (1:20) and protein expression was carried out using the following conditions: i) uninduced cultures were incubated at 37 °C for 5 h, ii) cold shock at OD_{600} = 1.0 followed by addition of 0.5 mM IPTG and 3 h protein expression at 30 ̊C, iii) induction at OD_{600} = 1.2 by addition of 0.5 mM IPTG, followed by 1 h protein expression at 37 °C, iv) induction at OD_{600} = 1.2 by addition of 0.5 mM IPTG, followed by 16 h expression at 20°C. 1 ml of bacterial cleared lysate was purified on TALON-IMAC resin using a small scale batch procedure (Materials and Methods 2.2.2.). Equivalent volumes of the first two elution fractions were analysed by 12% SDS-PAGE and Coomassie staining.

4.1.2. Large scale expression and purification of 6His:*Pf*TFIIB

Following expression as described in Section 4.1.1., selective lysis of the bacterial periplasm was achieved by adoption of an osmotic shock protocol established to increase the yield of metal-affinity purification of low abundance proteins expressed in *E.coli* (Materials and Methods, Section 2.2.2; Magnusdottir *et al.*, 2009). Initially, recombinant bacterially expressed 6His:*Pf*TFIIB protein was purified by IMAC, followed by ion exchange chromatography on Q-Sepharose and SP-Sepharose resins (Fig. 33A). Further purification through Q-Sepharose significantly increased the purity of IMAC-purified 6His:*Pf*TFIIB (Fig. 33B, compare lanes 2 and 3). However, the purity of the 6His:*Pf*TFIIB preparations were not significantly increased by the additional chromatography step on SP-Sepharose ion exchange resin (Fig. 33B, compare lanes 3 and 4). For this reason, subsequent large scale expression and purification of 6His:*Pf*TFIIB and 6His:*Pf*TFIIB deletion mutant variants involved only IMAC and Q-Sepharose chromatography steps.

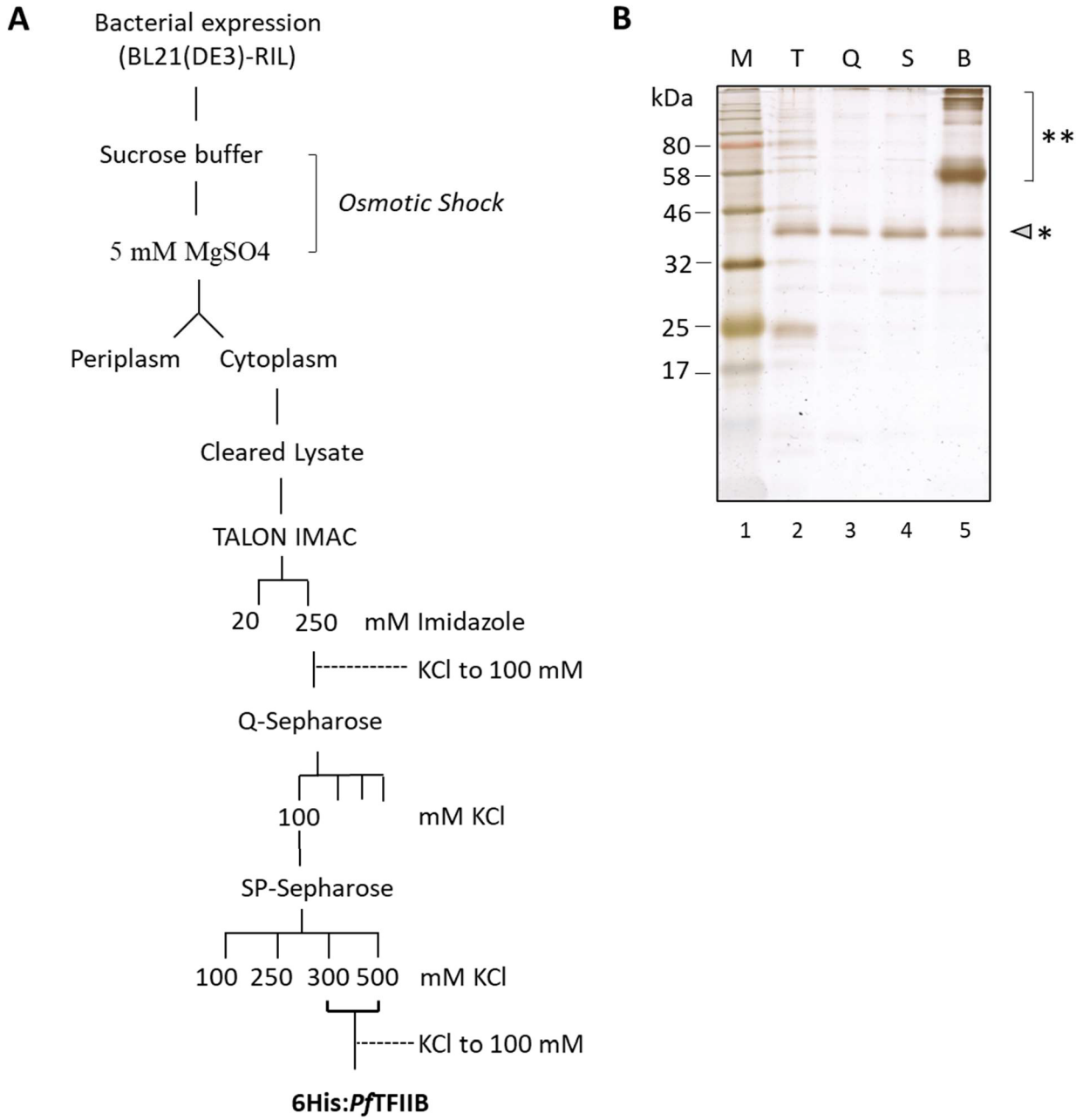


Figure 33. Purification of 6His:*Pf*TFIIB. (A) Schematic illustrating purification of 6His:*Pf*TFIIB by sequential TALON IMAC, Q-Sepharose and SP-Sepharose chromatography steps. **(B)** Analysis of fractions containing 6His:*Pf*TFIIB (44.3 kDa) following different purification steps. Fractions containing equivalent 6His:*Pf*TFIIB amounts were resolved by 12% SDS-PAGE and visualised by silver stain. T, TALON IMAC eluate; Q, Q-Sepharose eluate; S, SP-Sepharose eluate; B, final 6His:*Pf*TFIIB preparation following buffer exchange and addition of BSA. Bands corresponding to 6His:*Pf*TFIIB (*) and added BSA (**) are indicated.

4.2. Purification and characterisation of anti-*Pf*TFIIB antibody

Polyclonal antibodies were raised against the peptide sequence LKQKYLSEDKRKKN, located in the C-terminal disordered tail of *Pf*TFIIB (aa residues 353-367) using a commercial service (*BioGenes*, Berlin, Germany) and further purified by antigen affinity chromatography and ammonium sulphate precipitation (Materials and Methods, Section 2.1.5.). The purified antibody recognised recombinantly expressed 6His:*Pf*TFIIB when assessed by immunoblot analysis (Fig. 34A) and specifically recognised expressed *Pf*TFIIB in *E.coli* extracts (Talvik, unpublished).

To investigate whether *Pf*TFIIB expression could be detected in *P. falciparum* trophozoites, nuclear and cytoplasmic *P. falciparum* extracts were resolved by 12% SDS-PAGE, transferred to a PVDF membrane and probed with the purified anti-*Pf*TFIIB antibody (Fig. 34B). Unfortunately, detection with the anti-*Pf*TFIIB antibody did not reveal any protein bands at the expected molecular mass distinct from bands seen with the pre-immune control (Fig. 34B, compare lanes 1 and 2 with lanes 4 and 5). This experiment was repeated multiple times with different *P. falciparum* extract samples as well as whole trophozoites (data not shown). In all experiments the anti-*Pf*TFIIB antibody failed to detect *Pf*TFIIB in the tested samples. Considering that expression of *Pf*TFIIB has been detected by mass-spectrometry during the intraerythrocytic stages (Treeck et al., 2011), these results suggest that either *Pf*TFIIB protein levels in the parasite lysates was below the threshold for detection in these assays, or, alternatively, that this anti-*Pf*TFIIB antibody is unable to recognise the natural form of *Pf*TFIIB expressed in *P. falciparum*.

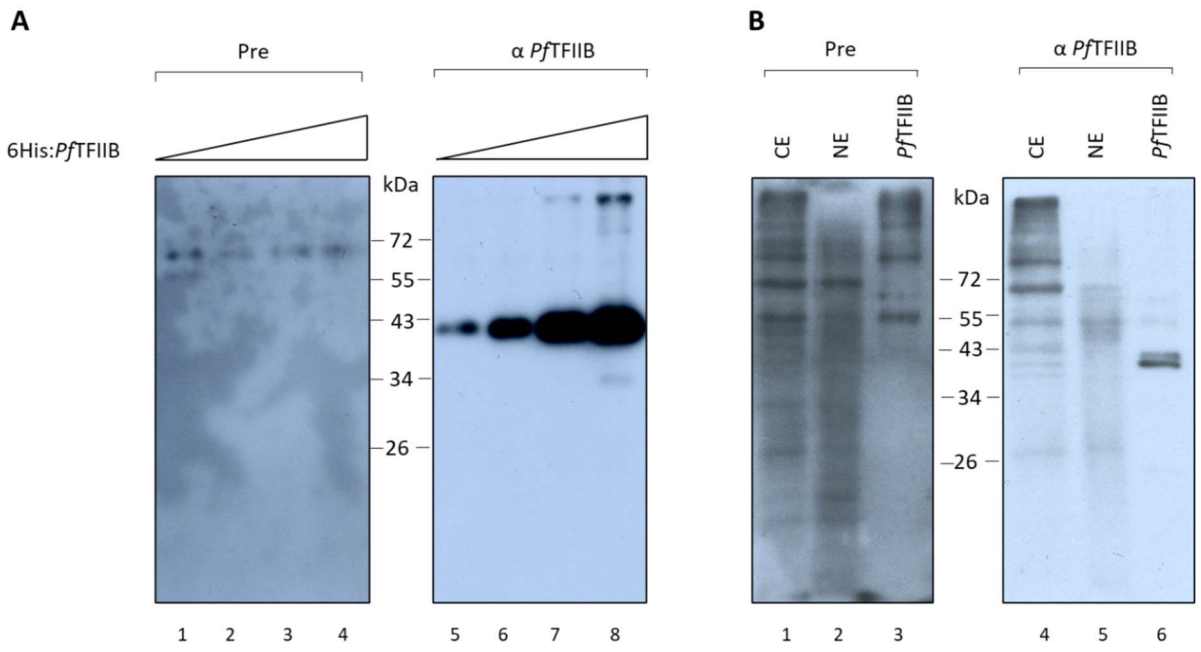


Figure 34. Immunoblot analysis of purified 6His:*Pf*TFIIB and *P. falciparum* extracts using anti-*Pf*TFIIB antibody. 2.5, 5, 10, or 20 ng 6His:*Pf*TFIIB (lanes 1-4 and 5-8) **(A)**, or 8 µg of cytoplasmic extract (CE) or 8 µg nuclear extract (NE) prepared from cultured blood-stage *P. falciparum* and 5 ng recombinant 6His:*Pf*TFIIB **(B)** were resolved by 12% PAGE, transferred to a PVDF membrane, and probed with affinity-purified anti-*Pf*TFIIB antibody or an equivalent amount of pre-immune serum as indicated.

4.3. TBP-independent DNA-binding activity of *Pf*TFIIB

4.3.1. Initial characterisation of *Pf*TFIIB TBP-independent DNA-binding

As a point of entry into the further characterisation of *Pf*TFIIB DNA-binding activity a series of EMSA experiments were carried out to determine whether *Pf*TFIIB-DNA interactions are a *Plasmodium*-specific feature or a general feature of TFIIB proteins under the given experimental conditions (i), to confirm that the observed DNA interaction is mediated by *Pf*TFIIB and not by a contaminant protein (ii), and to examine whether *Pf*TFIIB DNA interaction is sensitive to high ionic strength (iii).

To test whether *Pf*TFIIB-DNA interaction is a specific feature of *Pf*TFIIB, the DNA binding properties of 6His:*Pf*TFIIB and 6His:*Hs*TFIIB were compared by PAGE EMSA using a short 60 bp KHARP promoter DNA probe to which *Pf*TFIIB binding has previously been demonstrated (Talvik, 2016). As seen in Fig. 35A, incubation of the KAHRP DNA probe with increasing concentrations of *Pf*TFIIB resulted in a corresponding decrease in free probe mobility, demonstrating *Pf*TFIIB-DNA binding (Fig. 35A, lanes 5-7). In contrast, no DNA interaction was detected with equimolar amounts of *Hs*TFIIB, supporting the notion that the ability of *Pf*TFIIB to form stable complexes with DNA is a *Plasmodium*-specific feature (Fig. 35A, lanes 8-13).

Next, antibody experiments were performed in order to provide evidence that the nucleoprotein complex detected in EMSAs is *Pf*TFIIB dependent. To this end, *Pf*TFIIB protein was pre-incubated with purified anti-*Pf*TFIIB antibody before addition of DNA template and initiation of the DNA-binding reaction. This type of experiment can have one of two different outcomes. First, nucleoprotein complex formation may be inhibited if the DNA interaction site is occluded by binding of the antibody. Alternatively, antibody binding to the protein-DNA complex may cause a 'supershift' – a further decrease in the electrophoretic mobility of the nucleoprotein complex. As shown in Fig. 35B, pre-incubation of 6His:*Pf*TFIIB with 50 ng or 100 ng anti-*Pf*TFIIB antibody inhibited complex formation to near undetectable levels (Fig. 35B, compare lane 2 with lanes 3 and 4). In contrast, nucleoprotein complex formation was not inhibited by identical amounts of affinity-purified anti-NC2 (Fig. 35B, lanes 5-6) or anti-*Pf*TBP control antibodies (Fig. S5, compare lanes 3 and 4 with lanes 5 and 6). Finally, anti-*Pf*TFIIB antibody itself did not interact with the KAHRP probe (Fig. 32B, lane 7). Together, these results provide strong evidence that the observed nucleoprotein complex is formed by 6His:*Pf*TFIIB.

To examine the contribution of electrostatic interactions to *Pf*TFIIB-DNA interactions, and to determine if *Pf*TFIIB-DNA interactions can be observed at physiological salt concentrations

(~150 mM KCl), 6His:*Pf*TFIIB was incubated with 60 bp KAHRP DNA in the presence of increasing KCl concentrations. As shown in Fig. 35C, increasing the concentration of KCl ranging from 50 mM to 400 mM only moderately decreased *Pf*TFIIB-KAHRP complex formation (Fig. 35C, compare lanes 2-5), suggesting that the observed *Pf*TFIIB-DNA interactions extend beyond electrostatic interactions between *Pf*TFIIB and the negatively charged DNA phosphate backbone.

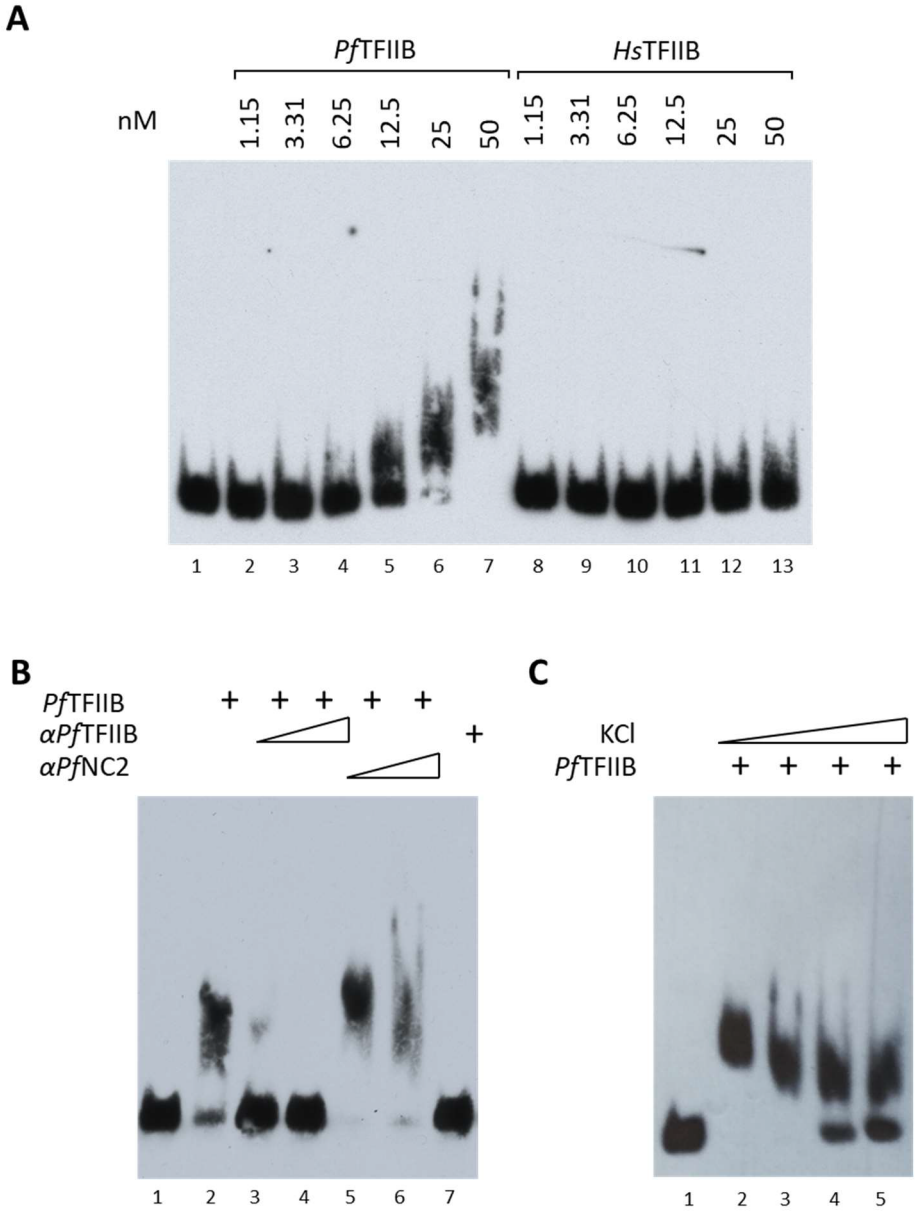


Figure 35. *Pf*TFIIB binds DNA independently of *Pf*TBP.
DNA-binding reactions (10 µl) contained 5 fmol 60 bp biotinylated KAHRP DNA probe and **(A)** DNA-binding with indicated concentrations of *Pf*TFIIB or *Hs*TFIIB, **(B)** 23 nM *Pf*TFIIB and 50 ng (lanes 3 and 5) or 100 ng (lanes 4 and 6) anti-*Pf*TFIIB or anti-NC2 antibody as indicated, or 100 ng anti-*Pf*TFIIB antibody in the absence of *Pf*TFIIB protein (lane 7), **(C)** 15 nM *Pf*TFIIB and 50, 100, 200, or 400 mM KCl (lanes 2-5). For the antibody experiments shown in **(B)** 6His:*Pf*TFIIB and antibody were preincubated for 30 min before addition of DNA and incubation of the DNA-binding reaction on ice for 45 min. Binding reactions were resolved by 5% (37:1) PAGE in the presence of 2 mM Mg^{2+}.

4.3.2. The *Pf*TFIIB structured core region, comprised of two cyclin-like domains, is sufficient for stable *Pf*TFIIB-DNA complex formation

Previous work in the laboratory showed that *Pf*TFIIB-DNA binding is significantly reduced upon deletion of a *Plasmodium*-specific basic amino acid sequence at the extreme N-terminus (aa residues 1-37; Talvik, 2016). However, *Pf*TFIIB also contains a *Plasmodium*-specific basic amino acid sequence at its C-terminal end (aa residues 350-367). To further delineate the role of intrinsically disordered *Pf*TFIIB N- and C-terminal basic regions and the structured cyclin-like core domain in *Pf*TFIIB-DNA interactions, a panel of 6His:*Pf*TFIIB deletion mutants (Fig. 36B) was recombinantly expressed and purified by IMAC and Q-Sepharose chromatography using the protocol established for full-length 6His:*Pf*TFIIB. These mutants included 6His:*Pf*TFIIBΔ37 (i), which lacks the basic N-terminus consisting of amino acid residues 1-37, 6His:*Pf*TFIIBΔ140 (ii), which lacks the mostly disordered N-terminal region (aa residues 1-140), (iii) 6His:*Pf*TFIIBΔ350, which lacks the highly basic *Plasmodium*-specific C-terminal region consisting of aa 350-367 (Bing, 2015; Talvik, 2016), and 6His:*Pf*TFIIBΔ140Δ350 (iv), consisting only of the minimal structured core region comprised of two cyclin-like domains. Purified full-length 6His:*Pf*TFIIB and mutant proteins were analysed by SDS-PAGE (Fig. 36C) and normalised by immunoblot using an anti-His antibody (Fig. 36D; Materials and Methods, Section 2.1.9).

Next, the DNA-binding activity of the panel of 6His:*Pf*TFIIB mutants was assessed by EMSA using the 60 bp KAHRP promoter DNA probe, previously shown to be bound by full-length *Pf*TFIIB (Fig. 35). As shown in Fig. 37, *Pf*TFIIB-DNA complexes were not well resolved in these experiments and full-length *Pf*TFIIB-DNA binding was detected by the appearance of a band of slightly lower mobility than free probe (Fig. 37A and B, lanes 2 and 3). At higher concentrations this band of slightly reduced mobility is replaced by a diffuse lower mobility complex (Fig. 37A and B, lanes 4 and 5), suggesting multiple binding of full-length *Pf*TFIIB to the DNA probe. Because *Pf*TFIIB-DNA complexes were not well resolved, DNA binding of full-length *Pf*TFIIB and deletion mutants was deduced from the reduction of the signal for the free probe rather than from the increase in *Pf*TFIIB-DNA complex intensity.

Comparison of the loss of free probe at equivalent concentrations of full-length *Pf*TFIIB and *Pf*TFIIBΔ37 and *Pf*TFIIBΔ140 suggested that both 6His:*Pf*TFIIBΔ37 and 6His:*Pf*TFIIBΔ140 had reduced DNA-binding activity compared to full-length *Pf*TFIIB (Fig. 37A). While only 3 nM full-length *Pf*TFIIB was required to bind approximately 50% of the free probe, 25 nM *Pf*TFIIBΔ37 in the DNA binding reaction resulted in very little decrease in the signal for free probe. Furthermore, a 8-fold higher concentration of 6His:*Pf*TFIIBΔ140 was required for detection of DNA-binding activity comparable to full-length 6His:*Pf*TFIIB (Fig. 37A, compare lanes 2-5 with lanes 6-10 and lanes 11-15). Thus the first

N-terminal 37 amino acid residues significantly contribute to overall *Pf*TFIIB DNA-binding activity, consistent with previous observations (Talvik, 2016).

While deletion of the *P. falciparum*-specific basic N-terminal region (aa 1-37) greatly reduced *Pf*TFIIB-DNA binding, deletion of the C-terminal *P. falciparum*-specific basic region (aa 350-367) decreased *Pf*TFIIB-DNA binding only slightly. Specifically, an approximately two-fold reduction in DNA-binding activity was observed for 6His:*Pf*TFIIBΔ350 relative to full-length 6His:*Pf*TFIIB, supporting that the basic C-terminus (aa 350-367) contributes little to overall *Pf*TFIIB DNA-binding affinity (Fig. 37B, compare lanes 2-5 with lanes 6-9). However, DNA-protein complexes formed with 6His:*Pf*TFIIBΔ350 complexes appeared less stable during electrophoresis compared to complexes formed with wildtype *Pf*TFIIB and could not be resolved (Fig. 37B, compare lanes 2-5 with lanes 6-10), suggesting a possible role for the basic C-terminal IDR in stabilising the *Pf*TFIIB-DNA complex.

Interestingly, 6His:*Pf*TFIIBΔ140Δ350, consisting of only the *Pf*TFIIB cyclin-like domains, bound DNA with similar affinity to 6His:*Pf*TFIIBΔ350 and with only slightly lower affinity compared to full-length 6His:*Pf*TFIIB (Fig. 37B, compare lanes 2-5 and 6-10 with lanes 11-15). This result suggests that the *Pf*TFIIB C-terminal core structured region, comprised of the two cyclin-like domains, is largely responsible for *Pf*TFIIB DNA-binding activity. The result further suggests that the N-terminus of *Pf*TFIIB may not play a direct role in *Pf*TFIIB DNA-interactions, as previously suggested (Talvik, 2016). Instead, the reduced affinity of *Pf*TFIIBΔ37 relative to full-length *Pf*TFIIB appears to arise from an inhibitory effect of the C-terminal basic region (aa 350-367) when the N-terminal basic patch (aa 1-37) is absent.

Together, the results shown in Fig. 37 demonstrate that the structured C-terminal core region of *Pf*TFIIB containing cyclin-like domains has intrinsic DNA-binding activity that may be modulated by the *Plasmodium*-specific N- and C-terminal basic regions.

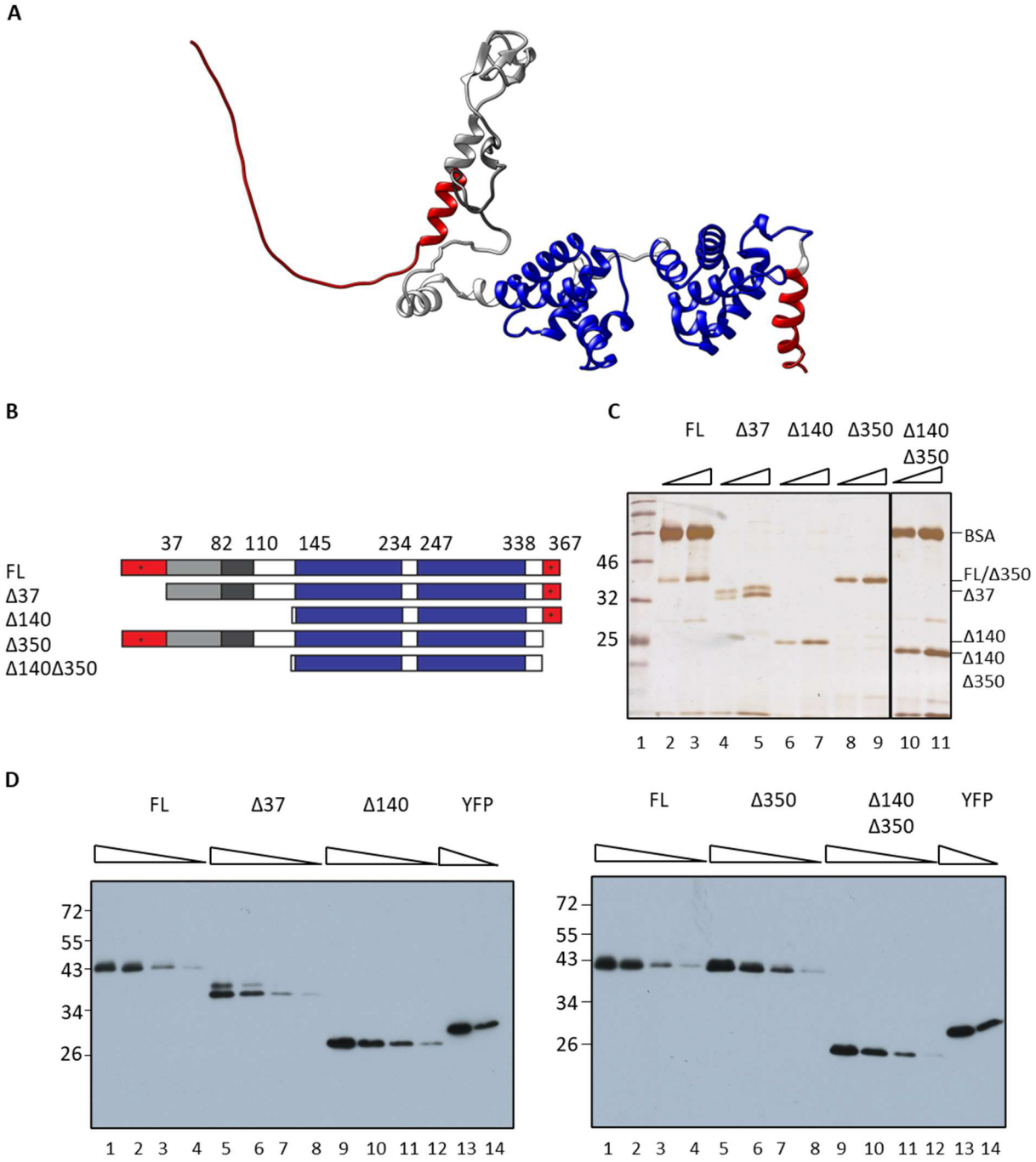


Figure 36. *Pf*TFIIB constructs used in this study. (A) Cartoon representation of the AlphaFold predicted *Pf*TFIIB structure. *Plasmodium*-specific basic N- and C-terminal regions are coloured red, the region corresponding to the B-ribbon region of *Hs*TFIIB is coloured light grey, the region corresponding to B-reader region of *Hs*TFIIB is coloured dark grey, the structured TFIIB core region comprising two cyclin-like domains is coloured blue. **(B)** Schematic of *Pf*TFIIB deletion mutant variants used in this study with regions coloured corresponding to panel **(A)**. **(C)** SDS-PAGE analysis of purified 6His:*Pf*TFIIB variants. 12.5 and 25 ng of each 6His:*Pf*TFIIB construct were resolved by 12% SDS-PAGE and visualised by silver staining. **(D)** Normalisation of 6His:*Pf*TFIIB constructs by immunoblot. Molar equivalent amounts of 6His:*Pf*TFIIB constructs were resolved by 12% SDS-PAGE, transferred to a PVDF membrane and probed with mouse anti-His antibody. FL, full-length 6His:*Pf*TFIIB; Δ37, 6His:*Pf*TFIIB lacking aa residues 1-37; Δ140, 6His:*Pf*TFIIB lacking aa residues 1-140; Δ350, 6His:*Pf*TFIIB lacking aa residues 350-367; Δ140Δ350 6His:*Pf*TFIIB consisting only of the structured core comprising cyclin-like domains.

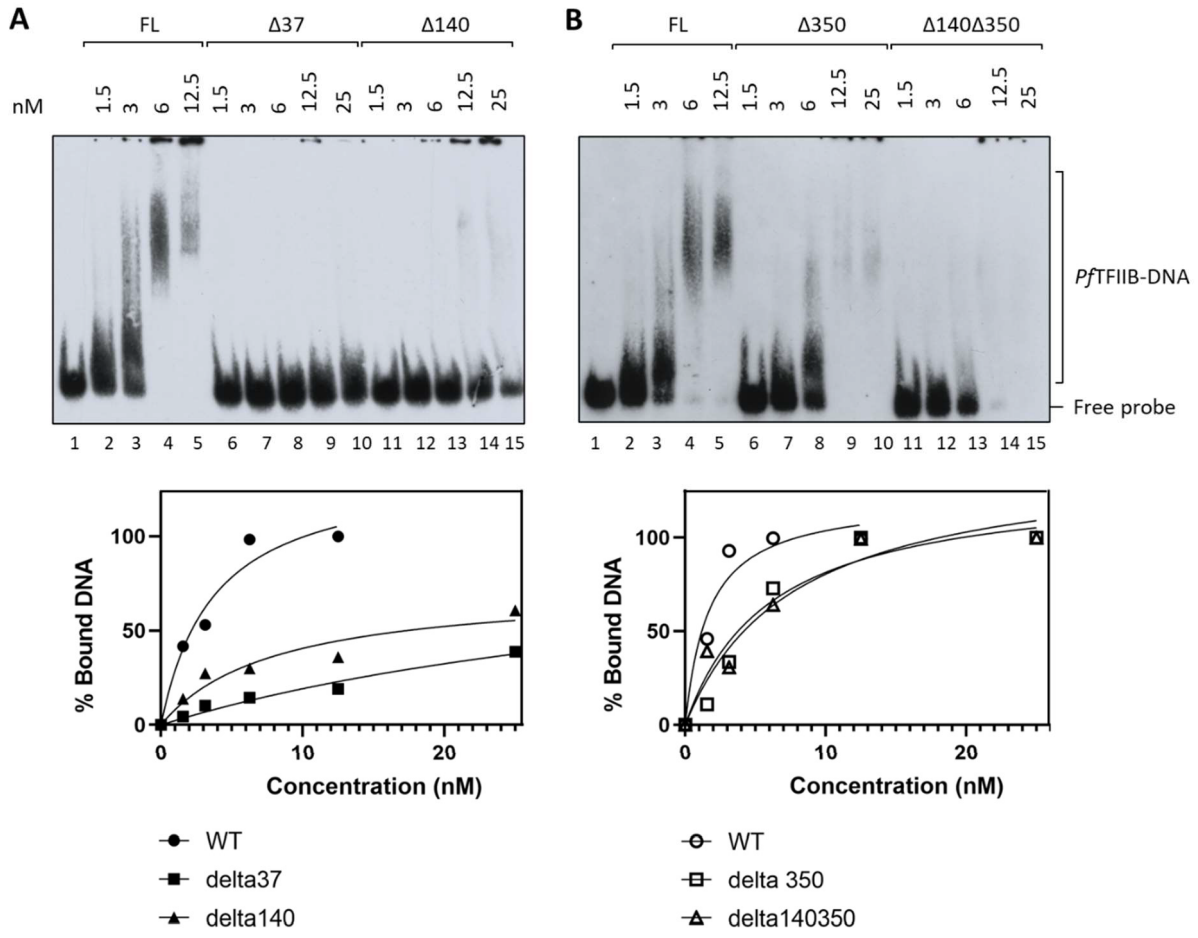


Figure 37. The *Pf*TFIIB structured core region comprising cyclin-like repeats is sufficient for 6His:*Pf*TFIIB-DNA interactions. DNA-binding reactions (10 µl) contained 5 *f*mol 60 bp biotinylated KAHRP DNA probe with indicated concentrations of full-length 6His:*Pf*TFIIB (FL) or 6His:*Pf*TFIIB deletion mutant variants. DNA-binding reactions were incubated on ice for 45 min and resolved by 5% (37:1) PAGE in the presence of 2 mM Mg^{2+}. DNA interactions were quantified by densitometry using loss of free-probe as a measure of DNA-binding activity. Non-linear regression and data plotting were performed using Graphpad Prism 9.5.1 for Windows, (GraphPad Software, San Diego, California USA, www.graphpad.com). FL, full-length 6His:*Pf*TFIIB; Δ37, 6His:*Pf*TFIIB lacking aa residues 1-37; Δ140, 6His:*Pf*TFIIB lacking aa residues 1-140; Δ350, 6His:*Pf*TFIIB lacking aa residues 350-367; Δ140Δ350 6His:*Pf*TFIIB consisting only of the minimal structured core comprising of cyclin-like domains.

4.3.3. *Pf*TFIIB does not recognise BREu or BREd core promoter sequence elements

*Hs*TFIIB has been shown to specifically interact with specific DNA sequences, designated B-recognition elements (BREs) located up- (BREu) and downstream (BREd) of the TATA sequence in TATA-containing promoters (Lagrange *et al.*, 1998; Deng and Roberts, 2005). To test if *Pf*TFIIB displays sequence selectivity towards these DNA sequence elements EMSAs were carried out designed to determine the relative affinity of *Pf*TFIIB for a 60 bp DNA probe containing the adenovirus 2 major late (Ad2ML) promoter compared to 60 bp A/T-rich putative *P. falciparum* promoter DNA regions, the KAHRP promoter (comprising positions -15 to -74 of the mapped TSS (Lanzer 1992)) and the *Pf*EMP1 promoter (Materials and Methods, section 2.2.3.). The Ad2ML promoter contains well-characterised near perfect consensus BREu (located upstream of TATA) and BREd (located downstream of TATA) sequence elements (Lagrange *et al.*, 1998; Deng and Roberts, 2005), that are absent in the KHARP and *Pf*EMP1 promoter regions.

As shown in Fig. 38, *Pf*TFIIB bound all three 60 bp DNA probes with similar affinity, suggesting that BRE sequences present within the Ad2ML promoter probe do not significantly contribute to *Pf*TFIIB DNA binding. Given that these probes also have vastly different A/T nucleotide content (Ad2ML 38,3 %, *Pf*EMP1 80 %, KAHRP 86,7 %), this result also demonstrates that, unlike *Pf*TBP and *Pf*TLP, *Pf*TFIIB does not bind preferentially to A/T-rich sequences. Instead, these results suggest that *Pf*TFIIB binds DNA in a largely sequence-independent manner.

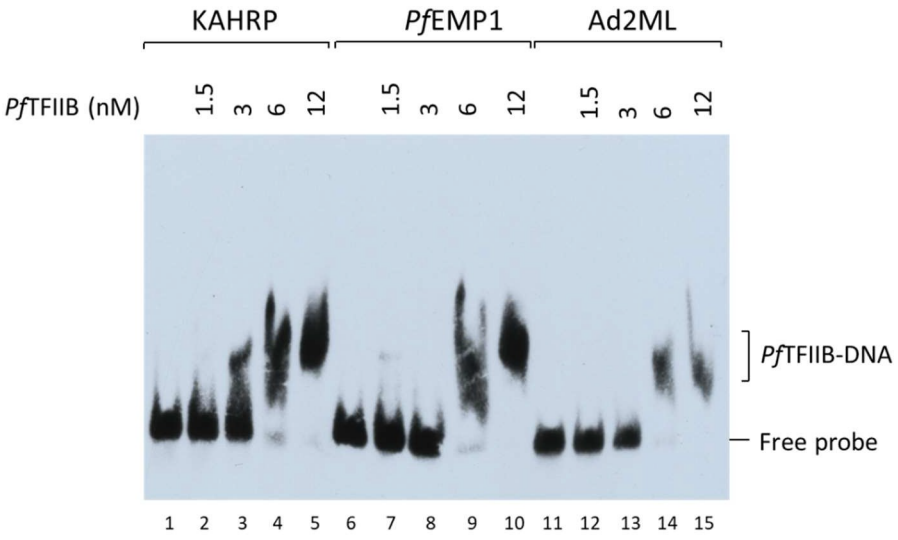


Figure 38. *Pf*TFIIB does not bind BRE sequences with high affinity.
DNA-binding reactions (10 µl) contained 5 fmol 60 bp biotinylated KAHRP, *Pf*EMP1 or Ad2ML DNA probe and indicated concentrations of *Pf*TFIIB DNA binding reactions were incubated on ice for 45 min, resolved by 5% (37:1) PAGE in the presence of 2 mM Mg^{2+}.

4.4. Intra- versus intermolecular interactions of *Pf*TFIIB

In well-studied eukaryotic systems, the TFIIB N-terminal B-finger domain and the second cyclin-like domain have been shown to engage in intramolecular interactions resulting in two TFIIB conformations that are in dynamic equilibrium. TFIIB is in either a 'closed conformation', where the B-finger domain and second cyclin-like domain are in contact, or an 'open conformation', where these regions are not in contact (Roberts and Green, 1994; Hayashi *et al.*, 1998; Bangur *et al.*, 1999; Wu and Hampsey, 1999; Hawkes, Evans and Roberts, 2000; Zhang *et al.*, 2000; Glossop, Dafforn and Roberts, 2004; Deng and Roberts, 2007). Importantly, these intramolecular interactions have been suggested to modulate, and to be modulated by interactions between TFIIB and transcriptional activators (Roberts and Green, 1994; Hawkes, Evans and Roberts, 2000; Zhang *et al.*, 2000). While these studies focused on intramolecular TFIIB interactions, their findings also raise the possibility of intermolecular interactions between TFIIB molecules, which could also play a role in transcriptional regulation. To examine intermolecular interactions of *Pf*TFIIB GST pull-down assays were performed (Fig. 39; Materials and Methods, Section 2.2.4.). Equivalent amounts of 6His:*Pf*TFIIB:GST and GST:6His control were immobilised on glutathione magnetic beads and incubated with 6His:*Pf*TFIIB. Following removal of unbound protein, bound proteins were eluted by addition of glutathione and analysed by immunoblot using an anti-His antibody. As shown in Fig. 39, 6His:*Pf*TFIIB bound to immobilised 6His:*Pf*TFIIB:GST, but not to immobilised GST:6His (Fig. 39, compare lanes 3 and 4). These results provide strong evidence for intermolecular interactions between different *Pf*TFIIB molecules under the conditions tested.

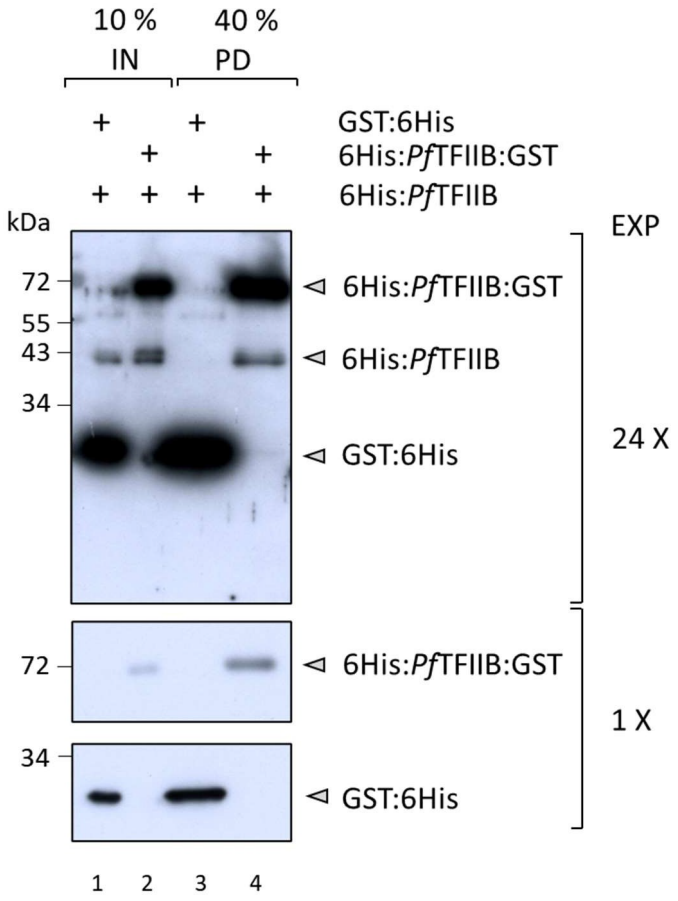


Figure 39. *Pf*TFIIB engages in intermolecular interactions. 250 nM GST:6His or 6His:*Pf*TFIIB:GST were immobilised on glutathione magnetic beads and incubated with 15 nM 6His:*Pf*TFIIB as indicated. Beads were washed twice with binding buffer to remove unbound proteins before elution with glutathione. 10% of the input (IN) or 40% of the eluted sample (PD) were resolved by 12% SDS-PAGE transferred to nitrocellulose membrane and probed with mouse anti-His antibody. Exposure times of individual panels are indicated. The position of protein MW standards is indicated on the left.

4.5. *Pf*TFIIB interacts with *Pf*TBP and *Pf*TLP in the absence of DNA

TFIIB-TBP interactions play an essential role in the assembly of the RNA Pol II transcription complex in higher eukaryotes. Previous *in silico* approaches studying *P. falciparum* TFIIB-TBP interactions have suggested that TBP amino acid residues responsible for interaction with TFIIB are highly conserved between *Hs*TBP and *Pf*TBP proteins (Buendía-Orozco, Guerrero and Pastor, 2005). This is consistent with the results of the bioinformatic analysis presented here, which noted a high degree of conservation between *Hs*TBP and *Pf*TBP and *Pf*TLP at amino acid residues predicted to play a key role in TBP-TFIIB interactions (Section 3.1.2). However, the TFIIB amino acid residues located at the TFIIB interface with TBP are less well conserved between *Hs*TFIIB and *Pf*TFIIB (Buendía-Orozco, Guerrero and Pastor, 2005). This moderate degree of divergence at the TFIIB-TBP interface raised the possibility of the existence of *Plasmodium*-specific interactions, which could be targeted by rational drug design (Buendía-Orozco, Guerrero and Pastor, 2005).

As a first step towards examining *Pf*TFIIB interactions with *Pf*TBP and *Pf*TLP, a GST-pulldown assay was performed. To this end, purified recombinant 26 kDa 6His:*Pf*TBP (Fig. 13), 6His:*Pf*TLP and 6His:*Hs*TBP were normalised by immunoblotting using anti-His antibody and assayed for interaction with 6His:*Pf*TFIIB:GST or GST:6His, immobilised on glutathione magnetic beads, using standard *in vitro* transcription conditions in the absence of DNA (Materials and Methods, Section 2.2.4.). In these assays, 6His:*Pf*TFIIB:GST interacted with both 6His:*Pf*TBP and 6His:*Pf*TLP to a comparable extent (Fig. 40, compare lane 4 with lane 8). In contrast, no binding was observed to the immobilised GST:6His control (Fig. 40, lanes 3 and 7).

Interestingly, *Hs*TBP was also found to interact with 6His:*Pf*TFIIB:GST (Fig. 40, lane 12), suggesting a degree of conservation in TBP-TFIIB interactions between *H. sapiens* and *P. falciparum*. However, given that residues located at the TBP interface with TFIIB are highly conserved between *Hs*TBP and *Pf*TBP and *Pf*TLP, this result is not entirely unexpected.

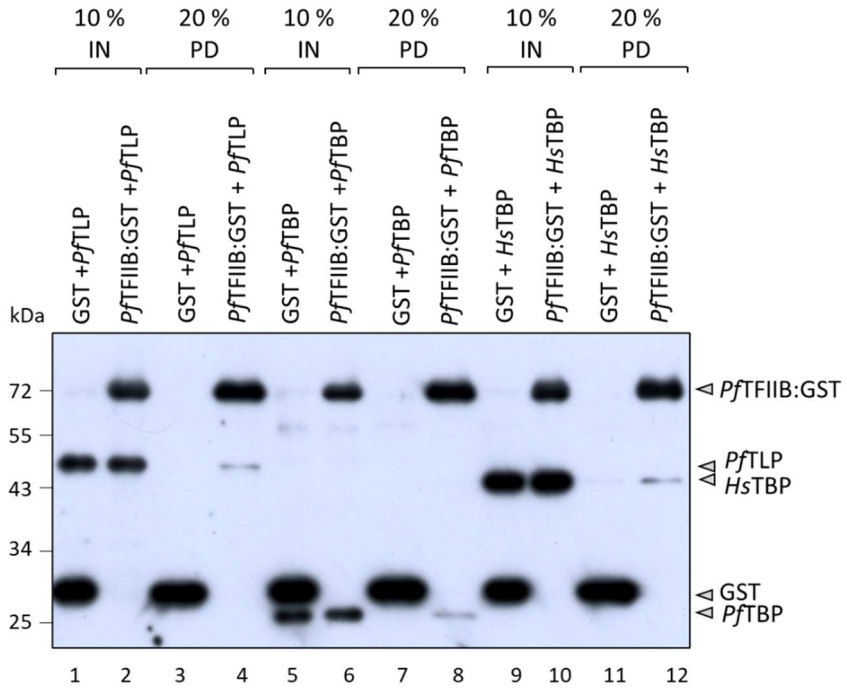


Figure 40*. Pf*TFIIB interacts with *Pf*TBP, *Pf*TLP and *Hs*TBP *in vitro*. 250 nM GST:6His (GST) or 6His:*Pf*TFIIB:GST (*Pf*TFIIB:GST) were immobilised on glutathione magnetic beads and incubated with 50 nM 6His:tagged *Pf*TLP, *Pf*TBP or *Hs*TBP as indicated. Immobilised proteins were washed twice with binding buffer to remove unbound protein before elution with glutathione. 10% of the input or 20% of the eluted sample were resolved by 12% SDS-PAGE, transferred to PVDF membrane and probed with mouse anti-His antibody. The position of protein molecular mass standards is indicated on the left.

4.6. *Pf*TFIIB stimulates DNA binding of both *Pf*TBP and *Pf*TLP

In higher eukaryotes, TFIIB has been shown to stabilise sequence-specific TBP-TATA interactions (Barberis *et al.*, 1993; Hisatake, Roeder and Horikoshi, 1993) and to contribute to promoter recognition through contacts with DNA up- and downstream of the TATA box (Lee and Hahn, 1995; Nikolov *et al.*, 1995). To test whether *Pf*TFIIB similarly stabilises or enhances sequence-specific *Pf*TBP- and/or *Pf*TLP-DNA interactions (Section 3.5), EMSA assays were performed (Fig. 41). In these assays the GBP-130 putative promoter DNA was chosen as labelled DNA probe as both *Pf*TBP and *Pf*TLP bind selectively to this A/T-rich *P. falciparum* promoter DNA (Sections 3.5.4 and 3.5.6). Of additional importance, *Pf*TFIIB intrinsic DNA-binding activity was suppressed by the presence of poly(dG-dC) competitor DNA. Under these conditions, *Pf*TFIIB significantly stimulated DNA binding by both *Pf*TBP and *Pf*TLP (Fig. 41, compare lanes 2-5 to lanes 7-10 in panel A and lanes 2-5 with lanes 6-9 in panel B). In contrast, *Pf*TFIIB did not bind DNA alone under these conditions (Fig. 41A lane 6 and Fig. 41B lane 10).

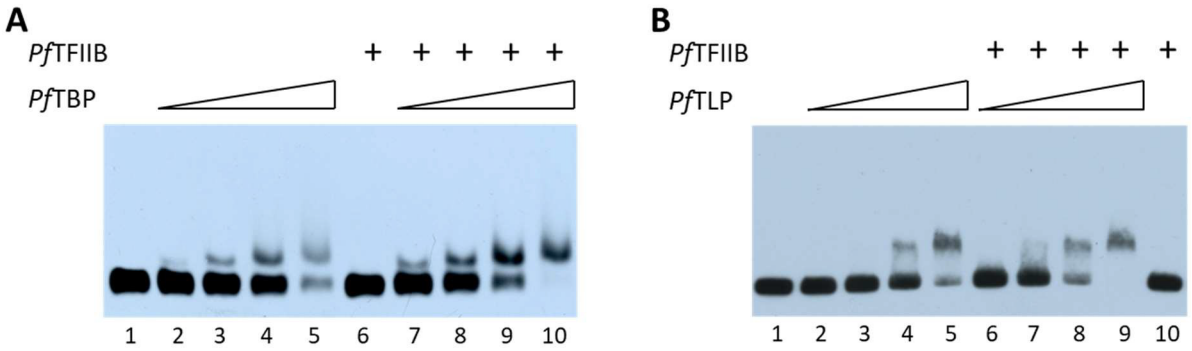


Figure 41. *Pf*TFIIB stimulates both *Pf*TBP and *Pf*TLP interactions with GBP-130 DNA.
DNA-binding reactions (10 µl) contained 5 fmol biotinylated 282 bp GBP-130 DNA probe, 30 ng poly(dG-dC), and two-fold increasing concentrations (30 – 240 nM) *Pf*TBP **(A)** or (1.75 -14 nM) *Pf*TLP **(B)** in the absence or presence of 10 nM 6His:*Pf*TFIIB (+). Reactions were incubated on ice for 45 min before being resolved by 1.4% agarose gel electrophoresis in the presence **(A)** or absence **(B)** of Mg^{2+}.

4.7. The *Pf*TFIIB structured core region stimulates *Pf*TLP-DNA interactions but not *Pf*TBP-DNA interactions

The TFIIB structured core region, comprised of two cyclin-like domains, has been shown to be sufficient for the stimulation of TATA box binding by prototypical TBP (Barberis *et al.*, 1993; Nikolov *et al.*, 1995). To delineate the regions of *Pf*TFIIB required for stimulation of *Pf*TBP and *Pf*TLP DNA binding, EMSAs were carried out using the panel of previously generated *Pf*TFIIB deletion mutants (Section 4.3.2.). In these assays, DNA was incubated with a fixed concentration of purified recombinant 26 kDa 6His:*Pf*TBP or 6His:*Pf*TLP in the absence or presence of equivalent molar concentrations of 6His:*Pf*TFIIB or 6His:tagged *Pf*TFIIB deletion mutant proteins. As shown in Fig. 42A, the *Pf*TFIIB deletion mutant lacking the mostly disordered N-terminal region (6His:*Pf*TFIIBΔ140) and the *Pf*TFIIB deletion mutant lacking the highly basic *Plasmodium*-specific C-terminal region (6His:*Pf*TFIIBΔ350) stimulated 6His:*Pf*TBP-DNA interactions to a lesser extent compared to full-length *Pf*TFIIB (Fig. 42A, compare lanes 3-5 with lanes 7-9 and lanes 11-13). Consistent with these results, the *Pf*TFIIB structured core region (6His:*Pf*TFIIBΔ140Δ350) did not significantly stimulate *Pf*TBP DNA binding activity (Fig. 42A, lanes 15-17). This result suggests an important role of both the unstructured N-terminal half of *Pf*TFIIB and the *P. falciparum*-specific basic C-terminal end of *Pf*TFIIB in modulating *Pf*TBP-DNA interactions.

In contrast, full-length *Pf*TFIIB (6His:*Pf*TFIIB) and the minimal *Pf*TFIIB structured core region (6His:*Pf*TFIIBΔ140Δ350) enhanced *Pf*TLP-DNA interactions in a comparable manner (Fig. 42B, compare lanes 3-5 with lanes 9-11), strongly suggesting that the structured *Pf*TFIIB structured core region, comprised of two cyclin-like domains, is sufficient for stimulation of *Pf*TLP-DNA interactions.

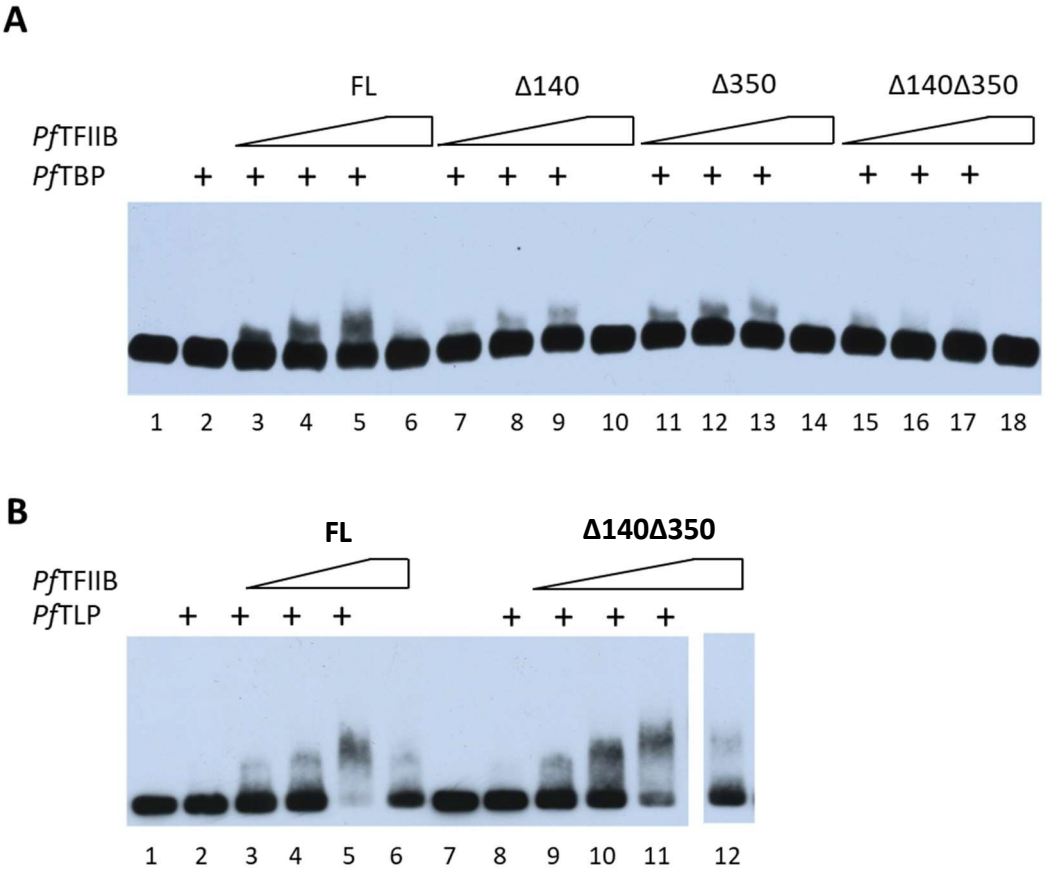


Figure 42. The structured *Pf*TFIIB structured core region is sufficient to stimulate DNA binding by *Pf*TLP but not *Pf*TBP. DNA-binding reactions (10 µl) contained 5 fmol biotinylated 282 bp GBP-130 DNA probe, 30 ng poly(dG-dC), and 60 nM 6His:*Pf*TBP **(A)** or 4 nM 6His:*Pf*TLP **(B)** with two-fold increasing concentrations (7.5-30 nM) of 6His:*Pf*TFIIB or 6His:tagged *Pf*TFIIB deletion mutants 6His:*Pf*TFIIBΔ140 (Δ140), 6His:*Pf*TFIIBΔ350 (Δ350), 6His:*Pf*TFIIBΔ140Δ350 (Δ140Δ350) as indicated. Proteins were incubated for 30 min on ice before addition of DNA and further incubation of the DNA-binding reaction on ice for 45 min. Protein-DNA complexes were resolved from free DNA by 1.4% agarose gel electrophoresis in the presence **(A)** or absence **(B)** of Mg^{2+}.

4.8. *Pf*TFIIB does not confer sequence-specific binding of *Pf*TBP or *Pf*TLP to a prototypical TATA box

In higher eukaryotes, TFIIB has been shown to stabilise sequence-specific TBP binding to the TATA box (Barberis *et al.*, 1993; Hisatake, Roeder and Horikoshi, 1993; Imbalanzo, Zaret and Kingston, 1993). It was thus possible that the presence of *Pf*TFIIB may selectively enhance binding of *Pf*TBP or *Pf*TLP to a prototypical TATA box DNA sequence and in this way confer sequence-specific TATA box-binding activity to *Pf*TBP or *Pf*TLP (Section 3.5.3). To test this possibility, EMSA competition assays were performed to compare the degree of competition exerted by unlabelled functional or mutated TATA-box dsDNA competitor on *Pf*TLP-*Pf*TFIIB-DNA and *Pf*TBP-TFIIB-DNA complexes. DNA-binding reactions using the GBP-130 282 bp DNA probe were set up under conditions at which a stimulatory effect of *Pf*TFIIB on the formation of *Pf*TBP or *Pf*TLP DNA complexes was observed (Fig. 43 and 44). The following pairs of 40 bp DNA competitors were chosen to test for TATA box sequence specificity: (i) Ad2ML-TATA and Ad2ML-GAGA, containing a consensus TATA box and a TATA sequence mutation that abolishes TBP sequence-specific DNA binding, and (ii) the G/C-rich mTdT TATA-less promoter sequence (mTdT-WT) and a mTdT promoter variant containing the Ad2ML TATA sequence (mTdT-TATA).

As shown in Fig. 43, the *Pf*TBP-*Pf*TFIIB-GBP-130 complex is outcompeted in the presence of either 800-fold molar excess Ad2ML-TATA competitor or 400-fold molar excess of Ad2ML-GAGA competitor (compare lanes 8 and 11). Consistent with this observation, *Pf*TBP-DNA complexes were outcompeted by the mTdT-TATA and mTdT-WT competitors to a comparable degree (Fig. 43, compare lanes 13-16 with lanes 17-20). Taken together, these data suggest that *Pf*TFIIB does not confer sequence-specific TATA box-binding activity to *Pf*TBP.

Interestingly, significantly higher concentrations of mTdT competitor DNAs compared to Ad2ML competitor DNAs were required to outcompete the *Pf*TBP-*Pf*TFIIB-GBP-130 complex (Fig. 43, compare lanes 5-12 with lanes 13-20). This result suggests a higher affinity of the *Pf*TBP-*Pf*TFIIB complex for the Ad2ML probe over the mTdT probe, unrelated to the presence of a consensus TATA box DNA sequence, that was not apparent in earlier EMSA competition assays addressing TBP sequence selectivity (Section 3.5.3). To test whether this difference in affinity was imparted by *Pf*TFIIB, the competition experiments were repeated for *Pf*TBP in the absence of *Pf*TFIIB. As shown in Fig. S6, isolated *Pf*TBP also bound the Ad2ML DNA probes with higher affinity compared to the mTdT DNA probes. Thus, the increased affinity of *Pf*TBP for Ad2ML competitors over mTdT competitors does not reflect a role of *Pf*TFIIB in altering the sequence-specificity of the *Pf*TBP-*Pf*TFIIB complex.

Instead, it seems more likely that the more stringent conditions used for these competition assays, specifically a lower *Pf*TBP concentration and higher concentration of poly(dG-dC), unmasked preferential binding of *Pf*TBP to the Ad2ML-TATA and Ad2ML-GAGA DNA probes over the mTdT-TATA and mTdT-WT DNA probes.

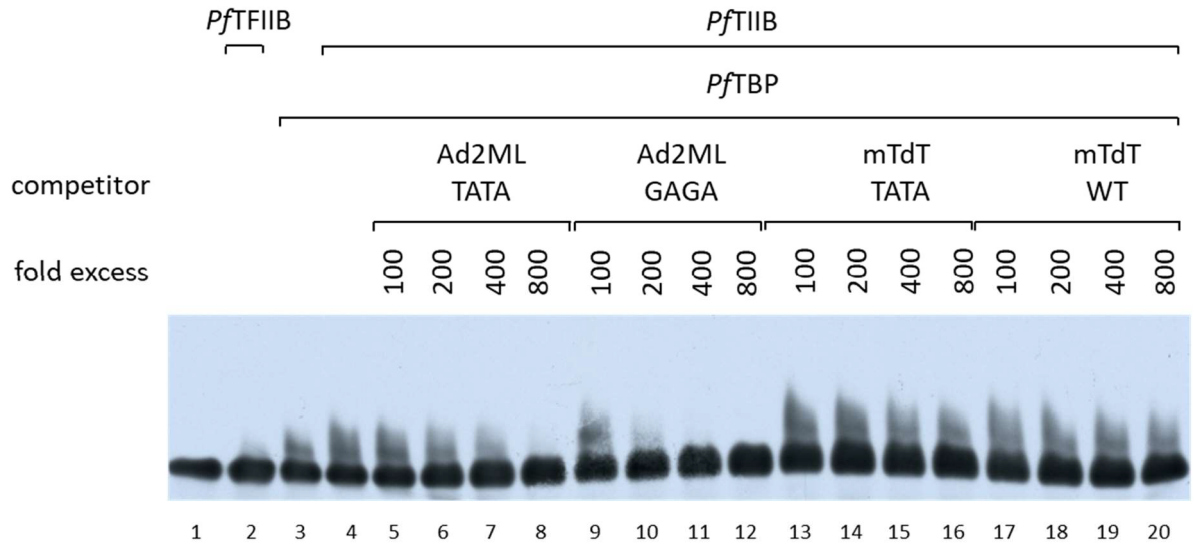


Figure 43*. Pf*TFIIB does not confer sequence-specific TATA box-binding activity to *Pf*TBP. DNA-binding reactions (10 µl) contained 5 fmol biotinylated 282 bp GBP-130 DNA probe, 50 ng poly(dG-dC), 45 nM *Pf*TBP and 30 nM *Pf*TFIIB as indicated, and 100 to 800-fold excess 40 bp unlabelled Ad2ML-TATA, Ad2ML-GAGA, mTdT-TATA or mTdT-WT competitor as indicated. Fold excess competitor relative to labelled GBP-130 probe was calculated in molar nucleotides. DNA-binding reactions were incubated on ice for 45 min and resolved by 1.4% agarose gel electrophoresis in the presence of 2 mM Mg^{2+}**.**

In contrast to *Pf*TBP, which in the presence of *Pf*TFIIB formed well-defined nucleoprotein complexes in EMSAs, *Pf*TLP nucleoprotein complexes formed in the presence of *Pf*TFIIB were not well resolved and could only be detected as a smear during agarose electrophoresis (Fig. 44). Nevertheless, stimulation of *Pf*TLP-DNA binding by *Pf*TFIIB could clearly be detected by the increased disappearance of free probe when both *Pf*TFIIB and *Pf*TLP were present in the DNA binding reaction (Fig. 44, compare lanes 2 and 3 to lane 4). Conversely, competition of *Pf*TLP/*Pf*TFIIB nucleoprotein complexes could be detected as recovery of free labelled DNA in response to increasing concentrations of unlabelled competitor DNA. As shown in Fig. 44, addition of Ad2ML-TATA or Ad2ML-GAGA unlabelled DNA competitor probes resulted in comparable recovery of free probe (Fig. 44, compare lanes 5-8 with lanes 9-12), suggesting that *Pf*TLP, together with *Pf*TFIIB, bound the Ad2ML probes containing a consensus TATA box sequence or the GAGA mutant TATA box with similar affinity. Similarly, unlabelled mTdT-WT and mTdT-TATA probes competed equally for DNA

binding of *Pf*TLP together with *Pf*TFIIB (Fig. 44, compare lanes 13-16 with lanes 17-20). These results suggest that, similar to the results obtained with *Pf*TBP, *Pf*TFIIB does not impart sequence-specific TATA box-binding activity to *Pf*TLP.

As seen for *Pf*TBP and the *Pf*TBP-*Pf*TFIIB complex, *Pf*TLP bound Ad2ML DNA competitors with higher affinity compared to the mTdT DNA competitors in these experiments (Fig. 44, compare lanes 5-12 with lanes 13-20). However, a similar difference in affinity was also observed for isolated *Pf*TLP (Fig. S7) and thus this degree of sequence selectivity is not imparted by *Pf*TFIIB.

Together, these results suggest that *Pf*TFIIB does not confer selectivity towards prototypical TATA-box DNA sequences to *Pf*TBP or *Pf*TLP. Whether *Pf*TFIIB modulates the sequence-selectivity of *Pf*TBP or *Pf*TLP at all needs to be determined in additional experiments using non-biased approaches such as a systematic evolution of ligands by exponential enrichment (SELEX) strategy.

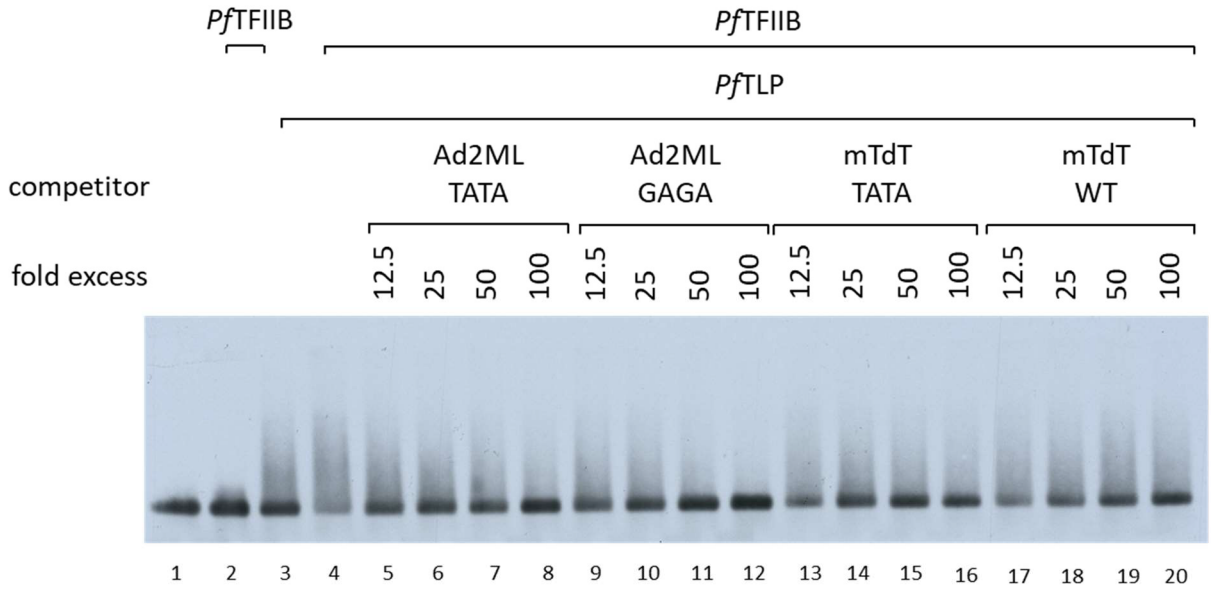


Figure 44. *Pf*TFIIB does not confer sequence-specific TATA box-binding activity to *Pf*TLP. DNA-binding reactions (10 µl) contained 5 fmol biotinylated 282 bp GBP-130 DNA probe, 30 ng poly(dG-dC), 7.5 nM *Pf*TLP and or 30 nM *Pf*TFIIB as indicated, and 100 to 800-fold excess 40 bp unlabelled Ad2ML-TATA, Ad2ML-GAGA, mTdT-TATA or mTdT-WT competitor as indicated. Fold excess competitor relative to labelled GBP-130 probe were calculated in molar nucleotides. DNA-binding reactions were incubated on ice for 45 min before being resolved by 1.4% gel electrophoresis.

Chapter 5: Results Part 3

Biomolecular condensation of *Pf*TFIIB and the potential role of *Pf*TFIIB in transcription condensate assembly

Biomolecular condensation refers to the formation of molecular assemblies enriched in certain biomolecules. These condensates can form through a variety of mechanisms, including, but not limited to, liquid-liquid phase separation (LLPS). LLPS refers to the spontaneous de-mixing of solutes in a liquid into a dilute and a concentrated phase, triggered by changes in environmental conditions. This process has recently emerged as an exciting lens under which to investigate transcriptional regulation. A growing body of work has revealed that components of the RNA Pol II transcription machinery, activators and co-activators, including the mediator complex, the RNA Pol II CTD and the transcription regulator BRD4, readily undergo LLPS *in vitro* and form condensates with liquid-like properties in cultured cells (Boehning *et al.*, 2018; Sabari *et al.*, 2018). Condensates formed by these "scaffold" transcription factors have been shown to recruit other transcription factors (Boija *et al.*, 2018) and to respond to modifications such as phosphorylation as well as environmental factors such as RNA abundance (Henninger *et al.*, 2021). These findings have led to the development of a model in which tuneable transcription condensates act to localise transcription factors into so-called transcription factories and assemble and disassemble in response to environmental cues (Fig. 9, Hnisz *et al.*, 2017; Henninger *et al.*, 2021). Ongoing work by our research group has demonstrated that many *P. falciparum* and human general transcription factors, including *Pf*TBP and *Pf*TLP, undergo *in vitro* LLPS (van der Linden, 2019, Karamanof, 2022, Michowicz, in preparation). This part of the thesis work aimed to examine the propensity of *Pf*TFIIB to undergo *in vitro* LLPS and to examine *Pf*TFIIB interactions with *Pf*TBP, *Pf*TLP and the *Pf*RPB1CTD within the context of our developing model of *P. falciparum* transcription condensates.

5.1 Condensation properties of *Pf*TFIIB

5.1.1. Intrinsically disordered regions within *Pf*TFIIB

Intrinsically disordered regions (IDRs) have been established to play a key role in driving liquid-liquid phase separation of many phase-separating proteins (Dignon, Best and Mittal, 2020). IDRs are regions within proteins which do not fold into a defined tertiary structure but instead are conformationally heterogenous and able to adopt an ensemble of conformations, which in turn allows them to engage with a variety of different interaction partners. To determine if *Pf*TFIIB harbours intrinsically disordered regions, the *Pf*TFIIB amino acid sequence was analysed using a variety of disorder prediction programs including two meta prediction servers, Dismeta (Huang, Acton and Montelione, 2014) and MobiDB (Piovesan *et al.*, 2021). Meta prediction servers combine different prediction programs rather than relying on the output of a single algorithm to improve predictive power and to reduce the risk of false negatives and false positives. Disorder predictors integrated into the consensus disorder prediction score generated by Dismeta include DISEMBL (Linding, Jensen, *et al.*, 2003), DISOPRED2 (Ward *et al.*, 2004), DISPro (Cheng, Sweredoski and Baldi, 2005), FoldIndex (Prilusky *et al.*, 2005), GlobPlot2 (Linding, Russell, *et al.*, 2003) , IUPred (Dosztányi, Csizmók, *et al.*, 2005), RONN (Yang *et al.*, 2005), and VL2 (Vucetic *et al.*, 2003). MobiDB on the other hand integrates predictions by IUPred (Dosztányi, Csizmok, *et al.*, 2005), ESpritz (Walsh *et al.*, 2012), DisEMBL (Linding, Jensen, *et al.*, 2003), JRONN (Yang *et al.*, 2005), PONDR® VSL2B (Obradovic *et al.*, 2005; Peng *et al.*, 2006), and GlobPlot (Linding, Russell, *et al.*, 2003).

MobiDB highlighted two regions of *Pf*TFIIB with a high propensity for disorder, spanning amino acid positions 1-53 and 38-119 (Fig. 45C). These regions are separated by four β-strands, predicted by Alphafold, which comprise the *Pf*TFIIB B-ribbon domain (Fig. 45A, see also Fig. 36A). The Dismeta consensus disorder prediction agrees largely with the prediction by MobiDB, highlighting the first 53 residues as having a high propensity for disorder (Fig. 45D). However, the second disordered region predicted by Dismeta is much smaller and comprises only aa residues 92-105. Dismeta also highlighted the extreme C-terminus of *Pf*TFIIB (aa 350-367) as having a high probability of being disordered. Overall, the disordered regions predicted by Dismeta and MobiDB are in good agreement with the AlphaFold structure prediction. None of these disordered regions were highlighted as low-complexity by the SEG algorithm (Wootton and Federhen, 1993).

Electrostatic interactions, determined by charge and charge patterning, can affect the conformation of disordered regions in solution (Das and Pappu, 2013), their propensity to undergo LLPS (Boyko *et al.*, 2019), their ability to engage in protein-protein and protein-nucleic acid interactions, as well as

their incorporation into transcriptional condensates (Lyons *et al.*, 2023). The net charge per residue was determined for the *Pf*TFIIB amino acid sequence using CIDER (Holehouse *et al.*, 2015). As shown in Fig. 45E, *Pf*TFIIB features multiple patches of positively and negatively charged amino acids which may contribute to its LLPS properties, its incorporation into transcriptional condensates and its protein-protein/protein-nucleic acid interactions.

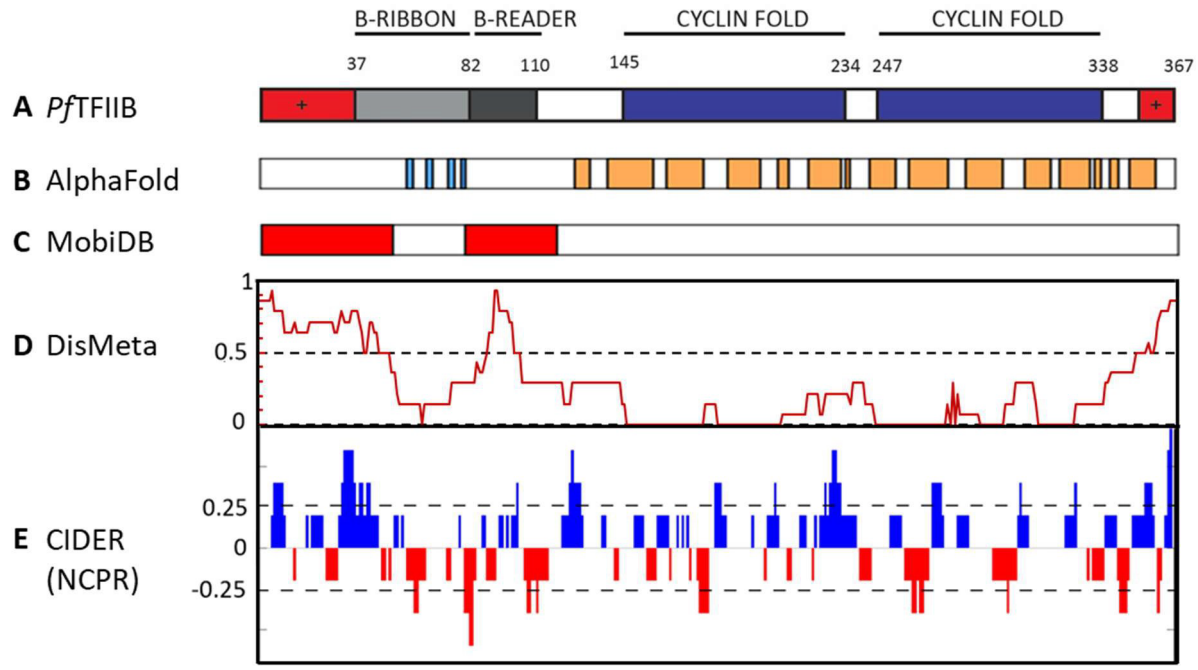


Figure 45. Bioinformatic analysis of *Pf*TFIIB structure.
Prediction of structural disorder and charge distribution within *Pf*TFIIB. **(A)** Schematic of predicted *Pf*TFIIB structure based on amino acid alignment with *Hs*TFIIB. *Plasmodium*-specific basic N- and C-terminal regions are shown in red and labelled with a '+'. Regions corresponding to the B-ribbon (light grey), B-reader (dark gray), and cyclin folds (dark blue) are indicated with amino acid residue positions shown. **(B)** AlphaFold predicted secondary structures, with a per-residue confidence score >70 % with β-strands shown in light blue and α-helices shown in orange. **(C)** MobiDB metaserver analysis of *Pf*TFIIB with disordered regions shown in red. **(D)** Consensus disorder output from the Dismeta metaserver. Regions with a DisMeta score ≥ 0.5 are predicted to be disordered. **(E)** CIDER analysis of net charge per residue (NCPR), calculated based on a sliding window of 5 amino acid residues. Positive charge is indicated in blue, negative charge in red.

5.1.2. Generation of *Pf*TFIIB variants fused to fluorescent protein

To facilitate the characterisation of *Pf*TFIIB LLPS behaviour, a panel of mCherry-*Pf*TFIIB fusion constructs were generated. These constructs were designed to examine *Pf*TFIIB biomolecular condensation, to assess the role of different *Pf*TFIIB regions in determining *Pf*TFIIB LLPS properties, and to investigate the properties of *Pf*TFIIB condensates formed in the presence of protein interaction partners and DNA. Constructs generated included 6His:mCherry fused to the N-terminus of full-length *Pf*TFIIB (6His:mCherry-*Pf*TFIIB), mCherry fused to the C-terminus of the *Pf*TFIIB N-terminal disordered region, comprising amino acid residues 1-140, (6His:*Pf*TFIIB1-140-mCherry), 6His:mCherry fused to the C-terminus of the *Pf*TFIIB lacking the basic N-terminus amino acid residues 1-37 (6His:*Pf*TFIIB38-140-mCherry), 6His:mCherry fused to the C-terminal half of *Pf*TFIIB containing the structured core domain with cyclin-fold repeats (6His:mCherry-*Pf*TFIIBΔ140) and a 6His:mCherry control (Fig. 46). Expression vectors for the bacterial expression of these proteins were generated as described in Materials and Methods, Section 2.3.2. For 6His:mCherry-*Pf*TFIIB and 6His:mCherry-*Pf*TFIIBΔ140, the expression protocol for the expression of 6His:*Pf*TFIIB was followed (Materials and Methods, Section 2.3.3.), while for the expression of 6His:*Pf*TFIIB1-140-mCherry, 6His:*Pf*TFIIB38-140-mCherry and 6His:mCherry a simplified expression protocol was used (Materials and Methods, Section 2.3.3.). All protein constructs were purified by IMAC, analysed by SDS-PAGE and normalised by a combination of Bradford protein determination assays and immunoblot analyses.

SDS-PAGE analysis of purified proteins showed prominent bands corresponding to the calculated molecular mass for all protein constructs (Fig. 46B). These bands were recognised by an anti-His antibody, confirming their identity (Fig. 46C). The *Pf*TFIIB-mCherry fusion protein preparations varied in purity and contained, to various extents, additional bands. In 6His:mCherry-*Pf*TFIIB preparations (calculated mass: 70.77 kDa), two bands with higher electrophoretic mobility were detected in addition to the prominent 70 kDa band corresponding to the intact protein construct, one at approximately 60 kDa and the other at 30 kDa. The 60 kDa band was not recognised by the anti-His antibody and likely represents an N-terminal truncation. The 30 kDa band was recognised by the anti-His antibody and can be identified as free 6His:mCherry (calculated mass 28.83 kDa), which is likely generated by premature translation termination during overexpression in *E.coli*. The 6His:mCherry-*Pf*TFIIBΔ140 (calculated mass 55.96 kDa) preparation showed an additional band at 45 kDa (Fig. 46B, lane 5), which likely results from N-terminal degradation of 6His:mCherry-*Pf*TFIIBΔ140 as it was not recognised by the anti-His antibody. Both 6His:*Pf*TFIIB1-140-mCherry and 6His:*Pf*TFIIB38-140-mCherry also contained additional bands which were not recognised by the anti-

His antibody. Given the predominantly disordered nature of these protein constructs products it is likely that they are particularly susceptible to proteolytic cleavage and that these bands reflect N-terminal truncations of the *Pf*TFIIB disordered region. Pilot experiments attempting to further purify these proteins by sequential Q-Sepharose and SP-Sepharose ion exchange chromatography resulted in a loss of material and did not significantly improve purity. Given the relatively low yield in the expression of these proteins and the high protein concentrations required for LLPS assays, further attempts to separate full-length protein constructs from truncated contaminants were deemed not to be feasible within the scope of this study.

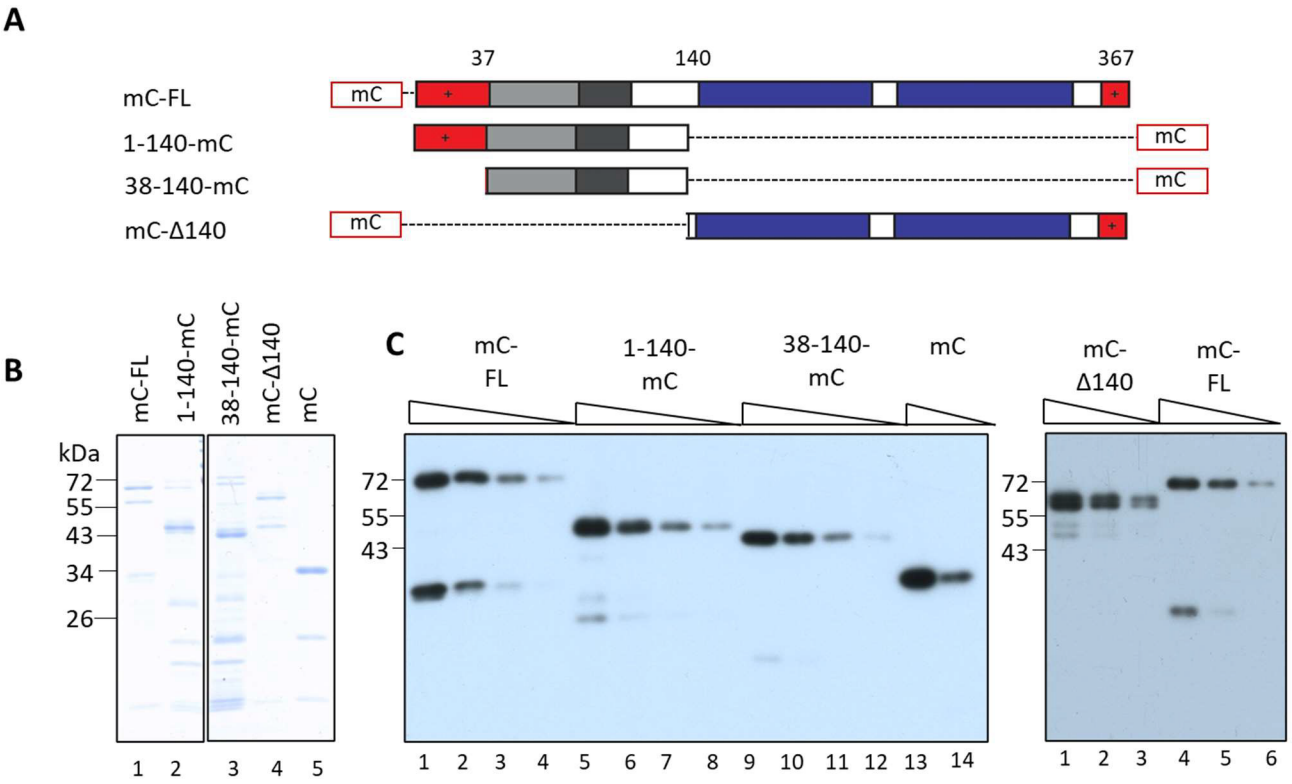


Figure 46. Generation of *Pf*TFIIB-mCherry fusion proteins to study *Pf*TFIIB biomolecular condensation.
(A) Schematic of 6His-tagged *Pf*TFIIB-mCherry fusion constructs, with N- or C-terminal location of mCherry indicated. **(B)** SDS-PAGE analysis of *Pf*TFIIB-mCherry fusion proteins. 200 ng of 6His:mCherry-*Pf*TFIIB, or 6His:tagged *Pf*TFIIB-mCherry fusion constructs were resolved by SDS-PAGE and stained with Coomassie. **(C)** Normalisation of *Pf*TFIIB-mCherry fusion proteins by immunoblot analysis. Equivalent molar amounts of *Pf*TFIIB-mCherry fusion constructs were resolved by 12% SDS-PAGE, transferred to a PVDF membrane and probed with mouse anti-His antibody. The position of protein molecular weight markers are indicated on the left. mC, mCherry; mC-FL, 6His:mCherry-*Pf*TFIIB; 1-140-mC, 6His:*Pf*TFIIB1-140-mCherry; 38-140-mC, 6His:*Pf*TFIIB38-140-mCherry; mC-Δ140 6His:mCherry-*Pf*TFIIB-141-367.

5.1.3. *Pf*TFIIB forms small dynamic assemblies in the presence of molecular crowder

To assess the condensation properties of *Pf*TFIIB in isolation, a microscopy-based approach was used in which *Pf*TFIIB condensate assembly was monitored under a range of different conditions by mixing purified 6His:mCherry-*Pf*TFIIB fusion proteins with a customisable condensation buffer. Condensation buffers used in classical LLPS assays are designed to mimic cellular conditions as closely as possible (Alberti, Gladfelter and Mittag, 2019). To this end condensation buffers typically include a near physiological concentration of monovalent ions (typically 150 mM NaCl or KCl), a buffering component and a molecular crowder such as polyethylene glycol (PEG) to mimic the crowded environment of the cell (Alberti, Gladfelter and Mittag, 2019). Additional factors such as nucleic acids have also been shown to significantly increase the *in vitro* LLPS propensity of a number of proteins (Larson *et al.*, 2017; Rhine, Vidaurre and Myong, 2020; Claeys Bouuaert *et al.*, 2021). If the experimental conditions are conducive to phase separation of the biomolecule under study, phase separation can be visualised by the appearance of assemblies enriched in the biomolecule of interest. Assemblies with liquid-like properties are spherical due to interfacial tension, fuse with each other and relax into larger spherical assemblies, and exhibit evidence of internal dynamics and rearrangement. A single component phase separating system is characterised by a defined concentration at and above the biomolecule of interest in solution de-mixes into two distinct phases and below which only a single homogenous phase is observed. This concentration is defined as C_{sat}. When the concentration of the biomolecule of interest in the system is above C_{sat}, further increasing the protein concentration within the system should increase the volume of the condensed phase but the concentrations of the biomolecule of interest in the dilute and concentrated phases should remain unchanged (Alberti, Gladfelter and Mittag, 2019).

Given that phase separation is known to be condition-dependent and highly sensitive to environmental factors, the microscopic properties of 6His:mCherry-*Pf*TFIIB were initially assessed in the presence of different molecular crowders. In brief, this involved the addition of 6His:mCherry-*Pf*TFIIB to a condensation buffer containing molecular crowder and microscopic visualisation of the reaction after a one hour incubation period on an inverted coverslip (Materials and Methods, Section 2.3.4.). Images from ten randomly chosen fields of view were then recorded, qualitatively assessed and the condensed fraction, defined as the proportion of each field of view occupied by biomolecular condensates, determined using a custom Python script adopted by a previous MSc student in the laboratory, Leo Karamanof (Karamanof, 2022).

As shown in Fig. 47A, in the absence of molecular crowder, 6His:mCherry-*Pf*TFIIB exists as a single phase at concentrations up to 1 µM, the maximum concentration of purified 6His:mCherry-*Pf*TFIIB that could be obtained in the condensation assays. In contrast, addition of condensation buffers containing the molecular crowder polyethylene glycol (PEG) with different molecular mass (3 000 kDa, 8 000 kDa and 20000 kDa) resulted in the formation of relatively small (<1 µm) assemblies, which contained higher concentrations of 6His:mCherry-*Pf*TFIIB than the surrounding dilute phase (Fig. 47A). These assemblies appeared not to be distinctly spherical when visualised after 1 h incubation and appeared to stick to each other upon contact, with no evidence for fusion and relaxation into larger assemblies. *Pf*TFIIB assemblies were formed across a broad range of KCl concentrations (Fig. 47C), indicating that they are moderately resistant to high ionic strength and unlikely to be driven solely by electrostatic interactions. Given the small size of the *Pf*TFIIB assemblies and the resolution limits of the microscope in use it was not possible to determine if the somewhat irregular assemblies represented irregular aggregates or whether these assemblies could be comprised of smaller, partially fused spherical condensates, similar to those that have been observed for initially liquid-like condensates, formed by LLPS, that transition into a gel- or solid-like state over time (Banani et al., 2017; see Section 6.6). The formation of these assemblies is driven by *Pf*TFIIB as no condensates were detected for the 6His:mCherry control (Fig. S8) – even at a final concentration of 50 µM.

Next, condensate formation was examined at a range of 6His:mCherry-*Pf*TFIIB concentrations. If the condensates were indeed formed through LLPS, one would expect a homogenous distribution of 6His:mCherry-*Pf*TFIIB in solution below a defined C_{sat}. However, small assemblies could still be detected at a 6His:mCherry-*Pf*TFIIB concentration as low as 0.125 µM, suggesting that either the C_{sat} for 6His:mCherry-*Pf*TFIIB condensation falls below this value or that *Pf*TFIIB assemblies are formed by a condensation process other than LLPS (Fig. 47B). It was not possible to examine 6His:mCherry-*Pf*TFIIB at concentrations below 0.125 µM because of rapid photobleaching within the timeframe required for visualisation.

To determine if 6His:mCherry-*Pf*TFIIB molecules in condensates could dynamically rearrange and exchange with the surrounding dilute phase, indicating liquid- or gel-like properties of 6His:mCherry-*Pf*TFIIB condensates, and to address whether these dynamic properties change with time, fluorescence recovery after photobleaching (FRAP) experiments were performed. In these experiments, 6His:mCherry-*Pf*TFIIB assemblies were photobleached with a high-intensity laser to less than 10 % of the original signal immediately after condensate formation or after further incubation for up to 3 h. Hereafter, recovery of fluorescence was measured over 8 min or until photobleaching during image acquisition prevented further imaging. Due to the relatively small size

of 6His:mCherry-*Pf*TFIIB condensates (diameter < 1µm) a high laser power and high magnification were required for visualisation. As a result, the extent of background photobleaching during image acquisition was relatively high, limiting the duration of image acquisition. The recovery after photobleaching for all conditions was analysed using the web-based FRAP analysis server EasyFRAP (Koulouras *et al.*, 2018) and results were normalised by double normalisation (Phair, Gorski and Misteli, 2003) to account for unintentional photobleaching during image acquisition. For reference, typical half times for full fluorescence recovery after photobleaching of biomolecular condensates can be highly variable even for different condensates formed by the same protein (McSwiggen *et al.*, 2019) and can range from a few seconds, for example for highly dynamic liquid-like Med1 condensates observed in cultured cells (T1/2 ~ 7 s; Sabari *et al.*, 2018) to several minutes, for example in case of the more gel-like Meg-3 condensates observed in *Caenorhabditis elegans* embryos (t½: 128 - 384 s; Putnam *et al*., 2019).

As shown in Fig. 48, 6His:mCherry-*Pf*TFIIB condensates recovered over 50% of the original droplet intensity within 166 s when bleached 15 - 30 min after condensate formation (mobile fraction: 0.86 ± 0.24; half time of recovery (t½): 114.75 ± 37.16 s). This result demonstrates that a large proportion of 6His:mCherry-*Pf*TFIIB molecules are dynamically exchanged between the condensed and dilute phases and suggests that these 6His:mCherry-*Pf*TFIIB assemblies represent liquid or gel-like condensates rather than solid aggregates or glasses. In contrast, the fluorescence of 6His:mCherry-*Pf*TFIIB assemblies photobleached 1-2 h after condensate formation recovered at a much slower rate and less than 20% of the original 6His:mCherry-*Pf*TFIIB condensate fluorescence intensity was recovered within 166 s of bleaching (mobile fraction: 1 ± 0.01; t½: 180.3 ± 0 s). Recovery rates were further reduced after a 2 - 3 h incubation of the condensation reaction (mobile fraction: 0.88 ± 0.21; t½: 456.81 ± 0.03 s), suggesting that the exchange of 6His:mCherry-*Pf*TFIIB between condensates and the dilute phase decreases over time. Together, these results suggest that *Pf*TFIIB assemblies are initially dynamic and that the irregular morphology of *Pf*TFIIB condensates imaged after longer incubation times may be caused by a transition from initially dynamic liquid- or gel-like assemblies into viscous gels with long-lived interactions and limited exchange of molecules between condensates and the surrounding solution. This change in material properties over time may result in incomplete fusion of small spherical condensates, resulting in the generation of assemblies with irregular appearance. Such changes in material properties, often termed hardening or maturation, have been observed for other phase separating proteins and are thought to occur through β-strand interactions, entanglement of polymers within the densely packed condensate and increased formation of intermolecular interactions characterised by slow dissociation rates (Banani *et al.*, 2017).

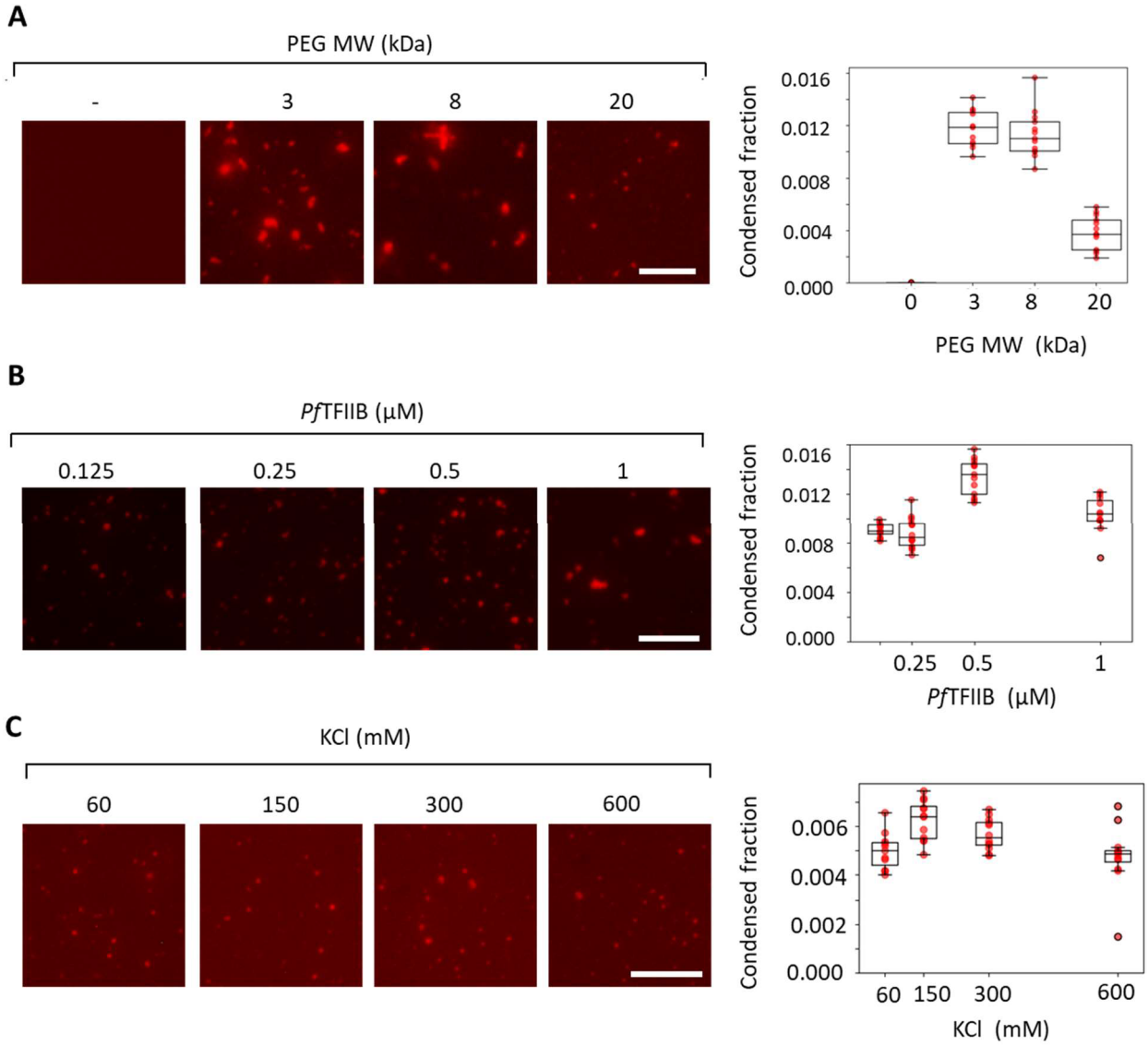


Figure 47. *Pf*TFIIB assembles into small condensates at sub-micromolar concentrations in *in vitro* LLPS assays. Representative widefield fluorescent microscopy images of 6His:mCherry-*Pf*TFIIB condensates and the corresponding quantification of condensed fraction. **(A)** 6His:mCherry-*Pf*TFIIB assembly formation in the presence of 10% PEG molecular crowder with different molecular weights (MW). Condensation reactions contained 600 nM 6His:mCherry-*Pf*TFIIB, 60 mM KCl, 5 mM magnesium acetate and no (-) or 10% PEG of the indicated MW. **(B)** 6His:mCherry-*Pf*TFIIB assembly formation at increasing *Pf*TFIIB concentrations. Condensation reactions were carried out as in (A) in the presence of 10% PEG 20000, 150 mM KCl and the indicated 6His:mCherry-*Pf*TFIIB concentrations. **(C)** 6His:mCherry-*Pf*TFIIB assembly formation in the presence of different KCl concentrations. Condensation reactions were carried out as in (B) with 400 nM 6His:mCherry-*Pf*TFIIB and the indicated KCl concentrations. Scale bar = 10 µm. Image analysis and quantification of the condensed fraction was performed for 10-15 fields of view per condition.

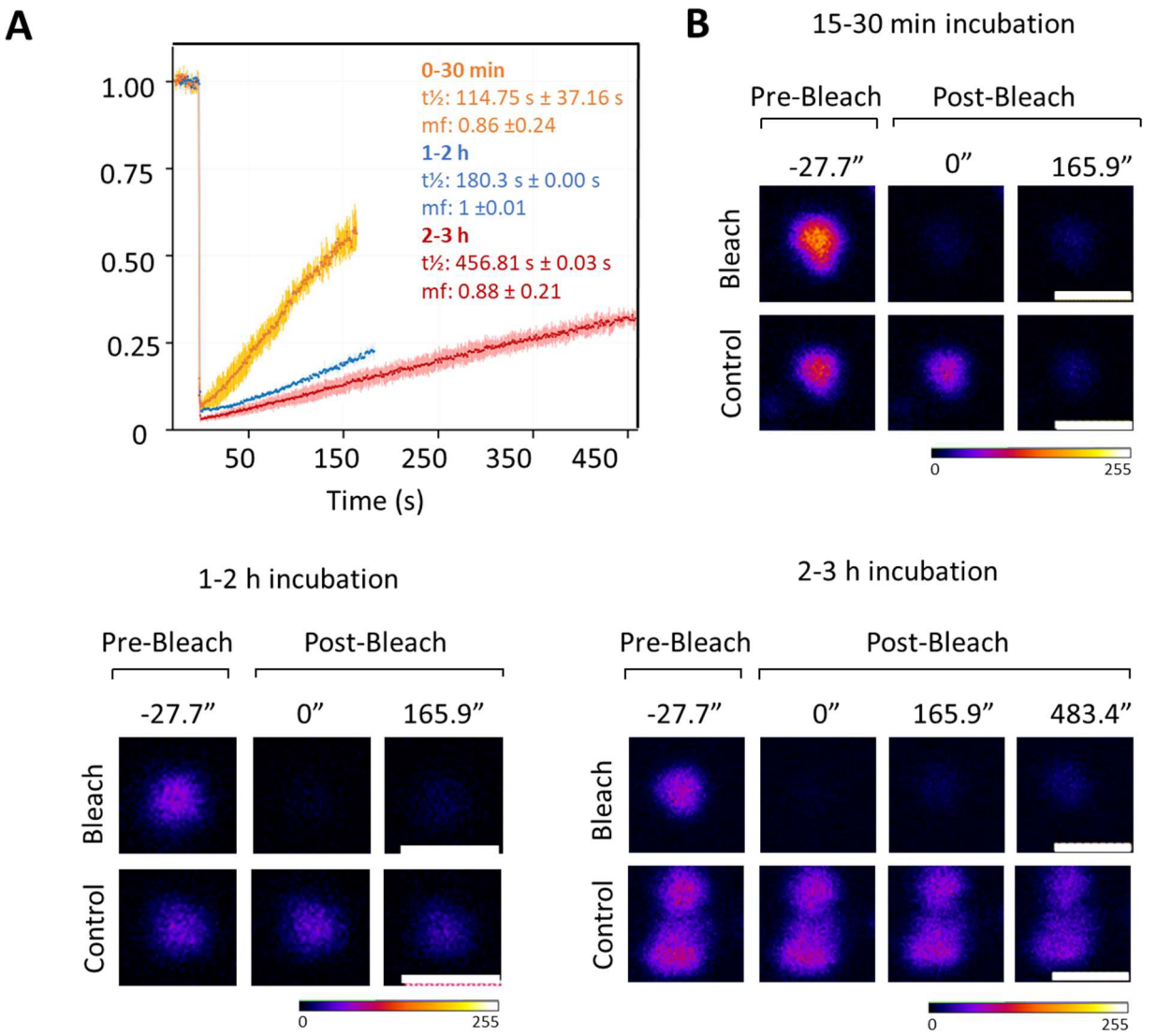


Figure 48. *Pf*TFIIB condensates exhibit reduced dynamic properties over time. (A) Normalised fluorescence recovery after photobleaching of 6His:mCherry-*Pf*TFIIB condensates formed with 1 µM 6His:mCherry-*Pf*TFIIB in the presence of 10% PEG 20000, 150 mM KCl and 5 mM magnesium acetate. Photobleaching was carried out 15 - 30 min (yellow), 1 - 2 h (blue) or 2 - 3 h (red) after inducing condensation. Data presented as mean ± the standard deviation (0 - 30 min/2 - 3 h: n=3; 1 - 2 h: n=4). Mobile fraction (mf) and half time for recovery of fluorescence after photobleaching (t½), estimated using EasyFrap (Koulouras *et al.*, 2018), are indicated. **(B)** Representative images of bleached (top) and control (bottom) condensates at indicated time points relative to completion of bleaching. Non-bleached control condensates in the same field of view were used for the calculation of normalised fluorescence intensity. Images are coloured using the fire look up table in ImageJ. Scale Bar = 1 µm.

5.1.4. The N-terminal disordered region and the C-terminal cyclin-fold repeats within *Pf*TFIIB are sufficient to drive biomolecular condensation

To delineate which regions of *Pf*TFIIB may be driving the formation of *Pf*TFIIB condensates, condensation reactions were performed with a series of *Pf*TFIIB deletion mutants fused to mCherry fluorescent protein, using conditions under which full-length 6His:mCherry-*Pf*TFIIB condensates were detected. These deletion mutants included 6His:*Pf*TFIIB1-140-mCherry, comprising only the predominantly disordered N-terminal half of *Pf*TFIIB, 6His:*Pf*TFIIB38-140-mCherry, comprising the disordered N-terminal half of *Pf*TFIIB lacking the *Plasmodium*-specific basic N-terminus, and 6His:mCherry-*Pf*TFIIBΔ140, comprising only of the largely structured C-terminal half of *Pf*TFIIB (see Figs. 45 and 46A).

As shown in Fig. 49, small assemblies similar to those seen with 6His:mCherry-*Pf*TFIIB were detected with all deletion variants tested at a concentration of 1 µM (Fig. 49). This result suggests that both the unstructured N-terminal region, comprising aa 38-140, and the structured *Pf*TFIIB core region, comprising aa 141-367, can engage in homotypic intermolecular interactions that drive biomolecular condensation at low concentrations. It should be noted that because preparations of *Pf*TFIIB deletion variants contained variable amounts of contaminating mCherry-fused truncated protein (Fig. 46B), a quantitative comparison of the condensed fraction of the reactions shown in Fig. 49 was not performed.

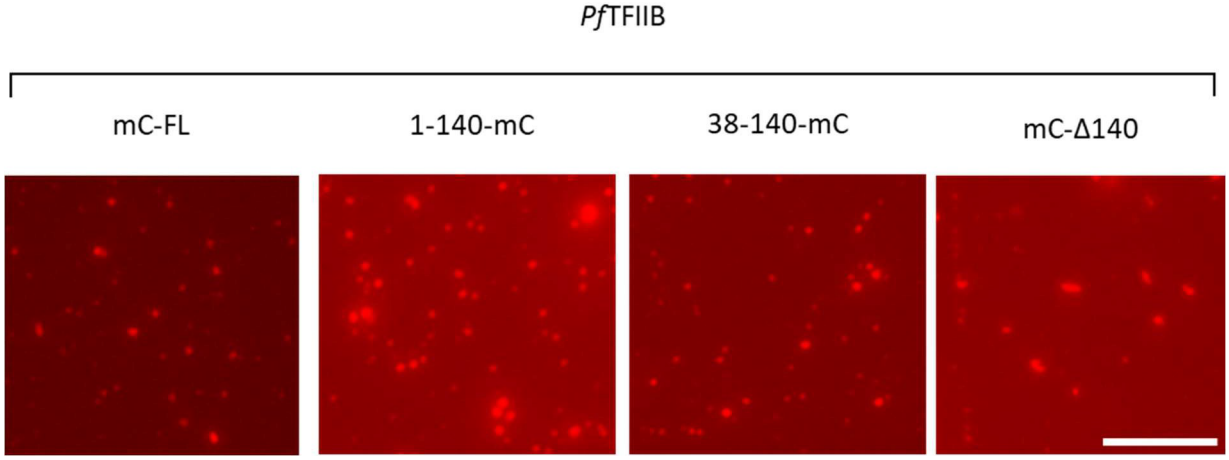


Figure 49. The disordered *Pf*TFIIB N-terminal region (aa 38-140) and the structured C-terminal region (aa 141-367) are sufficient to drive biomolecular condensation. Representative widefield fluorescent microscopy images of condensates formed with 1 µM 6His:mCherry-*Pf*TFIIB (mC-FL), 6His:*Pf*TFIIB1-140-mCherry (1-140-mC), 6His:*Pf*TFIIB38-140-mCherry (38-140-mC) or 6His:mCherry-*Pf*TFIIB-141-367 (mC-Δ140), in the presence of 10% PEG 20000, 150 mM KCl and 5 mM magnesium acetate. Imaging was performed 1 h after inducing condensation. Scale bar = 10 µm.

5.2. Impact of DNA on *Pf*TFIIB condensation

5.2.1. DNA undergoes biomolecular condensation

Given that *Pf*TFIIB binds DNA (Fig. 35), and considering that many DNA-binding proteins have been shown to undergo DNA-dependent condensation (Larson *et al.*, 2017; Zhou *et al.*, 2019; Claeys Bouuaert *et al.*, 2021), it was important to assess the impact of DNA on the formation and properties of *Pf*TFIIB condensates. However, DNA can undergo condensation by itself at favourable ionic conditions and in the presence of molecular crowders (Bloomfield, 1996; Cheng, Jia and Ran, 2015; Muzzopappa, Hertzog and Erdel, 2021). Therefore, the condensation behaviour of DNA was examined first, using the *in vitro* LLPS conditions established for *Pf*TFIIB. Poly(dI-dC), a double-stranded alternating copolymer that contains the chemical signatures of A-T base pairs in the DNA minor groove and G-C base pairs in the major groove (Fig. 26A), was chosen for DNA condensation assays as this synthetic DNA polymer can be efficiently labelled with DAPI, which preferentially binds to the minor groove of A/T-rich DNA (Kapuscinski, 1979). Furthermore, poly(dI-dC) is selectively bound by *Pf*TBP and *Pf*TLP (Section 3.5.5 and 3.5.7) and is therefore suitable to study the formation of mixed condensates containing *Pf*TFIIB, *Pf*TBP or *Pf*TLP, and DNA.

Consistent with the result of earlier studies that demonstrated PEG- and salt-induced (PSI) condensation of DNA (Muzzopappa, Hertzog and Erdel, 2021), poly(dI-dC) DNA formed spherical assemblies at concentrations above 50 ng/µl (Fig. 50A). Further increasing the concentration of poly(dI-dC) resulted in an increase in droplet size and a corresponding increase in the condensed fraction (Fig. 50A). Condensate formation increased in the presence of high concentrations of KCl (Fig. 50B) and was strongly dependent on the concentration of magnesium ions in the condensation reaction (Fig. 50C). These results suggest that positively charged ions present in the reaction shield the repulsive forces between the negatively charged DNA molecules at high concentrations during LLPS. GBP-130 (282 bp) and mTdT TATA (260 bp) dsDNA fragments also formed phase-separated condensates (Fig. S9) at similar concentrations, proving that the ability to undergo LLPS under these conditions is a general property of dsDNA and not a specific property of poly(dI-dC).

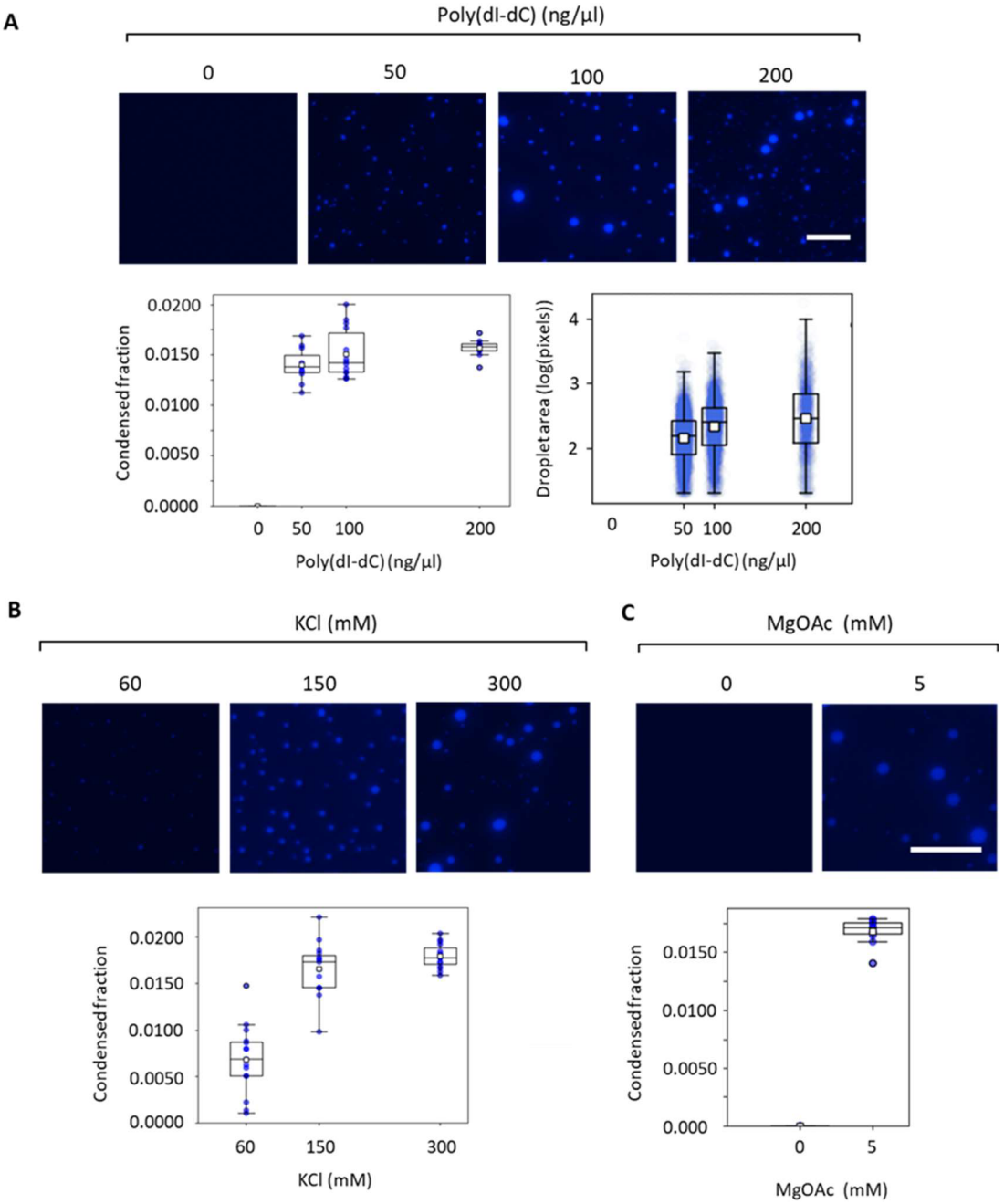


Figure 50. Poly(dI-dC) dsDNA undergoes biomolecular condensation in the presence of molecular crowder and magnesium. (A) *T*op: representative widefield fluorescent microscopy images showing poly(dI-dC) DNA condensation as a function of poly(dI-dC) concentration. Condensation reactions contained the indicated concentrations of poly(dI-dC), 10% PEG 20000, 150 mM KCl and 5 mM magnesium acetate. *Bottom:* quantification of condensed fraction and individual droplet areas in pixels. **(B)** *Top*: representative fluorescent microscopy images showing poly(dI-dC) condensation at different KCl concentrations. Condensation reactions were carried out as in (A) with 200 ng/µl poly(dI-dC) and the indicated concentration of KCl. *Bottom:* quantification of condensed fraction**. (C)** *Top:* representative fluorescent microscopy images showing poly(dI-dC) condensation in the absence or presence of 5 mM magnesium acetate (MgOAc). Condensation reactions were carried out as in (A) with 250 ng/µl poly(dI-dC) and 60 mM KCl. *Bottom:* quantification of condensed fraction. 50 µM 4',6-diamidino-2-phenylindole (DAPI) was included in the condensation buffer in all reactions to facilitate DNA visualisation. Scale bars = 10 µm. Image analysis and quantification of condensed fraction and droplet area were performed for 10-15 fields of view per condition except for the 0 ng/µl DNA condition where only one field of view was quantified. White squares denote mean value.

5.2.2. *Pf*TFIIB and DNA form mixed phase separated condensates

Next, to determine whether 6His:mCherry-*Pf*TFIIB and DNA co-assemble into mixed condensates, increasing concentrations of poly(dI-dC) were incubated with a fixed concentration of 6His:mCherry-*Pf*TFIIB before mixing with condensation buffer. Consistent with *Pf*TFIIB-DNA interactions detected in EMSAs (Fig. 35), poly(dI-dC) and 6His:mCherry-*Pf*TFIIB formed small mixed condensates, in which 6His:mCherry-*Pf*TFIIB and DNA were homogenously distributed, at low poly(dI-dC) concentrations (50 ng/µl; Fig. 51). Similarly to the *Pf*TFIIB condensates formed in the absence of DNA, condensates formed in the presence of low concentrations of poly(dI-dC) were not distinctly spherical and appeared not to fuse upon contact when visualised after 1 h, suggesting gel-like properties. In contrast, condensates formed by *Pf*TFIIB and poly(dI-dC) at higher poly(dI-dC) concentrations (100-200 ng/µl) were significantly larger (Fig. 51) and distinctly spherical in appearance. Importantly, isolated 6His:mCherry was not enriched in poly(dI-dC) condensates (Fig. S10), demonstrating that enrichment of 6His:mCherry-*Pf*TFIIB is driven by *Pf*TFIIB-DNA interactions and not by mCherry interacting with DNA.

Interestingly, the distribution of *Pf*TFIIB and poly(dI-dC) DNA within condensates formed at high poly(dI-dC) concentrations was not entirely homogenous but instead appeared to consist of a core of homogenously mixed 6His:mCherry-*Pf*TFIIB and poly(dI-dC) surrounded by a shell of poly(dI-dC) DNA (Fig 51). This result suggests the existence of different territories, dominated by either homotypic DNA-DNA interactions or heterotypic *Pf*TFIIB-DNA interactions, within mixed *Pf*TFIIB-poly(dI-dC) condensates.

To examine the mobility of 6His:mCherry-*Pf*TFIIB within the *Pf*TFIIB-poly(dI-dC) condensates, a fluorescence recovery after photobleaching experiment was performed (Fig. 52). *Pf*TFIIB-poly(dI-dC) mixed condensates that had been incubated on the coverslip for 3 h were bleached to less than 10% of the initial fluorescence signal. As shown in Fig. 52, 6His:mCherry-*Pf*TFIIB condensates recovered to over 50% of their initial fluorescence within the 500 s window of observation (mobile fraction: 0.85 ± 0.19; t½ = 297.35 ± 108.71 s), providing clear evidence that 6His:mCherry-*Pf*TFIIB is dynamically exchanged between 6His:mCherry-*Pf*TFIIB-poly(dIdC) mixed condensates and the surrounding dilute phase, even following longer incubation times. Given that 6His:mCherry-*Pf*TFIIB-only condensates only recovered to just over 25% of the initial fluorescence intensity in the same window of observation when bleached 3 h after condensate formation (Fig. 48), these results support that the presence of DNA alters the material properties of *Pf*TFIIB condensates, specifically increasing 6His:mCherry-*Pf*TFIIB exchange between the dilute and condensed phase.

Together, the spherical morphology of *Pf*TFIIB condensates and increased dynamic exchange of *Pf*TFIIB molecules between condensates and the dilute phase suggests that *Pf*TFIIB interactions with DNA compete with interactions between *Pf*TFIIB molecules that lead to transition of *Pf*TFIIB assemblies into a gel- or solid-like state.

In order to gain insights into the nature of interactions that mediate the formation of *Pf*TFIIB-poly(dI-dC) condensates, condensation assays were carried out at increasing ionic strength or in the presence of 1,6-hexanediol, an inhibitor of weak hydrophobic interactions (Düster *et al.*, 2021). As shown in Fig. 53, 1,6-hexanediol did not significantly affect 6His:mCherry-*Pf*TFIIB-poly(dI-dC) condensation. In contrast, increasing the ionic strength significantly impacted 6His:mCherry-*Pf*TFIIB-poly(dI-dC) condensate morphology (Fig. 54). At 60 mM and 150 mM KCl, 6His:mCherry-*Pf*TFIIB and poly(dI-dC) were homogenously distributed in mixed condensates. However, in the presence of 300 mM KCl 6His:mCherry-*Pf*TFIIB was excluded from the interior and, instead, localised to the periphery of the much larger poly(dI-dC) condensates. This exclusion to the periphery of poly(dI-dC) condensates was accompanied by a decrease in the condensed fraction of 6His:mCherry-*Pf*TFIIB. These results suggest that heterotypic *Pf*TFIIB-DNA interactions that mediate homogenous distribution of *Pf*TFIIB within *Pf*TFIIB-DNA condensates have a dominant electrostatic component. Increasing KCl concentrations appear to disrupt these interactions, to favour salt-resistant homotypic dsDNA (Fig. 50B) and/or homotypic *Pf*TFIIB (Fig. 47C) interactions, and, as a result, lead to de-mixing of dsDNA and 6His:mCherry-*Pf*TFIIB and subsequent expulsion of concentrated 6His:mCherry-*Pf*TFIIB on the surface of poly(dI-dC) condensates.

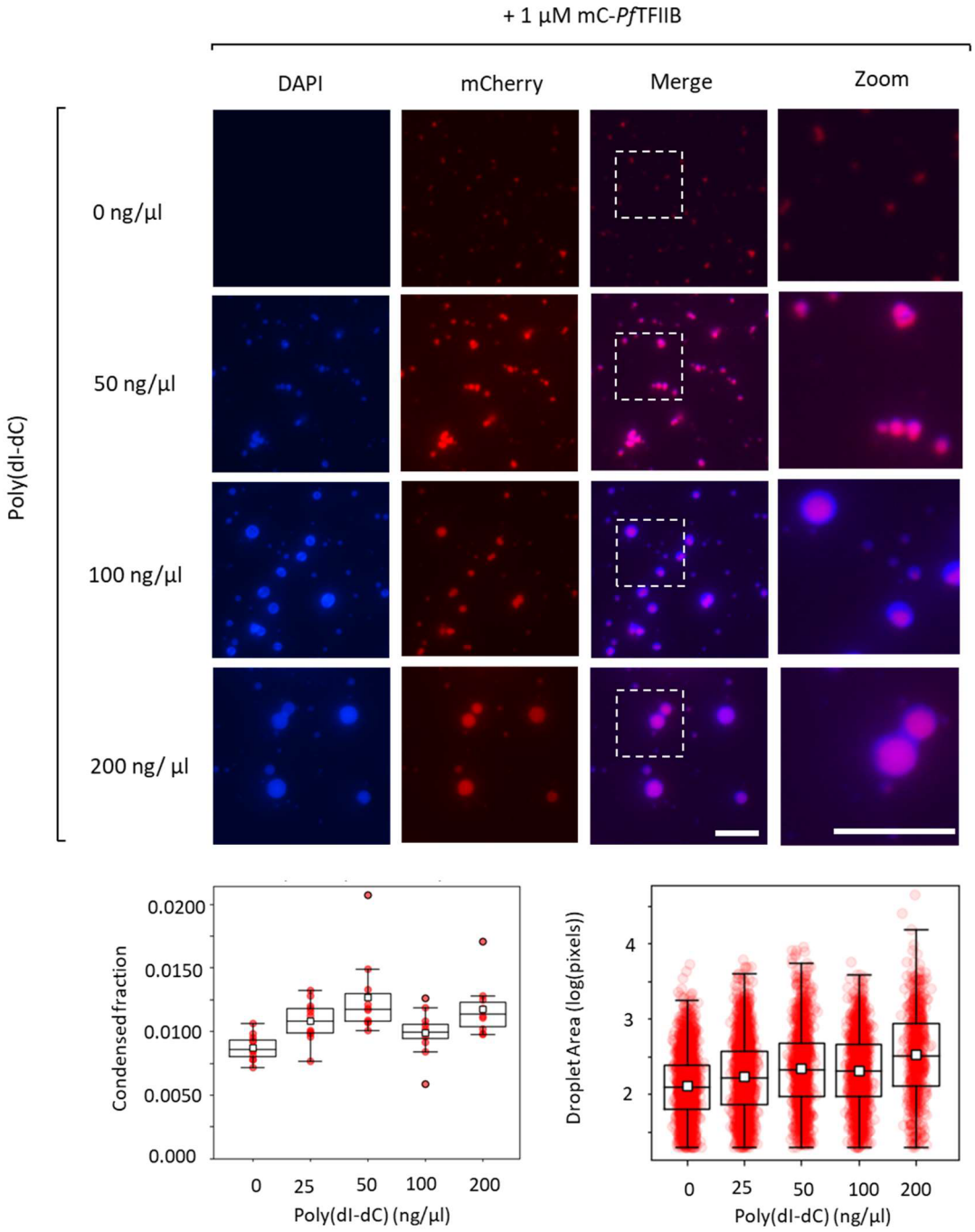


Figure 51. 6His:mCherry-*Pf*TFIIB and DNA form mixed condensates. *Top:* Representative widefield fluorescent microscopy images of condensate formation. Condensation reactions contained 1 µM 6His:mCherry-*Pf*TFIIB, the indicated concentrations of poly(dI-dC) DNA, 10% PEG 20000, 150 mM KCl, 5 mM magnesium acetate and 50 µM DAPI to facilitate DNA visualisation. Poly(dI-dC) and mCherry-*Pf*TFIIB were mixed thoroughly and incubated for 5 min before the addition of condensation buffer. Mixed assemblies were visualised using the relevant fluorescent channels. White dashed boxes indicate areas of the merged image enlarged on the right (ZOOM). Scale bar = 10 µm (white). *Bottom*: Box plot showing quantification of 6His:mCherry-*Pf*TFIIB condensed fraction and droplet area of 6His:mCherry-*Pf*TFIIB condensates as a function of increasing poly(dI-dC) concentrations. Mean values are denoted by white boxes. Image analysis and quantification of condensed fraction was performed for 10-15 fields of view per condition.

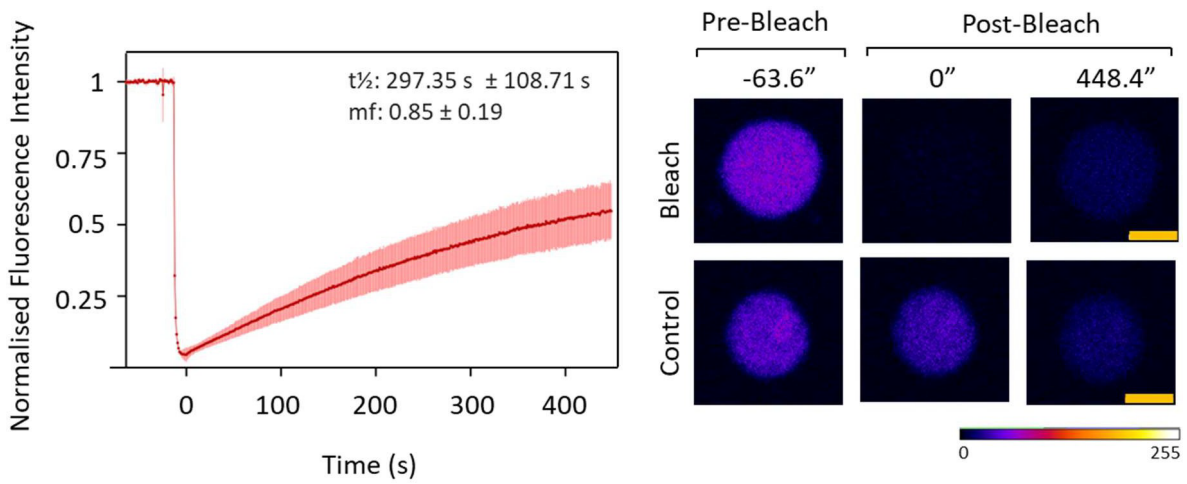


Figure 52. *Pf*TFIIB-DNA condensates display fluorescence recovery after photobleaching. *Left:* Normalised fluorescence recovery after photobleaching of *Pf*TFIIB condensates formed with 1 µM 6His:mCherry-*Pf*TFIIB, 200 ng/µl poly(dI-dC), 10 % PEG 20000, 5 mM magnesium acetate, 150 mM KCl. Bleaching was carried out 2-3 h after inducing condensation. Data presented as mean ± the standard deviation (n=4). Mobile fraction (mf) and half time for recovery of fluorescence after photobleaching (t½), were estimated using EasyFrap (Koulouras *et al*., 2018), are indicated. *Right:* Representative images of bleached (top) and control (bottom) condensates at indicated time points relative to completion of bleaching. Non-bleached control condensate in the same field of view were used for the calculation of normalised fluorescence intensity. Images are coloured using the fire look up table in ImageJ. Scale bar = 2 µm (orange).

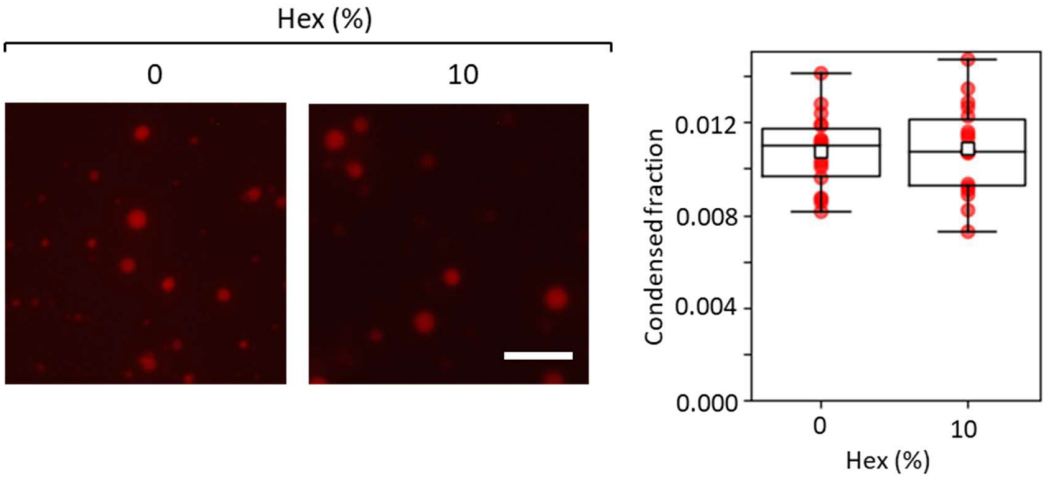


Figure 53. 6His:mCherry-*Pf*TFIIB poly(dI-dC) condensates are insensitive to 1,6-hexanediol. *Left:* Representative widefield fluorescent microscopy images of condensates formed by 6His:mCherry-*Pf*TFIIB and poly(dI-dC) DNA in the absence or presence of 10% 1,6-hexanediol (Hex). Condensation reactions contained 1 µM 6His:mCherry-*Pf*TFIIB, 200 ng/µl poly(dI-dC), 10% PEG 20000, 150 mM KCl and 5 mM magnesium acetate. Poly(dI-dC) and 6His:mCherry-*Pf*TFIIB were mixed thoroughly and incubated for 5 min before the addition of condensation buffer. Scale bar = 10 µm. *Right*: Boxplot showing quantification of condensed fraction of 6His:mCherry-*Pf*TFIIB in the presence or absence of 10% 1,6-Hexanediol. Mean values are denoted by white boxes. Image analysis and quantification of condensed fraction was performed for 15 fields of view per condition.

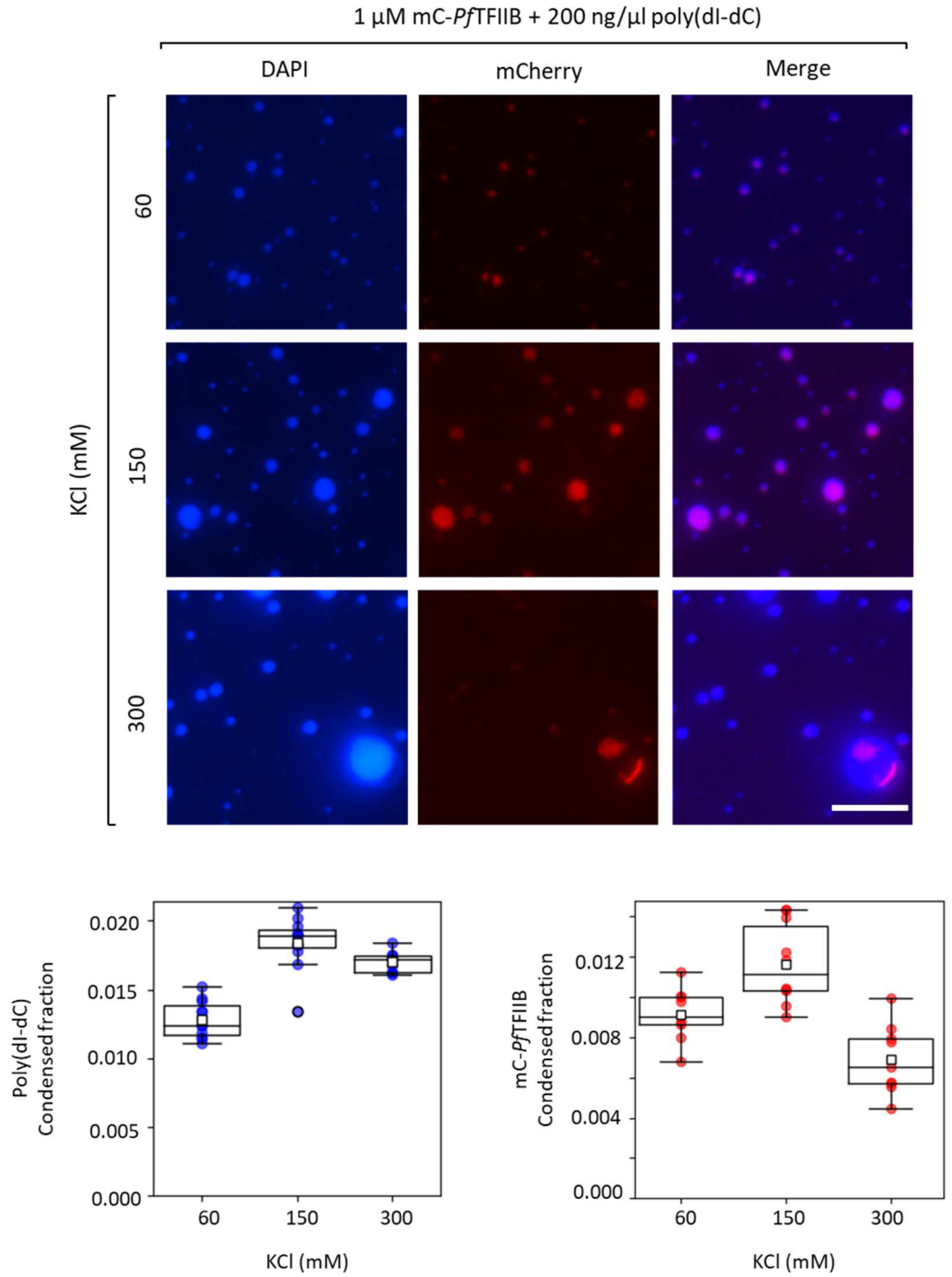


Figure 54 Increasing ionic strength results in exclusion of 6His:mCherry-*Pf*TFIIB from poly(dI-dC) condensates. *Top:* Representative widefield fluorescent microscopy images of condensates formed by 6His:mCherry-*Pf*TFIIB and poly(dI-dC) DNA at indicated KCl concentrations. Condensation reactions contained 400 nM 6His:mCherry-*Pf*TFIIB, 200 ng/µl poly(dI-dC), 10% PEG 20000, 5 mM magnesium acetate, indicated KCl concentrations and 50 µM DAPI to facilitate DNA visualisation. Poly(dI-dC) and 6His:mCherry-*Pf*TFIIB were mixed thoroughly and incubated for 5 min before the addition of condensation buffer. Mixed assemblies were visualised using the relevant fluorescent channels. Scale bar = 10 µm. *Bottom*: Boxplots showing quantification of condensed fraction of poly(dI-dC) or *Pf*TFIIB as a function of increasing KCl concentrations. Mean values are denoted by white boxes. Image analysis and quantification of condensed fraction was performed for 9-11 fields of view per condition.

5.2.3. Multiple regions within *Pf*TFIIB determine the properties of *Pf*TFIIB-DNA condensates

To delineate which regions of *Pf*TFIIB are required for the formation of mixed *Pf*TFIIB-DNA condensates, condensation reactions with *Pf*TFIIB deletion variants, including 6His:*Pf*TFIIB1-140-mCherry, lacking the structured C-terminal region, 6His:*Pf*TFIIB38-140-mCherry, lacking the *Plasmodium*-specific basic N-terminal 37 amino acid residues, or 6His:mCherry-*Pf*TFIIBΔ140, lacking the largely unstructured N-terminal portion of *Pf*TFIIB, were carried out. As shown in Fig. 55, all of the tested *Pf*TFIIB deletion mutants showed increased heterogeneity with regard to their distribution in the poly(dI-dC) condensates relative to full-length *Pf*TFIIB - forming complex multiphase condensates with a DNA-rich and a *Pf*TFIIB-rich phase. This heterogeneity was most evident in mixed condensates containing 6His:*Pf*TFIIB38-140-mCherry and 6His:mCherry-*Pf*TFIIBΔ140.

The differences in the morphology of condensates formed with full-length *Pf*TFIIB and the various N- and C-terminal *Pf*TFIIB deletion mutants, suggest that multiple regions of *Pf*TFIIB must contribute to interactions leading to the formation of homogenous *Pf*TFIIB-DNA condensates. Upon deletion of these regions, interactions between *Pf*TFIIB and DNA are weakened and are outcompeted by homotypic *Pf*TFIIB-*Pf*TFIIB and DNA-DNA interactions.

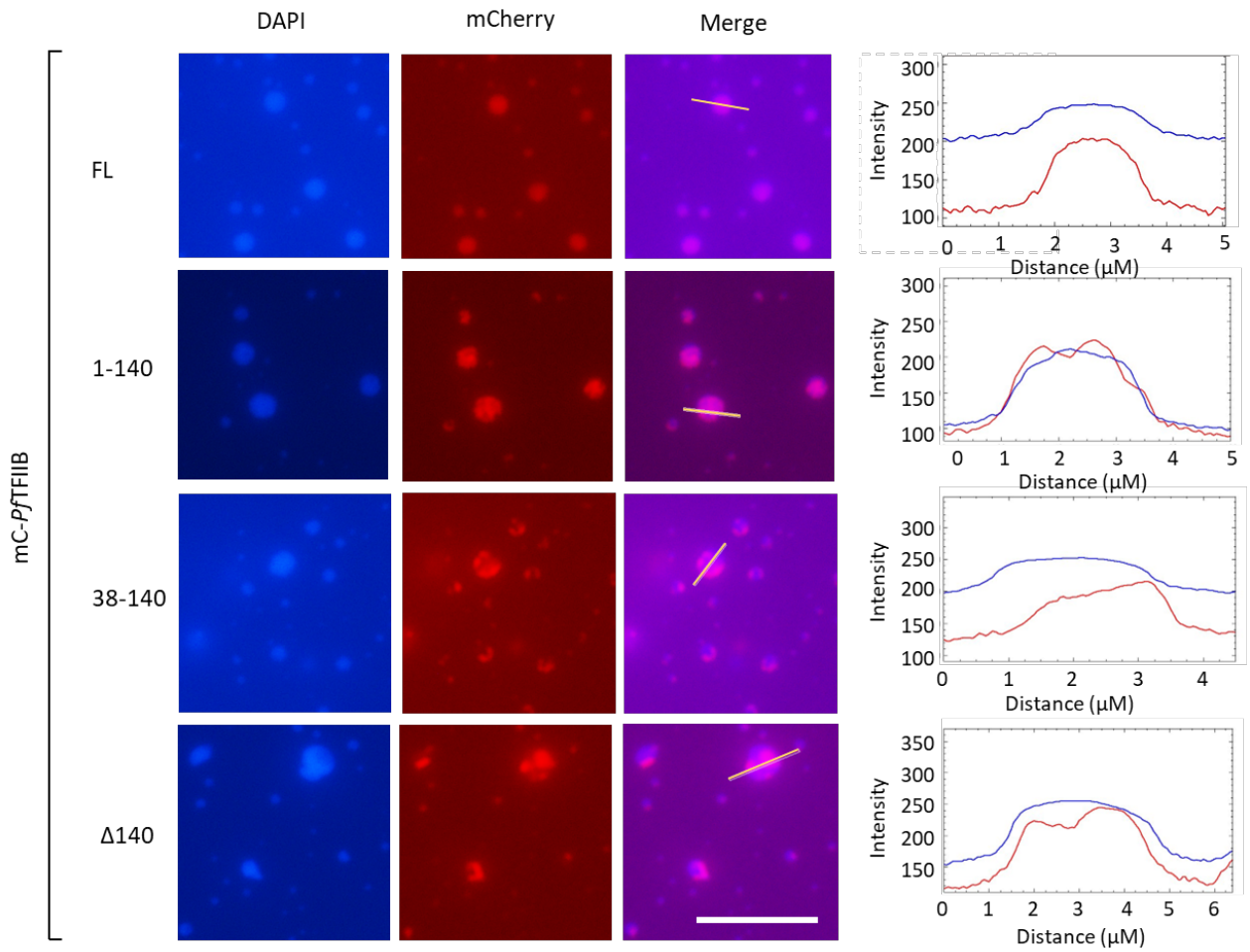


Figure 55. *Pf*TFIIB deletion variants form heterogenous condensates with poly(dI-dC) DNA. *Left:* Representative widefield fluorescent microscopy images of condensates formed by poly(dI-dC) DNA and 6His:mCherry-*Pf*TFIIB (FL), 6His:*Pf*TFIIB1-140-mCherry (1-140), 6His:*Pf*TFIIB38-140-mCherry (38-140) or 6His:mCherry-*Pf*TFIIBΔ140 as indicated. Condensation reactions contained 1 μM of the relevant mCherry-*Pf*TFIIB fusion protein, 200 ng/μl poly(dI-dC), 10% PEG 20000, 150 mM KCl, 5 mM magnesium acetate and 50 μM DAPI to facilitate DNA visualisation. DNA and protein were mixed thoroughly and incubated for 5 min before the addition of condensation buffer. Scale bar = 10 μm (white). *Right*: Intensity plot showing pixel intensity across the line indicated by the yellow bar. The blue line represents pixel intensity in the blue channel (DAPI-stained DNA), the red line indicates pixel intensity in the red channel (mCherry-*Pf*TFIIB).

5.3. RNA stimulates the formation of large *Pf*TFIIB assemblies with gel- or solid-like properties

Similarly to DNA, RNA has been shown to modulate the condensation behaviour of a variety of proteins (Shin and Brangwynne, 2017; Alemasova and Lavrik, 2022). RNA can impact the C_{sat} of a system by virtue of its ability to undergo phase separation and to partition interacting molecules into RNA condensates (Jain and Vale, 2017; Shin and Brangwynne, 2017; Alemasova and Lavrik, 2022). Alternatively, RNA and interacting proteins may undergo phase separation in a cooperative manner (Shin and Brangwynne, 2017; Alemasova and Lavrik, 2022). RNA can also modulate condensates properties such as size, composition, viscosity or propensity to form aggregates over time (Elbaum-Garfinkle *et al.*, 2015; Langdon *et al.*, 2018; Maharana *et al.*, 2018). Finally, RNA concentrations exceeding a favourable RNA-to-protein ratio can induce the dissolution of protein-RNA condensates due to an unfavourable charge balance (Banerjee *et al.*, 2017; Henninger *et al.*, 2021). Given that actively transcribed regions within the nucleus may accumulate very high local RNA concentrations, the potential regulatory role of RNA in modulating *Pf*TFIIB condensation properties was investigated.

To investigate the impact of RNA on *Pf*TFIIB condensation, 6His:mCherry-*Pf*TFIIB was incubated with increasing concentrations of total yeast RNA before addition of condensation buffer and assessment of 6His:mCherry-*Pf*TFIIB condensation by fluorescence microscopy. As shown in Fig. S11, total yeast RNA did not undergo condensation in isolation under these conditions. However, increasing concentrations of total yeast RNA in condensation reactions significantly increased the condensed fraction of 6His:mCherry-*Pf*TFIIB and resulted in the formation of large assemblies, several microns in length, composed of long strings of smaller spherical condensates (Fig. 56).

To assess whether 6His:mCherry-*Pf*TFIIB molecules within these condensates remain dynamic and can exchange with the surrounding dilute phase, FRAP experiments were carried out 1 h after condensate formation. As shown in Fig. 57, 6His:mCherry-*Pf*TFIIB condensates formed in the presence of RNA displayed minimal recovery after photobleaching, demonstrating that 6His:mCherry-*Pf*TFIIB molecules did not rapidly exchange with the surrounding dilute phase. This result is consistent with the morphology of the *Pf*TFIIB assemblies formed by *Pf*TFIIB on its own and suggests that spherical 6His:mCherry-*Pf*TFIIB condensates formed in the presence of RNA rapidly transition into a gel-like or solid state, thereby preventing the fusion of colliding small spheres into a single larger spherical condensate and instead resulting in the formation of assemblies that resemble beads on a string.

Together, these observations suggest that *Pf*TFIIB-RNA interactions stimulate *Pf*TFIIB condensation and, at the same time, promote transition of *Pf*TFIIB condensates into a gel- or solid-like state. Furthermore, the rapid hardening of *Pf*TFIIB condensates formed in the presence of RNA suggests that the interactions between *Pf*TFIIB and RNA are unlikely to compete with the interactions between *Pf*TFIIB molecules that drive condensate maturation, meaning that *Pf*TFIIB-RNA interactions and *Pf*TFIIB-*Pf*TFIIB interactions require distinct regions of *Pf*TFIIB.

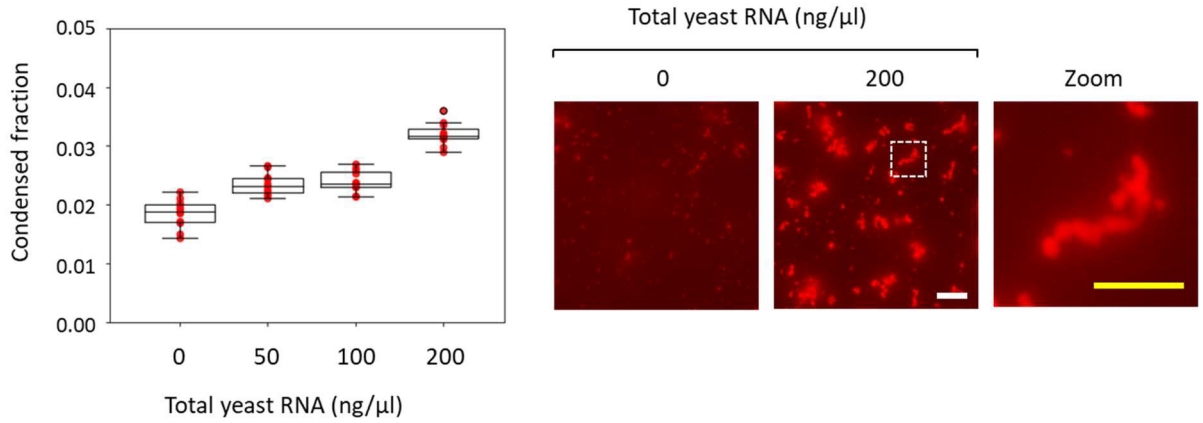


Figure 56. Total yeast RNA stimulates *Pf*TFIIB condensation and results in the formation of large condensate assemblies. *Left:* Boxplot showing quantification of condensed fraction of 6His:mCherry-*Pf*TFIIB in response to increasing concentrations of total yeast RNA. *Right*: Representative widefield fluorescent microscopy images of condensate assembled in the absence and presence of RNA. Condensation reactions contained 1 µM 6His:mCherry-*Pf*TFIIB, the indicated concentration of total yeast RNA, 10% PEG 3 000, 150 mM KCl and 5 mM magnesium acetate. Protein and RNA were incubated for 5 min before addition of condensation buffer. Scale bars = 10 µm (white), 5 µm (yellow). Image analysis and quantification of condensed fraction was performed for 11 fields of view per condition.

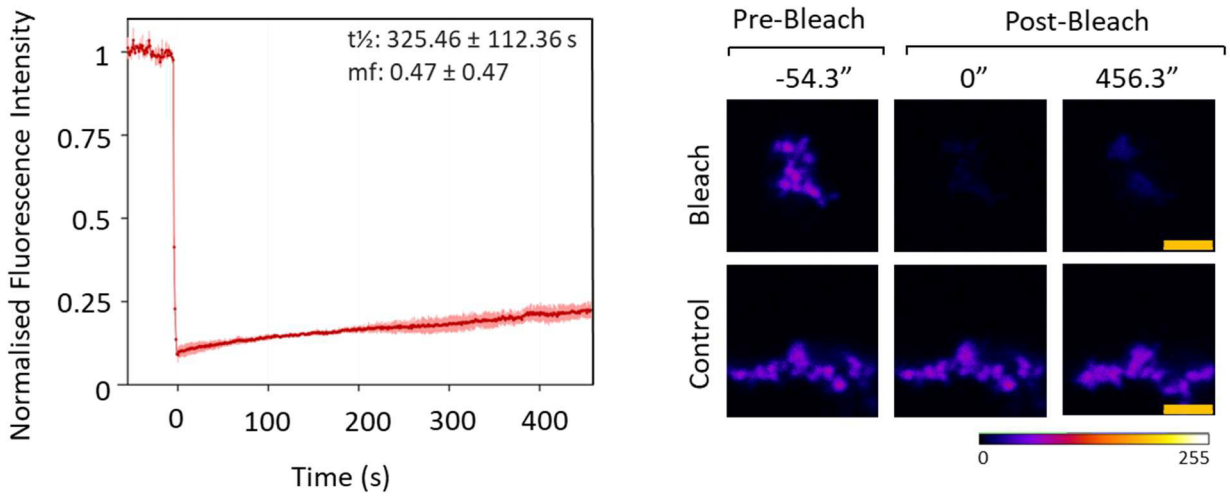


Figure 57. *Pf*TFIIB condensates formed in the presence of RNA display minimal fluorescence recovery after photobleaching. *Left:* Normalised fluorescence recovery after photobleaching of 6His:mCherry-*Pf*TFIIB condensates formed in the presence of total yeast RNA. Condensation reactions contained 1 µM 6His:mCherry-*Pf*TFIIB, 200 ng/µl total yeast RNA, 10% PEG 3000, 5 mM magnesium acetate and 150 mM KCl. Photobleaching was carried out 1-2 h after inducing condensation. Data presented as mean ± the standard deviation (n=3). Mobile fraction (mf) and half time for recovery of fluorescence after photobleaching (t½) was estimated using EasyFrap (Koulouras *et al*., 2018), are indicated. *Right*: Representative images of bleached (top) and control (bottom) condensates at indicated time points relative to completion of bleaching. Control: non-bleached condensate in the same field of view used for calculation of normalised fluorescence intensity. Images are coloured using the fire look up table in imageJ. Scale Bar = 2 µm (orange).

5.4. *Pf*TFIIB forms mixed condensates with *Pf*TBP and *Pf*TLP

Cellular condensates are comprised of tens to hundreds of proteins and RNA species and are often regulated by complex networks of interactions (Fong *et al.*, 2013; Banani *et al.*, 2016). Transcription condensates formed at super enhancers in Metazoans are no exception, and appear to be regulated by a combination of gene specific activators, coactivators, RNA Pol II and RNAs, which themselves are subject to temporal regulation and posttranslational modifications (Boehning *et al.*, 2018; Boija *et al.*, 2018; Sabari *et al.*, 2018; Guo *et al.*, 2019; Zamudio *et al.*, 2019; Henninger *et al.*, 2021). While it is not possible to recapitulate cellular conditions *in vitro*, the study of mixed protein condensate assembly using combinations of a small number of specific proteins can give first insights into heterotypic interactions between different molecular species that contribute to the formation of more complex biomolecular condensates.

Parallel work in the laboratory has shown that, like *Pf*TFIIB, *Pf*TBP and *Pf*TLP are strong drivers of biomolecular condensation *in vitro* (C_{sat} <1 µM; Pegram, 2022, Michowicz, in preparation). Furthermore, *Pf*TBP and *Pf*TLP localise to discrete puncta in blood stage *P. falciparum* parasites (Figs. 16 and 17), suggesting that *Pf*TBP and *Pf*TLP may be components of transcription condensates. Given that *Pf*TFIIB interacts with both *Pf*TBP and *Pf*TLP (Figs. 40 and 41), it was important to investigate assembly and properties of condensates formed by *Pf*TFIIB in combination with *Pf*TBP and *Pf*TLP.

5.4.1. *Pf*TFIIB and *Pf*TBP form mixed phase-separated condensates

To test if *Pf*TBP can form mixed condensates with *Pf*TFIIB, and to allow for independent visualisation of *Pf*TBP and *Pf*TFIIB in mixed condensates, mCherry-*Pf*TBP was combined with 6His:YFP-*Pf*TFIIB in condensation reactions.

As an important control, the condensation properties of 6His:YFP-*Pf*TFIIB were first determined and compared to the condensation properties of the 6His:mCherry-*Pf*TFIIB fusion protein. Similarly to 6His:mCherry-*Pf*TFIIB, 6His:YFP-*Pf*TFIIB formed small spherical assemblies at low concentrations (estimated C_{sat} = < 1 µM; see for example Fig. 58). However, compared to condensates formed with 6His:mCherry-*Pf*TFIIB, 6His:YFP-*Pf*TFIIB condensates displayed reduced fluorescence recovery in FRAP experiments within 30 min after condensate assembly (Fig. S13). This result suggests that, although YFP is unable to phase separate on its own under the conditions, specific YFP interactions contribute to 6His:YFP-*Pf*TFIIB condensate properties. Indeed, YFP has been reported to dimerise at high concentrations (Zacharias *et al.*, 2002), whereas mCherry appears to be a strictly monomeric

protein (Shemiakina *et al.*, 2012). Thus, the dimerization potential of YFP must be kept in mind when assessing the material properties of condensates formed by YFP fusion proteins.

As shown in Fig. 58, relatively small (< 1 μm) spherical condensates enriched in both 6His:YFP-*Pf*TFIIB and 6His:mCherry-*Pf*TBP were observed at a concentration at which both proteins undergo phase separation in isolation. In contrast, neither YFP or mCherry partitioned into *Pf*TBP or *Pf*TFIIB condensates on their own. This result confirms that interactions between *Pf*TBP and *Pf*TFIIB in the absence of DNA can drive partitioning of these proteins into mixed condensates. Interestingly, inclusion of both *Pf*TBP and *Pf*TFIIB in condensation reactions resulted in the formation of a small number of larger assemblies, consisting of one or more conjoined spherical condensates. These observations suggest that, while mixed *Pf*TBP-*Pf*TFIIB condensates were initially shaped by interfacial tension and therefore possessed liquid properties at the point of formation, they adopted a more gel- or solid-like material state over time. As a result, colliding condensates partially fused but did not to relax into larger spherical condensates.

FRAP experiments were performed to assess the mobility of 6His:YFP-*Pf*TFIIB and 6His:mCherry-*Pf*TBP within *Pf*TFIIB-*Pf*TBP condensates and their exchange with the dilute phase. As shown in Fig. 59, 6His:mCherry-*Pf*TBP fluorescence recovered slowly from 6% to an average of 41% within 180 s of photobleaching (mobile fraction: 0.82 ± 0.25; $t_{1/2}$: 134.84 ± 56.40 s). In contrast, a large proportion of 6His:YFP-*Pf*TFIIB fluorescence did not recover over the 180 s window of measurement, consistent with the material properties of condensates formed with isolated 6His:YFP-*Pf*TFIIB (Fig. S13). Thus, the hardening of mixed 6His:mCherry-*Pf*TBP-6His:YFP-*Pf*TFIIB condensates over time appears to be largely due to YFP-*Pf*TFIIB and contributing YFP-interactions. However, within these hardened condensates, 6His:mCherry-*Pf*TBP remains largely mobile and dynamically exchanges within the surrounding solution.

Given that both *Pf*TBP and *Pf*TFIIB bind to DNA (Sections 3.5 and 4.3), the formation of mixed condensates with *Pf*TFIIB and *Pf*TBP was next tested in the presence of poly(dI-dC) DNA, shown in EMSA experiments to be selectively bound by *Pf*TBP (Section 3.5.5). To this end 6His:YFP-*Pf*TFIIB, 6His:mCherry-*Pf*TBP and poly(dI-dC) were combined at concentrations at which they undergo condensation in isolation and at which they form mixed condensates with DNA. Under these conditions, spherical condensates were observed that were enriched in 6His:YFP-*Pf*TFIIB, 6His:mCherry-*Pf*TBP and poly(dI-dC) DNA (Fig. 60).

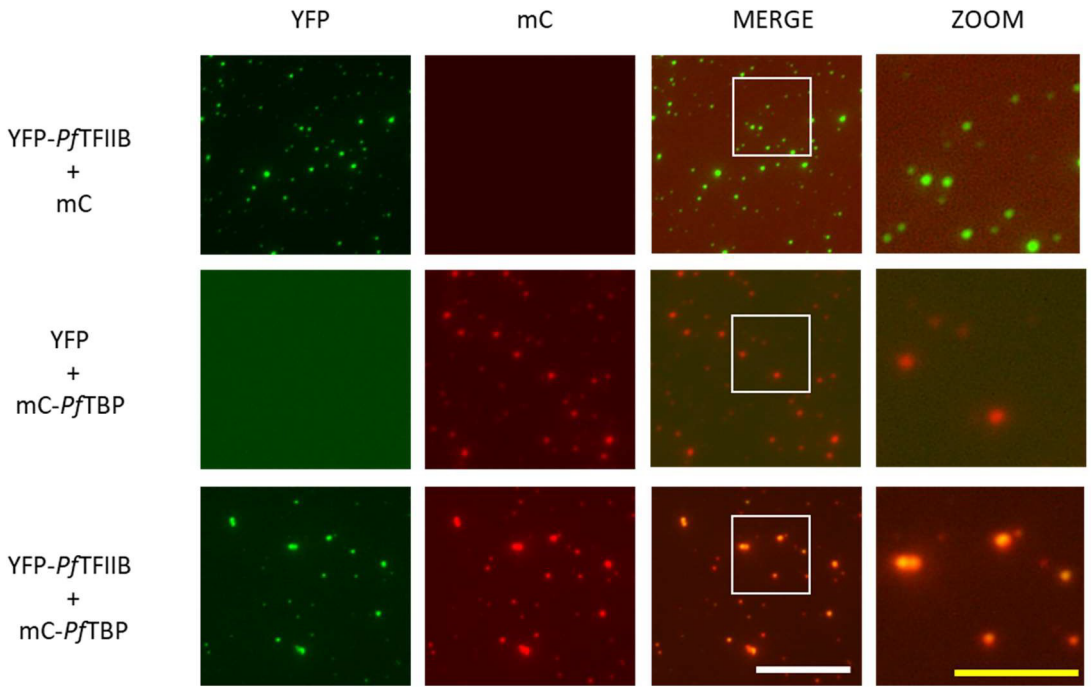


Figure 58. *Pf*TFIIB and *Pf*TBP partition into mixed spherical condensates. Representative widefield fluorescent microscopy images of condensation reactions containing either 0.5 µM 6His:YFP-*Pf*TFIIB and 1.3 µM 6His:mCherry (top row), 1.3 µM 6His:mCherry-*Pf*TBP and 0.5 µM 6His:YFP (middle row), or 0.5 µM YFP-*Pf*TFIIB and 1.3 µM mCherry-*Pf*TBP (bottom row). All reactions contained 10% PEG 20000, 150 mM KCl, and 5 mM magnesium acetate. Proteins were mixed thoroughly and incubated for 5 min before addition of condensation buffer. White boxes indicate areas of the merged image enlarged on the right (ZOOM). Scale bars = 10 µm (white), 5 µm (yellow).

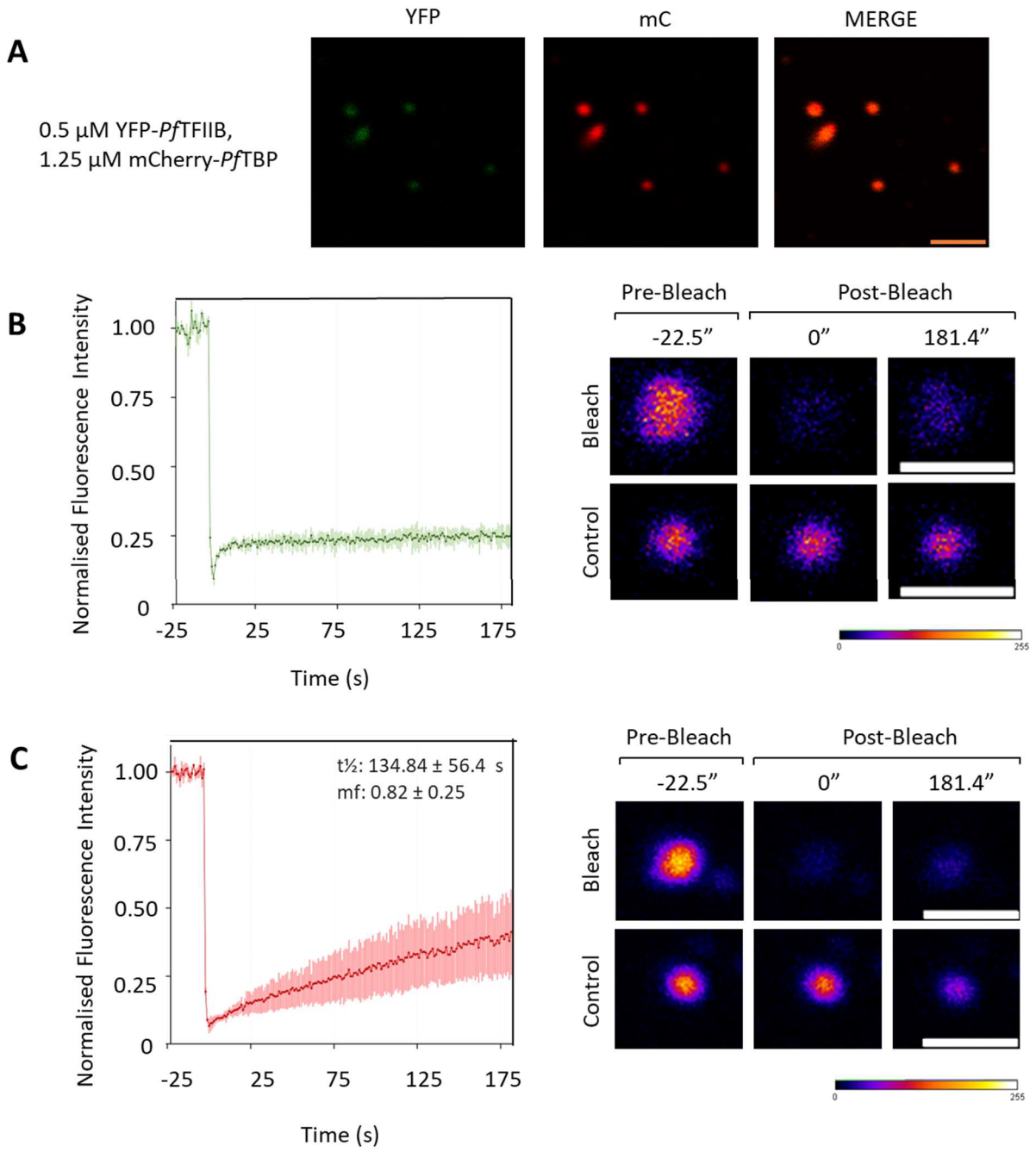


Figure 59. Fluorescence recovery after photobleaching in mixed condensates formed with *Pf*TFIIB and *Pf*TBP. (A) Representative confocal fluorescent microscopy images of *Pf*TFIIB-*Pf*TBP condensates formed with 0.5 µM 6His:YFP-*Pf*TFIIB and 1.25 µM 6His:mCherry-*Pf*TBP in the presence of 10% PEG 20000, 5 mM magnesium acetate and 150 mM KCl. **(B, C**) Normalised fluorescence recovery after photobleaching of 6His:YFP-*Pf*TFIIB **(B)** or 6His:mCherry-*Pf*TBP **(C)** in mixed condensates. Photobleaching was carried out 20-40 min after inducing condensation. Bleaching was carried out independently in each channel rather than in parallel in order to minimize photobleaching during imaging. *Left:* plot of normalised average fluorescence intensity. Data presented as mean ± the standard deviation (n=4). Mobile fraction (mf) and half time for recovery of fluorescence after photobleaching (t½) are indicated for values that could be reliably estimated using EasyFrap (Koulouras *et al.*, 2018). *Right:* Representative images of bleached (top) and control (bottom) condensates at indicated time points relative to completion of bleaching are shown. Non-bleached control condensates in the same field of view were used for the calculation of normalised fluorescence intensity. Images are coloured using the fire look up table in ImageJ. Scale bars = 1 µm (white), 2 µm (orange).

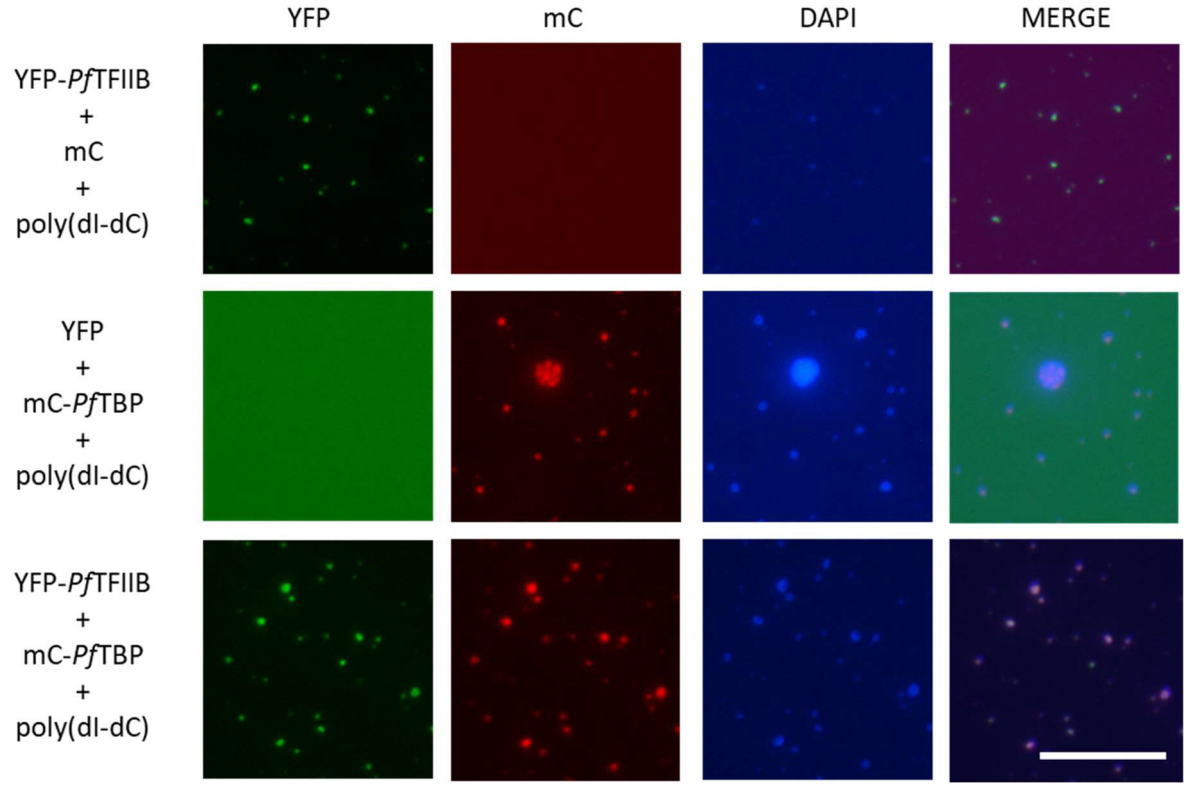


Figure 60. *Pf*TFIIB, *Pf*TBP and poly(dI-dC) DNA partition into mixed spherical condensates. Representative widefield fluorescent microscopy images of condensates. Condensation reactions contained 0.5 µM 6His:YFP-*Pf*TFIIB, 1.3 µM 6His:mCherry and 100 ng/µl poly(dI-dC) (top row), 1.3 µM 6His:mCherry-*Pf*TBP, 0.5 µM 6His:YFP and 100ng/µl poly(dI-dC) (middle row), or 0.5 µM 6His:YFP-*Pf*TFIIB, 1.3 µM 6His:mCherry-*Pf*TBP and 100 ng/µl poly(dI-dC) (bottom row). All reactions contained 10% PEG 20000, 150 mM KCl, 5 mM magnesium acetate. Proteins were mixed thoroughly and incubated for 5 min before addition of DNA. Protein/DNA mix was incubated for a further 5 min before the addition of condensation buffer. Scale bar = 10 µm (white).

5.4.2. The intrinsically disordered region *of Pf*TBP partitions into *Pf*TFIIB condensates

Recent work by our group has shown that the N-terminal *Pf*TBP IDR is sufficient to drive concentration-dependent biomolecular condensation *in vitro*, suggesting that condensation of *Pf*TBP may be driven, at least in part, by the *Pf*TBP IDR (Karamanof, 2022). To investigate whether *Pf*TFIIB and the *Pf*TBP IDR can assemble into mixed condensates, 6His:mCherry-*Pf*TFIIB and 6His:YFP-*Pf*TBP-IDR were added to condensation reactions at concentrations at which both proteins formed condensates in isolation. As shown in Fig. 61, 6His:mCherry-*Pf*TFIIB and 6His:YFP-*Pf*TBP-IDR formed small (~ 1 µm), homogenous mixed condensates upon addition of condensation buffer. In contrast, 6His:YFP did not partition into 6His:mCherry-*Pf*TFIIB condensates, nor did 6His:mCherry partition into 6His:*Pf*TBP-IDR-YFP condensates (Fig. 61). Thus, mixed condensate formation is dependent on interactions between the *Pf*TBP IDR and *Pf*TFIIB.

Mixed condensates formed with 6His:mCherry-*Pf*TFIIB and 6His:*Pf*TBP-IDR-YFP were spherical, suggesting they were shaped by interfacial tension and thus possessed liquid properties of condensates at the point of formation (Fig. 61). However, some strings composed of partially fused small spherical assemblies were present 1 h after initiation of condensation. This suggests that, similarly to condensates formed with *Pf*TFIIB and full-length *Pf*TBP (Fig. 58), condensates formed by *Pf*TFIIB and *Pf*TBP IDR transition into a gel- or solid-like state over time.

To further assess the material properties of these mixed condensates, FRAP experiments were performed. As shown in Fig. 62, 6His:*Pf*TBP-IDR-YFP recovered only just over 20% of its initial fluorescence intensity within just over 100 s of condensate bleaching (mobile fraction: 0.19 ± 0.08, estimated $t_{1/2}$: 13.96 ± 10.38 s), indicating that a large fraction of 6His:*Pf*TBP-IDR-YFP molecules are immobile within mixed condensates. In contrast, 6His:mCherry-*Pf*TFIIB fluorescence recovered to just under 50% of its initial fluorescence intensity within the same time period (mobile fraction: 0.96 +-0.08, estimated $t_{1/2}$: 103.31 ± 3.85 s). Interestingly the rate of *Pf*TFIIB fluorescence recovery following mixed condensate bleaching is slightly higher than the rate of fluoresce recovery for *Pf*TFIIB condensates formed by *Pf*TFIIB alone (mobile fraction: 1 ± 0.01; t½: 180.3 ± 0 s, Fig. 48), suggesting that *Pf*TBP-IDR interactions with *Pf*TFIIB may compete with *Pf*TFIIB-*Pf*TFIIB interactions that drive condensate hardening, even though *Pf*TBP-IDR itself is largely immobile within mixed condensates. Overall, the slow recovery of *Pf*TFIIB and a large fraction of immobile *Pf*TBP-IDR molecules in mixed condensates are consistent with the observation of stalled fusion events between condensates and suggest that mixed condensates rapidly transition into a gel-like state.

Condensation of 6His:*Pf*TBP-IDR-YFP required significantly higher protein concentrations (C_{sat} > 4 µM) compared to 6His:mCherry-*Pf*TFIIB (C_{sat} < 0.125 µM; compare Fig. 63A and 63B). This presented an opportunity to test whether the *Pf*TBP IDR partitions into *Pf*TFIIB condensates at a concentration at which the *Pf*TBP IDR does not phase separate on its own. To this end, condensation reactions were carried out with 6His:mCherry-*Pf*TFIIB at a concentration above its C_{sat} (0.5 µM, Fig. 63A) and 6His:*Pf*TBP-IDR-YFP at a concentration below its $C_{sat.}$ (4 µM, Fig. 63B). As shown in Fig. 63C, 6His:*Pf*TBP-IDR-YFP indeed partitioned into, and was significantly enriched in, 6His:mCherry-*Pf*TFIIB condensates. This result suggests that *Pf*TFIIB can function as a scaffold protein to drive partitioning of interacting factors into transcription condensates.

Unlike full-length *Pf*TBP (Fig. 60) and *Pf*TFIIB (Fig. 64, top row), the *Pf*TBP-IDR, lacking the structured *Pf*TBP DNA-binding domain, does not partition into DNA condensates (Figs. 64, panels A and B). To examine whether the *Pf*TBP-IDR partitions into condensates assembled with *Pf*TFIIB and dsDNA, condensation reactions were carried out with 6His:mCherry-*Pf*TFIIB, 6His:*Pf*TBP-IDR-YFP and poly(dI-dC) DNA. As shown in Fig. 64, 6His:*Pf*TBP-IDR-YFP was indeed enriched in condensates formed by poly(dI-dC) in the presence of 6His:mCherry-*Pf*TFIIB (Fig. 64, panels A and B). However, within these condensates, 6His:mCherry-*Pf*TFIIB, 6His:*Pf*TBP-IDR-YFP and poly(dI-dC) were not homogenously mixed. Instead, 6His:mCherry-*Pf*TFIIB, 6His:*Pf*TBP-IDR-YFP and poly(dI-dC) were heterogeneously distributed within mixed condensates, which possessed a central DNA-rich core encapsulated by a *Pf*TFIIB-*Pf*TBP-IDR-rich shell. (Fig. 64). This result suggests that, upon concentration of the *Pf*TBP-IDR into condensates formed by *Pf*TFIIB and DNA, *Pf*TFIIB-*Pf*TBP-IDR interactions compete with *Pf*TFIIB-DNA interactions, resulting in partial de-mixing and the formation of two condensate compartments, one enriched in poly(dI-dC) DNA and another enriched in *Pf*TFIIB and the *Pf*TBP IDR.

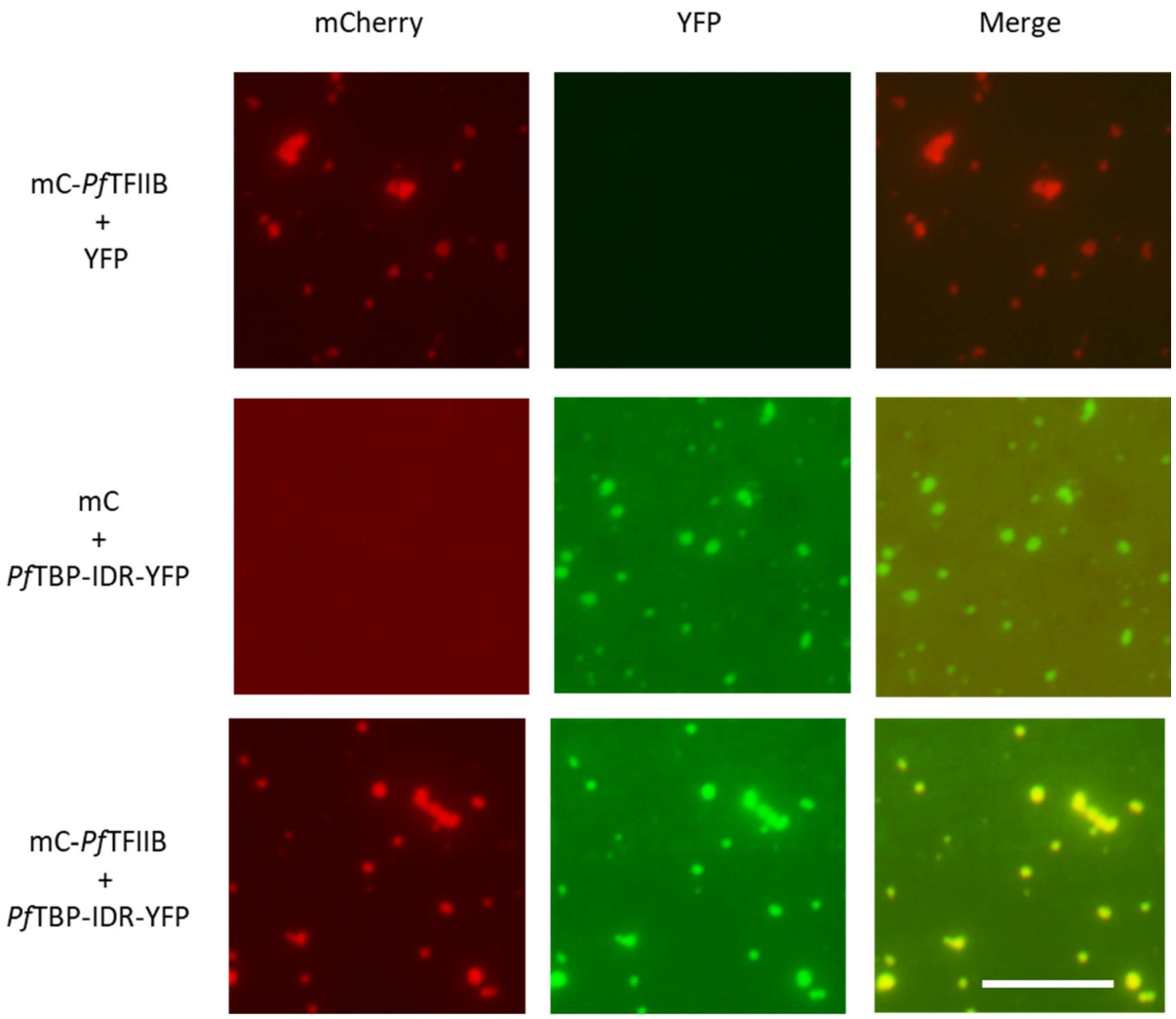


Figure 61. *Pf*TFIIB and the N-terminal *Pf*TBP IDR form homogenous mixed spherical condensates. Representative widefield fluorescent microscopy images of condensates. Condensation reactions contained 0.5 µM 6His:mCherry-*Pf*TFIIB and 7 µM 6His:YFP (top row), 7 µM 6His:*Pf*TBP-IDR-YFP and 0.5 µM 6His:mCherry (middle row), or 0.5 µM 6His:mCherry-*Pf*TFIIB and 7 µM 6His:*Pf*TBP-IDR-YFP (bottom row). All reactions contained 10% PEG 20000, 150 mM KCl, and 5 mM magnesium acetate. Proteins were mixed thoroughly and incubated for 5 min before the addition of condensation buffer. Scale bar = 10 µm (white).

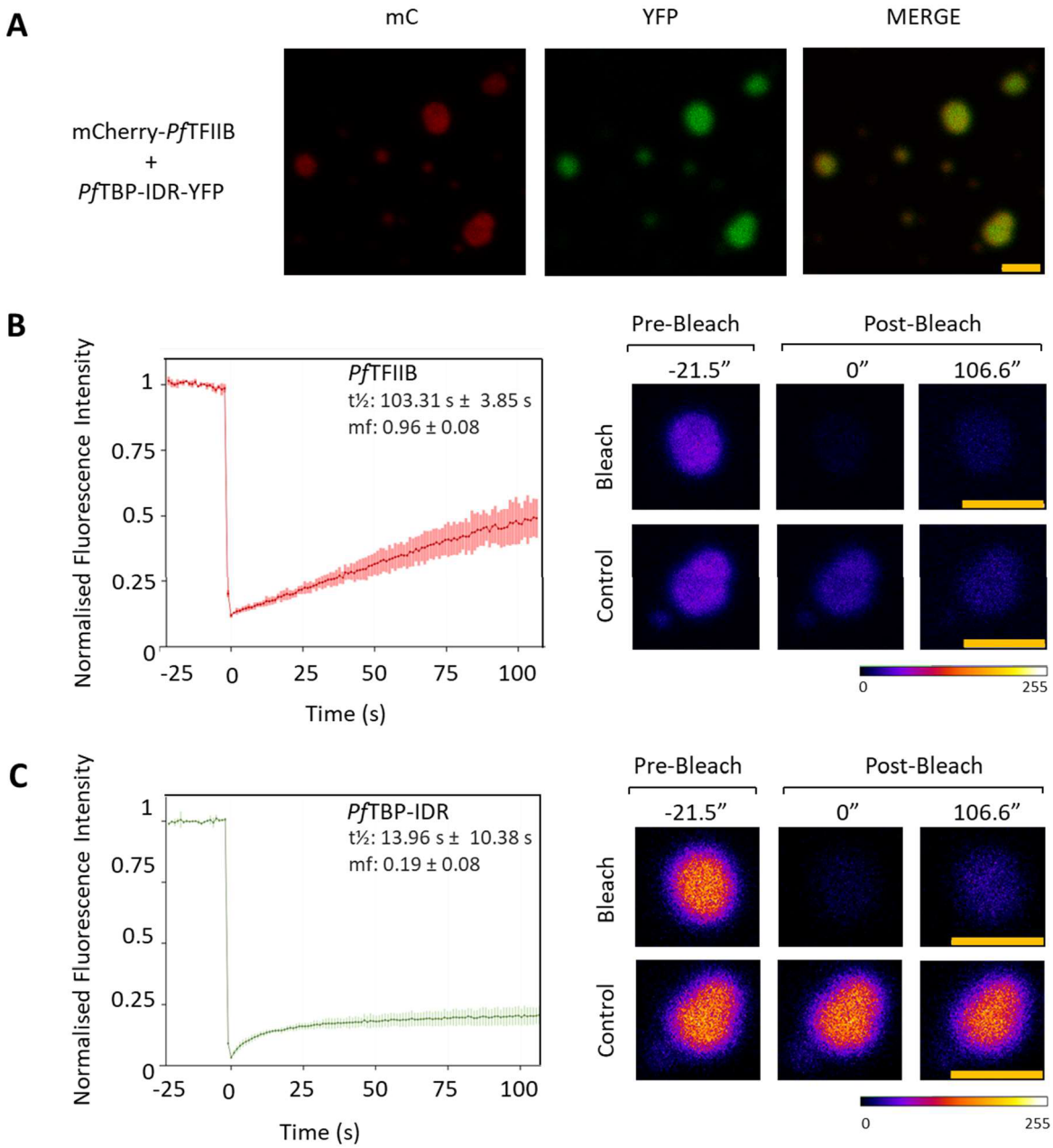


Figure 62. Fluorescence recovery after photobleaching of mixed condensates formed with *Pf*TFIIB and the *Pf*TBP IDR. (A) Representative images of *Pf*TFIIB-*Pf*TBP-IDR mixed condensates formed with 0.5 µM 6His:mCherry-*Pf*TFIIB and 7 µM 6His:*Pf*TBP-IDR-YFP in the presence of 10 % PEG 20000, 5 mM magnesium acetate and 150 mM KCl. **(B, C)** Normalised fluorescence recovery after photobleaching of 6His:mCherry-*Pf*TFIIB **(B)** or 6His:*Pf*TBP-IDR-YFP **(C)** in mixed condensates. Photobleaching was carried out 1 - 2 h after inducing condensation. 20 pre-bleach measurements were recorded before the condensate was bleached at a wavelength of 561 nm (mC-*Pf*TFIIB, panel B) or 488 nm (*Pf*TBP-IDR, panel C). *Left:* plots of normalised average fluorescence intensity of bleached condensates over time. Data presented as mean ± the standard deviation (n=3). Mobile fraction (mf) and half time for recovery of fluorescence after photobleaching (t½) are indicated for values that could be reliably estimated using EasyFrap (Koulouras *et al.,* 2018). *Right:* Representative images of bleached (top) and control (bottom) condensates at indicated time points relative to completion of bleaching. Non-bleached control condensate in the same field of view were used for the calculation of normalised fluorescence intensity. Images are coloured using the fire look up table in ImageJ. Scale bar = 2 µm (orange).

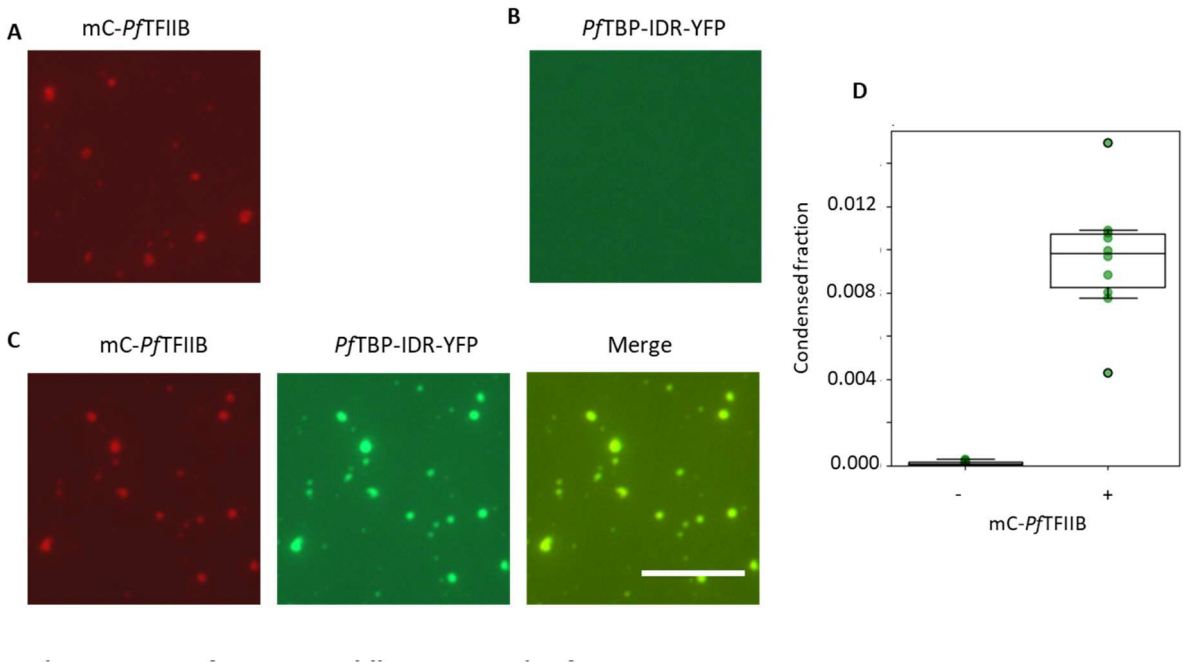


Figure 63. Partitioning of the *Pf*TBP IDR into *Pf*TFIIB condensates. Representative widefield fluorescent microscopy images of condensates. Condensation reactions contained 0.5 µM 6His:mCherry-*Pf*TFIIB **(A)**, 4 µM 6His:*Pf*TBP-IDR-YFP **(B)**, or 0.5 µM 6His:mCherry-*Pf*TFIIB and 4 µM 6His:*Pf*TBP-IDR-YFP **(C)**. All reactions contained 10% PEG 20000, 150 mM KCl and 5 mM magnesium acetate. Proteins were mixed thoroughly and incubated for 5 min before the addition of condensation buffer. Scale bar = 10 µm. Boxplot showing quantification of the 6His:*Pf*TBP-IDR-YFP condensed fraction in the presence or absence of 6His:mCherry-*Pf*TFIIB. Image analysis and quantification of condensed fraction was performed for 10 fields of view per condition.

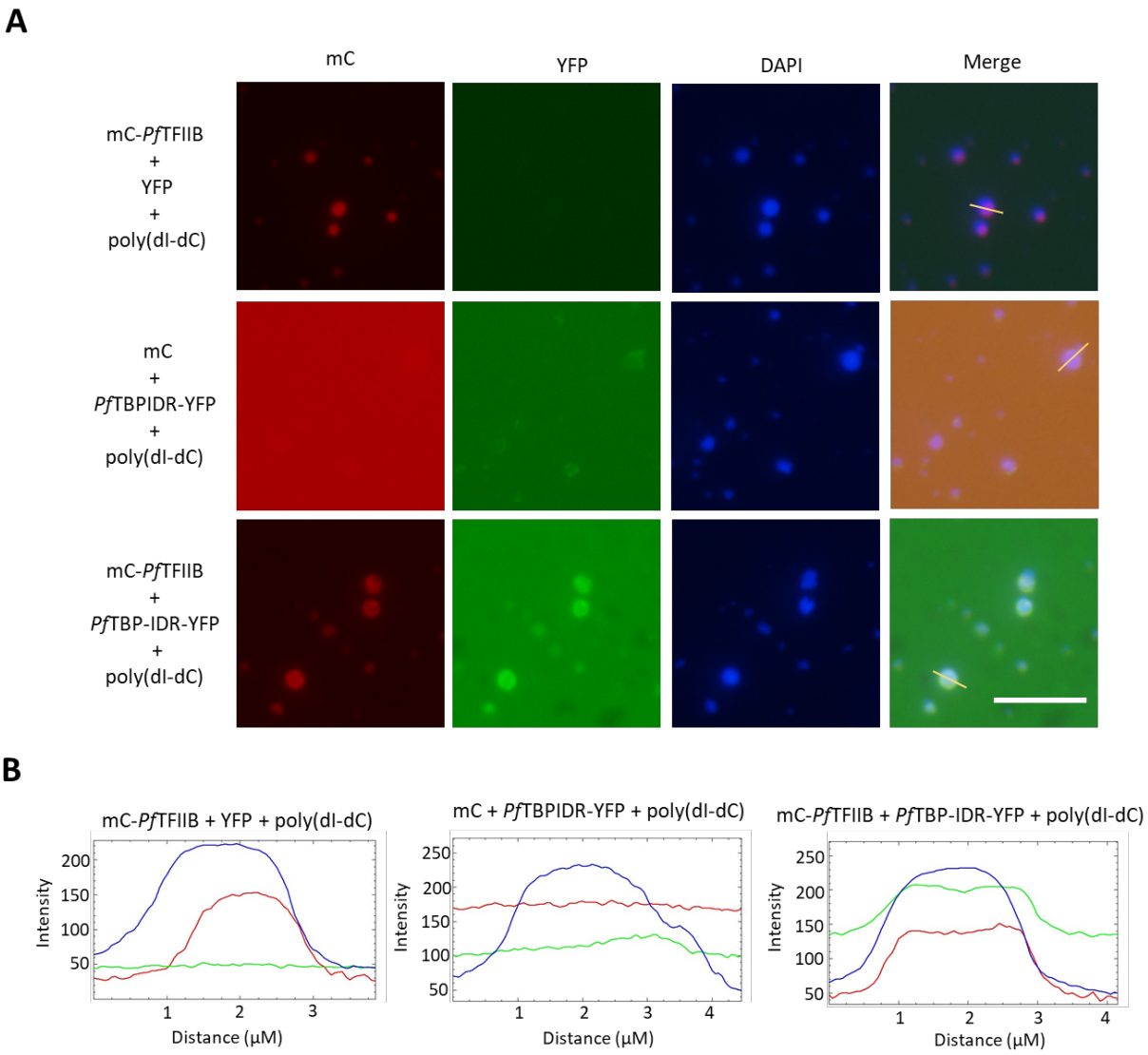


Figure 64. *Pf*TFIIB facilitates partitioning of the *Pf*TBP IDR into poly(dI-dC) condensates. **(A)** Representative images of condensates. Condensation reactions contained 0.5 µM 6His:mCherry-*Pf*TFIIB, 7 µM 6His:YFP and 100 ng/µl poly(dI-dC) (top row), *7* µM 6His:*Pf*TBP-IDR-YFP, 4 µM 6His:mCherry and 100 ng/µl poly(dI-dC) (middle row), or 0.5 µM 6His:mCherry-*Pf*TFIIB, 4 µM 6His:*Pf*TBP-IDR-YFP and 100 ng/µl poly(dI-dC) (bottom row). All reactions contained 10% PEG 20000, 150 mM KCl, 5 mM magnesium acetate. Proteins were mixed thoroughly and incubated for 5 min before addition of DNA. The protein/DNA mix was incubated for a further 5 min before addition of condensation buffer. Scale bar= 10 µm (white). **(B)** Intensity plots showing pixel intensity measured for the excitation of DAPI-stained DNA (blue), mCherry or mCherry-*Pf*TFIIB (red) and *Pf*TBP-IDR-YFP or YFP (green). Measurements were taken across selected condensates along the yellow line shown in the merged images on the right.

5.4.3. *Pf*TFIIB and *Pf*TLP form mixed phase-separated condensates

*Pf*TLP is also a strong driver of biomolecular condensation (C_{sat} < 1 μM) and forms condensates with gel-like properties (Michowicz, in preparation). To examine whether *Pf*TFIIB and *Pf*TLP form mixed phase-separated condensates, 6His:mCherry-*Pf*TFIIB and 6His:YFP-*Pf*TLP were incubated at the maximum achievable concentration in the presence or absence of poly(dI-dC) dsDNA before the addition of condensation buffer containing molecular crowder. In the absence of DNA, 6His:mCherry-*Pf*TFIIB and 6His:YFP-*Pf*TLP concentrated in assemblies that appeared to be comprised of smaller conjoined condensates, likely formed by incomplete fusion after collision. Unfortunately, due to the inherent challenges associated with expressing *P. falciparum* proteins (Birkholtz *et al.*, 2008), the concentration of 6His:mCherry-*Pf*TFIIB and 6His:YFP-*Pf*TLP preparations were very low. This limited the achievable concentration of 6His:mCherry-*Pf*TFIIB and 6His:YFP-*Pf*TLP in condensation assays and, consequently, the size of mixed condensates, which were too small to determine whether 6His:mCherry-*Pf*TFIIB and 6His:YFP-*Pf*TLP were homogenously mixed.

FRAP experiments revealed slow rates of 6His:mCherry-*Pf*TFIIB exchange between mixed condensates and the dilute phase (mobile fraction= 1 ± 0.0; t½ = 175.52 ± 32.1 s), similar to the exchange rates for condensates formed by 6His:mCherry-*Pf*TFIIB in isolation (compare Figs. 48 and Fig. 66B). Furthermore, very little 6His:YFP-*Pf*TLP was exchanged between condensates and the surrounding solution, indicating that, at the time of the experiment, a large proportion of 6His:YFP-*Pf*TLP within the condensates was immobile. The results of FRAP experiments are consistent with the irregular morphology of *Pf*TFIIB-*Pf*TLP mixed condensates and suggest that mixed condensates have gel-like properties at the time point of observation.

Both *Pf*TFIIB and *Pf*TLP are DNA-binding proteins (Sections 3.5.6., 3.5.7. and 4.3). It was therefore important to test whether mixed condensate assembly is affected by the presence of DNA. To this end, 6His:mCherry-*Pf*TFIIB, 6His:YFP-*Pf*TLP and poly(dI-dC) dsDNA were combined in condensation reactions at concentrations at which each component undergoes phase separation in isolation. As shown in Fig. 65B and Fig. 67A, 6His:mCherry-*Pf*TFIIB, 6His:YFP-*Pf*TLP and poly(dI-dC) partitioned into phase-separated assemblies with a distinctly spherical morphology, enriched in all three components. Analysis of fluorescence recovery following *Pf*TFIIB-*Pf*TLP condensate bleaching in FRAP experiments showed that 6His:mCherry-*Pf*TFIIB exchanged at a higher rate with the dilute phase in poly(dI-dC)-containing condensates (mobile fraction= 1 ± 0.0; t½ = 81.47 ± 42.43 s) compared to those formed in the absence of poly(dI-dC) (compare Fig. 66B with Fig. 67B), similar to condensates formed with *Pf*TFIIB in the absence of *Pf*TLP (compare Fig.48 and Fig. 52). However, the exchange of 6His:YFP-*Pf*TLP between mixed condensates and the dilute phase was not increased in

the presence of poly(dI-dC) (compare Fig. 66C with Fig. 67C). Thus, DNA interactions compete with interactions between *Pf*TFIIB molecules that lead to transition of *Pf*TFIIB assemblies into a gel- or solid-like state but do not affect corresponding interactions between *Pf*TLP molecules or between *Pf*TLP molecules and *Pf*TFIIB molecules within the context of phase-separated condensates.

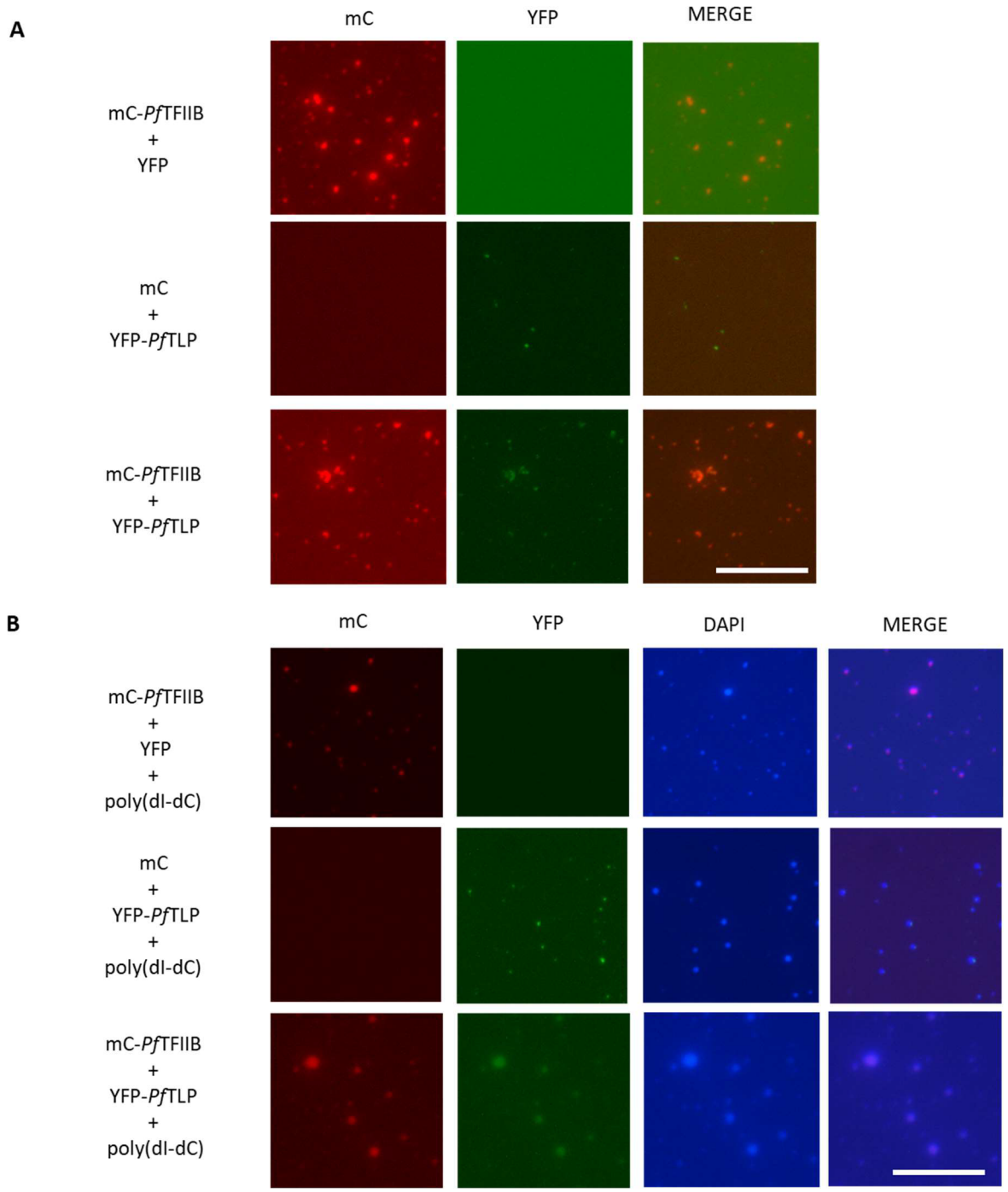


Figure 65. *Pf*TFIIB and *Pf*TLP form mixed condensates in the absence and presence of poly(dI-dC) dsDNA. Representative widefield fluorescence microscopy images of condensates formed in the absence **(A)** or presence **(B)** of 100 ng/µl poly(dI-dC). Condensation reactions contained 0.5 µM 6His:mCherry-*Pf*TFIIB and 0.1 µM 6His:YFP (top row), 0.1 µM 6His:YFP-*Pf*TLP and 0.5 µM 6His:mCherry (middle row), or 0.5 µM 6His:mCherry-*Pf*TFIIB and 0.1 µM 6His:YFP-*Pf*TLP (bottom row). All reactions contained 10% PEG 20000, 150 mM KCl and 5 mM magnesium acetate. Proteins were mixed thoroughly and pre-incubated for 5 min before addition of poly(dI-dC). The protein/DNA mix was incubated for a further 5 min before addition of condensation buffer. Scale bar = 10 µm (white).

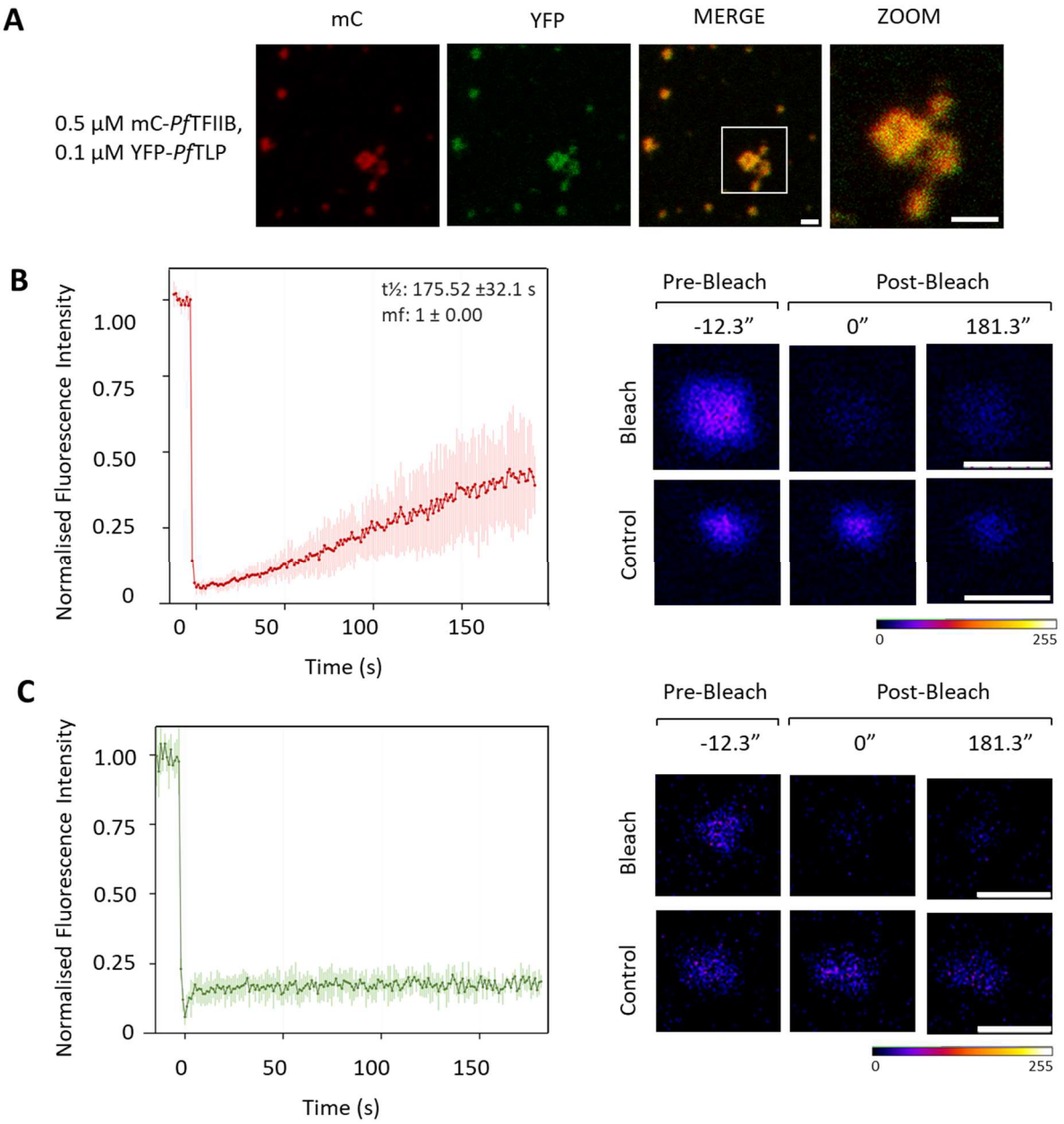


Figure 66. Fluorescence recovery after photobleaching of mixed condensates formed with *Pf*TFIIB and *Pf*TLP in the absence of DNA. (A) Representative confocal fluorescent microscopy images of *Pf*TFIIB-*Pf*TLP mixed condensates formed with 0.5 µM 6His:mCherry-*Pf*TFIIB and 0.1 µM 6His-YFP-*Pf*TLP in the presence of 10% PEG 20000, 5 mM magnesium acetate and 150 mM KCl. White boxes indicate areas of the merged image enlarged on the right (ZOOM). **(B, C)** Normalised fluorescence recovery after photobleaching of 6His:mCherry-*Pf*TFIIB (B) or 6His:YFP-*Pf*TLP (C) in mixed condensates 1-2 h after initiation of condensation. *Left*: Quantification of normalised average fluorescence intensity. Data presented as mean ± the standard deviation (n=4). Mobile fraction (mf) and half time for recovery of fluorescence after photobleaching (t½) are indicated for values that could be reliably estimated using EasyFrap (Koulouras *et al.,* 2018). *Right:* Representative images of bleached (top) and control (bottom) condensates at indicated time points relative to completion of bleaching. Non-bleached control condensate in the same field of view were used for the calculation of normalised fluorescence intensity. Images are coloured using the fire look up table in ImageJ. Scale Bar = 1 µm (white).

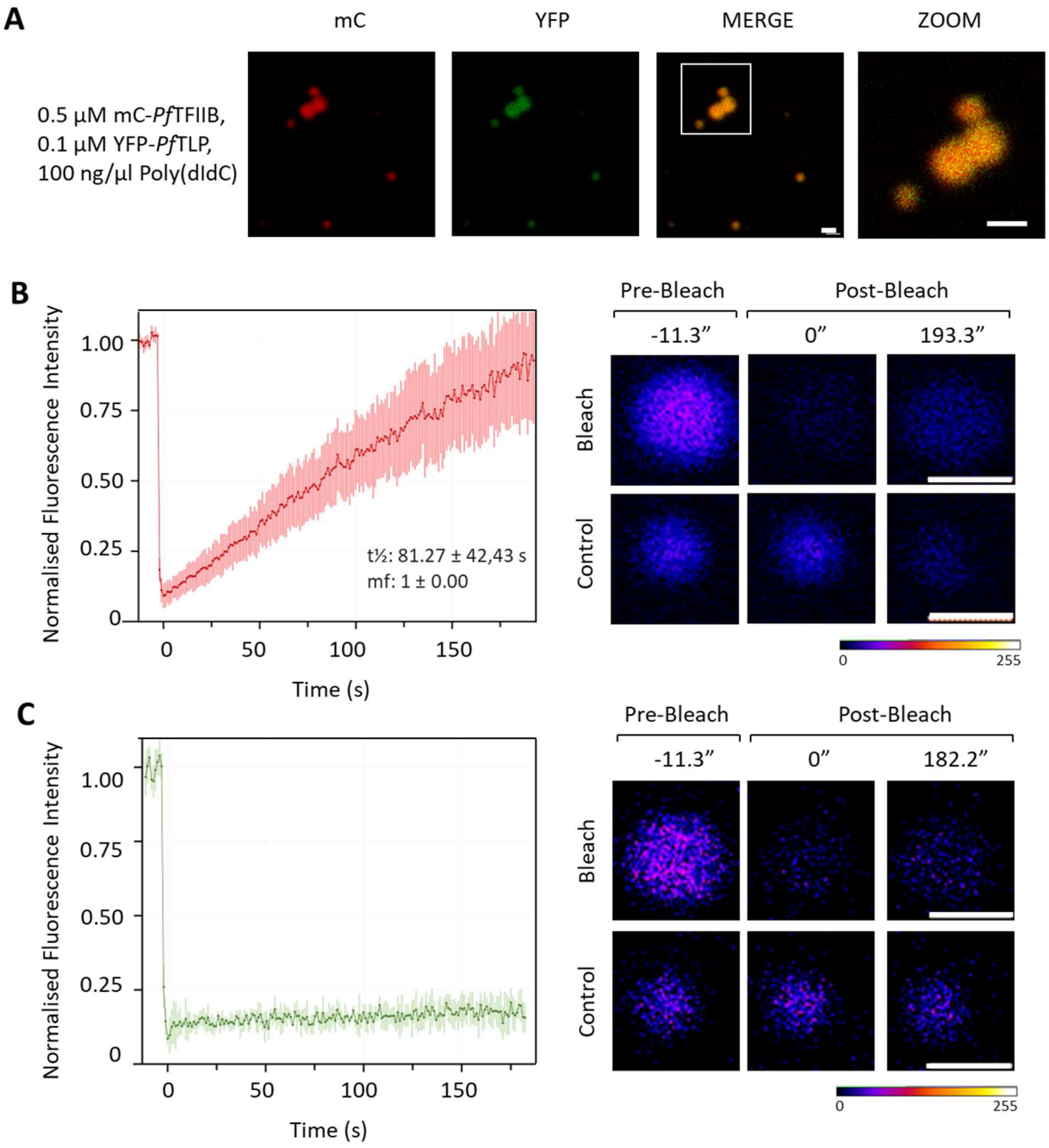


Figure 67. Fluorescence recovery after photobleaching of *Pf*TFIIB-*Pf*TLP mixed condensates in the presence of DNA. (A) Representative confocal fluorescent microscopy images of *Pf*TFIIB-*Pf*TLP mixed condensates formed with 0.5 µM 6His:mCherry-*Pf*TFIIB, 0.1 µM 6His:YFP-*Pf*TLP and 100 ng/µl poly(dI-dC) at 10% PEG 20000, 5 mM magnesium acetate, 150 mM KCl. White boxes indicate areas of the merged image enlarged on the right (ZOOM) **(B, C)** Normalised fluorescence recovery after photobleaching of 6His:mCherry-*Pf*TFIIB (B) or 6His:YFP-*Pf*TLP (C) in mixed condensates following 1-2 h incubation of condensation reactions. *Left*: Quantification of normalised average fluorescence intensity. Data presented as mean ± the standard deviation (n=6). Mobile fraction (mf) and half time for recovery of fluorescence after photobleaching (t½) are indicated for values that could be reliably estimated using EasyFrap (Koulouras et al., 2018). *Right:* Representative images of bleached (top) and control (bottom) condensates at indicated time points relative to completion of bleaching. Non-bleached control condensate in the same field of view were used for the calculation of normalised fluorescence intensity. Images are coloured using the fire look up table in ImageJ. Scale Bar = 1 µm (white).

5.4.4. An intrinsically disordered region within the DNA-binding domain of *Pf*TLP (*Pf*TLP-IDR1) is sufficient to drive partitioning into *Pf*TFIIB condensates

Unlike *Pf*TBP, *Pf*TLP does not possess a long intrinsically disordered N-terminal domain but contains two intrinsically disordered regions, *Pf*TLP-IDR1 and *Pf*TLP-IDR2, inserted between the S3 and S4 and S3' an S4' β-strands, respectively, of the *Pf*TLP saddle-shaped DNA-binding domain (Sections 3.1.2 and 3.1.3.). Previous work in the laboratory identified *Pf*TLP-IDR1 as a potent driver of LLPS *in vitro* (van der Linden, 2019, Michowicz, in preparation). To test whether *Pf*TFIIB and the *Pf*TLP-IDR1 form mixed phase-separated condensates *in vitro*, 6His:mCherry-*Pf*TFIIB and 6His:YFP-*Pf*TLP-IDR1 were combined at concentrations at which both proteins form condensates in isolation. As shown in Figs. 68 and Fig. 69A, 6His:YFP-*Pf*TLP-IDR1 and 6His:mCherry-*Pf*TFIIB formed spherical assemblies, in which both protein components were enriched and homogenously distributed, whereas the negative controls 6His:YFP and 6His:mCherry did not partition into condensates and 6His:mCherry did not partition into 6His:YFP-*Pf*TLP-IDR1 condensates (Fig. 68).

Photobleaching of mixed condensates resulted in rapid recovery of 6His:YFP-*Pf*TLP-IDR1 fluorescence to approximately 50% of the initial fluorescence within 37 s of condensate bleaching (mobile fraction = 0.6 ± 0.1, t½ = 15.39 s ± 4.99 s, Fig. 69C). In contrast, 6His:mCherry-*Pf*TFIIB fluorescence recovered more slowly to approximately 50% of the original fluorescence intensity within 140 s of condensate bleaching (estimated mobile fraction = 0.99 ± 0.02, t½ = 157.27 s ± 32.39 s, Fig. 69B), similar to recovery rate of *Pf*TFIIB fluorescence in condensates formed with *Pf*TFIIB by itself (Fig. 48, see fluorescence recovery 1-2 h after inducing condensation). These results suggest liquid-like properties of condensates assembled with *Pf*TFIIB and the *Pf*TLP-IDR1. In addition, these results show that while *Pf*TLP-IDR molecules themselves are highly dynamic, *Pf*TFIIB-*Pf*TLP-IDR1 interactions do not compete with interactions between *Pf*TFIIB molecules that dictate exchange of *Pf*TFIIB molecules between condensates and the dilute phase.

*Pf*TLP-IDR1 partitions, like *Pf*TFIIB, with poly(dI-dC) into homogenously mixed phase-separated condensates (Fig. 70, Michowicz, in preparation). It was therefore not surprising that 6His:mCherry-*Pf*TFIIB and 6His:YFP-*Pf*TLP-IDR1 also formed homogenous assemblies containing poly(dI-dC) (Fig. 70). Taken together, these observations demonstrate that the *Pf*TLP-IDR1 is sufficient to drive partitioning into *Pf*TFIIB condensates, both in the absence and presence of DNA. Furthermore, the homogenous distribution of 6His:mCherry-*Pf*TFIIB, 6His:YFP-*Pf*TLP-IDR1 and DNA in mixed condensates, suggests that *Pf*TFIIB-DNA, *Pf*TLP-DNA and *Pf*TLP-*Pf*TFIIB interactions are not competitive and are likely mediated by distinct protein regions.

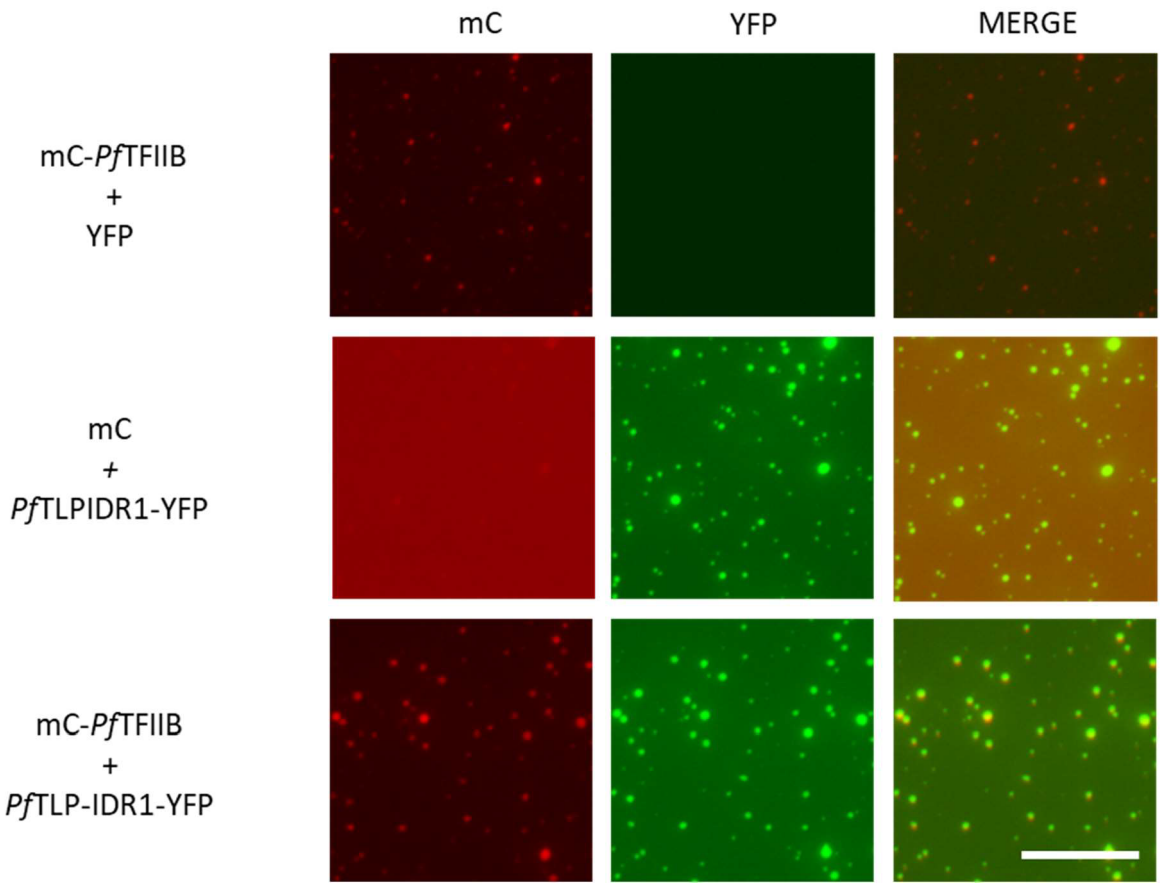


Figure 68. *Pf*TFIIB and the *Pf*TLP-IDR1 form homogenously mixed condensates. Representative widefield fluorescent microscopy images of condensates. Condensation reactions contained 0.5 µM 6His:mCherry-*Pf*TFIIB and 2 µM 6His:YFP (top row), *2* µM 6His:*Pf*TLP-IDR1-YFP and 4 µM 6His:mCherry (middle row), or 0.5 µM 6His:mCherry-*Pf*TFIIB and 4 µM 6His:*Pf*TLP-IDR1-YFP (bottom row). All reactions contained and 10% PEG 20000, 150 mM KCl and 5 mM magnesium acetate. Proteins were mixed thoroughly and incubated for 5 min before addition of condensation buffer. Scale bar = 10 µm (white).

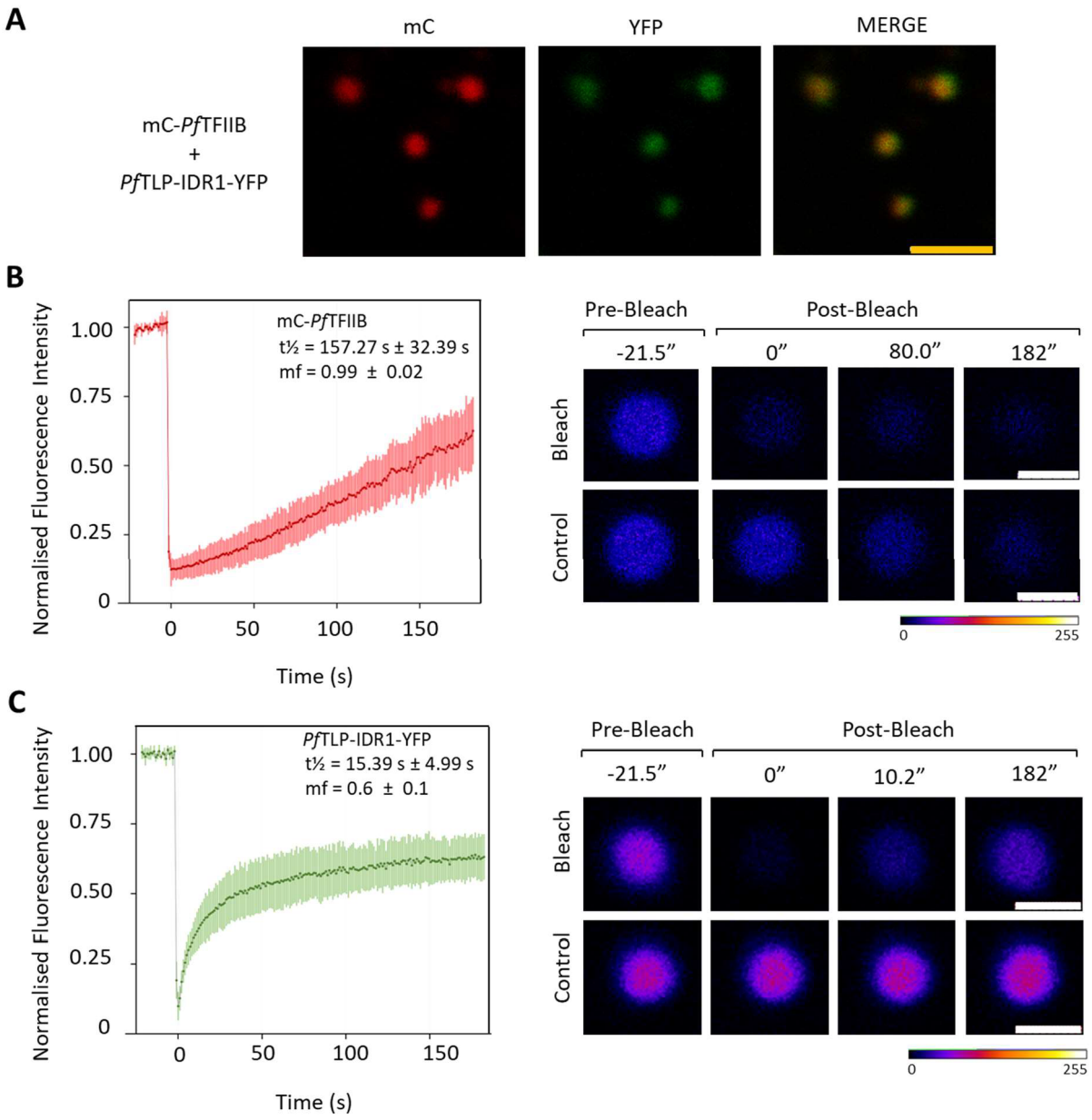


Figure 69. Fluorescence recovery after photobleaching of mixed condensates formed with *Pf*TFIIB and *Pf*TLP-IDR1. (A) Representative confocal fluorescent microscopy images of *Pf*TFIIB-*Pf*TLPIDR1 mixed condensates formed with 0.5 µM 6His:mCherry-*Pf*TFIIB and 2 µM 6His:YFP-*Pf*TLP-IDR1 at 10% PEG 20000, 5 mM magnesium acetate and 150 mM KCl. **(B, C)** Normalised fluorescence recovery after photobleaching of 6His:mCherry-*Pf*TFIIB **(B)** or 6His:*Pf*TLP-IDR1-YFP **(C)** in mixed condensates. Photobleaching was carried out 1-2 h after initiation of condensation. *Left*: Quantification of normalised fluorescence intensity. Data presented as mean ± the standard deviation (n=4). Mobile fraction (mf) and half time for recovery of fluorescence after photobleaching (t½) are indicated where these for values that could be reliably estimated using by EasyFrap (Koulouras *et al.*, 2018). *Right:* Representative images of bleached (top) and control (bottom) condensates at indicated time points relative to completion of bleaching. Non-bleached control condensate in the same field of view were used for the calculation of normalised fluorescence intensity. Images are coloured using the fire look up table in ImageJ. Scale bar: 2 µm (orange), 1 µm (white).

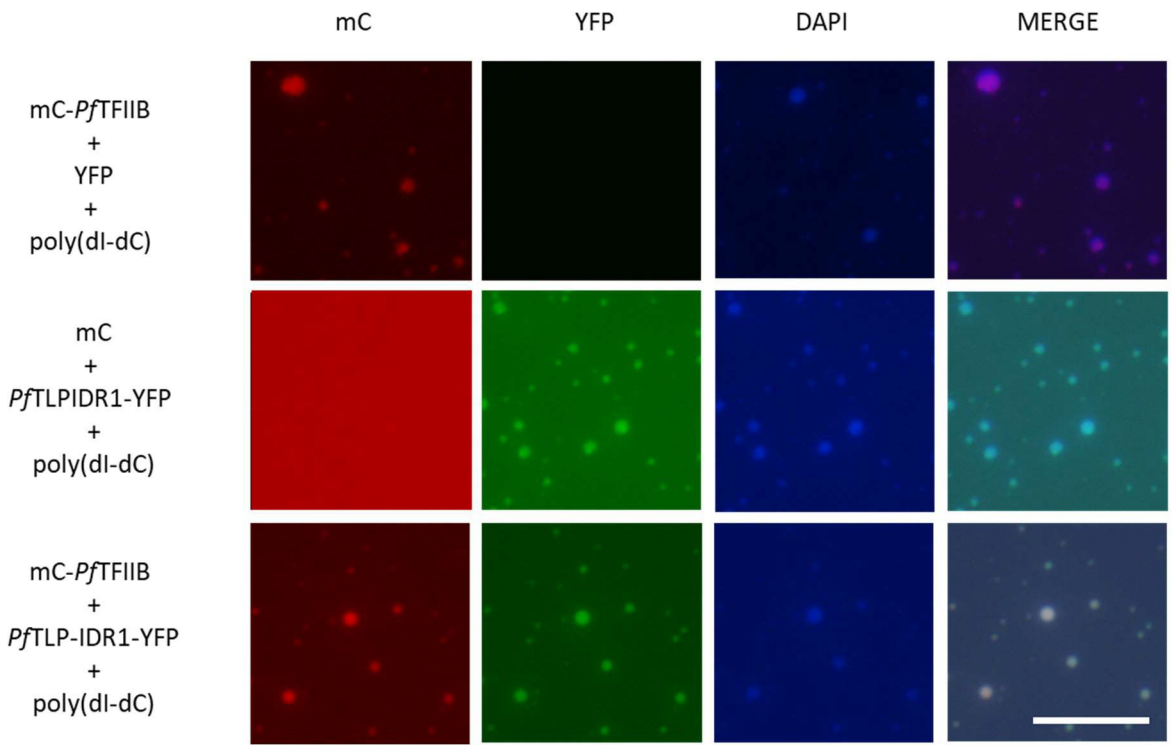


Figure 70. *Pf*TFIIB, the *Pf*TLP-IDR1 and dsDNA form homogenously mixed condensates. Representative widefield fluorescent microscopy images of condensates. Condensation reactions contained 0.5 µM 6His:mCherry-*Pf*TFIIB, 2 µM 6His:YFP and 100 ng/µl poly(dI-dC) (top row), *2* µM 6His:*Pf*TLP-IDR1-YFP, 4 µM 6His:mCherry and 100 ng/µl poly(dI-dC) (middle row), or 0.5 µM 6His:mCherry-*Pf*TFIIB, 4 µM 6His:*Pf*TLP-IDR1-YFP and 100 ng/µl poly(dI-dC) (bottom row). All reactions contained 10% PEG 20000, 150 mM KCl and 5 mM magnesium acetate. Proteins were mixed thoroughly and incubated for 5 min before addition of DNA. The protein/DNA mix was incubated for a further 5 min before addition of condensation buffer. Scale bar = 10 µm (white).

*5.5. Pf*TFIIB and the C-terminal domain of *Pf*RPB1 form multiphasic condensates in the presence of DNA

In well-studied eukaryotes, the disordered C-terminal domain of the largest RNA pol II subunit, RBP1, has been shown to be a potent driver of LLPS *in vitro* and to regulate partitioning of RNA Pol II into transcriptional condensates through homotypic interactions and interactions with transcriptional activators (Boehning *et al.*, 2018). Consistent with these observations, recent work by our group has shown the C-terminal domain of the *P. falciparum* RPB1 ortholog (*Pf*RBP1-CTD) also readily undergoes LLPS *in vitro*, supporting the notion that the *Pf*RPB1-CTD may also act as a potential scaffold protein in the formation of *P. falciparum* transcription condensates (Karamanof, 2022).

In well-studied eukaryotes, TFIIB directly interacts with RNA Pol II and plays key roles in the recruitment of RNA Pol II during transcription initiation complex assembly and during transcription initiation (Zheng *et al.*, 1987; Buratowski *et al.*, 1989; Barberis *et al.*, 1993; Buratowski and Zhou, 1993; Ha *et al.*, 1993; Malik, Lee and Roeder, 1993; Chen and Hahn, 2003; Thomas and Chiang, 2006). As a point of entry into assessing a potential role of the *Pf*RPB1-CTD in the assembly of *P. falciparum* transcription condensates, condensation assays were carried out using 6His:mCherry-*Pf*TFIIB and 6His:mEGFP-*Pf*RPB1-CTD fusion proteins. As shown in Fig. 71A, 6His:mCherry-*Pf*TFIIB and 6His:mEGFP-*Pf*RPB1-CTD formed irregular assemblies composed of small spherical conjoined condensates that were unable to fully fuse upon contact, reminiscent of beads on a string (Figs. 71A and 72A). FRAP experiments suggested that fluorescence recovery of 6His:mCherry*Pf*TFIIB in mixed condensates (mobile fraction= 0.86 ± 0.15; t½= 81.61 ± 16.00 s, Fig. 72) was slightly faster than the recovery of 6His:mCherry*Pf*TFIIB condensates formed by 6His:mCherry*Pf*TFIIB alone (mobile fraction: 1 ± 0.01; t½: 180.3 ± 0 s, Fig. 48), but was still slow relative to highly dynamic condensates formed by other transcription factors characterised by our group, such as those formed by *Pf*TLP-IDR1 (Michowicz, in preparation). In contrast, 6His:mEGFP-*Pf*RPB1-CTD fluorescence did not recover at all following condensate bleaching (Fig. 72C). These observations are consistent with the observation of stalled fusions events between mixed condensates and suggest that *Pf*RPB1-CTD molecules concentrated in mixed condensates form a largely immobile scaffold within which *Pf*TFIIB molecules retain the ability to rearrange dynamically.

Next, condensation reactions were carried out to assess whether the material properties of *Pf*TFIIB-*Pf*RPB1-CTD mixed condensates are altered by the presence of dsDNA. 6His:mEGFP-*Pf*RPB1-CTD formed heterogenous phase-separated condensates with poly(dI-dC) in the absence of *Pf*TFIIB, composed of small sub-compartments enriched in 6His:mEGFP-*Pf*RPB1-CTD associated with larger compartments enriched in poly(dI-dC). When included in the condensation reaction, 6His:mCherry-

*Pf*TFIIB mixed homogenously with poly(dI-dC), whereas the CTD of RPB1 remained enriched in associated sub-compartments, suggesting that interactions between *Pf*TFIIB and DNA disrupt the interactions between *Pf*TFIIB and the RPB1 CTD that drive the formation of homogenously mixed *Pf*TFIIB-RPB1-CTD condensates in the absence of DNA.

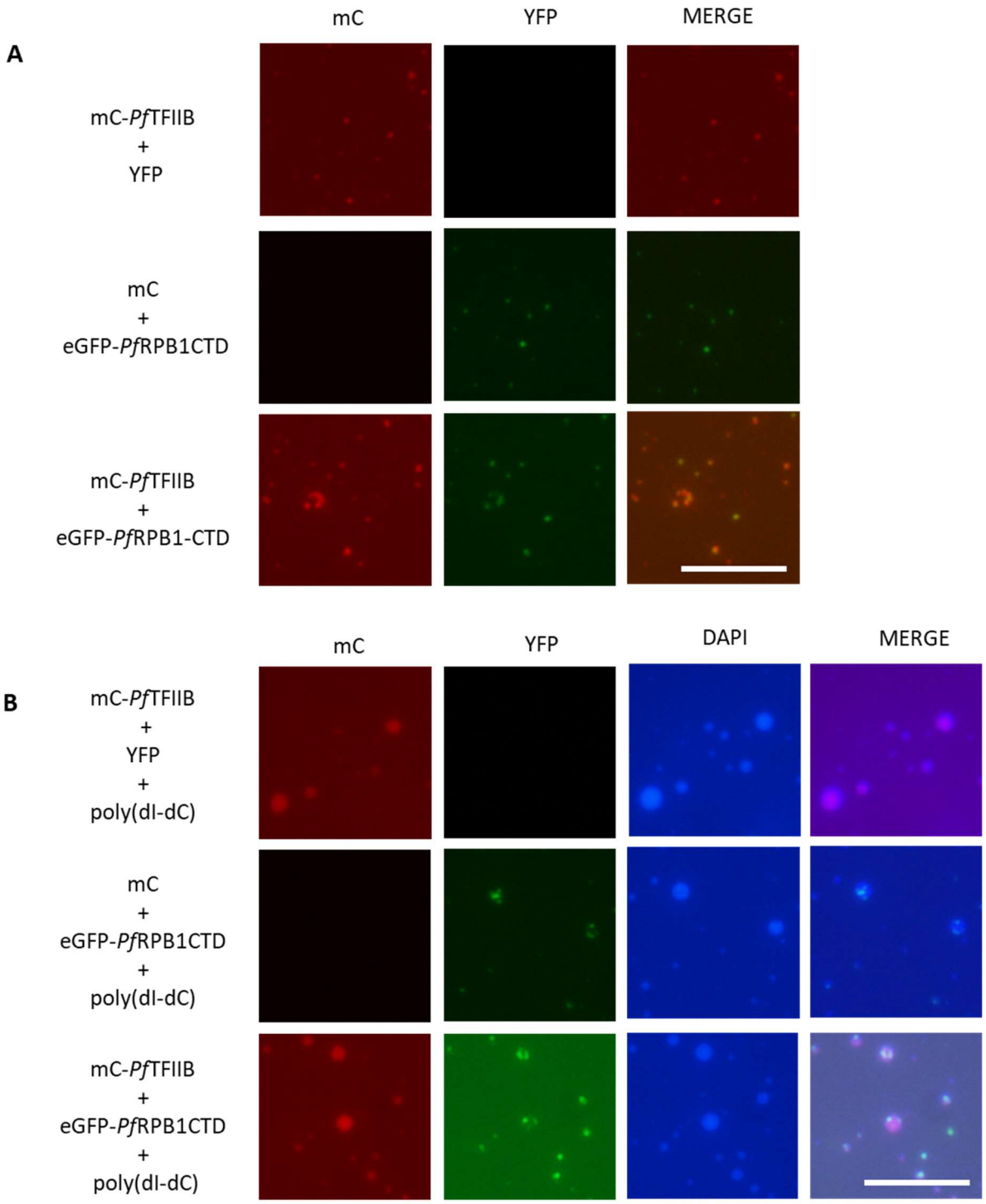


Figure 71. *Pf*TFIIB and the *Pf*RPB1-CTD form homogenously mixed condensates in the absence of DNA but heterogenous, multiphasic condensates in the presence of DNA. Representative widefield fluorescent microscopy images of condensates formed in the absence **(A)** or presence **(B)** of 100 ng/µl poly(dI-dC). Condensation reactions contained 0.5 µM 6His:mCherry-*Pf*TFIIB and 1 µM 6His:YFP (top row), 1 µM 6His:eGFP-*Pf*RPB1-CTD and 0.5 µM 6His:mCherry (middle row), 0.5 µM 6His:mCherry-*Pf*TFIIB and 1 µM 6His:eGFP-*Pf*RPB1-CTD (bottom row). All reactions contained 10% PEG 20000, 150 mM KCl and 5 mM magnesium acetate. Proteins were mixed thoroughly and incubated for 5 min before addition of DNA. The protein/DNA mix was incubated for a further 5 min before addition of condensation buffer. Scale bar = 10 µm.

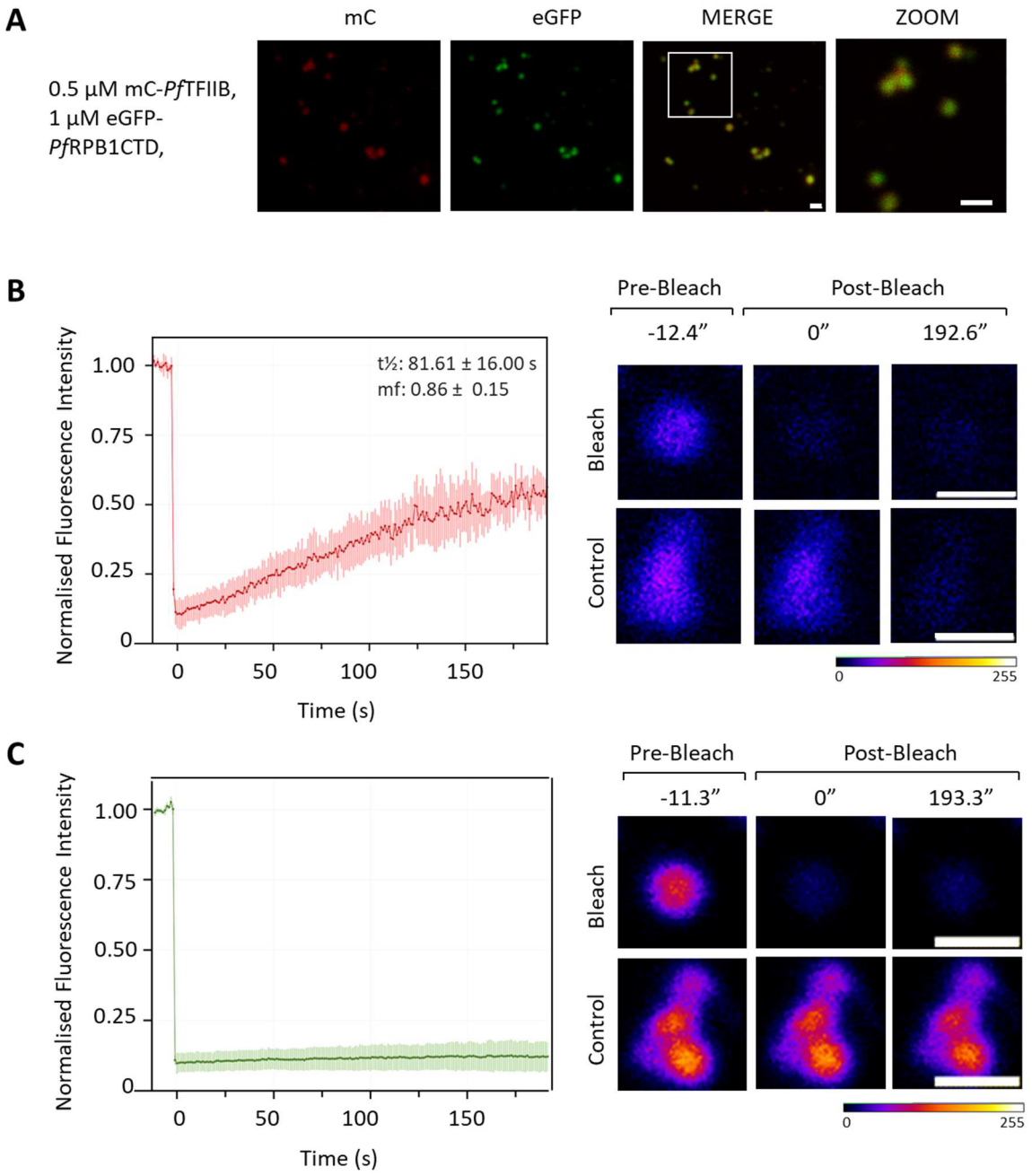


Figure 72. Fluorescence recovery after photobleaching in mixed condensates formed with *Pf*TFIIB and the *Pf*RPB1-CTD. (A) Representative confocal fluorescent microscopy images of *Pf*RPB1-CTD-*Pf*TFIIB mixed condensates formed with 0.5 µM 6His:mCherry-*Pf*TFIIB, 1 µM 6His:eGFP-*Pf*RPB1-CTD and 100 ng/µl poly(dI-dC) at 10% PEG 20000, 5 mM magnesium acetate, 150 mM KCl. White boxes indicate areas of the merged image enlarged on the right (ZOOM). **(B and C)** Normalised fluorescence recovery after photobleaching of 6His:mCherry-*Pf*TFIIB (B) or eGFP-*Pf*RPB1-CTD (C) in mixed condensates 1-2 h after initiation of condensation. *Left*: Quantification of normalised fluorescence intensity. Data presented as mean ± the standard deviation (*Pf*TFIIB: n=6; *Pf*RPB1-CTD: n=9). Mobile fraction (mf) and half time for recovery of fluorescence after photobleaching (t½) are indicated for values that could be reliably estimated using EasyFrap (Koulouras *et al*., 2018). *Right:* Representative images of bleached (top) and control (bottom) condensates at indicated time points relative to completion of bleaching. Non-bleached control condensate in the same field of view were used for the calculation of normalised fluorescence intensity. Images are coloured using the fire look up table. Scale Bar = 1 µm (white).

Chapter 6 Discussion

*6.1 Pf*TLP and a 38 kDa *Pf*TBP protein are expressed during the *P. falciparum* intraerythrocytic cycle

The *P. falciparum* genome encodes two TBP family proteins, designated *P. falciparum* TATA-binding protein (*Pf*TBP; PlasmoDB ID PF3D7_ 0506200) and *P. falciparum* TBP-like protein (*Pf*TLP; PlasmoDB ID PF3D7_1428800). To date, there are no published studies which have functionally characterised *Pf*TLP. In addition, published basic properties of *Pf*TBP, such as the molecular mass of *Pf*TBP expressed during the *P. falciparum* intraerythrocytic cycle as well as the function of *Pf*TBP in promoter recognition, remain a matter of contention. The first part of this thesis aimed to firmly determine the properties of these *P. falciparum* TBP family proteins in order to provide a solid basis for understanding their role in the assembly of the RNA Pol II transcription initiation complex in *P. falciparum.*

The gene encoding *Pf*TBP was initially identified in 1993 by a study that used highly conserved motifs in the structured C-terminal DNA-binding domain of TBP as a basis for the design of PCR primers that allowed sequential amplification of genomic DNA fragments containing the *Pf*TBP open reading frame (McAndrew *et al.*, 1993). Expression of *Pf*TBP was initially confirmed at the mRNA level by northern blot analysis of total RNA from mixed asexual *P. falciparum* parasites (McAndrew *et al.*, 1993). Further evidence for the expression of *Pf*TBP has been obtained in subsequent studies at both the RNA and protein levels using DNA microarray, northern blot analysis, RNA sequencing, immunoblotting and proteomic mass spectrometry analyses (McAndrew *et al.*, 1993; Bozdech *et al.*, 2003; Ruvalcaba-Salazar *et al.*, 2005; Gopalakrishnan *et al.*, 2009; Treeck *et al.*, 2011).

Importantly, current literature is inconsistent regarding the molecular mass of the expressed *Pf*TBP protein. Initially, the *Pf*TBP open reading frame was reported to encode a 26 kDa *Pf*TBP protein (McAndrew *et al.*, 1993). However, sequencing of the *P. falciparum* genome revealed the presence of a longer open reading frame beginning upstream of the start codon originally identified (The Plasmodium Genome Consortium, 2001). This open reading frame encodes a 38 kDa *Pf*TBP protein, with a much longer intrinsically disordered N-terminal region. The first study that investigated expression of *Pf*TBP by immunoblot analysis of blood-stage parasites extracts, used a newly developed polyclonal anti-*Pf*TBP antibody, which detected a 26 kDa *Pf*TBP protein (Ruvalcaba-Salazar et al., 2005). However, a later study probing *P. falciparum* whole cell extract using a different custom-made anti-*Pf*TBP antibody reported the detection of a band with apparent molecular mass of approximately 42 kDa (Gopalakrishnan et al., 2009). Curiously, this 42 kDa band was also

recognised by a commercially available *Hs*TBP antibody (Santa-Cruz Biotechnology, sc-273) reported to be cross-reactive with *Pf*TBP by the first study (Ruvalcaba-Salazar *et al.*, 2005; Gopalakrishnan *et al.*, 2009). Finally, mass spectrometry analysis of the phospho-proteome of *P. falciparum* trophozoites detected a fragment corresponding to amino acid residues E48-K74 in the predicted 38 kDa *Pf*TBP protein (Treeck *et al.,* 2011). Given that these residues are absent in the predicted 26 kDa *Pf*TBP protein, detection of this peptide suggests that the 38kDa *Pf*TBP protein must be expressed in the *P. falciparum* blood stages. However, it remaines formally possible that a 26 kDa *Pf*TBP protein may be expressed, potentially by alternative translation initiation. Alternative translation initiation has been demonstrated to play a role in the regulation of other general transcription factor proteins such as TAF12, which is expressed as a more abundant 20 kDa isoform and a less abundant 15 kDa isoform (Hoffmann and Roeder, 1996; Perletti, Dantonel and Davidson, 1999; Hamard *et al.*, 2005).

In this study, we used a custom-made rabbit polyclonal anti-*Pf*TBP antibody raised against an amino acid sequence in the N-terminal disordered region (aa 125 -138) that is present in both the predicted 26 kDa and 38 kDa *Pf*TBP proteins. Immunoblot analysis of bacterially expressed, purified recombinant 6His-tagged 26 kDa and 38 kDa *Pf*TBP proteins revealed reduced mobility of the 38 kDa *Pf*TBP in SDS-PAGE gels, migrating at an apparent molecular mass of about 50 kDa (Fig. 14A). The reduced mobility of recombinant 38 kDa *Pf*TBP is not seen with recombinant 26 kDa *Pf*TBP (Fig. 14A) and can be explained by the highly acidic nature of the additional N-terminal amino acid sequence present in 38 kDa *Pf*TBP (calculated pI: 4.37, Fig. S2). Consistent with the immunoblot analysis results using purified 38 kDa *Pf*TBP, our immunoaffinity-purified anti-*Pf*TBP antibody recognised a prominent band in *P. falciparum* whole cell extracts migrating at an apparent molecular mass of ~ 50 kDa (Fig. 14B). A 26 kDa *Pf*TBP isoform could not be detected, suggesting that it may not be expressed in *P. falciparum* blood-stage parasites, or that its expression levels are significantly lower compared to the 38 kDa *Pf*TBP protein.

*Pf*TLP was identified *in silico* by bioinformatics studies using hidden Markov model profile searches (Coulson, Hall and Ouzounis, 2004). Expression of *Pf*TLP has been detected across *P. falciparum* life stages by DNA microarray analysis, RNA sequencing experiments and proteomic mass-spectrometry (Le Roch *et al.*, 2003; Bischoff and Vaquero, 2010; Otto *et al.*, 2010; Silvestrini *et al.*, 2010; Treeck *et al.*, 2011; Pease *et al.*, 2013; Lasonder *et al.*, 2016; Toenhake *et al.*, 2018; Chappell *et al.*, 2020). Here we make use of a custom-made polyclonal anti-*Pf*TLP antibody raised against an amino acid sequence (aa 294-307) within the intrinsically disordered region (IDR1) located between S3' and S4'. The affinity-purified anti-*Pf*TLP antibody recognised a 44 kDa band in immunoblots of *P. falciparum* whole trophozoite extracts (Fig. 15B), consistent with the apparent molecular mass of purified recombinant 6His:*Pf*TLP in SDS-PAGE (Fig. 15A).

Together, immunoblot analyses presented in this study using highly reactive custom-made polyclonal rabbit antibodies raised for *Pf*TLP and *Pf*TBP confirm that both proteins are expressed in *P. falciparum* blood stage parasites. Our observations suggest expression of a 38 kDa *Pf*TBP protein during the intraerythrocytic cycle consistent with the *P. falciparum* genome sequence analysis (The Plasmodium Genome Consortium, 2001). However, our results do not fully exclude that the 26 kDa *Pf*TBP isoform, reported earlier (Ruvalcaba-Salazar *et al.*, 2005), could be expressed at a different point in the *P. falciparum* life cycle or at low levels as a result of alternative translation events.

*6.2 Pf*TBP and *Pf*TLP localise to discrete foci in cultured *P. falciparum* parasites

Immunofluorescence microscopy using our antibodies directed against *Pf*TBP and *Pf*TLP, detected puncta in *P. falciparum* trophozoites and schizonts, suggesting that these proteins are concentrated in discrete foci or condensates (Figs. 16, 17). Previous studies have shown that TBP also localises to puncta in cultured murine cells (Basu *et al.*, 2020; Shao *et al.*, 2022). These studies also demonstrated that isolated recombinant *Hs*TBP can undergo phase separation *in vitro,* suggesting that *Hs*TBP is a component of transcription condensates (Basu *et al.*, 2020; Shao *et al.*, 2022). Interestingly, research within our laboratory has shown that both *Pf*TBP and *Pf*TLP are strong drivers of biomolecular condensation *in vitro* (Van der Linden, 2020, Karamanof, 2022; Pegram 2022, Michowicz, in preparation). Thus, the observation that both *Pf*TBP and *Pf*TLP localise to discrete puncta can be taken as supporting evidence for the existence of transcription condensates in *P. falciparum*.

A limitation of the immunofluorescence microscopy results presented here is the use of widefield epifluorescence microscopy as opposed to confocal microscopy or an appropriate super resolution microscopy modality. Furthermore, we could not obtain an appropriate marker to clearly distinguish the nuclear periphery from the surrounding cytoplasm in isolated parasites. As explained in Section 3.4, the nucleus of *P. falciparum* parasites has a complex structure and consists of a transcriptionally active central region, which is stained by DAPI, surrounded by a perinuclear zone not stained by DAPI. A number of genes, including members of the *var* gene family, localise to distinct sites at the nuclear periphery, suggesting that the nuclear periphery is an important location for *Plasmodium* transcription regulation (Duraisingh *et al.*, 2005; Ralph, Scheidig-Benatar and Scherf, 2005; Volz *et al.*, 2010). At present, our data do not allow definite conclusions as to whether *Pf*TBP and *Pf*TLP puncta within *P. falciparum* parasites are strictly nuclear, cytoplasmic, or both.

This question could be addressed by future experiments using confocal microscopy or super resolution microscopy in combination with antibodies for proteins indicative of the nuclear

periphery. Proteins previously used as markers for the nuclear periphery include the nuclear pore protein nucleoporin 100 and CenPA, a histone H3 centromere-specific protein variant (Volz *et al.*, 2010). Additional experiments assessing *Pf*TBP and *Pf*TLP localisation in transgenic *P. falciparum* parasites expressing fluorescent protein-tagged general transcription factors should also be considered to complement these studies and allow for the direct visulaisation of these transcription factors in the absence of potential fixation artifacts or artifacts caused by non-specifc antibody binding.

*6.3 Pf*TBP and *Pf*TLP are not prototypical TATA-binding proteins

In silico analysis of *Pf*TBP and *Pf*TLP performed in this study showed that both proteins display a high level of divergence in overall primary amino acid sequence relative to well-studied eukaryotes and contain amino acid substitutions at several key residues involved in sequence-specific recognition of the TATA box by prototypical TBP in higher eukaryotes. Interestingly, previously characterised TBP proteins from other protozoa that display a similar degree of overall divergence in amino acid sequence (Fig. 10), have been shown to possess altered DNA-binding activities. For example, TBP proteins in *Trichomonas* and *Crypthecodinium*, two non-Apicomplexan protozoa, have similar overall primary sequence divergence from well-characterised TBP family proteins compared to members of the Api-TBP and Api-TLP clades identified in this study (Fig. 10). *Trichomonas* TBP proteins appear to bind DNA non-specifically (Parra-Marín *et al.*, 2019) and *Crypthecodinium* TBP has been shown to bind a TTTT box motif with much higher affinity than a prototypical TATA-box sequence (Guillebault *et al.*, 2002).

Consistent with the presence of non-conservative amino acid substitutions and insertions in *Pf*TBP and *Pf*TLP at amino acid positions that mediate TBP-TATA interactions in prototypical TATA-binding proteins, the *in vitro* characterisation of *Pf*TBP- and *Pf*TLP-DNA interactions presented here demonstrates that neither *Pf*TBP nor *Pf*TLP exhibit sequence-specific binding to the consensus TATA box DNA sequence (Section 3.5.3). Both *Pf*TBP and *Pf*TLP bind with equivalent affinity to dsDNA probes containing wild-type or mutated TATA box sequences and dsDNA probes lacking a TATA box altogether (Figs. 21, 22). Furthermore, while both *Pf*TBP and *Pf*TLP can bind stably to the Ad2ML-TATA in EMSAs, this binding is equally outcompeted by the Ad2ML TATA box sequence and poly(dG-dC) dsDNA consisting entirely of alternating GC dinucleotide steps (Fig. 22).

Additional evidence that *Pf*TBP interacts with the consensus TATA box DNA sequence in a manner distinct from prototypical TBP is provided by the fact that *Pf*TBP-Ad2ML-TATA complexes are destabilised in the presence of magnesium ions (Fig. 20), whereas prototypical TBP-TATA interactions are stabilised in the presence of magnesium ions (Fig. 20, Horikoshi *et al*., 1989; Kuddus & Schmidt, 1993; Reddy & Hahn, 1991). Indeed, non-specific DNA interactions of the cAMP response element binding protein CREB basic leucine zipper domain were shown to be selectively destabilised in EMSAs by the presence of magnesium ions in the resolving gel (Moll, 2002). Thus, the destabilisation of

*Pf*TBP-TATA box DNA complexes may occur as a consequence of the loss of TATA-specificity in *P. falciparum* TBP family proteins.

While neither *Pf*TBP nor *Pf*TLP are prototypical TATA-binding proteins, they do display sequence-selectivity towards DNA with a high A/T-nucleotide content. Attempts to delineate high affinity DNA motifs preferentially bound by *Pf*TBP and *Pf*TLP in the context of the A/T-rich *P. falciparum* GBP-130 and KAHRP promoter regions were unsuccessful, suggesting that *Pf*TBP and *Pf*TLP bind a diverse array of A/T-rich sequences with comparable affinity. Slightly different affinities towards some GBP-130 and KAHRP regions could be observed (Figs. 24, 25, 28), but these differences can be explained by differences in A/T-nucleotide content (Tables 2 and 3). For example, the GB (163 bp in length, 81.6% A/T) and KB DNA probes (157 bp in length, 78.3 % A/T), bound with greater affinity than GA (143 bp in length, 74.8 % A/T) and KA DNA probes (138 bp in length, 73.9 % A/T; Figs. 24, 28), have an increased A/T-content relative to GA and KA DNA probes and are also slightly longer (Table 2), which increases the number of DNA-binding sites. We were unable to reproduce evidence for the previously reported sequence-selective *Pf*TBP interactions with TGTAA or TATAA DNA motifs in the GBP-130 and KAHRP putative promoter regions (Ruvalcaba-Salazar *et al.*, 2006). Such sequence-selective interactions would also be inconsistent with the observed non-specific binding of *Pf*TBP to the Ad2ML sequence containing a consensus TATA box (Figs. 21 and 22). Finally, our experimental results corroborate previous attempts in our group to identify high affinity *Pf*TBP binding sites using a systematic evolution of ligands by exponential enrichment (SELEX) approach, which showed a general enrichment for A/T-nucleotides in *Pf*TBP-bound DNA sequences but could not identify any specific DNA sequence motifs (Milton, 2017).

Interestingly, while *Pf*TBP and *Pf*TLP are unable to bind to the TATA box in a sequence-specific manner, they interact with dsDNA in a manner characteristic for TBP family proteins. Prototypical TATA-binding proteins recognise the TATA DNA sequence through interactions in the DNA minor groove (Starr and Hawley, 1991; Kim *et al.*, 1993; Kim, Nikolov and Burley, 1993; Patikoglou *et al.*, 1999). Similarly, the results presented here demonstrate that sequence-selective DNA binding to A/T-rich DNA sequences by *Pf*TBP and *Pf*TLP does not require DNA contacts in the DNA major groove and are mediated by DNA minor groove interactions (Figs. 26, 29). Furthermore, substitution of conserved intercalating phenylalanine amino acid residues, required for stable interactions between prototypical TBP proteins and TATA box DNA, does not significantly affect the affinity of *Pf*TLP for A/T-rich DNA sequences (Fig. 30), but increases the electrophoretic mobility of *Pf*TLP-DNA complexes, when DNA binding reactions are carried out on ice (Fig. 30). This observation suggests that, although these residues (F60 and F283 in *Pf*TLP) are not required for stable DNA complex formation, they may be involved

in mediating a conformational change in the *Pf*TLP-DNA nucleoprotein complex, perhaps similar to the DNA bending observed in classical TBP-TATA box complexes (Horikoshi *et al.*, 1992; Kim *et al.*, 1993; Kim, Nikolov and Burley, 1993; Starr, Hoopes and Hawley, 1995; Nikolov *et al.*, 1996).

Taken together, the results presented here suggest a model in which, in the absence of other general transcription factors, DNA binding of *Pf*TBP or *Pf*TLP within the A/T-rich DNA sequence context of the *P. falciparum* genome is unlikely to contribute to the recognition of specific promoter regions during RNA Pol II transcription initiation.

In broad agreement with the findings presented here, recent studies that mapped transcription start sites in *P. falciparum* genome-wide showed that transcription of *P. falciparum* genes generally initiate concomitantly from multiple clusters of TSS sites that can span broad genomic regions (Adjalley *et al.*, 2016; Chappell *et al.*, 2020). Interestingly, one of these studies also suggested that there is an increased A/T-content of genomic DNA around the most frequently utilised clusters of TSSs (Adjalley *et al.*, 2016). It is tempting to speculate that such local changes in A/T-content modulate *Pf*TBP or *Pf*TLP binding to regions close to frequently used TSSs, even in the absence of a defined core promoter, and in this way influence the assembly of the RNA Pol II transcription intitation complex.

In mammals, broad promoters with dispersed patterns of transcription initiation are generally associated with nucleosome-depleted regions bordered by precisely positioned nucleosomes marked by histone H3 trimethylation at lysine 4 and histone H3 acetylation at lysine 27, rather than specific DNA sequence elements located near the TSS (Haberle and Stark, 2018). Our results support the notion that *in P. falciparum*, transcription initiation may be largely regulated through regulation of chromatin structure and dynamics, which in turn determine promoter accessibility. Within this context, specific transcription factors, such as APiAP2 transcription factors, may also guide the assembly of the RNA Pol II transcription intitation complex either through direct contacts or indirectly by contributing to the chromatin landscape. At this point, it is not clear whether *Pf*TBP or *Pf*TLP are functionally interchangeable and whether either one of these proteins plays a more dominant role in the assembly of RNA Pol II transcription initiation complexes. Given that TBP paralogs in metazoans have important roles in cell-type-specific and/or gene-specific transcription regulation (Jacobi *et al.*, 2007; Akhtar and Veenstra, 2011; Duttke, 2015), it is tempting to speculate that *Pf*TBP and *Pf*TLP may have distinct functions in the regulation of transcription initiation of specific genes during different *P. falciparum* life stages. Further research is required to determine the relative expression levels of *Pf*TBP and *Pf*TLP and their promoter occupancy genome-wide throughout the *P. falciparum* life cycle, and to uncover interactions with other regulatory factors.

The anti-*Pf*TBP and anti-*Pf*TLP antibodies generated for this study should be an invaluable resource for future immunolocalization experiments, co-immunoprecipitation and chromatin immunoprecipitation (ChIP) experiments, that aim to provide further insights into *Pf*TBP and *Pf*TLP cellular localisation and function.

*6.4 Pf*TFIIB has unusual DNA-binding properties

The results presented in this thesis suggest that *Pf*TBP and *Pf*TLP, in isolation, are unable to mediate recognition of a defined core promoter region during transcription initiation in *P. falciparum*, raising the possibility that other general transcription factors contribute to promoter binding during the assembly of the RNA Pol II transcription intitation complex at defined genomic locations. Previous studies by our group have demonstrated that *Pf*TFIIB can bind DNA in the absence of other transcription factors and that *Pf*TFIIB-DNA interactions are stable enough to be resolved in EMSAs. In well-studied eukaryotes, sequence-specific TFIIB interactions with the BREd core promoter element are only observed in the context of TBP bound to a TATA core promoter element (Deng and Roberts, 2005). *Hs*TFIIB interactions with the BREu are TBP-independent, but have only been detected in solution using fluorescence anisotropy experiments and site-specific protein-DNA photocrosslinking analysis (Lagrange *et al.*, 1998). A TFIIB-BREu complex stable enough to be resolved in typical DNA-binding experiments such as EMSAs or DNAseI footprinting has only been observed in the presence of *Hs*TBP (Lagrange *et al.*, 1998). *Pf*TFIIB is, to our knowledge, the first TFIIB ortholog demonstrated to form a TFIIB-DNA complex that can be resolved by EMSAs. The experiments presented in Chapter 4 of this thesis aimed to further characterise the DNA-binding properties of *Pf*TFIIB in order to provide insight into a potential role of *Pf*TFIIB in core promoter recognition during transcription initiation in *P. falciparum*.

The results presented here corroborate previous findings (Talvik, 2016) and demonstrate stable *Pf*TFIIB interactions with a variety of DNA probes even at conditions of high ionic strength, suggesting that *Pf*TFIIB DNA binding involves hydrophobic interactions in addition to electrostatic interactions between positively charged *Pf*TFIIB regions and the DNA backbone (Figs. 35, 38). Furthermore, our results show that *Hs*TFIIB does not form stable *Hs*TFIIB-DNA complexes under identical conditions - confirming that the observed DNA-binding activity of *Pf*TFIIB is a unique property and not an artifact of our experimental conditions (Fig. 35).

*Pf*TFIIB contains additional highly basic N- and C-terminal amino acid sequences that are absent in TFIIB orthologues in well-studied eukaryotes (Fig. 36; Bing, 2015, Talvik, 2016). Surprisingly, these regions were found not to be the main drivers of *Pf*TFIIB-DNA binding (Fig. 37). Instead, experiments designed to delineate the protein regions that mediate *Pf*TFIIB DNA-binding activity show that the

structured C-terminal core region of *Pf*TFIIB, comprised of two cyclin-like domains, is sufficient for TBP-independent DNA binding. Interestingly, the C-terminal basic region (aa 351-367) negatively affected the DNA-binding activity of the structured *Pf*TFIIB core region in the absence, but not in the presence, of the largely intrinsically disordered N-terminal region (aa 1-140; Fig. 37). Thus, the *Plasmodium*-specific basic regions and the N-terminal intrinsically disordered region appear to regulate *Pf*TFIIB DNA-binding activity through a complex network of interactions. Specific amino acid residues within the structured *Pf*TFIIB region responsible for TBP-independent DNA-binding activity remain to be identified.

With the exception of a few charged amino acid residues, key interacting amino acid residues in the helix-turn-helix region of *Hs*TFIIB involved in the recognition of the BREu core promoter element are not conserved in *Pf*TFIIB (Lagrange et al., 1998; Bing, 2015). There is also significant divergence in the amino acid residues that form the loop between H2 and H3 of the first direct repeat in *Hs*TFIIB, which make DNA contacts with the BREd (Bing, 2015). Most notably, G153 in *Hs*TFIIB, a key amino acid residue for specific *Hs*TFIIB-BREd interactions (Tsai and Sigler, 2000; Deng and Roberts, 2005), has been substituted by asparagine (N184) in *Pf*TFIIB. Taken together, these observations suggest that the molecular interactions that mediate TBP-independent binding of *Pf*TFIIB are completely distinct from the interactions that mediate DNA interactions of TFIIB in higher eukaryotes.

So far, there is no evidence for sequence-selective binding of *Pf*TFIIB to a specific DNA motif. *Pf*TFIIB DNA-binding also does not appear to be guided by nucleotide composition as *Pf*TFIIB binds to A/T-rich and G/C-rich DNA probes with comparable affinity (Fig. 38 and data not shown). However, at this point, we cannot exclude that *Pf*TFIIB can engage in sequence-selective interactions with a sequence absent in the DNA probes used here, nor can we exclude that *Pf*TFIIB contributes to sequence-selective DNA-binding in the context of other *P. falciparum* transcription factors. Future research using unbiased/untargeted approaches, such as a systematic evolution of ligands by exponential enrichment (SELEX) experiments with isolated *Pf*TFIIB, and *Pf*TFIIB in combination with other DNA-binding *P. falciparum* general transcription factors, may be able to further address this important question. In addition, chromatin immunoprecipitation (ChIP) experiments could be used to identify genomic regions preferentially occupied by *Pf*TFIIB in cultured parasites.

6.5 *Pf*TFIIB interacts with both *Pf*TBP and *Pf*TLP

TFIIB plays a crucial role in the assembly of the RNA Pol II transcription intitation complex in well-studied eukaryotes. According to the canonical model for the assembly of the RNA Pol II machinery at TATA-containing promoters, based on extensive biochemical studies in well-characterised

eukaryotes, binding of TFIIB to TATA-bound TBP facilitates the subsequent recruitment RNA Pol II and TFIIF, which in turn facilitate the recruitment of the remaining components of the RNA Pol II pre-initiation complex (Thomas and Chiang, 2006). Critically, formation of a TFIIB-TBP-TATA ternary complex appears to require significant distortion of the TATA box to allow TFIIB to interact simultaneously with DNA elements upstream and downstream of the TATA-box (Tsai and Sigler, 2000).

While the formation of a stable TFIIB-TBP-TATA complex appears dependent on the formation of a bent TBP-DNA complex, TFIIB and TBP can also interact directly in the absence of DNA in the context of GST pull-down experiments (Ha *et al.*, 1993; Lin *et al.*, 1997; Pan, Czarnecka-Verner and Gurley, 2000; Zhang *et al.*, 2000; Saxena *et al.*, 2005). The data presented in this thesis demonstrate that *Pf*TFIIB stably interacts with both *Pf*TBP and *Pf*TLP in the absence of DNA (Fig. 40), consistent with observations in well-studied eukaryotes. Interestingly, we found that *Pf*TFIIB also binds to *Hs*TBP in the absence of DNA (Fig. 40), suggesting that the amino acid residues that are important to *Pf*TFIIB-*Pf*TBP interactions may be functionally conserved in *Hs*TBP. This result corroborates previous *in silico* modelling of the *Pf*TBP structured domain that suggested a high degree of amino acid sequence conservation between *Hs*TBP and *Pf*TBP at the interface that mediates *Hs*TBP-*Hs*TFIIB interactions within the ternary TBP/IIB/TATA-box nucleoprotein complex (Buendía-Orozco, Guerrero and Pastor, 2005; Bing, 2015), and is also in agreement with the result of multiple sequence alignment of TBP family proteins performed in this study (Fig. 11, Table 6), which noted a high degree of conservation between *Pf*TBP and *Hs*TBP at amino acid residues involved in *Hs*TBP-TFIIB interactions and the presence of conservative substitutions at these residues in *Pf*TLP (Fig. 11, Table 6). However, it should be noted that the corresponding interaction surface within *Pf*TFIIB is less conserved compared to *Hs*TFIIB (Buendía-Orozco, Guerrero and Pastor, 2005; Bing, 2015). It is therefore possible that binding of *Pf*TFIIB to *Hs*TBP is mediated by different interaction surfaces than the canonical binding of *Hs*TFIIB to *Hs*TBP.

TFIIB interactions with the bent TBP-TATA complex have been shown to greatly stimulate and to stabilise specific TBP interactions with TATA-containing promoters (Imbalzano, Zaret and Kingston, 1994). Our results show that *Pf*TFIIB stimulates *Pf*TBP- and *Pf*TLP-DNA interactions with the A/T-rich GBP-130 promoter and forms ternary *Pf*TFIIB-TBP-DNA and *Pf*TFIIB-TLP-DNA complexes (Figs. 41, 42). In higher eukaryotes, the TFIIB C-terminal structured region, comprised of two cyclin-like domains, is sufficient to stimulate TBP-TATA interactions (Barberis *et al.*, 1993; Nikolov *et al.*, 1995). Interestingly, the structured C-terminal region of *Pf*TFIIB is also sufficient to stimulate *Pf*TLP DNA-binding activity and to form ternary *Pf*TLP-*Pf*TFIIB-DNA complexes (Fig. 42B), but exerts only a very weak stimulatory effect on *Pf*TBP DNA binding compared to the full-length *Pf*TFIIB protein (Fig. 42A).

These results suggest that *Pf*TFIIB engages in distinct interactions with *Pf*TBP and *Pf*TLP and that the *Pf*TFIIB N-terminal region (aa1-140) and *P. falciparum*-specific C-terminal (aa 350-367) basic region (Fig. 36) may have additional functions in stabilising *Pf*TBP-DNA interactions.

*Pf*TBP only possesses three of the four conserved phenylalanine residues involved in distortion of the DNA-helix in prototypical TBP-TATA interactions (Fig. 11, Table. 5). Thus, *Pf*TBP interactions with A/T-rich DNA might not result in bending of the DNA helix, at least not to the extent seen within the prototypical TBP-TATA box nucleoprotein complex. In contrast, all four of the conserved phenylalanine residues are present in *Pf*TLP (Fig. 11, Table 5) and mutation of F60 and/or F283 significantly alters the mobility of *Pf*TLP-DNA complexes when binding reactions were incubated on ice (Fig. 30). These results suggest that *Pf*TLP DNA binding induces a conformational change in the DNA probe that involves the action of F60 and F283. Given that the formation of a TFIIB-TBP-TATA complex requires DNA bending by TBP in classical systems, I propose that the differences in *Pf*TFIIB interactions with *Pf*TBP and *Pf*TLP in the presence of DNA may be related to differences in the distortion of the DNA helix by *Pf*TBP and *Pf*TLP.

The results presented here do not provide any indication of enhanced specificity of *Pf*TFIIB-*Pf*TBP or *Pf*TFIIB-*Pf*TLP complexes compared to *Pf*TBP and *Pf*TLP in isolation (Figs. 43, 44, S6 and S7). Thus, it appears likely that *Pf*TFIIB contributes to the stability and affinity of *Pf*TBP-/*Pf*TLP-DNA interactions without impacting sequence specificity, in agreement with the lack of evidence for sequence-selective binding of *Pf*TFIIB in isolation (Fig. 38 and data not shown). While it is still possible that an unbiased approach could uncover high affinity sequences selectively bound by *Pf*TFIIB-*Pf*TBP and *Pf*TFIIB-*Pf*TLP complexes, our combined results suggest that *P. falciparum* genes may not possess a single well-defined promoter or, alternatively, that transcription factors other than *Pf*TBP, *Pf*TLP and *Pf*TLP drive sequence-specific core promoter recognition.

As noted in Section 6.3, genome-wide TSS mapping in *P. falciparum* suggests that transcription of many genes initiates from dispersed clusters of TSSs rather than a well-defined nucleotide position (Adjalley *et al.*, 2016; Chappell *et al.*, 2020). These results are consistent with a model in which the assembly of the RNA Pol II transcription initiation complex is not driven by defined core promoter elements but is instead regulated by alternative mechanisms such as chromatin structure and dynamics that regulate the accessibility of DNA regions.

6.6. Transcription condensates in *Plasmodium falciparum*: a potential role for *Pf*TFIIB

Transcription condensates that selectively concentrate RNA Pol II and positive regulators of transcription play a key role in transcription regulation in metazoans and bacteria (Hnisz *et al.*, 2017; Boehning *et al.*, 2018; Boija *et al.*, 2018; Sabari *et al.*, 2018). In *P. falciparum*, transcription has been shown to occur at distinct developmentally regulated foci (Moraes *et al.*, 2013), suggesting that transcription condensates may also play an important role in regulating gene expression in the malaria parasite. Previous work by past and present lab members in our group has demonstrated that several components of the *P. falciparum* basal transcription machinery, including *Pf*TBP, *Pf*TLP and the CTD of *Pf*RPB1, undergo liquid-liquid phase separation *in vitro* (van der Linden 2019, Karamanof, 2022, Pegram, 2022, Michowicz, in preparation). Consistent with these observations, *Pf*TBP and *Pf*TLP localise to foci in *P. falciparum* trophozoites and schizonts, in support for the existence of transcription condensates in *P. falciparum* (Figs. 16 and 17). A key focus of this PhD research work was to study *in vitro* condensation of *Pf*TFIIB in both isolation and in combination with nucleic acids and other *P. falciparum* general transcription factors in order to investigate a potential role of *Pf*TFIIB in the assembly and regulation of transcription condensates in *P. falciparum*.

One of the major challenges in this study was the generation of *Pf*TFIIB protein preparations at concentrations suited for biomolecular condensation studies. It is well-known that heterologous expression of soluble recombinant *P. falciparum* proteins in *E.coli* can be extremely challenging, due to highly divergent codon usage in protein-coding genes resulting from the unusually high A/T-content of the *P. falciparum* genome, and due to the highly divergent amino acid composition and unusual features of *P. falciparum* proteins that are characterised by a high prevalence of low complexity amino acid sequence insertions with a high propensity for intrinsic disorder (Birkholtz *et al.*, 2008). Recombinant *P. falciparum proteins* are often toxic when expressed in *E. coli* bacteria and tend to accumulate in insoluble form in *E.coli* inclusion bodies (Birkholtz *et al.*, 2008). Protocols have been established in the laboratory to express *P. falciparum* general transcription factors in soluble form in *E.coli*. However, because of their toxicity, these proteins can only be expressed at relatively low levels and large volumes of bacterial culture are required to obtain appreciable quantities of proteins to allow for biochemical characterisation (Bing, 2015, Talvik, 2016, Milton, 2017, van der Linden, 2019, Michowicz, in preparation, Section 3.2). In this work, I made use of an osmotic shock protocol (Section 2.1.5), previously demonstrated to increase the yield of metal-affinity purification of proteins expressed in *E.coli* at low levels. Although this protocol did indeed significantly increase the yield of the proteins of interest (Section 3.2), the maximal yields that could be achieved for purified recombinant *Pf*TFIIB proteins were still relatively low (≤ 50 µg per l culture). Furthermore,

*Pf*TFIIB proteins could not be kept in solution at concentrations > 4 µM. For this reason, the final concentration of *Pf*TFIIB proteins in most condensation assays was limited to ≤ 1 µM, a much lower concentration compared to biomolecular condensation assays commonly reported in the literature (Boehning *et al.*, 2018; Boija *et al.*, 2018b; Sabari *et al.*, 2018). Indeed, the formation of micron sized condensates of many phase separating transcription factors and coactivators, such as Med1 and BRD4, typically requires protein concentrations above 10 µM under similar conditions to those used here (Sabari *et al.*, 2018). The size of phase separated condensates is intimately linked with the concentration of phase separating molecules and their C_{sat} (Alberti, Gladfelter and Mittag, 2019). Thus, protein condensates formed at the low protein concentrations used in this study were quite small. Fluorescence microscopy studies to determine condensate properties such as morphology, homogeneity and fluorescence recovery following photobleaching were therefore limited by the resolution limits of the microscopes used, as well as the high rate of photobleaching during imaging as a result of the high magnification and, consequently, the high illumination intensity required to visualise small condensates.

Despite the challenges associated with the characterisation of condensates formed at low protein concentrations, the results presented here clearly demonstrate that *Pf*TFIIB is a potent driver of *in vitro* biomolecular condensation, with a C_{sat} in the nanomolar range. Due to their small size (≤ 1 µm), it was not possible to determine whether *Pf*TFIIB protein condensates possess spherical morphology, characteristic of liquid condensates. Irregularly shaped condensates were observed (Fig. 47), but these could have been formed by incomplete fusion of individual small spherical condensates. Condensates formed by fluorescent protein-tagged *Pf*TFIIB exhibited much slower rates of fluorescence recovery following condensate bleaching in FRAP experiments compared to protein condensates for which liquid-like properties could be clearly established, for example condensates formed by *Pf*TLP-IDR1 (Fig. 69C; Michowicz, in preparation). These results suggest that *Pf*TFIIB condensates had gel-like material properties at the time of observation. However, the rate of fluorescence recovery of *Pf*TFIIB condensates further decreased over time (Fig. 48), suggesting that *Pf*TFIIB condensates may have been more dynamic and liquid-like at the point of formation, but then hardened over time into gel-like or solid-like assemblies characterised by long-lived rather than dynamic intermolecular interactions and slow or even negligible exchange of molecules with the surrounding solution.

Such a transition of condensates formed by liquid-liquid phase separation into gel-like or solid assemblies over time is not unusual and is well-documented for many proteins, in particular for RNA-binding proteins with a high propensity to undergo phase separation, including the fused-in sarcoma (FUS) protein (Lin *et al.*, 2015; Patel *et al.*, 2015), heterogeneous nuclear ribonucleoprotein A1

(hnRNPA1; Lin *et al.*, 2015; Molliex *et al.*, 2015), the yeast developmental regulatory protein Whi3 (Zhang *et al.*, 2015), and fibrillarin (Feric *et al.*, 2016, reviewed in Shin and Brangwynne, 2017). This process of hardening or maturation is thought to be brought about by the high local concentration of the phase-separated molecules within condensates and to involve a variety of mechanisms, which are not mutually exclusive. These include enhanced β-strand interactions between intrinsically disordered domains, an increase in the number of interactions characterised by high association rate constants and low dissociation rate constants ('kinetic vitrification') and entanglement of intrinsically disordered polymer chains (Fig. 8 ; Banani *et al.*, 2017). Importantly, the transition of condensates from a liquid state into gel- or solid like material states is frequently observed when biomolecules undergo phase separation in the absence of physiological interaction partners (Lin *et al.*, 2015; Molliex *et al.*, 2015; Maharana *et al.*, 2018). In the physiological context of the cell, the material properties of condensates formed by phase-separating proteins can be tuned by the concentration of physiological interactors, such as other proteins or nucleic acids, by enzymatic modifications of condensate components, by changes in environmental conditions such as temperature and salinity, and by ATP-dependent processes, such as the action of disaggregases and molecular chaperones (Zhang *et al.*, 2015; Banani *et al.*, 2017; Guo *et al.*, 2018; Maharana *et al.*, 2018). In some cases, aberrant condensate hardening or maturation has been shown to cause disease. For example, aberrant phase separation of RNA-binding proteins caused by mutation of key amino acid residues and/or mis-localisation in the cell, and, as a consequence, the absence of important interactors, correlates with the formation of pathological protein aggregates linked to neurodegenerative disease, such as amyotrophic lateral sclerosis (ALS; Patel *et al.*, 2015; Harrison and Shorter, 2017; Maharana *et al.*, 2018).

In line with the potential role of additional interacting molecules in modulating condensate maturation, the results presented here show that the presence of nucleic acids affects assembly and properties of *Pf*TFIIB condensates. In the presence of either RNA or dsDNA, *Pf*TFIIB assembled into much larger condensates, resulting in a significantly increased *Pf*TFIIB condensed fraction, compared to condensation reactions containing isolated *Pf*TFIIB at the same concentration (Figs. 51 and 56). Mixed *Pf*TFIIB condensates formed in the presence of either dsDNA or RNA were clearly spherical, suggesting condensate assembly by LLPS, but displayed very different material properties at the time of observation (Figs. 51 and 52 and Figs. 56 and 57). In condensates formed with dsDNA, *Pf*TFIIB was clearly highly dynamic and exchanged more rapidly with the surrounding dilute phase compared to condensates formed in the absence of DNA (compare Fig. 52 with 48). This result suggests that dynamic *Pf*TFIIB-DNA interactions are competitive with interactions between concentrated *Pf*TFIIB molecules responsible for the rapid condensate vitrification into a gel- or solid-like state (Fig. 48). It

follows that these interactions may involve (a) common *Pf*TFIIB region(s), or alternatively, that *Pf*TFIIB-DNA interactions induce conformational changes within *Pf*TFIIB that disfavour *Pf*TFIIB-*Pf*TFIIB interactions.

In contrast, *Pf*TFIIB condensates assembled in the presence of RNA formed large assemblies composed of chains of spherical condensates, resembling beads on a string (Figs. 56 and 57), suggesting rapid hardening of *Pf*TFIIB-RNA condensates into a gel- or solid-like material state, similar to *Pf*TFIIB condensates formed in the absence of nucleic acids (Figs. 51 and 52). Consistent with condensates with gel-like or solid-like properties, condensates formed in the presence of RNA exhibited minimal exchange of *Pf*TFIIB molecules between condensates and the dilute phase in FRAP experiments (Fig. 57). These observations suggest that, while *Pf*TFIIB interactions with RNA facilitate *Pf*TFIIB condensate formation, they are not competitive with the interactions between *Pf*TFIIB molecules that cause hardening of *Pf*TFIIB condensates over time. Thus, *Pf*TFIIB interactions with RNA and dsDNA must be functionally distinct and are likely mediated by different *Pf*TFIIB regions.

RNA has been established to be a component of many cellular condensates and often plays important roles in condensate assembly and dissolution (Boeynaems *et al.*, 2019; Henninger *et al.*, 2021). In the transcription condensate model (Fig. 9), derived from studies investigating the molecular mechanisms underlying the function of super-enhancers in cultured mammalian cells, low RNA concentrations, resulting from the production of short RNAs during the initial phase of transcription activation, stimulate partitioning of transcription activators, co-activators and components of the general RNA Pol II transcription machinery into transcription condensates, thus increasing transcription activity (Henninger *et al.*, 2021). However, productive transcription and the resulting increase of RNA concentrations within the transcription condensate eventually leads to unfavourable RNA-to-protein ratios and a charge imbalance that ultimately causes condensate dissolution (Henninger *et al.*, 2021).

While RNA plays a stimulatory role in *Pf*TFIIB condensation and results in an increase in condensed fraction (Fig. 56), the results presented here do not provide any evidence for an inhibitory effect of high RNA concentrations on *Pf*TFIIB condensation. This is surprising given the high concentrations of total yeast RNA used in our assays. It may be that such an inhibitory effect requires RNA of a certain type or length, which is not abundant in the total yeast RNA preparation used here. Indeed results of a recent computational study using multiscale simulations suggest that the maximum RNA concentration tolerated within protein-RNA condensates is strongly dependent on RNA length, with longer RNAs facilitating phase separation at greater RNA-to-protein ratios (Sanchez-Burgos *et al.*, 2023). These *in silico* results are consistent with several studies demonstrating the importance of

RNA features such as length, sequence, and structure in promoting or inhibiting condensate formation and determining the properties of RNA-rich condensates (Roden and Gladfelter, 2017; Langdon *et al.*, 2018; Maharana *et al.*, 2018; Boeynaems *et al.*, 2019).

Intriguingly, transcription activation in *P. falciparum* is characterised by a high prevalence of bidirectional transcription, leading to the generation of coregulated coding and non-coding transcripts (Adjalley *et al.*, 2016; Chappell *et al.*, 2020). Furthermore, long non-coding RNAs appear to play a key role *P. falciparum* gene regulation (Batugedara *et al.*, 2023). For example, a family of G/C-rich non-coding RNAs colocalise with the site of active *var* gene expression and play an essential role in *var* gene activation (Guizetti, Barcons-Simon and Scherf, 2016; Barcons-Simon *et al.*, 2020). Such long RNAs have been shown to bind to upstream promoter DNA regions and have been proposed to recruit gene regulatory factors, including members of the RNA Pol II transcription machinery (Batugedara *et al.*, 2023). Given that transcription of at least some *P. falciparum* genes occurs in localised foci, often associated with the nuclear periphery (Duraisingh *et al.*, 2005; Ralph, Scheidig-Benatar and Scherf, 2005; Volz *et al.*, 2010; Moraes *et al.*, 2013, see Section 1.4.6), it is tempting to speculate that non-coding RNAs may also regulate transcription condensate assembly in *P. falciparum* and perhaps tether these condensates to specific genomic and/or nuclear locations. This hypotheses could be tested by further *in vitro* experiments using physiological RNA species, such as the family of G/C-rich non-coding RNAs mentioned above (Guizetti, Barcons-Simon and Scherf, 2016; Barcons-Simon *et al.*, 2020), and by investigating the localisation of these RNA species relative to *Pf*TFIIB and other general transcription factors in *P. falciparum* parasites, for example by using a combination of RNA-FISH and immunofluorescence approaches.

Our investigation of *Pf*TFIIB condensation in the presence of *Pf*TBP and *Pf*TLP demonstrates that *Pf*TFIIB forms homogenously mixed condensates with both *Pf*TBP and *Pf*TLP *in vitro* (Figs. 58-60, 65-67). These results are consistent with interactions between *Pf*TFIIB and *Pf*TBP or *Pf*TLP seen in the absence and presence of DNA, in GST-pull down experiments and EMSAs, respectively (Figs. 40 and 41). It seems likely that the same interactions also play a role in the partitioning of *Pf*TBP and *Pf*TLP into mixed phase-separated condensates with *Pf*TFIIB.

In agreement with the established general role of IDRs as common drivers of LLPS and determinants of selective partitioning into mixed condensates (Elbaum-Garfinkle *et al.*, 2015; Nott *et al.*, 2015; Smith *et al.*, 2016; Shin and Brangwynne, 2017; Wurdeman *et al.*, 2017; Boehning *et al.*, 2018; Boija *et al.*, 2018; Alberti, Gladfelter and Mittag, 2019; Dignon, Best and Mittal, 2020; Lyons *et al.*, 2023), the *Pf*TBP IDR and *Pf*TLP-IDR1 are sufficient to facilitate the formation of homogenous mixed

condensates with *Pf*TFIIB in the absence of DNA (Figs. 61-64 and 68-70). Similarly, the *Pf*RBP1 CTD readily forms mixed condensates with *Pf*TFIIB (Figs. 71 and 72).

However, it is important to note that the *P. falciparum* IDRs characterised here do not fully recapitulate the properties of full-length proteins within mixed condensates. For example, *Pf*TLP-IDR1 is highly mobile in *Pf*TFIIB condensates, whereas full-length *Pf*TLP is largely immobile (Figs. 66 and 69), suggesting the PfTLP structured DNA-binding domain mediates additional intermolecular interactions that contribute to hardening of *Pf*TLP in mixed condensates. In addition, there were notable differences in the properties of mixed condensates formed by full-length proteins or IDRs in the presence of DNA. Specifically, the N-terminal IDRs of *Pf*TFIIB and *Pf*TBP were more heterogeneously distributed in DNA condensates than their full-length counterparts (Figs. 55, 60 and 64). Thus, while the IDRs appear to play an important role in the recruitment of *Pf*TBP or *Pf*TLP into *Pf*TFIIB condensates, interactions mediated by the structured DNA-binding domains of *Pf*TBP and *Pf*TLP and *Pf*TFIIB also contribute to mixed condensate properties.

In well-studied eukaryotes, the majority of research into the regulation of transcription condensates has focused on the role of coactivators, such as the mediator complex, and gene-specific transcription factors (Boija *et al.*, 2018; Sabari *et al.*, 2018; Zamudio *et al.*, 2019; Han *et al.*, 2020; Zhang *et al.*, 2022). In these studies, the mediator complex, specifically the Med1 mediator subunit, has been highlighted as a key scaffold protein with vital roles in condensate formation and the recruitment of additional positive regulators of transcription (Boija *et al.*, 2018; Sabari *et al.*, 2018; Zamudio *et al.*, 2019). The potential function of general transcription factors in the formation of transcription condensates, beyond their role as client proteins that are selectively partitioned to facilitate robust transcription, has so far not been addressed.

Importantly, the mediator subunit Med1 appears absent in *P. falciparum* (Bischoff and Vaquero, 2010; Iyer *et al.*, 2022). Furthermore, *P. falciparum* is characterised by a general paucity of gene-specific transcription factors compared to other organisms (Bischoff and Vaquero, 2010). Thus, it seems likely that the assembly of transcription condensates in *P. falciparum* is regulated by different protein components compared to well-studied eukaryotes.

Previous research by our group demonstrated that *P. falciparum* general transcription factors such as *Pf*TBP and *Pf*TLP and the RPB1 CTD possess a high propensity to undergo LLPS *in vitro* compared to their human counterparts (Karamanof, 2022, Pegram, 2022, Michowicz, in preparation). The results presented here demonstrate that *Pf*TFIIB is also a strong driver of *in vitro* biomolecular condensation and that *Pf*TFIIB can partition other *P. falciparum* general transcription factors into mixed condensates. These findings raise the interesting possibility that in *P. falciparum* the general

transcription factors play a more predominant role in transcription condensate assembly, perhaps compensating for the absence of important human transcription condensate scaffold proteins such as Med1. Following this model, it stands to reason that there may be fundamental differences in the interactions that regulate transcription condensate formation in *P. falciparum* and human cells.

Recent studies have shown that small molecule drugs can selectively partition into, and affect the properties of, specific cellular condensate compartments (Fang *et al.*, 2019; Klein *et al.*, 2020). For example, the anticancer drug cisplatin selectively partitions into mediator condensates at super enhancers, resulting in platinated DNA and, ultimately, the dissolution of these condensates (Klein *et al.*, 2020). Thus, cellular biomolecular condensates have emerged as highly attractive targets for drug discovery (Mitrea *et al.*, 2022). Selective disruption of *P. falciparum* transcription condensate function constitutes a promising avenue for future drug discovery research. The success of such drug discovery efforts is contingent on a deep understanding of the interactions that govern *P. falciparum* transcription condensates formation and composition, as well as the identification of *P. falciparum*-specific features that could be selectively targeted. The characterisation of condensates formed by *P. falciparum* transcription factors described here and in previous work by our group represent an important first step towards this goal.

While *in vitro* LLPS experiments provide vital insight into the properties of phase separating proteins and their role in the assembly of cellular condensates, they cannot fully recapitulate the complexities of biomolecular condensation within living cells. In the physiological context, additional interactors (Dignon, Best and Mittal, 2020; Shin, 2022), the timing and localisation of protein and RNA synthesis (Lin *et al.*, 2023), changes in environmental conditions (Dignon, Best and Mittal, 2020), and even the presence of small molecules and metabolites, such as adenosine triphosphate (Patel *et al.*, 2017; Li, Wang and Lai, 2023), significantly impact phase separation of biomolecules and the assembly and properties of biomolecular condensates. With this in mind, an important next step will be to assess the role of *Pf*TFIIB and other *P. falciparum* general transcription factors in the assembly of phase-separated biomolecular condensates in more complex protein mixtures, composed of either purified recombinant proteins or their natural counterparts within *P. falciparum* nuclear extracts and, ultimately, in cultured *P. falciparum* parasites.

The data presented here should provide a solid basis for future biochemical investigations into the biomolecular condensation of *Pf*TFIIB in combination with other *P. falciparum* interaction partners and with physiological RNA species. Furthermore, the immunofluorescence data for *Pf*TBP and *Pf*TLP, showing a punctate localisation of *Pf*TBP and *Pf*TLP within blood-stage trophozoites and schizonts, provides a useful starting point for more detailed fluorescence microscopy studies that

combine immunofluorescence, using antibodies characterised in this study, with RNA- and DNA-FISH to localise specific RNAs and DNA regions. These experiments could be complemented by studies using transgenic *P. falciparum* strains expressing fluorescently tagged transcription factors. Live cell imaging of endogenously expressed proteins helps to avoid potential pitfalls associated with immunocytochemistry methods. For example, cell fixation prior to immunostaining has been shown to affect the detection and size of puncta containing specific biomolecules (Irgen-Gioro *et al.*, 2022). The use of transgenic *P. falciparum* strains expressing fluorescent protein-tagged *P. falciparum* general transcription factors at levels comparable to their physiological counterparts would thus allow greater confidence in their localisation and also provide a platform to delineate the functional regions within individual condensate components and to interrogate the properties of transcription condensates *in vivo,* their colocalization with sites of active transcription, and their impact on transcriptional output.

It is important to point out that *in vitro* biomolecular condensation reactions produce far larger assemblies than those observed in the context of living cells. Furthermore, the much smaller transcription condensates in live cells have so far mostly been studied in the context of super-enhancers (Hnisz *et al.*, 2017; Boija *et al.*, 2018; Sabari *et al.*, 2018; Zamudio *et al.*, 2019). A localised concentration of positive-acting transcription factors during regular transcription activation occurs at much smaller scale, probably in form of nanoscale transcription factories at gene regulatory regions (Chong *et al.*, 2018, 2022). There is much discussion in the field as to whether or not these nanoscale transcription factories represent phase separated compartments or hubs of positive-acting transcription factors that form as a result of multivalent interactions below the threshold concentration for phase separation (Section 1.4.3; Chong *et al.*, 2018, 2022; Palacio and Taatjes, 2022). Interestingly, a recent *in vitro* study demonstrated the formation of quasi-spherical pre-percolation clusters of FUS and other RNA-binding proteins at concentrations below the saturation concentration for phase separation (Kar *et al.*, 2022). These pre-percolation clusters consist of only tens to thousands of molecules that associate in the absence of a phase boundary (Kar *et al.*, 2022). Importantly, the study further suggests that pre-percolation clusters are formed through the same multivalent intermolecular interactions that drive phase separation of condensate components. Thus, the extensive functional characterisation of homotypic interactions between *Pf*TFIIB molecules as well as heterotypic interactions between *Pf*TFIIB and DNA, *Pf*TBP and *Pf*TLP presented in this thesis, provides invaluable insight into the type of multivalent interactions that may facilitate not just the formation of phase separated condensates, but also the formation of smaller nanoscale transcription factor hubs in *P. falciparum*.

Chapter 7: Summary and Outlook

This thesis presents a functional characterisation of *Pf*TFIIB, *Pf*TBP and *Pf*TLP, focusing on the potential role of these factors in the assembly of the RNA Pol II transcription initiation complex in *P. falciparum*. The work demonstrates *P. falciparum* expresses two highly divergent TBP family proteins with DNA-binding activity that is clearly distinct from prototypical TATA-binding proteins characterised in well-studied eukaryotes. Although *Pf*TBP and *Pf*TLP preferentially bind A/T-rich DNA through DNA minor groove interactions, they are unable to interact with the prototypical TATA box element in a sequence-specific manner and show no detectable preference for specific sequence motifs within the context of A/T-rich DNA regions. These results support that, in the absence of other transcription factors, *Pf*TBP or *Pf*TLP DNA binding does not drive the recognition of a defined promoter within the A/T-rich context of the *P. falciparum* genome. The work presented here further demonstrates that the C-terminal structured region of *Pf*TFIIB, comprised of cyclin-like repeats, is sufficient to bind dsDNA to form stable nucleoprotein complexes. The results suggest that *Pf*TFIIB binds to DNA in an unspecific manner, without detectable preference for specific nucleotide sequences or overall nucleotide content. Interestingly, *Pf*TFIIB interacts with *Pf*TBP and *Pf*TLP both in the absence and presence of DNA, stimulates the DNA-binding activity of *Pf*TBP and *Pf*TLP and forms stable *Pf*TBP-*Pf*TFIIB-DNA and *Pf*TLP-*Pf*TFIIB-DNA nucleoprotein complexes. However, *Pf*TFIIB interactions do not detectably enhance the sequence-selectivity of *Pf*TBP- or *Pf*TLP-DNA interactions. Thus, recognition of genomic regions at which transcription is preferentially initiated must involve additional factors. Alternatively, transcription initiation in *P. falciparum* may not involve the recognition of specific promoter regions at all and may instead be regulated largely at the level of chromatin structure and dynamics that determine accessibility of specific DNA regions.

The *in vitro* characterisation of phase-separated condensates formed by *Pf*TFIIB presented here provides first insights into a potential role of *Pf*TFIIB in the formation of transcription condensates or transcriptional hubs in *P. falciparum*. *Pf*TFIIB is found to be a strong driver of *in vitro* biomolecular condensation, with a C_{sat} in the nanomolar range. Interestingly, *Pf*TFIIB interacts with *Pf*TBP, *Pf*TLP and the CTD of the largest *Pf*RNA Pol II subunit *Pf*RPB1 to form mixed condensates *in vitro*. Furthermore, *Pf*TFIIB condensate formation is modulated by DNA and RNA. Together with the finding that *Pf*TBP and *Pf*TLP localise to discrete foci in cultured *P. falciparum* blood stage parasites, these observations strongly support the notion that *Pf*TFIIB may constitute an important component of transcription condensates in *P. falciparum*.

Taken together, the results of the studies presented here provide novel insights into the regulation of transcription initiation in *P. falciparum* and should provide an excellent basis for future studies.

Given that that key regulators of *H. sapiens* transcription condensates are either entirely absent or highly divergent in *P. falciparum*, it seems likely that there are significant differences in the molecular interactions that govern transcription condensate assembly and in the physicochemical properties of phase separated transcription condensates or transcription hubs between *P. falciparum* and the human host. The work presented here presents a starting point for future efforts to pinpoint *P. falciparum*-specific features, and to identify compounds that selectively target *P. falciparum* transcription.

References

Adebolajo, D (2016) *Functional characterisation of the Plasmodium falciparum TBP-like protein (PfTLP) DNA-binding activity.* Honours Thesis. University of Cape Town.

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. Available at: https://doi.org/10.1038/nrg3293.

Adjalley, S.H. *et al.* (2016) 'Landscape and Dynamics of Transcription Initiation in the Malaria Parasite *Plasmodium falciparum*', *Cell Reports*, 14(10), pp. 2463–2475. Available at: https://doi.org/10.1016/j.celrep.2016.02.025.

Akhtar, W. and Veenstra, G.J.C. (2009) 'TBP2 is a substitute for TBP in *Xenopus* oocyte transcription', 2, pp. 1–11. Available at: https://doi.org/10.1186/1741-7007-7-45.

Akhtar, W. and Veenstra, G.J.C. (2011) 'TBP-related factors: A paradigm of diversity in transcription initiation', *Cell and Bioscience*, 1(1), p. 23. Available at: https://doi.org/10.1186/2045-3701-1-23.

Alberti, S. *et al.* (2018) 'A User's Guide for Phase Separation Assays with Purified Proteins', *Journal of Molecular Biology*, 430(23), pp. 4806–4820. Available at: https://doi.org/10.1016/j.jmb.2018.06.038.

Alberti, S. and Dormann, D. (2019) 'Liquid-Liquid Phase Separation in Disease', *Annual Review of Genetics*, 53, pp. 171–194. Available at: https://doi.org/10.1146/annurev-genet-112618-043527.

Alberti, S., Gladfelter, A. and Mittag, T. (2019) 'Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates', *Cell*, 176(3), pp. 419–434. Available at: https://doi.org/10.1016/j.cell.2018.12.035.

Alemasova, E.E. and Lavrik, O.I. (2022) 'Critical Reviews and Perspectives A sePARate phase ? Poly (ADP-ribose) versus RNA in the organization of biomolecular condensates', 50(19), pp. 10817–10838.

Allepuz-Fuster, P. *et al.* (2019) 'RNA polymerase II plays an active role in the formation of gene loops through the Rpb4 subunit', *Nucleic Acids Research*, 47(17), pp. 8975–8987. Available at: https://doi.org/10.1093/nar/gkz597.

Altobelli, G., Poch, O. and Gazdag, E. (2009) 'TBP2 is essential for germ cell development by regulating transcription and chromatin condensation in the oocyte', pp. 2210–2223. Available at: https://doi.org/10.1101/gad.535209.associated.

Altschul, S.F. *et al.* (1990) 'Basic Local Alignment Search Tool', *Journal of Molecular Biology*, pp. 403–

410. Available at: https://doi.org/10.1016/S0022-2836(05)80360-2.

Amino, R. *et al.* (2006) 'Quantitative imaging of *Plasmodium* transmission from mosquito to mammal', *Nature Medicine*, 12(2), pp. 220–224. Available at: https://doi.org/10.1038/nm1350.

Aravind, L. *et al.* (2003) 'Plasmodium Biology : Genomic Gleanings', 115, pp. 771–785.

Armougom, F. *et al.* (2006) 'Expresso: Automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee', *Nucleic Acids Research*, 34(WEB. SERV. ISS.), pp. 604–608. Available at: https://doi.org/10.1093/nar/gkl092.

Ashley, E.A. and Phyo, A.P. (2018) 'Drugs in Development for Malaria', *Drugs*, 78(9), pp. 861–879. Available at: https://doi.org/10.1007/s40265-018-0911-9.

Ay, F. *et al.* (2014) 'Three-dimensional modeling of the *P. falciparum* genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression', *Genome Research*, 24(6), pp. 974–988. Available at: https://doi.org/10.1101/gr.169417.113.

Bailly, C. *et al.* (1996) 'PCR-based development of DNA substrates containing modified bases: An efficient system for investigating the role of the exocyclic groups in chemical and structural recognition by minor groove binding drugs and proteins', *Proceedings of the National Academy of Sciences of the United States of America*, 93(24), pp. 13623–13628. Available at: https://doi.org/10.1073/pnas.93.24.13623.

Balduf, C. *et al.* (2004) 'TBP2 , a Vertebrate-Specific Member of the TBP Family , Is Required in Embryonic Development of Zebrafish', 14, pp. 593–598. Available at: https://doi.org/10.1016/j.

Banani, S.F. *et al.* (2016) 'Compositional Control of Phase-Separated Cellular Bodies', *Cell*, 166(3), pp. 651–663. Available at: https://doi.org/10.1016/j.cell.2016.06.010.Compositional.

Banani, S.F. *et al.* (2017) 'Biomolecular condensates: Organizers of cellular biochemistry', *Nature Reviews Molecular Cell Biology*, 18(5), pp. 285–298. Available at: https://doi.org/10.1038/nrm.2017.7.

Banerjee, P.R. *et al.* (2017) 'Reentrant Phase Transition Drives Dynamic Substructure Formation in Ribonucleoprotein Droplets', *Angewandte Chemie - International Edition*, 56(38), pp. 11354–11359. Available at: https://doi.org/10.1002/anie.201703191.

Bangur, C.S. *et al.* (1999) 'An interaction between the N-terminal region and the core domain of yeast TFIIB promotes the formation of TATA-binding protein-TFIIB-DNA complexes', *Journal of Biological Chemistry*, 274(33), pp. 23203–23209. Available at:

https://doi.org/10.1074/jbc.274.33.23203.

Banks, G.C., Mohr, B. and Reeves, R. (1999) 'The HMG-I(Y) A · T-hook peptide motif confers DNA-binding specificity to a structured chimeric protein', *Journal of Biological Chemistry*, 274(23), pp. 16536–16544. Available at: https://doi.org/10.1074/jbc.274.23.16536.

Barberis, A. *et al.* (1993) 'Delineation of two functional regions of transcription factor TFIIB', *Proceedings of the National Academy of Sciences of the United States of America*, 90(12), pp. 5628–5632. Available at: https://doi.org/10.1073/pnas.90.12.5628.

Barcons-Simon, A. *et al.* (2020) 'CRISPR interference of a clonally variant GC-rich noncoding RNA family leads to general repression of var genes in *Plasmodium falciparum*', *mBio*, 11(1). Available at: https://doi.org/10.1128/mBio.03054-19.

Bareket-Samish, A., Cohen, I. and Haran, T.E. (2000) 'Signals for TBP/TATA box recognition', *Journal of Molecular Biology*, 299(4), pp. 965–977. Available at: https://doi.org/10.1006/jmbi.2000.3797.

Basehoar, A.D., Zanton, S.J. and Pugh, B.F. (2004) 'Identification and distinct regulation of yeast TATA box-containing genes', *Cell*, 117(6), p. 847. Available at: https://doi.org/10.1016/j.cell.2004.05.022.

Basu, S. *et al.* (2020) 'Unblending of Transcriptional Condensates in Human Repeat Expansion Disease', *Cell*, 181(5), pp. 1062–1079. Available at: https://doi.org/10.1016/j.cell.2020.04.018.

Batugedara, G. *et al.* (2023) 'Novel insights into the role of long non-coding RNA in the human malaria parasite, *Plasmodium falciparum*', *Nature Communications*, 14(1). Available at: https://doi.org/10.1038/s41467-023-40883-w.

Baumann, D.G. and Gilmour, D.S. (2017) 'A sequence-specific core promoter-binding transcription factor recruits TRF2 to coordinately transcribe ribosomal protein genes', 45(18), pp. 10481–10491. Available at: https://doi.org/10.1093/nar/gkx676.

Beier, J.C. (1998) 'MALARIA PARASITE DEVELOPMENT'.

Beri, D., Balan, B. and Tatu, U. (2018) 'Commit, hide and escape: The story of *Plasmodium* gametocytes', *Parasitology*, 145(13), pp. 1772–1782. Available at: https://doi.org/10.1017/S0031182018000926.

Berry, J., Brangwynne, C.P. and Haataja, M. (2018) 'Physical principles of intracellular organization via active and passive phase transitions', *Reports on Progress in Physics*, 81(4). Available at: https://doi.org/10.1088/1361-6633/aaa61e.

Biggs, J.R., Ahn, N.G. and Kraft, A.. (1998) 'Activation of the mitogen-activated protein kinase

pathway in U937 leukemic cells induces phosphorylation of the amino terminus of the TATA-binding protein', *Cell Growth and Differentiation*, 9(August), pp. 667–676.

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

Birkholtz, L.M. *et al.* (2008) 'Heterologous expression of plasmodial proteins for structural studies and functional annotation', *Malaria Journal*, 7, pp. 1–20. Available at: https://doi.org/10.1186/1475-2875-7-197.

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). Available at: https:// doi.org/10.1186/1471-2164-11-34.

Bloomfield, V.A. (1996) 'DNA condensation', *Current Opinion in Structural Biology*, 6(3), pp. 334–341. Available at: https://doi.org/10.1016/S0959-440X(96)80052-2.

Boehning, M. *et al.* (2018) 'RNA polymerase II clustering through carboxy-terminal domain phase separation', *Nature Structural and Molecular Biology*, 25(9), pp. 833–840. Available at: https://doi.org/10.1038/s41594-018-0112-y.

Boeynaems, S. *et al.* (2019) 'Spontaneous driving forces give rise to protein−RNA condensates with coexisting phases and complex material properties', *Proceedings of the National Academy of Sciences of the United States of America*, 116(16), pp. 7889–7898. Available at: https://doi.org/10.1073/pnas.1821038116.

Böhme, U. *et al.* (2019) 'Progression of the canonical reference malaria parasite genome from 2002–2019 [version 1; peer review: 2 approved, 1 approved with reservations]', *Wellcome Open Research*, 4(May), pp. 1–22. Available at: https://doi.org/10.12688/wellcomeopenres.15194.1.

Boija, A. *et al.* (2018) 'Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains', *Cell*, 175(7), pp. 1842-1855.e16. Available at: https://doi.org/10.1016/j.cell.2018.10.042.

Bonam, S.R. *et al.* (2021) '*Plasmodium falciparum* malaria vaccines and vaccine adjuvants', *Vaccines*, 9(10), pp. 1–35. Available at: https://doi.org/10.3390/vaccines9101072.

Boyko, S. *et al.* (2019) 'Liquid – liquid phase separation of tau protein : The crucial role of electrostatic interactions', 294, pp. 11054–11059. Available at: https://doi.org/10.1074/jbc.AC119.009198.

Bozdech, Z. *et al.* (2003) 'The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*', *PLoS Biology*, 1(1), pp. 85–100. Available at: https://doi.org/10.1371/journal.pbio.0000005.

Brancorsini, S. *et al.* (2008) 'TIPT, a male germ cell-specific partner of TRF2, is chromatin associated and interacts with HP1', 4101. Available at: https://doi.org/10.4161/cc.7.10.5835.

Brancucci, N.M.B. *et al.* (2014) 'Heterochromatin protein 1 secures survival and transmission of malaria parasites', *Cell Host and Microbe*, 16(2), pp. 165–176. Available at: https://doi.org/10.1016/j.chom.2014.07.004.

Bratkowski, M. *et al.* (2018) 'Structural dissection of an interaction between transcription initiation and termination factors implicated in promoter-terminator cross-talk', *Journal of Biological Chemistry*, 293(5), pp. 1651–1665. Available at: https://doi.org/10.1074/jbc.M117.811521.

Bryant, J.M. *et al.* (2020) ' Exploring the virulence gene interactome with CRISPR / dC as9 in the human malaria parasite ', *Molecular Systems Biology*, 16(8), pp. 1–22. Available at: https://doi.org/10.15252/msb.20209569.

Bucher, P. (1990) 'Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences', *Journal of Molecular Biology*, 212(4), pp. 563–578. Available at: https://doi.org/10.1016/0022-2836(90)90223-9.

Buendía-Orozco, J., Guerrero, A. and Pastor, N. (2005) 'Model of the TBP-TFIIB complex from *Plasmodium falciparum*: Interface analysis and perspectives as a new target for antimalarial design', *Archives of Medical Research*, 36(4), pp. 317–330. Available at: https://doi.org/10.1016/j.arcmed.2005.03.020.

Bunnik, E.M. *et al.* (2014) 'DNA-encoded nucleosome occupancy is associated with transcription levels in the human malaria parasite *Plasmodium falciparum*', *BMC Genomics*, 15(1), pp. 1–15. Available at: https://doi.org/10.1186/1471-2164-15-347.

Buratowski, S. *et al.* (1989) 'Five intermediate complexes in transcription initiation by RNA polymerase II', *Cell*, 56(4), pp. 549–561. Available at: https://doi.org/10.1016/0092-8674(89)90578-3.

Buratowski, S. and Zhou, H. (1993) 'Functional domains of transcription factor TFIIB', *Proceedings of the National Academy of Sciences of the United States of America*, 90(12), pp. 5633–5637. Available at: https://doi.org/10.1073/pnas.90.12.5633.

Callebaut, I. *et al.* (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, pp. 1–20. Available at: https://doi.org/10.1186/1471-2164-6-100.

Carey, M., Peterson, C. and Smale, S. (1998) *Transcriptional Regulation in Eukaryotes Concepts, Strategies, and Techniques*. 2nd edn. COLD SPRING HARBOR LABORATORY PRESS.

Castañon-Sanchez, C.A. *et al.* (2010) 'Entamoeba histolytica: A unicellular organism containing two active genes encoding for members of the TBP family', *Protein Expression and Purification*, 70(1), pp. 48–59. Available at: https://doi.org/10.1016/j.pep.2009.12.007.

Chalkley, G.E. and Verrijzer, C.P. (1999) 'DNA binding site selection by RNA polymerase II TAFs: A TAF(II)250-TAF(II)150 complex recognizes the initiator', *EMBO Journal*, 18(17), pp. 4835–4845. Available at: https://doi.org/10.1093/emboj/18.17.4835.

Chappell, L. *et al.* (2020) 'Refining the transcriptome of the human malaria parasite *Plasmodium falciparum* using amplification-free RNA-seq', *BMC Genomics*, 21(1), pp. 1–19. Available at: https://doi.org/10.1186/s12864-020-06787-5.

Chen, H.T. and Hahn, S. (2003) 'Binding of TFIIB to RNA polymerase II: Mapping the binding site for the TFIIB zinc ribbon domain within the preinitiation complex', *Molecular Cell*, 12(2), pp. 437–447. Available at: https://doi.org/10.1016/S1097-2765(03)00306-X.

Chen, X. and Xu, Y. (2022) 'Structural insights into assembly of transcription preinitiation complex', *Current Opinion in Structural Biology*, 75(June), p. 102404. Available at: https://doi.org/10.1016/j.sbi.2022.102404.

Cheng, C., Jia, J.L. and Ran, S.Y. (2015) 'Polyethylene glycol and divalent salt-induced DNA reentrant condensation revealed by single molecule measurements', *Soft Matter*, 11(19), pp. 3927–3935. Available at: https://doi.org/10.1039/c5sm00619h.

Cheng, J., Sweredoski, M.J. and Baldi, P. (2005) 'Accurate prediction of protein disordered regions by mining protein structure data', *Data Mining and Knowledge Discovery*, 11(3), pp. 213–222. Available at: https://doi.org/10.1007/s10618-005-0001-y.

Cho, E.J. and Buratowski, S. (1999) 'Evidence that transcription factor IIB is required for a post-assembly step in transcription initiation', *Journal of Biological Chemistry*, 274(36), pp. 25807–25813. Available at: https://doi.org/10.1074/jbc.274.36.25807.

Choi, J.M., Holehouse, A.S. and Pappu, R. V. (2020) 'Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions', *Annual Review of Biophysics*, 49, pp. 107–133. Available at: https://doi.org/10.1146/annurev-biophys-121219-081629.

Chong, S. *et al.* (2018) 'Imaging dynamic and selective low-complexity domain interactions that control gene transcription', *Science*, 361(6400). Available at: https://doi.org/10.1126/science.aar2555.

Chong, S. *et al.* (2022) 'Tuning levels of low-complexity domain interactions to modulate endogenous oncogenic transcription', *Molecular Cell*, 82(11), pp. 2084-2097.e5. Available at: https://doi.org/10.1016/j.molcel.2022.04.007.

Claeys Bouuaert, C. *et al.* (2021) 'DNA-driven condensation assembles the meiotic DNA break machinery', *Nature*, 592(7852), pp. 144–149. Available at: https://doi.org/10.1038/s41586-021-03374-w.

Cogswell, F.B. (1992) 'The hypnozoite and relapse in primate malaria', *Clinical Microbiology Reviews*, 5(1), pp. 26–35. Available at: https://doi.org/10.1128/CMR.5.1.26.

Coleman, R.A. and Pugh, B.F. (1995) 'Evidence for functional binding and stable sliding of the TATA binding protein on nonspecific DNA', *Journal of Biological Chemistry*, 270(23), pp. 13850–13859. Available at: https://doi.org/10.1074/jbc.270.23.13850.

Coulson, R.M.R., 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. Available at: https://doi.org/10.1101/gr.2218604.

Cowman, A.F. *et al.* (2016) 'Review Malaria : Biology and Disease', *Cell*, 167(3), pp. 610–624. Available at: https://doi.org/10.1016/j.cell.2016.07.055.

Craig, A. and Scherf, A. (2001) 'Molecules on the surface of the *Plasmodium falciparum* infected erythrocyte and their role in malaria pathogenesis and immune evasion', *Molecular and Biochemical Parasitology*, 115(2), pp. 129–143. Available at: https://doi.org/10.1016/S0166-6851(01)00275-4.

Crowley, T. *et al.* (1993) 'A New factor related to TATA-binding protein has highly restricted expression patterns in Drosophila', *Nature*, 361, pp. 557–561.

Dantonel, J.C. *et al.* (1999) 'The TBP-like factor: An alternative transcription factor in Metazoa?', *Trends in Biochemical Sciences*, 24(9), pp. 335–339. Available at: https://doi.org/10.1016/S0968-0004(99)01436-X.

Das, D. and Deniz, A.A. (2023) 'Topological Considerations in Biomolecular Condensation', *Biomolecules*, 13(1). Available at: https://doi.org/10.3390/biom13010151.

Das, R.K. and Pappu, R. V. (2013) 'Conformations of intrinsically disordered proteins are influenced

by linear sequence distributions of oppositely charged residues', *Proceedings of the National Academy of Sciences of the United States of America*, 110(33), pp. 13392–13397. Available at: https://doi.org/10.1073/pnas.1304749110.

Davies. G. (2022) *Phase Separation Characteristics of Homo sapiens General Transcription Factor IIB.* Honours Thesis. University of Cape Town.

Delgadillo, R.F. *et al.* (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. Available at: https://doi.org/10.1021/bi8018724.

Deng, W. and Roberts, S.G.E. (2005) 'A core promoter element downstream of the TATA box that is recognized by TFIIB', *Genes and Development*, 19(20), pp. 2418–2423. Available at: https://doi.org/10.1101/gad.342405.

Deng, W. and Roberts, S.G.E. (2007) 'TFIIB and the regulation of transcription by RNA polymerase II', *Chromosoma*, 116(5), pp. 417–429. Available at: https://doi.org/10.1007/s00412-007-0113-9.

Deniz, A.A. (2022) 'Percolation physics and density transition frameworks converge in biomolecular condensation', *Proceedings of the National Academy of Sciences of the United States of America*, 119(32), pp. 1–3. Available at: https://doi.org/10.1073/pnas.2210177119.

DePristo, M.A., Zilversmit, M.M. and Hartl, D.L. (2006) 'On the abundance, amino acid composition, and evolutionary dynamics of low-complexity regions in proteins', *Gene*, 378(1–2), pp. 19–30. Available at: https://doi.org/10.1016/j.gene.2006.03.023.

Dichtl, B. *et al.* (2002) 'A role for SSU72 in balancing RNA polymerase II transcription elongation and termination', *Molecular Cell*, 10(5), pp. 1139–1150. Available at: https://doi.org/10.1016/S1097-2765(02)00707-4.

Dignon, G.L., Best, R.B. and Mittal, J. (2020) 'Biomolecular phase separation: From molecular driving forces to macroscopic properties', *Annual Review of Physical Chemistry*, 71, pp. 53–75. Available at: https://doi.org/10.1146/annurev-physchem-071819-113553.

Donczew, R. and Hahn, S. (2018) 'Mechanistic Differences in Transcription Initiation at TATA-Less and TATA-Containing Promoters', *Molecular and Cellular Biology*, 38(1). Available at: https://doi.org/10.1128/mcb.00448-17.

Dosztányi, Z., Csizmok, V., *et al.* (2005) 'IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content', *Bioinformatics*, 21(16), pp.

3433–3434. Available at: https://doi.org/10.1093/bioinformatics/bti541.

Dosztányi, Z., Csizmók, V., *et al.* (2005) 'The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins', *Journal of Molecular Biology*, 347(4), pp. 827–839. Available at: https://doi.org/10.1016/j.jmb.2005.01.071.

Duffy, M.F. *et al.* (2014) 'Epigenetic regulation of the *Plasmodium falciparum* genome', *Briefings in Functional Genomics*, 13(3), pp. 203–216. Available at: https://doi.org/10.1093/bfgp/elt047.

Duraisingh, M.T. *et al.* (2005) 'Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*', *Cell*, 121(1), pp. 13–24. Available at: https://doi.org/10.1016/j.cell.2005.01.036.

Düster, R. *et al.* (2021) '1,6-Hexanediol, commonly used to dissolve liquid-liquid phase separated condensates, directly impairs kinase and phosphatase activities', *Journal of Biological Chemistry*, 296, p. 100260. Available at: https://doi.org/10.1016/J.JBC.2021.100260.

Duttke, S.H.C. *et al.* (2014) 'TRF2 and the evolution of the bilateria', *Genes and Development*, 28(19), pp. 2071–2076. Available at: https://doi.org/10.1101/gad.250563.114.

Duttke, S.H.C. (2015) 'Evolution and diversification of the basal transcription machinery', *Trends in Biochemical Sciences*, 40(3), pp. 127–129. Available at: https://doi.org/10.1016/j.tibs.2015.01.005.

Eelu, H. (2016) *Transcription regulation in Plasmodium falciparum: Characterisation of P. falciparum transcription factor IIB (PfTFIIB) PfTBP-independent DNA-binding activity*. Honours Thesis. University of Cape Town.

Elbaum-Garfinkle, S. *et al.* (2015) 'The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics', *Proceedings of the National Academy of Sciences of the United States of America*, 112(23), pp. 7189–7194. Available at: https://doi.org/10.1073/pnas.1504822112.

Evans, R., Fairley, J.A. and Roberts, S.G.E. (2001) 'Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB', *Genes and Development*, 15(22), pp. 2945–2949. Available at: https://doi.org/10.1101/gad.206901.

Faiger, H., Ivanchenko, M. and Haran, T.E. (2007) 'Nearest-neighbor non-additivity versus long-range non-additivity in TATA-box structure and its implications for TBP-binding mechanism', *Nucleic Acids Research*, 35(13), pp. 4409–4419. Available at: https://doi.org/10.1093/nar/gkm451.

Faitar, S.L., Brodie, S.A. and Ponticelli, A.S. (2001) 'Promoter-Specific Shifts in Transcription Initiation

Conferred by Yeast TFIIB Mutations Are Determined by the Sequence in the Immediate Vicinity of the Start Sites', *Molecular and Cellular Biology*, 21(14), pp. 4427–4440. Available at: https://doi.org/10.1128/mcb.21.14.4427-4440.2001.

Fang, M.Y. *et al.* (2019) 'Small-Molecule Modulation of TDP-43 Recruitment to Stress Granules Prevents Persistent TDP-43 Accumulation in ALS/FTD', *Neuron*, 103(5), pp. 802-819.e11. Available at: https://doi.org/10.1016/j.neuron.2019.05.048.

Fang, S.M. and Burton, Z.F. (1996) 'RNA polymerase II-associated protein (RAP) 74 binds transcription factor (TF) IIB and blocks TFIIB-RAP30 binding', *Journal of Biological Chemistry*, 271(20), pp. 11703–11709. Available at: https://doi.org/10.1074/jbc.271.20.11703.

Feric, M. *et al.* (2016) 'Coexisting Liquid Phases Underlie Nucleolar Subcompartments', *Cell*, 165(7), pp. 1686–1697. Available at: https://doi.org/10.1016/j.cell.2016.04.047.

Fisher, C.K. and Stultz, C.M. (2011) 'Constructing ensembles for intrinsically disordered proteins', *Current Opinion in Structural Biology*, 21(3), pp. 426–431. Available at: https://doi.org/10.1016/j.sbi.2011.04.001.

Flueck, C. *et al.* (2009) '*Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors', *PLoS Pathogens*, 5(9). Available at: https://doi.org/10.1371/journal.ppat.1000569.

Fong, K. wing *et al.* (2013) 'Whole-genome screening identifies proteins localized to distinct nuclear bodies', *Journal of Cell Biology*, 203(1), pp. 149–164. Available at: https://doi.org/10.1083/jcb.201303145.

Francis, S.E., Sullivan, D.J. and Goldberg, and D.E. (1997) ' Hemoglobin Metabolism in the Malaria Parasite *Plasmodium falciparum* ', *Annual Review of Microbiology*, 51(1), pp. 97–123. Available at: https://doi.org/10.1146/annurev.micro.51.1.97.

Freitas-Junior, L.H. *et al.* (2000) 'Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*', *Nature*, 407(6807), pp. 1018–1022. Available at: https://doi.org/10.1038/35039531.

Galloway, C.A., Sowden, M.P. and Smith, H.C. (2003) 'Increasing the yield of soluble recombinant protein expressed in *E. coli* by Induction during late log phase', *BioTechniques*, 34(3), pp. 524–530. Available at: https://doi.org/10.2144/03343st04.

Gardner, M.J. *et al.* (2002) 'Genome sequence of the human malaria parasite *Plasmodium falciparum*', *Nature*, 419(6906), pp. 498–511. Available at: https://doi.org/10.1038/nature01097.

Geiger, J.H. *et al.* (1996a) 'Crystal structure of the yeast TFIIA/TBP/DNA complex', *Science*, pp. 830–836. Available at: https://doi.org/10.1126/science.272.5263.830.

Geiger, J.H. *et al.* (1996b) 'Crystal structure of the yeast TFIIA/TBP/DNA complex', *Science*, 272(5263), pp. 830–836. Available at: https://doi.org/10.1126/science.272.5263.830.

Ghamari, A. *et al.* (2013) 'In vivo live imaging of RNA polymerase II transcription factories in primary cells', *Genes and Development*, 27(7), pp. 767–777. Available at: https://doi.org/10.1101/gad.216200.113.

Gietl, A. *et al.* (2014) 'Eukaryotic and archaeal TBP and TFB/TF(II)B follow different promoter DNA bending pathways', *Nucleic Acids Research*, 42(10), pp. 6219–6231. Available at: https://doi.org/10.1093/nar/gku273.

Gilfillan, S. *et al.* (2005) 'Efficient binding of NC2·TATA-binding protein to DNA in the absence of TATA', *Journal of Biological Chemistry*, 280(7), pp. 6222–6230. Available at: https://doi.org/10.1074/jbc.M406343200.

Glossop, J.A., Dafforn, T.R. and Roberts, S.G.E. (2004) 'A conformational change in TFIIB is required for activator-mediated assembly of the preinitiation complex', *Nucleic Acids Research*, 32(5), pp. 1829–1835. Available at: https://doi.org/10.1093/nar/gkh504.

Gopalakrishnan, A.M. *et al.* (2009) '*Plasmodium falciparum*: Preinitiation complex occupancy of active and inactive promoters during erythrocytic stage', *Experimental Parasitology*, 121(1), pp. 46–54. Available at: https://doi.org/10.1016/j.exppara.2008.09.016.

Greenwood, B. *et al.* (2015) 'Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: Final results of a phase 3, individually randomised, controlled trial', *The Lancet*, 386(9988), pp. 31–45. Available at: https://doi.org/10.1016/S0140-6736(15)60721-8.

Guillebault, D. *et al.* (2002) 'A new class of transcription initiation factors, intermediate between TATA box-binding proteins (TBPs) and TBP-like factors (TLFs), is present in the marine unicellular organism, the dinoflagellate *Crypthecodinium cohnii*', *Journal of Biological Chemistry*, 277(43), pp. 40881–40886. Available at: https://doi.org/10.1074/jbc.M205624200.

Guizetti, J., Barcons-Simon, A. and Scherf, A. (2016) 'Trans-acting GC-rich non-coding RNA at var expression site modulates gene counting in malaria parasite', *Nucleic Acids Research*, 44(20), pp. 9710–9718. Available at: https://doi.org/10.1093/nar/gkw664.

Guo, L. *et al.* (2018) 'Nuclear-Import Receptors Reverse Aberrant Phase Transitions of RNA-Binding

Proteins with Prion-like Domains', *Cell*, 173(3), pp. 677-692.e20. Available at: https://doi.org/10.1016/j.cell.2018.03.002.

Guo, Y.E. *et al.* (2019) 'Pol II phosphorylation regulates a switch between transcriptional and splicing condensates', *Nature* [572(7770), pp. 543–548. Available at: https://doi.org/10.1038/s41586-019-1464-0.

Gupta, A.P. *et al.* (2017) 'Histone 4 lysine 8 acetylation regulates proliferation and host-pathogen interaction in *Plasmodium falciparum'*, *Epigenetics and Chromatin*, 10(1), pp. 1–17. Available at: https://doi.org/10.1186/s13072-017-0147-z.

Gupta, S. *et al.* (2007) 'DNA and protein footprinting analysis of the modulation of DNA binding by the N-terminal domain of the *Saccharomyces cerevisiae* TATA binding protein', *Biochemistry*, 46(35), pp. 9886–9898. Available at: https://doi.org/10.1021/bi7003608.

Ha, I. *et al.* (1993) 'Multiple functional domains of human transcription factor IIB: Distinct interactions with two general transcription factors and RNA polymerase II', *Genes and Development*, 7(6), pp. 1021–1032. Available at: https://doi.org/10.1101/gad.7.6.1021.

Haberle, V. and Stark, A. (2018) 'Eukaryotic core promoters and the functional basis of transcription initiation', *Nature Reviews Molecular Cell Biology*, 19(10), pp. 621–637. Available at: https://doi.org/10.1038/s41580-018-0028-8.

Hahn, S. *et al.* (1989) 'Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences', *Proceedings of the National Academy of Sciences of the United States of America*, 86(15), pp. 5718–5722. Available at: https://doi.org/10.1073/pnas.86.15.5718.

Hamard, P.J. *et al.* (2005) 'A functional interaction between ATF7 and TAF12 that is modulated by TAF4', *Oncogene*, 24(21), pp. 3472–3483. Available at: https://doi.org/10.1038/sj.onc.1208565.

Hamilton, W.L. *et al.* (2017) 'Extreme mutation bias and high AT content in *Plasmodium falciparum*', *Nucleic acids research*, 45(4), pp. 1889–1901. Available at: https://doi.org/10.1093/nar/gkw1259.

Han, X. *et al.* (2020) 'Roles of the BRD4 short isoform in phase separation and active gene transcription', *Nature Structural and Molecular Biology*, 27(4), pp. 333–341. Available at: https://doi.org/10.1038/s41594-020-0394-8.

Hansen, S.K. *et al.* (1997) 'Transcription Properties of a Cell Type – Specific TATA-Binding Protein , TRF', 91, pp. 71–83.

Harlow, E. and Lane, D. (eds) (1988) *Antibodies: A Laboratory Manual*. New York: Cold Spring Harbor: Cold Spring Harbor Laboratory.

Harris, C.R. *et al.* (2020) 'Array programming with NumPy', *Nature*, 585(7825), pp. 357–362. Available at: https://doi.org/10.1038/s41586-020-2649-2.

Harrison, A.F. and Shorter, J. (2017) 'RNA-binding proteins with prion-like domains in health and disease', *Biochemical Journal*, 474(8), pp. 1417–1438. Available at: https://doi.org/10.1042/BCJ20160499.

Hart, D.O. *et al.* (2007) 'Initiation of zebrafish haematopoiesis by the TATA-box-binding protein-related factor Trf3', *Nature*, 450(December), pp. 1082–1085. Available at: https://doi.org/10.1038/nature06349.

Hart, D.O. *et al.* (2009) 'Selective interaction between Trf3 and Taf3 required for early development and hematopoiesis.', *Developmental dynamics : an official publication of the American Association of Anatomists*, 238(10), pp. 2540–2549. Available at: https://doi.org/10.1002/dvdy.22083.

Hawkes, N.A., Evans, R. and Roberts, S.G.E. (2000) 'The conformation of the transcription factor TFIIB modulates the response to transcriptional activators in vivo', *Current Biology*, 10(5), pp. 273–276. Available at: https://doi.org/10.1016/S0960-9822(00)00363-8.

Hawking, F., Wilson, M. and Kenneth, G. (1971) 'Evidence for cyclic development and short-lived maturity in the gametocytes of *Plasmodium falciparum*', *Tropical Medicine*, 65(5), pp. 549–559.

Hayashi, F. *et al.* (1998) 'Human general transcription factor TFIIB: Conformational variability and interaction with VP16 activation domain', *Biochemistry*, 37(22), pp. 7941–7951. Available at: https://doi.org/10.1021/bi9801098.

He, Y. *et al.* (2016) 'Near-atomic resolution visualization of human transcription promoter opening', *Nature*, 533(7603), pp. 359–365. Available at: https://doi.org/10.1038/nature17970.

Heckman, K.L. and Pease, L.R. (2007) 'Gene splicing and mutagenesis by PCR-driven overlap extension', *Nature Protocols*, 2(4), pp. 924–932. Available at: https://doi.org/10.1038/nprot.2007.132.

Hellman, L.M. and Fried, M.G. (2007) 'Electrophoretic Mobility Shift Assay (EMSA) for Detecting ProteinNucleic Acid Interactions', *Nature Protocols*, 2(8), pp. 1849–1861.

Henninger, J.E. *et al.* (2021) 'RNA-Mediated Feedback Control of Transcriptional Condensates', *Cell*, 184(1), pp. 207-225.e24. Available at: https://doi.org/10.1016/j.cell.2020.11.030.

Henriques, T. *et al.* (2012) 'Transcription termination between polo and snap, two closely spaced tandem genes of *D. melanogaster*.', *Transcription*, 3(4), pp. 198–212. Available at:

https://doi.org/10.4161/trns.21967.

Hessler, T (2014) *Towards the identification of Plasmodium falciparum core promoter elements; establishment of an electrophoretic mobility assay*. Honours Thesis. University of Cape Town.

Hisatake, K., Roeder, R.G. and Horikoshi, M. (1993) 'Functional dissection of TFIIB domains required for TFIIB-TFIID-promoter complex formation and basal transcription activity', *Nature*, 363(6431), pp. 744–747. Available at: https://doi.org/10.1038/363744a0.

Hnisz, D. *et al.* (2017) 'A Phase Separation Model for Transcriptional Control', *Cell*, 169(1), pp. 13–23. Available at: https://doi.org/10.1016/j.cell.2017.02.007.

Hobbs, N.K. *et al.* (2006) 'Removing the Vertebrate-Specific TBP N Terminus Disrupts Placental β2m-Dependent Interactions with the Maternal Immune System Nicole', *Bone*, 23(1), pp. 1–7.

Hochheimer, A. *et al.* (2002) 'TRF2 associates with DREF and directs promoter-selective gene expression in *Drosophila'*, *Nature*, 420(6914), pp. 439–445. Available at: https://doi.org/10.1038/nature01167.

Hoeijmakers, W.A.M. *et al.* (2013) 'H2A.Z/H2B.Z double-variant nucleosomes inhabit the AT-rich promoter regions of the *Plasmodium falciparum* genome', *Molecular Microbiology*, 87(5), pp. 1061–1073. Available at: https://doi.org/10.1111/mmi.12151.

Hoffmann, A. and Roeder, R.G. (1991) 'Purification of his-tagged proteins in non-denaturing conditions suggests a convenient method for protein interaction studies', *Nucleic Acids Research*, 19(22), pp. 6337–6338. Available at: https://doi.org/10.1093/nar/19.22.6337.

Hoffmann, A. and Roeder, R.G. (1996) 'Cloning and characterization of human TAF20/15. Multiple interactions suggest a central role in TFIID complex formation', *Journal of Biological Chemistry*, 271(30), pp. 18194–18202. Available at: https://doi.org/10.1074/jbc.271.30.18194.

Holehouse, A.S. *et al.* (2015) 'CIDER: Classification of Intrinsically Disordered Ensemble Regions', *Biophysical Journal*, 108(2), p. 228a. Available at: https://doi.org/10.1016/j.bpj.2014.11.1260.

Hollin, T., Chahine, Z. and Le Roch, K.G. (2023) 'Epigenetic Regulation and Chromatin Remodeling in Malaria Parasites', *Annual Review of Microbiology*, 77, pp. 255–276. Available at: https://doi.org/10.1146/annurev-micro-032521-041554.

Hollin, T. and Le Roch, K.G. (2020) 'From Genes to Transcripts, a Tightly Regulated Journey in *Plasmodium*', *Frontiers in Cellular and Infection Microbiology*, 10(December), pp. 1–13. Available at: https://doi.org/10.3389/fcimb.2020.618454.

Holmes, M.C. and Tjian, R. (2000) 'Promoter-Selective Properties of the TBP-Related Factor TRF1', 288(MAY), pp. 867–871.

Hoo, R. *et al.* (2016) 'Integrated analysis of the *Plasmodium* species transcriptome', *EBioMedicine*, 7, pp. 255–266. Available at: https://doi.org/10.1016/j.ebiom.2016.04.011.

Horikoshi, M. *et al.* (1989) 'Purification of a yeast TATA box-binding protein that exhibits human transcription factor IID activity', *Proceedings of the National Academy of Sciences of the United States of America*, 86(13), pp. 4843–4847. Available at: https://doi.org/10.1073/pnas.86.13.4843.

Horikoshi, M. *et al.* (1992) 'Transcription factor TFIID induces DNA bending upon binding to the TATA element', *Proceedings of the National Academy of Sciences of the United States of America*, 89(3), pp. 1060–1064. Available at: https://doi.org/10.1073/pnas.89.3.1060.

Horrocks, P. *et al.* (2009) 'Control of gene expression in *Plasmodium falciparum* - Ten years on', *Molecular and Biochemical Parasitology*, 164(1), pp. 9–25. Available at: https://doi.org/10.1016/j.molbiopara.2008.11.010.

Huang, Y.J., Acton, T.B. and Montelione, G.T. (2014) 'DisMeta: A meta server for construct design and optimization', *Methods in Molecular Biology*, 1091(9), pp. 3–16. Available at: https://doi.org/10.1007/978-1-62703-691-7_1.

Hunter, J.D. (2007) 'MATPLOTLIB: A 2D GRAPHICS ENVIRONMENT', *Computing in Science and Engineering*, 9(3), pp. 90–95.

Al Husini, N., Kudla, P. and Ansari, A. (2013) 'A Role for CF1A 3' End Processing Complex in Promoter-Associated Transcription', *PLoS Genetics*, 9(8). Available at: https://doi.org/10.1371/journal.pgen.1003722.

Iborra, F.J. *et al.* (1996) 'Active RNA polymerases are localized within discrete transcription "factories" in human nuclei', *Journal of Cell Science*, 109(6), pp. 1427–1436. Available at: https://doi.org/10.1242/jcs.109.6.1427.

Imbalanzo, A., Zaret, K. and Kingston, R. (1993) 'Transcription factor (TF) IIB and TFIIA Can Independently Increase the affinity of the TATA-binding protein for DNA', *The Journal of Biological Chemistry*, 269(11), pp. 8280–8286.

Inoue, H., Nojima, H. and Okayama, H. (1990) 'High efficiency transformation of *Escherichia coli* with plasmids', *Gene*, 96(1), pp. 23–28. Available at: https://doi.org/10.1016/0378-1119(90)90336-P.

Irgen-Gioro, S. *et al.* (2022) 'Fixation can change the appearance of phase separation in living cells', *eLife*, 11, pp. 1–24. Available at: https://doi.org/10.7554/eLife.79903.

Isogai, Y. *et al.* (2007) 'Transcription of histone gene cluster by differential core-promoter factors', *Genes and Development*, 21(22), pp. 2936–2949. Available at: https://doi.org/10.1101/gad.1608807.

Iyer, U.B. *et al.* (2022) 'Mediator Complex of the Malaria Parasite *Plasmodium falciparum* Associates with Evolutionarily Novel Subunits', *ACS Omega*, 7(17), pp. 14867–14874. Available at: https://doi.org/10.1021/acsomega.2c00368.

Jackson, D.A. *et al.* (1993) 'Visualization of focal sites of transcription within human nuclei', *EMBO Journal*, 12(3), pp. 1059–1065. Available at: https://doi.org/10.1002/j.1460-2075.1993.tb05747.x.

Jacobi, U.G. *et al.* (2007) 'TBP paralogs accommodate metazoan- and vertebrate-specific developmental gene regulation', *EMBO Journal*, 26(17), pp. 3900–3909. Available at: https://doi.org/10.1038/sj.emboj.7601822.

Jain, A. and Vale, R.D. (2017) 'RNA phase transitions in repeat expansion disorders', *Nature*, 546(7657), pp. 243–247. Available at: https://doi.org/10.1038/nature22386.

Jallow, Z. *et al.* (2004) 'Specialized and redundant roles of TBP and a vertebrate-specific TBP paralog in embryonic gene regulation in *Xenopus'*, 101(37), pp. 13525–13530.

Jefferson A. Vaughan, Bruce H. Noden, J.C.B. (1994) 'Sporogonic development of cultured *Plasmodium falciparum* in six species of laboratory-reared Anopheles Mosquitos', *The American Society of Tropical Medicine and Hygiene*, 51(2), pp. 233–243.

Jeninga, M.D., Quinn, J.E. and Petter, M. (2019) 'Apiap2 transcription factors in apicomplexan parasites', *Pathogens*, 8(2), pp. 1–24. Available at: https://doi.org/10.3390/pathogens8020047.

Jiang, L. *et al.* (2013) 'PfSETvs methylation of histone H3K36 represses virulence genes in *Plasmodium falciparum*', *Nature*, 499(7457), pp. 223–227. Available at: https://doi.org/10.1038/nature12361.

Jumper, J. *et al.* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596(7873), pp. 583–589. Available at: https://doi.org/10.1038/s41586-021-03819-2.

El Kaderi, B. *et al.* (2009) 'Gene looping is conferred by activator-dependent interaction of transcription initiation and termination machineries', *Journal of Biological Chemistry*, 284(37), pp.

25015–25025. Available at: https://doi.org/10.1074/jbc.M109.007948.

Kapli, P., Yang, Z. and Telford, M.J. (2020) 'Phylogenetic tree building in the genomic age', *Nature Reviews Genetics*, pp. 1–17. Available at: https://doi.org/10.1038/s41576-020-0233-0.

Kapuscinski, J. (1979) 'DAPI : a DNA-Specific Fluorescent Probe', *Biotechnic and histochemistry*, 70(5), pp. 220–233.

Kar, M. *et al.* (2022) 'Phase-separating RNA-binding proteins form heterogeneous distributions of clusters in subsaturated solutions', *Proceedings of the National Academy of Sciences of the United States of America*, 119(28), pp. 1–30. Available at: https://doi.org/10.1073/pnas.2202222119.

Karamanof, L. (2022) *Biomolecular condensation of intrinsically disordered regions in P. falciparum and H. sapiens transcription factors*. Masters Thesis. University of Cape Town.

Karlin, S. *et al.* (2002) 'Amino acid runs in eukaryotic proteomes and disease associations', *Proceedings of the National Academy of Sciences of the United States of America*, 99(1), pp. 333–338. Available at: https://doi.org/10.1073/pnas.012608599.

Kato, M. *et al.* (2012) 'Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels', *Cell*, 149(4), pp. 753–767. Available at: https://doi.org/10.1016/j.cell.2012.04.017.

Kaufmann, J. *et al.* (1998) 'CIF150, a Human Cofactor for Transcription Factor IID-Dependent Initiator Function', *Molecular and Cellular Biology*, 18(1), pp. 233–239. Available at: https://doi.org/10.1128/mcb.18.1.233.

Kedmi, A. *et al.* (2014) '*Drosophila* TRF2 is a preferential core promoter regulator', *Genes and Development*, 28(19), pp. 2163–2174. Available at: https://doi.org/10.1101/gad.245670.114.

Kensche, P.R. *et al.* (2015) 'The nucleosome landscape of *Plasmodium falciparum* reveals chromatin architecture and dynamics of regulatory sequences', *Nucleic Acids Research*, 44(5), pp. 2110–2124. Available at: https://doi.org/10.1093/nar/gkv1214.

Khrapunov, S. and Brenowitz, M. (2007) 'Influence of the N-terminal domain and divalent cations on self-association and DNA binding by the *Saccharomyces cerevisiae* TATA binding protein', *Biochemistry*, 46(16), pp. 4876–4887. Available at: https://doi.org/10.1021/bi061651w.

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), pp. 520–527. Available at: https://doi.org/10.1038/365520a0.

Kim, T.K. *et al.* (1994) 'Effects of activation-defective TBP mutations on transcription initiation in yeast', *Nature*, 369(6477), pp. 252–255. Available at: https://doi.org/10.1038/369252a0.

Kim, Y. *et al.* (1993) 'Crystal structure of a yeast TBP/TATA-box complex', *Nature*, 365(6446), pp. 512–520. Available at: https://doi.org/10.1038/365512a0.

Klein, I.A. *et al.* (2020) 'Partitioning of cancer therapeutics in nuclear condensates', *Science (New York, N.Y.)*, 368(6497), pp. 1386–1392. Available at: https://doi.org/10.1126/science.aaz4427.

Klejman, M.P. *et al.* (2005) 'Mutational analysis of BTAF1-TBP interaction: BTAF1 can rescue DNA-binding defective TBP mutants', *Nucleic Acids Research*, 33(17), pp. 5426–5436. Available at: https://doi.org/10.1093/nar/gki850.

Kostrewa, D. *et al.* (2009) 'RNA polymerase II-TFIIB structure and mechanism of transcription initiation', *Nature*, 462(7271), pp. 323–330. Available at: https://doi.org/10.1038/nature08548.

Koulouras, G. *et al.* (2018) 'EasyFRAP-web: A web-based tool for the analysis of fluorescence recovery after photobleaching data', *Nucleic Acids Research*, 46(W1), pp. W467–W472. Available at: https://doi.org/10.1093/nar/gky508.

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(12), p. 2962. Available at: https://doi.org/10.1093/nar/21.12.2962-a.

Kumar, S. *et al.* (2018) 'MEGA X: Molecular evolutionary genetics analysis across computing platforms', *Molecular Biology and Evolution*, 35(6), pp. 1547–1549. Available at: https://doi.org/10.1093/molbev/msy096.

Kurshakova, M.M. *et al.* (2019) 'TRF4, the novel TBP-related protein of *Drosophila melanogaster*, is concentrated at the endoplasmic reticulum and copurifies with proteins participating in the processes associated with endoplasmic reticulum', *Journal of Cellular Biochemistry*, 120(5), pp. 7927–7939. Available at: https://doi.org/10.1002/jcb.28070.

Kwon, I. *et al.* (2013) 'Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains', *Cell*, 155(5), p. 1049. Available at: https://doi.org/10.1016/j.cell.2013.10.033.

Lagrange, T. *et al.* (1996) 'High-resolution mapping of nucleoprotein complexes by site-specific protein-DNA photocrosslinking: Organization of the human TBP-TFIIA-TFIIB-DNA quaternary complex', *Proceedings of the National Academy of Sciences of the United States of America*, 93(20), pp. 10620–10625. Available at: https://doi.org/10.1073/pnas.93.20.10620.

Lagrange, T. *et al.* (1998) 'New core promoter element in RNA polymerase II-dependent transcription: Sequence-specific DNA binding by transcription factor IIB', *Genes and Development*, 12(1), pp. 34–44. Available at: https://doi.org/10.1101/gad.12.1.34.

Langdon, E.M. *et al.* (2018) 'mRNA structure determines specificity of a polyQ-driven phase separation', *Science*, 360(6391), pp. 922–927. Available at: https://doi.org/10.1126/science.aar7432.

Lanzer, M., De Bruin, D. and Ravetch, J. V. (1992a) 'A sequence element associated with the *Plasmodium falciparum* KAHRP gene is the site of developmentally regulated protein-DNA interactions', *Nucleic Acids Research*, 20(12), pp. 3051–3056. Available at: https://doi.org/10.1093/nar/20.12.3051.

Lanzer, M., de Bruin, D. and Ravetch, J.V. (1992b) 'Transcription mapping of a 100 kb locus of *Plasmodium falciparum* identifies an intergenic region in which transcription terminates and reinitiates.', *The EMBO Journal*, 11(5), pp. 1949–1955. Available at: https://doi.org/10.1002/j.1460-2075.1992.tb05248.x.

Larson, A.G. *et al.* (2017) 'Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin', *Nature*, 547(7662), pp. 236–240. Available at: https://doi.org/10.1038/nature22822.

Laskowski, R.A. *et al.* (1993) 'PROCHECK: a program to check the stereochemical quality of protein structures', *Journal of Applied Crystallography*, 26(2), pp. 283–291. Available at: https://doi.org/10.1107/s0021889892009944.

Lasonder, E. *et al.* (2016) 'Integrated transcriptomic and proteomic analyses of P. Falciparum gametocytes: Molecular insight into sex-specific processes and translational repression', *Nucleic Acids Research*, 44(13), pp. 6087–6101. Available at: https://doi.org/10.1093/nar/gkw536.

Le, S.Q. and Gascuel, O. (2008) 'An improved general amino acid replacement matrix', *Molecular Biology and Evolution*, 25(7), pp. 1307–1320. Available at: https://doi.org/10.1093/molbev/msn067.

Lee, M. and Struhl, K. (2001) 'Multiple functions of the nonconserved N-terminal domain of yeast TATA-binding protein', *Genetics*, 158(1), pp. 87–93. Available at: https://doi.org/10.1093/genetics/158.1.87.

Lee, S. and Hahn, S. (1995) 'Model for binding of transcription factor TFIIB to the TBP-DNA complex', *Nature*, pp. 609–612. Available at: https://doi.org/10.1038/376609a0.

Lee, T.I. and Young, R.A. (2013) 'Transcriptional regulation and its misregulation in disease', *Cell*, 152(6), pp. 1237–1251. Available at: https://doi.org/10.1016/j.cell.2013.02.014.

Li, S., Wang, Y. and Lai, L. (2023) 'Small molecules in regulating protein phase separation', *Acta Biochimica et Biophysica Sinica*, 55(7), pp. 1075–1083. Available at: https://doi.org/10.3724/abbs.2023106.

Liang, H. *et al.* (2020) 'TATA box-binding protein-related factor 3 drives the mesendoderm specification of human embryonic stem cells by globally interacting with the TATA box of key mesendodermal genes', *Stem Cell Research and Therapy*, 11(1), pp. 1–16. Available at: https://doi.org/10.1186/s13287-020-01711-w.

Lieberman, P.M. *et al.* (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. Available at: https://doi.org/10.1128/mcb.11.1.63-74.1991.

Lieberman, P.M. and Berk, A.J. (1994) 'A mechanism for TAFs in transcriptional activation: Activation domain enhancement of TFIID-TFIIA-promoter DNA complex formation', *Genes and Development*, 8(9), pp. 995–1006. Available at: https://doi.org/10.1101/gad.8.9.995.

Lim, C.Y. *et al.* (2004) 'The MTE, a new core promoter element for transcription by RNA poymerase II', *Genes and Development*, 18(13), pp. 1606–1617. Available at: https://doi.org/10.1101/gad.1193404.

Lin, A.Z. *et al.* (2023) 'Dynamical control enables the formation of demixed biomolecular condensates', *bioRxiv*, p. 2023.01.04.522702. Available at: https://www.biorxiv.org/content/10.1101/2023.01.04.522702v1%0Ahttps://www.biorxiv.org/content/10.1101/2023.01.04.522702v1.abstract.

Lin, Y. *et al.* (1997) 'Hepatitis B virus X protein is a transcriptional modulator that communicates with transcription factor IIB and the RNA polymerase II subunit 5', *Journal of Biological Chemistry*, 272(11), pp. 7132–7139. Available at: https://doi.org/10.1074/jbc.272.11.7132.

Lin, Y. *et al.* (2015) 'Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins', *Molecular Cell*, 60(2), pp. 208–219. Available at: https://doi.org/10.1016/j.molcel.2015.08.018.

Linding, R., Russell, R.B., *et al.* (2003) 'GlobPlot: Exploring protein sequences for globularity and disorder', *Nucleic Acids Research*, 31(13), pp. 3701–3708. Available at: https://doi.org/10.1093/nar/gkg519.

Linding, R., Jensen, L.J., *et al.* (2003) 'Protein disorder prediction: Implications for structural proteomics', *Structure*, 11(11), pp. 1453–1459. Available at:

https://doi.org/10.1016/j.str.2003.10.002.

Lindner, S.E. *et al.* (2013) 'Total and putative surface proteomics of malaria parasite salivary gland sporozoites', *Molecular and Cellular Proteomics*, 12(5), pp. 1127–1143. Available at: https://doi.org/10.1074/mcp.M112.024505.

Liu, W. *et al.* (2015) 'IBS: An illustrator for the presentation and visualization of biological sequences', *Bioinformatics*, 31(20), pp. 3359–3361. Available at: https://doi.org/10.1093/bioinformatics/btv362.

López-Barragán, M.J. *et al.* (2011) 'Directional gene expression and antisense transcripts in sexual and asexual stages of Plasmodium falciparum', *BMC Genomics*, 12. Available at: https://doi.org/10.1186/1471-2164-12-587.

Lopez-Rubio, J.J. *et al.* (2007) '5' Flanking Region of Var Genes Nucleate Histone Modification Patterns Linked To Phenotypic Inheritance of Virulence Traits in Malaria Parasites', *Molecular Microbiology*, 66(6), pp. 1296–1305. Available at: https://doi.org/10.1111/j.1365-2958.2007.06009.x.

Loschi, M. *et al.* (2009) 'Dynein and kinesin regulate stress-granule and P-body dynamics', *Journal of Cell Science*, 122(21), pp. 3973–3982. Available at: https://doi.org/10.1242/jcs.051383.

Louder, R.K. *et al.* (2016) 'Structure of promoter-bound TFIID and model of human pre-initiation complex assembly', *Nature*, 531(7596), pp. 604–609. Available at: https://doi.org/10.1038/nature17394.

Luthy, R., Bowie, J. and Eisenberg, D. (1992) 'Assessment of protein models with three-dimensional profiles', *Nature*, 356, pp. 83–85.

Lyons, H. *et al.* (2023) 'Functional partitioning of transcriptional regulators by patterned charge blocks.', *Cell*, pp. 327–345. Available at: https://doi.org/10.1016/j.cell.2022.12.013.

Magnusdottir, A. *et al.* (2009) 'Enabling IMAC purification of low abundance recombinant proteins from *E. coli* lysates', *Nature Methods*, 6(7), pp. 477–478. Available at: https://doi.org/10.1038/nmeth0709-477.

Maharana, S. *et al.* (2018) 'RNA buffers the phase separation behavior of prion-like RNA binding proteins', *Science*, 360(6391), pp. 918–921. Available at: https://doi.org/10.1126/science.aar7366.

Malecová, B. *et al.* (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. Available at: https://doi.org/10.1074/jbc.M702776200.

Malik, S., Lee, D.K. and Roeder, R.G. (1993) 'Potential RNA Polymerase II-Induced Interactions of Transcription Factor TFIIB', *Molecular and Cellular Biology*, 13(10), pp. 6253–6259. Available at: https://doi.org/10.1128/mcb.13.10.6253-6259.1993.

Martianov, I. *et al.* (2016) 'TRF2 is recruited to the pre- initiation complex as a testis- specific subunit of TFIIA / ALF to promote haploid cell gene expression', *Nature Publishing Group*, (July), pp. 1–13. Available at: https://doi.org/10.1038/srep32069.

Martinez, E. *et al.* (1994) 'TATA-binding protein-associated factor(s) in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter', *EMBO Journal*, 13(13), pp. 3115–3126. Available at: https://doi.org/10.1002/j.1460-2075.1994.tb06610.x.

Martinez, E. *et al.* (1995) 'Core promoter-specific function of a mutant transcription factor TFIID defective in TATA-box binding', *Proceedings of the National Academy of Sciences of the United States of America*, 92(25), pp. 11864–11868. Available at: https://doi.org/10.1073/pnas.92.25.11864.

Martinez, E. *et al.* (1998) ' Novel Cofactors and TFIIA Mediate Functional Core Promoter Selectivity by the Human TAF II 150-Containing TFIID Complex ', *Molecular and Cellular Biology*, 18(11), pp. 6571–6583. Available at: https://doi.org/10.1128/mcb.18.11.6571.

McAndrew, M.B. *et al.* (1993) 'Characterisation of the gene encoding an unusually divergent TATA-Binding Protein (TBP) from the extremely A+T-rich human malaria parasite *Plasmodium falciparum*', *Gene*, 124(2), pp. 165–171. Available at: https://doi.org/10.1016/0378-1119(93)90390-O.

McGinnis, S. and Madden, T.L. (2004) 'BLAST: At the core of a powerful and diverse set of sequence analysis tools', *Nucleic Acids Research*, 32(WEB SERVER ISS.), pp. 20–25. Available at: https://doi.org/10.1093/nar/gkh435.

McSwiggen, David T. *et al.* (2019) 'Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences', *Genes & development*, 33(23–24), pp. 1619–1634. Available at: https://doi.org/10.1101/gad.331520.119.

McSwiggen, David Trombley *et al.* (2019) 'Evidence for DNA-mediated nuclear compartmentalization distinct from phase separation', *eLife*, 8, pp. 1–31. Available at: https://doi.org/10.7554/eLife.47098.

Medler, S. *et al.* (2011) 'Evidence for a complex of transcription factor IIB (TFIIB) with poly(A) polymerase and cleavage factor 1 subunits required for gene looping', *Journal of Biological Chemistry*, 286(39), pp. 33709–33718. Available at: https://doi.org/10.1074/jbc.M110.193870.

Menard, R. *et al.* (2013) 'Looking under the skin: the first steps in malarial infection and immunity.pdf', *Nature Reviews Microbiology*, 11, pp. 701–712.

Metcalf, C.E. and Wassarman, D.A. (2006) 'DNA binding properties of TAF1 isoforms with two AT-hooks', *Journal of Biological Chemistry*, 281(40), pp. 30015–30023. Available at: https://doi.org/10.1074/jbc.M606289200.

Michowicz, J. MSc Thesis in preparation. University of Cape Town.

Miranda, K. *et al.* (2010) 'Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells', *Nature Genetics*, 42(1), pp. 53–56. Available at: https://doi.org/10.1038/ng.496.Preferential.

Milton, R. (2017) *Regulation of transcription in Plasmodium falciparum, the causative agent of severe malaria: initial characterisation of PfTBP and PfTFIIA.* Doctoral Thesis. University of Cape Town.

Mitchell, J.A. and Fraser, P. (2008) 'Transcription factories are nuclear subcompartments that remain in the absence of transcription', *Genes and Development*, 22(1), pp. 20–25. Available at: https://doi.org/10.1101/gad.454008.

Mitrea, D.M. *et al.* (2022) 'Modulating biomolecular condensates: a novel approach to drug discovery', *Nature Reviews Drug Discovery*. Available at: https://doi.org/10.1038/s41573-022-00505-4.

Mittag, T. and Pappu, R. V. (2022) 'A conceptual framework for understanding phase separation and addressing open questions and challenges', *Molecular Cell*, 82(12), pp. 2201–2214. Available at: https://doi.org/10.1016/j.molcel.2022.05.018.

Moll, J.R. (2002) 'Magnesium is required for specific DNA binding of the CREB B-ZIP domain', *Nucleic Acids Research*, 30(5), pp. 1240–1246. Available at: https://doi.org/10.1093/nar/30.5.1240.

Moll, K. *et al.* (eds) (2013) *Methods in Malaria Research*. sixth edit. EVIMalaR.

Molliex, A. *et al.* (2015) 'Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization', *Cell*, 163(1), pp. 123–133. Available at: https://doi.org/10.1016/j.cell.2015.09.015.

Moore, P.A. *et al.* (1999) 'A Human TATA Binding Protein-Related Protein with Altered DNA Binding Specificity Inhibits Transcription from Multiple Promoters and Activators', *Molecular and Cellular Biology*, 19(11), pp. 7610–7620. Available at: https://doi.org/10.1128/mcb.19.11.7610.

Moorthy, V. and Binka, F. (2021) 'R21/Matrix-M: a second malaria vaccine?', *The Lancet*, 397(10287), pp. 1782–1783. Available at: https://doi.org/10.1016/S0140-6736(21)01065-5.

Moraes, C.B. *et al.* (2013) 'Transcription Sites Are Developmentally Regulated during the Asexual Cycle of *Plasmodium falciparum*', *PLoS ONE*, 8(2), pp. 1–10. Available at: https://doi.org/10.1371/journal.pone.0055539.

Mota, M.M. *et al.* (2001) 'Migration of Plasmodium Sporozoites Through Cells Before Infection', 291(January), pp. 141–145.

Muzzopappa, F., Hertzog, M. and Erdel, F. (2021) 'DNA length tunes the fluidity of DNA-based condensates', *Biophysical Journal*, 120(7), pp. 1288–1300. Available at: https://doi.org/10.1016/j.bpj.2021.02.027.

Nakamura, K. (2001) 'SCA17, a novel polyglutamine disease caused by the expansion of polyglutamine tracts in TATA-binding protein', *Clinical Neurology*, 41(12), pp. 1123–1125.

ngoc, L.V. *et al.* (2017) 'The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters', *Genes and Development*, 31(1), pp. 6–11. Available at: https://doi.org/10.1101/gad.293837.116.

ngoc, L.V. *et al.* (2019) 'The punctilious RNA polymerase II core promoter', *Genetics*, 212(1), pp. 13–24. Available at: https://doi.org/10.1534/genetics.119.302021.

Nikolov, D.B. *et al.* (1995) 'Crystal structure of the TFIIB-TBP-TATA-element ternary complex', *Nature*, 377(6545), pp. 119–128. Available at: http://www.nature.com/doifinder/10.1038/377119a0%0Apapers3://publication/doi/10.1038/377119a0.

Nikolov, D.B. *et al.* (1996) 'Crystal structure of a human TATA box-binding protein/TATA element complex', *Proceedings of the National Academy of Sciences of the United States of America*, 93(10), pp. 4862–4867. Available at: https://doi.org/10.1073/pnas.93.10.4862.

Nitika, N., Nema, S. and Bharti, P. (2023) 'R21 vaccine: A ray of hope for malaria elimination', *Asian Pacific Journal of Tropical Medicine*, 16(6), pp. 243–244. Available at: https://doi.org/10.4103/1995-7645.378560.

Nott, T.J. *et al.* (2015) 'Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles', *Molecular Cell*, 57(5), pp. 936–947. Available at: https://doi.org/10.1016/j.molcel.2015.01.013.

Nott, T.J., Craggs, T.D. and Baldwin, A.J. (2016) 'Membraneless organelles can melt nucleic acid duplexes and act as biomolecular filters', *Nature Chemistry*, 8(6), pp. 569–575. Available at: https://doi.org/10.1038/nchem.2519.

O'Brien, M.J. and Ansari, A. (2022) 'Beyond the canonical role of TFIIB in eukaryotic transcription', *Current Genetics*, 68(1), pp. 61–67. Available at: https://doi.org/10.1007/s00294-021-01223-x.

O'Brien, M.J. and Ansari, A. (2024) 'Protein interaction network revealed by quantitative proteomic analysis links TFIIB to multiple aspects of the transcription cycle', *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1872(1), p. 140968. Available at: https://doi.org/10.1016/j.bbapap.2023.140968.

Oakley, M.S. *et al.* (2011) 'Clinical and molecular aspects of malaria fever', *Trends in Parasitology*, 27(10), pp. 442–449. Available at: https://doi.org/10.1016/j.pt.2011.06.004.

Obradovic, Z. *et al.* (2005) 'Exploiting heterogeneous sequence properties improves prediction of protein disorder', *Proteins: Structure, Function and Genetics*, 61(SUPPL. 7), pp. 176–182. Available at: https://doi.org/10.1002/prot.20735.

Oelgeschlager, T., Chiang, C.M. and Roeder, R.G. (1996) 'Topology and reorganization of a human TFIID-promoter complex', *Nature*, pp. 735–738. Available at: https://doi.org/10.1038/382735a0.

Oguike, M.C. *et al.* (2011) '*Plasmodium ovale curtisi* and *Plasmodium ovale wallikeri* circulate simultaneously in African communities', *International Journal for Parasitology*, 41(6), pp. 677–683. Available at: https://doi.org/10.1016/j.ijpara.2011.01.004.

Ohbayashi, T., Makino, Y. and Tamura, T.A. (1999) 'Identification of a mouse TBP-like protein (TLP) distantly related to the *Drosophila* TBP-related factor', *Nucleic Acids Research*, 27(3), pp. 750–755. Available at: https://doi.org/10.1093/nar/27.3.750.

Osborne, C.S. *et al.* (2004) 'Active genes dynamically colocalize to shared sites of ongoing transcription', *Nature Genetics*, 36(10), pp. 1065–1071. Available at: https://doi.org/10.1038/ng1423.

Otto, T.D. *et al.* (2010) 'New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq', *Molecular Microbiology*, 76(1), pp. 12–24. Available at: https://doi.org/10.1111/j.1365-2958.2009.07026.x.

Painter, H.J. *et al.* (2018) 'Genome-wide real-time in vivo transcriptional dynamics during *Plasmodium falciparum* blood-stage development', *Nature Communications*, 9(1), pp. 1–12. Available at: https://doi.org/10.1038/s41467-018-04966-3.

Pal, M., Ponticelli, A.S. and Luse, D.S. (2005) 'The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II', *Molecular Cell*, 19(1), pp. 101–110. Available at: https://doi.org/10.1016/j.molcel.2005.05.024.

Palacio, M. and Taatjes, D.J. (2022) 'Merging Established Mechanisms with New Insights : Condensates , Hubs , and the Regulation of RNA Polymerase II Transcription', *Journal of Molecular Biology*, 434(1), p. 167216. Available at: https://doi.org/10.1016/j.jmb.2021.167216.

Pan, S., Czarnecka-Verner, E. and Gurley, W.B. (2000) 'Role of the TATA binding protein-transcription factor IIB interaction in supporting basal and activated transcription in plant cells', *Plant Cell*, 12(1), pp. 125–135. Available at: https://doi.org/10.2307/3871034.

Pardee, T.S., Bangur, C.S. and Ponticelli, A.S. (1998) 'The N-terminal region of yeast TFIIB contains two adjacent functional domains involved in stable RNA polymerase II binding and transcription start site selection', *Journal of Biological Chemistry*, 273(28), pp. 17859–17864. Available at: https://doi.org/10.1074/jbc.273.28.17859.

Parra-Marín, O. *et al.* (2019) 'An in vitro characterisation of the *Trichomonas vaginalis* TATA box-binding proteins (TBPs)', *Parasitology Research*, 118(10), pp. 3019–3031. Available at: https://doi.org/10.1007/s00436-019-06438-z.

Patel, A. *et al.* (2015) 'A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation', *Cell*, 162(5), pp. 1066–1077. Available at: https://doi.org/10.1016/j.cell.2015.07.047.

Patel, A. *et al.* (2017) 'Biochemistry: ATP as a biological hydrotrope', *Science*, 356(6339), pp. 753–756. Available at: https://doi.org/10.1126/science.aaf6846.

Patikoglou, G.A. *et al.* (1999) 'TATA element recognition by the TATA box-binding protein has been conserved throughout evolution', *Genes and Development*, 13(24), pp. 3217–3230. Available at: https://doi.org/10.1101/gad.13.24.3217.

Pease, B.N. *et al.* (2013) 'Global analysis of protein expression and phosphorylation of three stages of *Plasmodium falciparum* intraerythrocytic development', *Journal of Proteome Research*, 12(9), pp. 4028–4045. Available at: https://doi.org/10.1021/pr400394g.

Pegram, L. (2022). *Transcriptional condensates: a comparison of the phase separation properties of H. sapiens and P. falciparum TATA-binding proteins*. Honours Thesis. University of Cape Town.

Peng, P. *et al.* (2006) 'Length-dependent prediction of protein intrinsic disorder', *BMC Bioinformatics*, 208(7), pp. 208–212. Available at: https://doi.org/10.1186/1471-2105-7-208.

Perletti, L., Dantonel, J.C. and Davidson, I. (1999) 'The TATA-binding protein and its associated factors are differentially expressed in adult mouse tissues', *Journal of Biological Chemistry*, 274(22), pp. 15301–15304. Available at: https://doi.org/10.1074/jbc.274.22.15301.

Petter, M. *et al.* (2013) 'H2A.Z and H2B.Z double-variant nucleosomes define intergenic regions and dynamically occupy var gene promoters in the malaria parasite *Plasmodium falciparum*', *Molecular Microbiology*, 87(6), pp. 1167–1182. Available at: https://doi.org/10.1111/mmi.12154.

Pettersen, E.F. *et al.* (2004) 'UCSF Chimera - A visualization system for exploratory research and analysis', *Journal of Computational Chemistry*, 25(13), pp. 1605–1612. Available at: https://doi.org/10.1002/jcc.20084.

Phadtare, S., Alsina, J. and Inouye, M. (1999) 'Cold-shock response and cold-shock proteins', *Current Opinion in Microbiology*, 2(2), pp. 175–180. Available at: https://doi.org/10.1016/S1369-5274(99)80031-9.

Phair, R.D., Gorski, S.A. and Misteli, T. (2003) 'Measurement of Dynamic Protein Binding to Chromatin In Vivo, Using Photobleaching Microscopy', *Methods in Enzymology*, 375(2000), pp. 393–414. Available at: https://doi.org/10.1016/S0076-6879(03)75025-3.

Pinto, I. *et al.* (1994) 'Characterization of sua7 mutations defines a domain of TFIIB involved in transcription start site selection in yeast', *Journal of Biological Chemistry*, 269(48), pp. 30569–30573. Available at: https://doi.org/10.1016/s0021-9258(18)43851-3.

Piovesan, D. *et al.* (2021) 'MobiDB: Intrinsically disordered proteins in 2021', *Nucleic Acids Research*, 49(D1), pp. D361–D367. Available at: https://doi.org/10.1093/nar/gkaa1058.

Ponts, N. *et al.* (2010) 'Nucleosome landscape and control of transcription in the human malaria parasite', *Genome Research*, 20(2), pp. 228–238. Available at: https://doi.org/10.1101/gr.101063.109.

Ponts, N. *et al.* (2011) 'Nucleosome occupancy at transcription start sites in the human malaria parasite: A hard-wired evolution of virulence?', *Infection, Genetics and Evolution*, 11(4), pp. 716–724. Available at: https://doi.org/10.1016/j.meegid.2010.08.002.

Press, A. (1992) 'Population Dynamics of *Plasmodium falciparum* Sporogony in Laboratory-Infected Anopheles gambiae Author (s): Jefferson A . Vaughan , Bruce H . Noden and John C . Beier Published by : Allen Press on behalf of The American Society of Parasitologists Stable', *American Society of Parasitologists*, 78(4), pp. 716–724.

Prilusky, J. *et al.* (2005) 'FoldIndex©: A simple tool to predict whether a given protein sequence is intrinsically unfolded', *Bioinformatics*, 21(16), pp. 3435–3438. Available at: https://doi.org/10.1093/bioinformatics/bti537.

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. Available at: https://doi.org/10.1038/nrmicro1529.

Putnam, A. *et al.* (2019) 'A gel phase promotes condensation of liquid P granules in *Caenorhabditis elegans* embryos', 26(March).

Quadt, K.A. *et al.* (2020) '*Plasmodium falciparum* parasites exit the infected erythrocyte after haemolysis with saponin and streptolysin O', *Parasitology Research*, 119(12), pp. 4297–4302. Available at: https://doi.org/10.1007/s00436-020-06932-9.

Rabenstein, M.D. *et al.* (1999) 'TATA box-binding protein (TBP)-related factor 2 (TRF2), a third member of the TBP family', *Proceedings of the National Academy of Sciences of the United States of America*, 96(9), pp. 4791–4796. Available at: https://doi.org/10.1073/pnas.96.9.4791.

Rai, R. *et al.* (2014) 'Genome-wide analysis in *Plasmodium falciparum* reveals early and late phases of RNA polymerase II occupancy during the infectious cycle', *BMC Genomics*, 15(1). Available at: https://doi.org/10.1186/1471-2164-15-959.

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', *Proceedings of the National Academy of Sciences of the United States of America*, 102(15), pp. 5414–5419. Available at: https://doi.org/10.1073/pnas.0408883102.

Reddy, P. and Hahn, S. (1991) 'Dominant negative mutations in yeast TFIID define a bipartite DNA-binding region', *Cell*, 65(2), pp. 349–357. Available at: https://doi.org/10.1016/0092-8674(91)90168-X.

Rhine, K., Vidaurre, V. and Myong, S. (2020) 'RNA Droplets', *Annual Review of Biophysics*, 49(1), pp. 247–265. Available at: https://doi.org/10.1146/annurev-biophys-052118-115508.

Roberts, S.G.E. and Green, M.R. (1994) 'Activator-induced conformational change in general

transcription factor TFIIB', *Nature*, pp. 717–720. Available at: https://doi.org/10.1038/371717a0.

Le Roch, K.G. *et al.* (2003) 'Discovery of gene function by expression profiling of the malaria parasite life cycle', *Science*, 301(5639), pp. 1503–1508. Available at: https://doi.org/10.1126/science.1087025.

Le Roch, K.G. *et al.* (2004) 'Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle', *Genome Research*, 14(11), pp. 2308–2318. Available at: https://doi.org/10.1101/gr.2523904.

Roden, C. and Gladfelter, A.S. (2017) 'RNA contributions to the form and function of biomolecular condensates', *Physiology & behavior*, 176(3), pp. 139–148. Available at: https://doi.org/10.1159/000444169.Carotid.

Rojas, D.A. *et al.* (2022) 'Initiator-Directed Transcription: Fission Yeast Nmtl Initiator Directs Preinitiation Complex Formation and Transcriptional Initiation', *Genes*, 13(2). Available at: https://doi.org/10.3390/genes13020256.

Rosano, G.L. and Ceccarelli, E.A. (2014) 'Recombinant protein expression in *Escherichia coli*: Advances and challenges', *Frontiers in Microbiology*, 5(APR), pp. 1–17. Available at: https://doi.org/10.3389/fmicb.2014.00172.

Rosenberg, R., Wirtz, R. A., Schneider, I. & Burge, R. (1990) 'An estimation of the number of malaria sporozoites ejected by a feeding mosquito.', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, (84), pp. 209–212.

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. Available at: https://doi.org/10.1016/j.tibs.2015.01.007.

Ruan, J. -p. *et al.* (2004) 'Functional Characterization of a *Trypanosoma brucei* TATA-Binding Protein-Related Factor Points to a Universal Regulator of Transcription in Trypanosomes', *Molecular and Cellular Biology*, 24(21), pp. 9610–9618. Available at: https://doi.org/10.1128/mcb.24.21.9610-9618.2004.

Ruiz, J.L. *et al.* (2018) 'Characterization of the accessible genome in the human malaria parasite *Plasmodium falciparum*', *Nucleic Acids Research*, 46(18), pp. 9414–9431. Available at: https://doi.org/10.1093/nar/gky643.

Russel, A. MSc Thesis in preparation. University of Cape Town.

Ruvalcaba-Salazar, O.K. *et al.* (2005) 'Recombinant and native *Plasmodium falciparum* TATA-binding-

protein binds to a specific TATA box element in promoter regions', *Molecular and Biochemical Parasitology*, 140(2), pp. 183–196. Available at: https://doi.org/10.1016/j.molbiopara.2005.01.002.

Ruvalcaba-Salazar, O.K. *et al.* (2006) 'Preparation and characterization of monoclonal antibody specific to PfCP-2.9 chimeric protein of *Plasmodium falciparum*', *HYBRIDOMA*, 25(6), pp. 367–371.

Sabari, B.R. *et al.* (2018) 'Coactivator condensation at super-enhancers links phase separation and gene control', *Science*, 361(6400), p. eaar3958. Available at: https://doi.org/10.1126/science.aar3958.

Sabari, B.R., Dall'Agnese, A. and Young, R.A. (2020) 'Biomolecular Condensates in the Nucleus', *Trends in Biochemical Sciences*, 45(11), pp. 961–977. Available at: https://doi.org/10.1016/j.tibs.2020.06.007.

Sainsbury, S., Bernecky, C. and Cramer, P. (2015) 'Structural basis of transcription initiation by RNA polymerase II', *Nature Reviews Molecular Cell Biology*, 16(3), pp. 129–143. Available at: https://doi.org/10.1038/nrm3952.

Salcedo-Amaya, A.M. *et al.* (2009) 'Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of *Plasmodium falciparum*', *Proceedings of the National Academy of Sciences of the United States of America*, 106(24), pp. 9655–9660. Available at: https://doi.org/10.1073/pnas.0902515106.

Sanchez-Burgos, I. *et al.* (2023) 'Surfactants or scaffolds? RNAs of varying lengths control the thermodynamic stability of condensates differently', *Biophysical Journal*, 122(14), pp. 2973–2987. Available at: https://doi.org/10.1016/j.bpj.2023.03.006.

Santana, J.F. *et al.* (2022) 'Differential dependencies of human RNA polymerase II promoters on TBP, TAF1, TFIIB and XPB', *Nucleic Acids Research*, 50(16), pp. 9127–9148. Available at: https://doi.org/10.1093/nar/gkac678.

Saraf, A. *et al.* (2016) 'Dynamic and combinatorial landscape of histone modifications during the intraerythrocytic developmental cycle of the malaria parasite', *Journal of Proteome Research*, 15(8), pp. 2787–2801. Available at: https://doi.org/10.1021/acs.jproteome.6b00366.

Sato, S. (2021) '*Plasmodium*—a brief introduction to the parasites causing human malaria and their basic biology (Journal of Physiological Anthropology, (2021), 40, 1, (1), 10.1186/s40101-020-00251-9)', *Journal of Physiological Anthropology*, 40(1), pp. 1–13. Available at: https://doi.org/10.1186/s40101-021-00254-0.

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. Available at: https://doi.org/10.1038/nrm1981.

Sawadogo, M. and Roeder, R.G. (1985) 'Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region', *Cell*, 43(1), pp. 165–175. Available at: https://doi.org/10.1016/0092-8674(85)90021-2.

Saxena, A. *et al.* (2005) 'Structure-Function Analysis of the Human TFIIB-Related Factor II Protein Reveals an Essential Role for the C-Terminal Domain in RNA Polymerase III Transcription', *Molecular and Cellular Biology*, 25(21), pp. 9406–9418. Available at: https://doi.org/10.1128/mcb.25.21.9406-9418.2005.

Schmidt, E.E. *et al.* (2003) 'Fundamental cellular processes do not require vertebrate-specific sequences within the TATA-binding protein', *Journal of Biological Chemistry*, 278(8), pp. 6168–6174. Available at: https://doi.org/10.1074/jbc.M211205200.

Schneider, C., Rasband, W. and Eliceiri, K. (2010) 'NIH Image to ImageJ: 25 years of Image Analysis Caroline', *Nature Methods*, 9(7), pp. 185–188. Available at: https://doi.org/10.1007/978-1-84882-087-6_9.

Segal, E. and Widom, J. (2009) 'Poly(dA:dT) tracts: major determinants of nucleosome organization', *Current Opinion in Structural Biology*, 19(1), pp. 65–71. Available at: https://doi.org/10.1016/j.sbi.2009.01.004.

Shao, W. *et al.* (2022) 'Phase separation of RNA-binding protein promotes polymerase binding and transcription', *Nature Chemical Biology*, 18(1), pp. 70–80. Available at: https://doi.org/10.1038/s41589-021-00904-5.

Sharp, P.A. *et al.* (2022) 'RNA in formation and regulation of transcriptional condensates', *Rna*, 28(1), pp. 52–57. Available at: https://doi.org/10.1261/rna.078997.121.

Shemiakina, I.I. *et al.* (2012) 'A monomeric red fluorescent protein with low cytotoxicity', *Nature Communications*, 3, pp. 1–7. Available at: https://doi.org/10.1038/ncomms2208.

Shin, Y. (2022) 'Rich Phase Separation Behavior of Biomolecules', *Molecules and Cells*, 45(1), pp. 6–15. Available at: https://doi.org/10.14348/molcells.2021.0204.

Shin, Y. and Brangwynne, C.P. (2017) 'Liquid phase condensation in cell physiology and disease', *Science*, 357(6357). Available at: https://doi.org/10.1126/science.aaf4382.

Silberhorn, E. *et al.* (2016) '*Plasmodium falciparum* Nucleosomes Exhibit Reduced Stability and Lost Sequence Dependent Nucleosome Positioning', *PLoS Pathogens*, 12(12), pp. 1–29. Available at:

https://doi.org/10.1371/journal.ppat.1006080.

Silvestrini, F. *et al.* (2010) 'Protein export marks the early phase of gametocytogenesis of the human malaria parasite *Plasmodium falciparum*', *Molecular and Cellular Proteomics*, 9(7), pp. 1437–1448. Available at: https://doi.org/10.1074/mcp.M900479-MCP200.

Singh, B.N. and Hampsey, M. (2007) 'A Transcription-Independent Role for TFIIB in Gene Looping', *Molecular Cell*, 27(5), pp. 806–816. Available at: https://doi.org/10.1016/j.molcel.2007.07.013.

Smale, S.T. and Baltimore, D. (1989) 'The "initiator" as a transcription control element', *Cell*, 57(1), pp. 103–113. Available at: https://doi.org/10.1016/0092-8674(89)90176-1.

Smale, S.T. and Kadonaga, J.T. (2003) 'The RNA polymerase II core promoter', *Annual Review of Biochemistry*, 72, pp. 449–479. Available at: https://doi.org/10.1146/annurev.biochem.72.121801.161520.

Smith, J. *et al.* (2016) 'Spatial patterning of P granules by RNA-induced phase separation of the intrinsically-disordered protein MEG-3', *eLife*, 5(DECEMBER2016), pp. 1–18. Available at: https://doi.org/10.7554/eLife.21337.

Smith, J.D. *et al.* (2001) 'Decoding the language of var genes and *Plasmodium falciparum* sequestration', *Trends in Parasitology*, 17(11), pp. 538–545. Available at: https://doi.org/10.1016/S1471-4922(01)02079-7.

Starr, D.B. and Hawley, D.K. (1991) 'TFIID binds in the minor groove of the TATA box', *Cell*, 67(67), pp. 1231–1240.

Starr, D.B., Hoopes, B.C. and Hawley, D.K. (1995) 'DNA bending is an important component of site-specific recognition by the TATA binding protein', *Journal of Molecular Biology*, 250(4), pp. 434–446. Available at: https://doi.org/10.1006/jmbi.1995.0388.

Su, X.Z. *et al.* (1996) 'Reduced extension temperatures required for PCR amplification of extremely A+T-rich DNA', *Nucleic Acids Research*, 24(8), pp. 1574–1575. Available at: https://doi.org/10.1093/nar/24.8.1574.

Takada, S. *et al.* (2000) 'A TRF1 : BRF Complex Directs Drosophila RNA Polymerase III Transcription', 101, pp. 459–469.

Talman, A.M. *et al.* (2004) 'Gametocytogenesis: The puberty of *Plasmodium falciparum*', *Malaria Journal*, 3, pp. 1–14. Available at: https://doi.org/10.1186/1475-2875-3-24.

Talvik, G. (2016) *Transcription regulation in Plasmodium falciparum: Functional characterisation of general transcription factor IIB.* Masters Thesis. University of Cape Town.

Tan-Wong, S.M. *et al.* (2012) 'Gene Loops enhance Transcriptional Directionality', *Science*, (November), pp. 671–676.

Tang, H. *et al.* (1996) 'Protein-protein interactions in eukaryotic transcription initiation: Structure of the preinitiation complex', *Proceedings of the National Academy of Sciences of the United States of America*, 93(3), pp. 1119–1124. Available at: https://doi.org/10.1073/pnas.93.3.1119.

Thanh, N.V. *et al.* (2017) 'Rapid decline in the susceptibility of *Plasmodium falciparum* to dihydroartemisinin-piperaquine in the south of Vietnam', *Malaria Journal*, 16(1), pp. 1–10. Available at: https://doi.org/10.1186/s12936-017-1680-8.

The Plasmodium Genome Consortium (2001) 'PlasmoDB: An integrative database of the *Plasmodium falciparum* genome. Tools for accessing and analyzing finished and unfinished sequence data', *Nucleic Acids Research*, 29(1), pp. 66–69. Available at: https://doi.org/10.1093/nar/29.1.66.

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. Available at: https://doi.org/10.1080/10409230600648736.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) 'CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice', *Nucleic Acids Research*, 22(22), pp. 4673–4680. Available at: https://doi.org/10.1093/nar/22.22.4673.

Tillo, D. and Hughes, T.R. (2009) 'G+C content dominates intrinsic nucleosome occupancy', *BMC Bioinformatics*, 10, pp. 1–13. Available at: https://doi.org/10.1186/1471-2105-10-442.

Toenhake, C.G. *et al.* (2018) 'Chromatin Accessibility-Based Characterization of the Gene Regulatory Network Underlying *Plasmodium falciparum* Blood-Stage Development', *Cell Host and Microbe*, 23(4), pp. 557-569.e9. Available at: https://doi.org/10.1016/j.chom.2018.03.007.

Toenhake, C.G. and Bártfai, R. (2019) 'What functional genomics has taught us about transcriptional regulation in malaria parasites', *Briefings in Functional Genomics*, 18(5), pp. 290–301. Available at: https://doi.org/10.1093/bfgp/elz004.

Van Der Walt, S. *et al.* (2014) 'Scikit-image: Image processing in python', *PeerJ,* 2014(1). Available at: https://doi.org/10.7717/peerj.453.

Van Treeck, B. *et al.* (2018) 'RNA self-assembly contributes to stress granule formation and defining the stress granule transcriptome', *Proceedings of the National Academy of Sciences of the United*

States of America, 115(11), pp. 2734–2739. Available at: https://doi.org/10.1073/pnas.1800038115.

Treeck, M. *et al.* (2011) 'The Phosphoproteomes of *Plasmodium falciparum* and *Toxoplasma gondii* Reveal Unusual Adaptations Within and Beyond the Parasites' Boundaries', *Cell Host & Microbe*, 10(4), pp. 410–419. Available at: https://doi.org/10.1016/j.chom.2011.09.004.

Trojanowski, J. *et al.* (2022) 'Transcription activation is enhanced by multivalent interactions independent of phase separation', *Molecular Cell*, 82(10), pp. 1878-1893.e10. Available at: https://doi.org/10.1016/j.molcel.2022.04.017.

Tsai, F.T.F. and Sigler, P.B. (2000) 'Structural basis of preinitiation complex assembly on human Pol II promoters', *EMBO Journal*, 19(1), pp. 25–36. Available at: https://doi.org/10.1093/emboj/19.1.25.

Turoverov, K.K., Kuznetsova, I.M. and Uversky, V.N. (2010) 'The protein kingdom extended: Ordered and intrinsically disordered proteins, their folding, supramolecular complex formation, and aggregation', *Progress in Biophysics and Molecular Biology*, 102(2–3), pp. 73–84. Available at: https://doi.org/10.1016/j.pbiomolbio.2010.01.003.

Uversky, V.N. (2013) 'Unusual biophysics of intrinsically disordered proteins', *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1834(5), pp. 932–951. Available at: https://doi.org/10.1016/j.bbapap.2012.12.008.

van der Linden, L. (2019) *Functional Characterization of Plasmodium falciparum TATA-box binding-like Protein*. Masters Thesis. University of Cape Town.

Varadi, M. *et al.* (2022) 'AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models', *Nucleic Acids Research*, 50(D1), pp. D439–D444. Available at: https://doi.org/10.1093/nar/gkab1061.

Vaughan, J., Noden, B. and Beier, J. (2002) 'Proteoglycans mediate malaria sporozoite targeting to the liver', 45, pp. 637–651.

Vaughan, J., Noden, B. and Beier, J. (2007) 'Novel TRF1/BRF target genes revealed by genome-wide analysis of *Drosophila* Pol III transcription', *EMBO Journal*, 26(1), pp. 79–89. Available at: https://doi.org/10.1038/sj.emboj.7601448.

Verma, N. *et al.* (2013) 'Differential utilization of TATA box-binding protein (TBP) and TBP-related factor 1 (TRF1) at different classes of RNA polymerase III promoters', *Journal of Biological Chemistry*, 288(38), pp. 27564–27570. Available at: https://doi.org/10.1074/jbc.C113.503094.

Virtanen, P. *et al.* (2020) 'SciPy 1.0: fundamental algorithms for scientific computing in Python',

Nature Methods, 17(3), pp. 261–272. Available at: https://doi.org/10.1038/s41592-019-0686-2.

Volz, J. *et al.* (2010) 'Potential epigenetic regulatory proteins localise to distinct nuclear sub-compartments in *Plasmodium falciparum*', *International Journal for Parasitology*, 40(1), pp. 109–121. Available at: https://doi.org/10.1016/j.ijpara.2009.09.002.

Voss, T.S. *et al.* (2002) '*Plasmodium falciparum* possesses a cell cycle-regulated short type replication protein A large subunit encoded by an unusual transcript', *Journal of Biological Chemistry*, 277(20), pp. 17493–17501. Available at: https://doi.org/10.1074/jbc.M200100200.

Voß, Y. *et al.* (2023) 'Malaria parasite centrins can assemble by Ca2+-inducible condensation', *PLOS Pathogens*, 19(12), p. e1011899. Available at: https://doi.org/10.1371/journal.ppat.1011899.

Vucetic, S. *et al.* (2003) 'Flavors of protein disorder', *Proteins: Structure, Function and Genetics*, 52(4), pp. 573–584. Available at: https://doi.org/10.1002/prot.10437.

Wahlgren, M., Goel, S. and Akhouri, R.R. (2017) 'Variant surface antigens of *Plasmodium falciparum* and their roles in severe malaria', *Nature Reviews Microbiology*, 15(8), pp. 479–491. Available at: https://doi.org/10.1038/nrmicro.2017.47.

Walsh, I. *et al.* (2012) 'Espritz: Accurate and fast prediction of protein disorder', *Bioinformatics*, 28(4), pp. 503–509. Available at: https://doi.org/10.1093/bioinformatics/btr682.

Wang, Y., Fairley, J.A. and Roberts, S.G.E. (2010) 'Phosphorylation of TFIIB Links Transcription Initiation and Termination', *Current Biology*, 20(6), pp. 548–553. Available at: https://doi.org/10.1016/j.cub.2010.01.052.

Wang, Y.L. *et al.* (2014) 'TRF2, but not TBP, mediates the transcription of ribosomal protein genes', *Genes and Development*, 28(14), pp. 1550–1555. Available at: https://doi.org/10.1101/gad.245662.114.

Wansink, D.G. *et al.* (1993) 'Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus', *Journal of Cell Biology*, 122(2), pp. 283–293. Available at: https://doi.org/10.1083/jcb.122.2.283.

Ward, J.J. *et al.* (2004) 'Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life', *Journal of Molecular Biology*, 337(3), pp. 635–645. Available at: https://doi.org/10.1016/j.jmb.2004.02.002.

Watzlowik, M.T. *et al.* (2021) 'Peculiarities of *Plasmodium falciparum* gene regulation and chromatin structure', *International Journal of Molecular Sciences*, 22(10). Available at:

https://doi.org/10.3390/ijms22105168.

Weideman, C.A. *et al.* (1997) 'Dynamic interplay of TFIIA, TBP and TATA DNA', *Journal of Molecular Biology*, 271(1), pp. 61–75. Available at: https://doi.org/10.1006/jmbi.1997.1152.

Wellems, T.E. *et al.* (1990) 'Chloroquine resistance not linked to mdr-like genes in a *Plasmodium falciparum* cross', *Nature*, 345(6272), pp. 253–255. Available at: https://doi.org/10.1038/345253a0.

Wells, T.N.C., Burrows, J.N. and Baird, J.K. (2010) 'Targeting the hypnozoite reservoir of *Plasmodium vivax* : the hidden obstacle to malaria elimination', (February), pp. 145–151. Available at: https://doi.org/10.1016/j.pt.2009.12.005.

White-cooper, H. and Davidson, I. (2011) 'Unique Aspects of Transcription Regulation in Male Germ Cells', *Cold Spring Harbour Perspectives in Biology*, (3), p. a002626.

Witkowski, B. *et al.* (2017) 'A surrogate marker of piperaquine-resistant *Plasmodium falciparum* malaria: a phenotype–genotype association study', *The Lancet Infectious Diseases*, 17(2), pp. 174–183. Available at: https://doi.org/10.1016/S1473-3099(16)30415-7.

Wobbe, C.R. and Struhl, K. (1990) 'Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription in vitro.', *Molecular and Cellular Biology*, 10(8), pp. 3859–3867. Available at: https://doi.org/10.1128/mcb.10.8.3859.

Wootton, J.C. and Federhen, S. (1993) 'Statistics of local complexity in amino acid sequences and sequence databases', *Computers and Chemistry*, 17(2), pp. 149–163. Available at: https://doi.org/10.1016/0097-8485(93)85006-X.

World Health Organisation (2022) *World malaria report 2022*, *World Health Organization*. Available at: https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2021.

World Health Organization (2020) *World Malaria Report: 20 years of global progress and challenges*, *World Health Organization*. Available at: https://www.who.int/publications/i/item/9789240015791.

Wu, W.H. *et al.* (1999) 'Mutational analysis of yeast TFIIB: A functional relationship between Ssu72 and Sub1/Tsp1 defined by allele-specific interactions with TFIIB', *Genetics*, 153(2), pp. 643–652. Available at: https://doi.org/10.1093/genetics/153.2.643.

Wu, W.H. and Hampsey, M. (1999) 'An activation-specific role for transcription factor TFIIB in vivo', *Proceedings of the National Academy of Sciences of the United States of America*, 96(6), pp. 2764–2769. Available at: https://doi.org/10.1073/pnas.96.6.2764.

Wurdeman, S.R. *et al.* (2017) 'Coactivator condensation at super-enhancers links phase separation

and gene control', *Physiology & behavior*, 176(5), pp. 498–503. Available at: https://doi.org/10.1126/science.aar3958.Coactivator.

Xu, M. *et al.* (2011) 'Core promoter-selective function of HMGA1 and Mediator in Initiator-dependent transcription', *Genes and Development*, 25(23), pp. 2513–2524. Available at: https://doi.org/10.1101/gad.177360.111.

Yang, Z.R. *et al.* (2005) 'RONN: The bio-basis function neural network technique applied to the detection of natively disordered regions in proteins', *Bioinformatics*, 21(16), pp. 3369–3376. Available at: https://doi.org/10.1093/bioinformatics/bti534.

Zacharias, D.A. *et al.* (2002) 'Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells', *Science*, 296(5569), pp. 913–916. Available at: https://doi.org/10.1126/science.1068539.

Zamudio, A. V *et al.* (2019) 'Mediator condensates localize signaling factors to key cell identity genes', 76(5), pp. 753–766. Available at: https://doi.org/10.1016/j.molcel.2019.08.016.Mediator.

Zerby, D. and Lieberman, P. (1997) 'Functional Analysis of TFIID-Activator Interaction by Magnesium-Agarose Gel Electrophoresis.', *Methods*, 3(12), pp. 217–23. Available at: https://doi.org/10.1006/meth.1997.0474.

Zhang, D.Y. *et al.* (2000) 'Intramolecular interaction of yeast TFIIB in transcription control', *Nucleic Acids Research*, 28(9), pp. 1913–1920. Available at: https://doi.org/10.1093/nar/28.9.1913.

Zhang, D.Y., Carson, D.J. and Ma, J. (2002) 'The role of TFIIB-RNA polymerase II interaction in start site selection in yeast cells', *Nucleic Acids Research*, 30(14), pp. 3078–3085. Available at: https://doi.org/10.1093/nar/gkf422.

Zhang, F. *et al.* (2022) 'Dynamic phase separation of the androgen receptor and its coactivators to regulate gene expression', *Cancer Research*, 82(12_Supplement), pp. 5729–5729. Available at: https://doi.org/10.1158/1538-7445.am2022-5729.

Zhang, H. *et al.* (2015) 'RNA Controls PolyQ Protein Phase Transitions', *Molecular Cell*, 60(2), pp. 220–230. Available at: https://doi.org/10.1016/j.molcel.2015.09.017.

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. Available at: https://doi.org/10.1016/S0092-8674(02)00648-7.

Zheng, X.M. *et al.* (1987) 'A general transcription factor forms a stable complex with RNA

polymerase B (II)', *Cell*, 50(3), pp. 361–368. Available at: https://doi.org/10.1016/0092-8674(87)90490-9.

Zhou, H. *et al.* (2019) 'Mechanism of DNA-Induced Phase Separation for Transcriptional Repressor VRN1', *Angewandte Chemie*, 131(15), pp. 4912–4916. Available at: https://doi.org/10.1002/ange.201810373.

Appendix

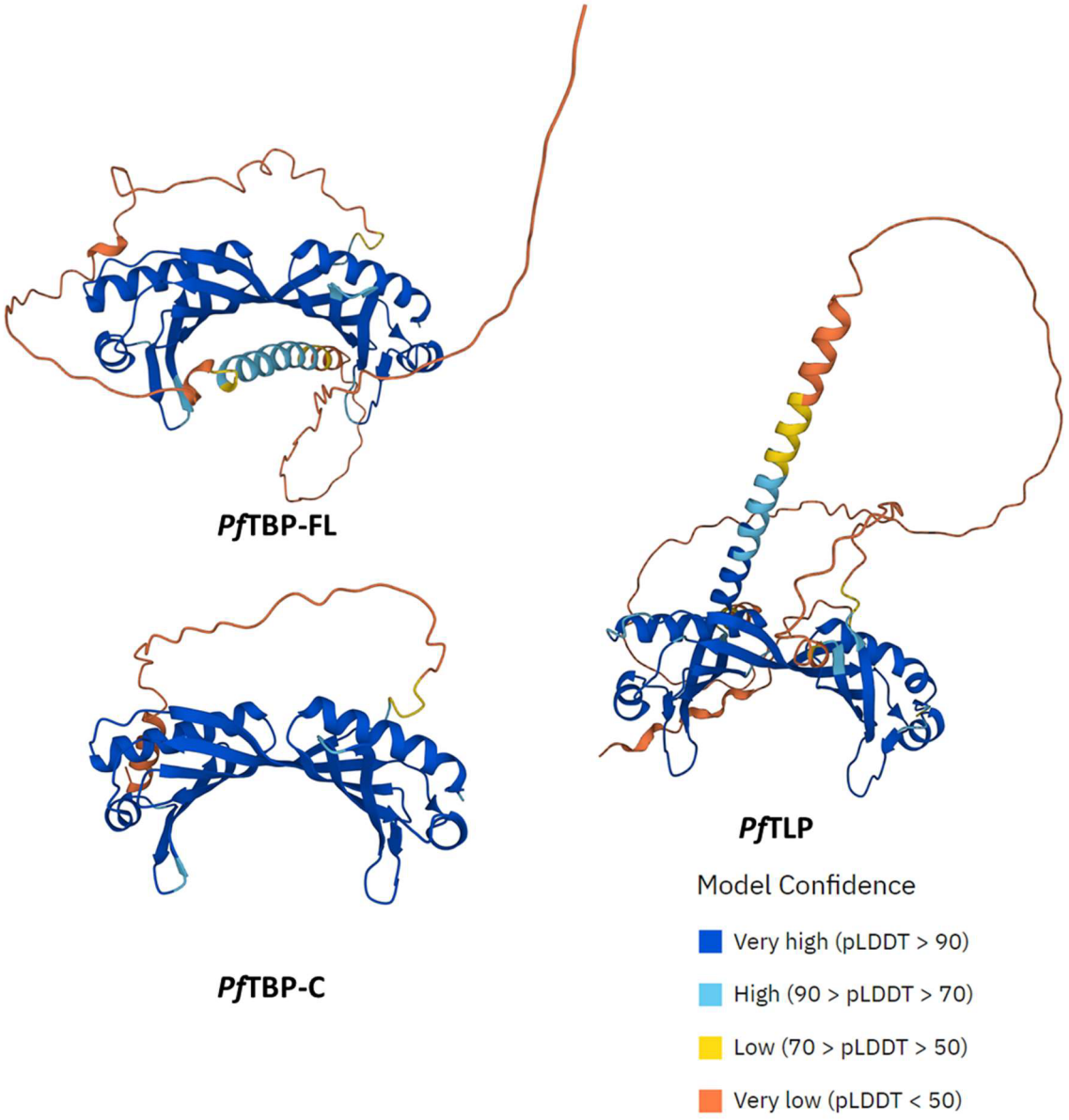


Figure S1. Intrinsically disordered regions in *Pf*TLP and *Pf*TBP could not be modelled with high confidence. Cartoon representations of predicted structures of 38 kDa *Pf*TBP (*Pf*TBP-FL), 26 kDa *Pf*TBP (*Pf*TBP-C) and *Pf*TLP. Secondary structures are coloured according to Alphafold per residue confidence scores. Regions likely to be intrinsically disordered have low per residue confidence scores and are thus shown in yellow and orange.

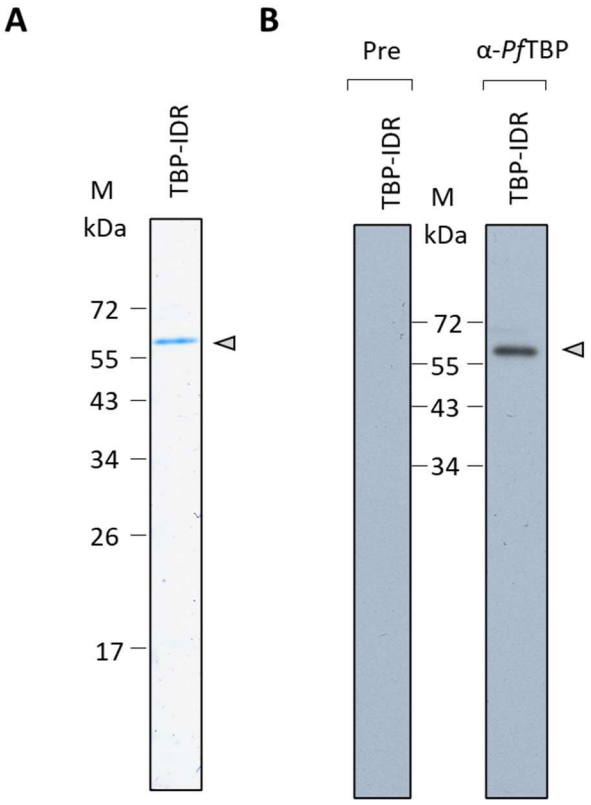


Figure S2. The *Pf*TBP-IDR is a low mobility protein in SDS-PAGE.
(A) SDS-PAGE analysis of purified 6His:*Pf*TBPIDR-YFP. 500 ng purified recombinant 6His:*Pf*TBP-IDR-YFP was resolved by 12% SDS-PAGE and visualised by Coomassie staining. **(B)** 250 fmol purified 6His:*Pf*TBP-IDR-YFP was resolved by 12% SDS-PAGE and analysed by immunoblotting purified anti-*Pf*TBP antibody or an equivalent amount of pre-immune serum (PRE). 6His:*Pf*TBP-IDR-YFP has a calculated molecular mass of 47 kDa but migrates with an apparent molecular mass of ~ 62 kDa.

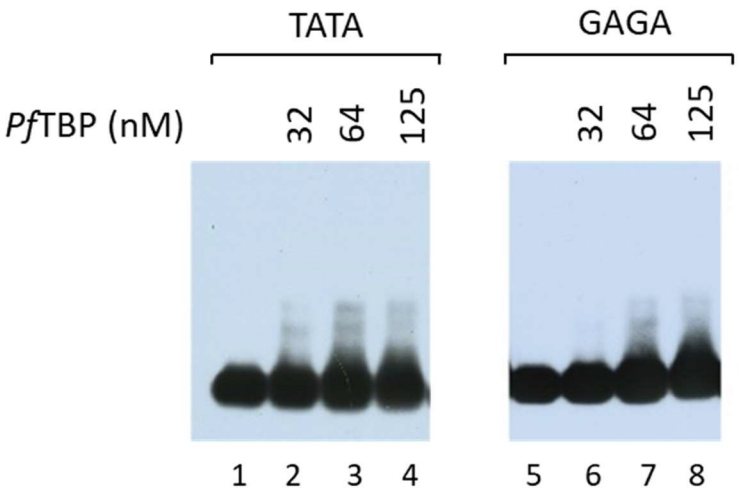


Figure S3*. Pf*TBP does not bind a prototypical TATA-box DNA in a sequence-specific manner.
DNA binding reactions (10 µl) contained 5 fmol Ad2ML-TATA or GAGA DNA probe and indicated concentrations of *Pf*TBP. Reactions were incubated at 27°C for 45 min and resolved by 1.4% agarose electrophoresis in the absence of Mg^{2+}.

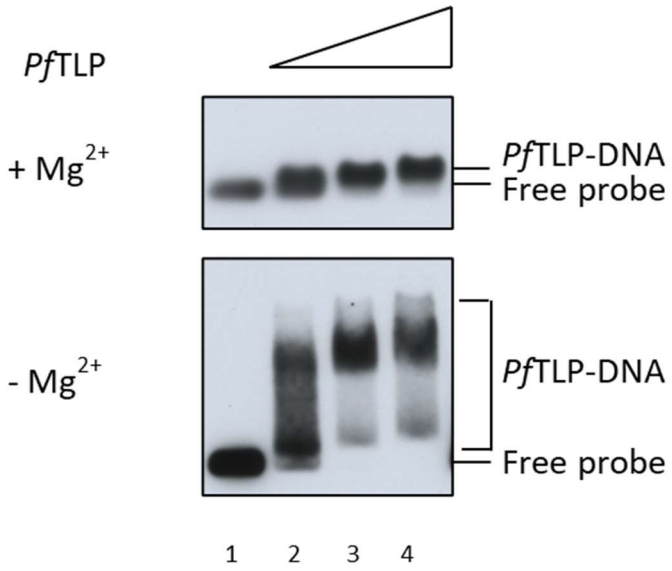


Figure S4. Improved resolution of *Pf*TLP-DNA complexes in agarose EMSA in the absence of magnesium ions. DNA-binding reactions (10 µl) contained 5 fmol GBP-130 282 bp biotinylated DNA probe and 3 nM (lane 2), 6 nM (lane 4), or 12nM (lane 5) 6His:*Pf*TLP . Reactions were incubated on ice for 45 min before being resolved by 1.4 % agarose gel electrophoresis in the absence presence or absence of 5 mM magnesium acetate in the gel and 2 mM magnesium acetate in the electrophoresis buffer.

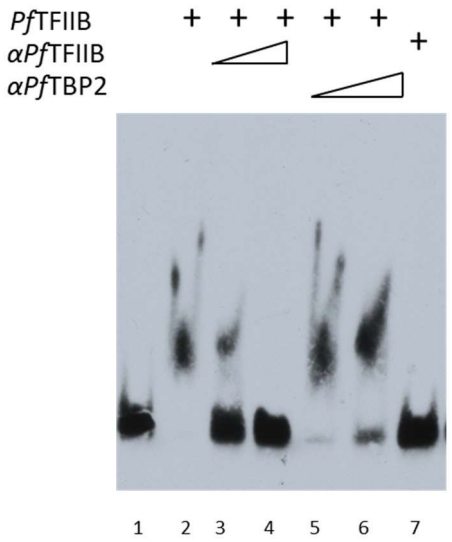


Figure S5. Selective inhibition of *Pf*TFIIB-DNA interactions by anti-*Pf*TFIIB antibody.
DNA binding reactions (10 µl) contained 5 fmol 60 bp biotinylated KAHRP DNA probe, 23 nM *Pf*TFIIB and 50 ng (lanes 3 and 5) or 100 ng (lanes 4 and 6) anti-*Pf*TFIIB or anti-*Pf*TBP antibody as indicated. Lane 7 shows a DNA-binding reaction containing only DNA and 100ng anti-PfTFIIB antibody. 6His:*Pf*TFIIB and antibody were preincubated for 30 min before addition of DNA. DNA-binding reactions were incubated on ice for 45 min, and resolved by 5% (37:1) PAGE in the presence of 2 mM Mg^{2+}.

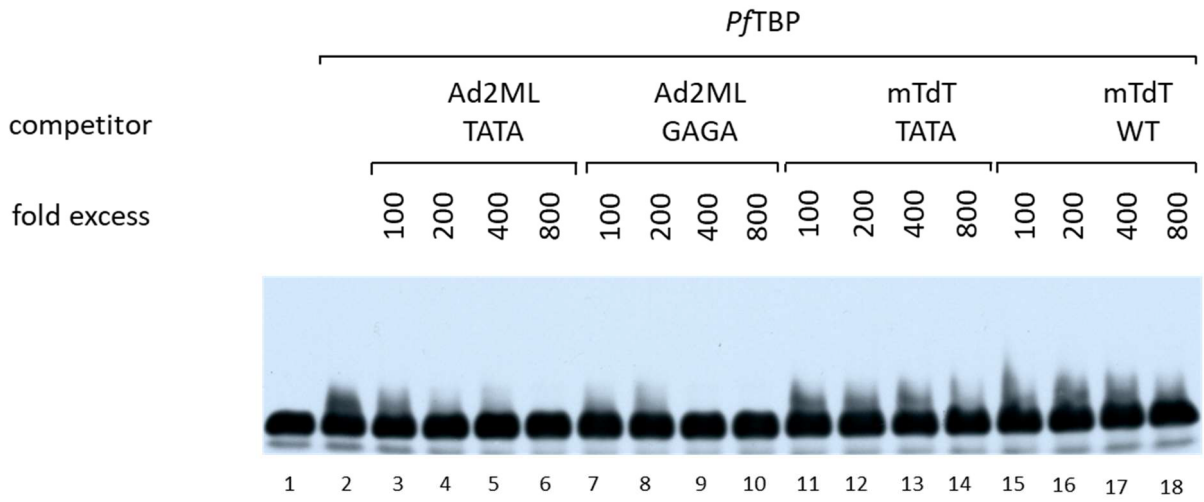


Figure S6. *Pf*TBP does not possess sequence-specific TATA box-binding activity. DNA-binding reactions (10 µl) contained 5 fmol biotinylated 282 bp GBP-130 DNA probe, 50 ng poly(dG-dC), 60 nM *Pf*TBP, and 100 – 800 fold 40 bp unlabelled Ad2ML-TATA, Ad2ML-GAGA, mTdT-TATA or mTdT-WT competitor DNA as indicated. Fold excess competitor relative to the labelled GBP-130 probe was calculated in molar nucleotides. DNA binding reactions were incubated on ice for 45 min and being resolved by 1.4 % gel electrophoresis in the presence of 2 mM Mg^{2+}.

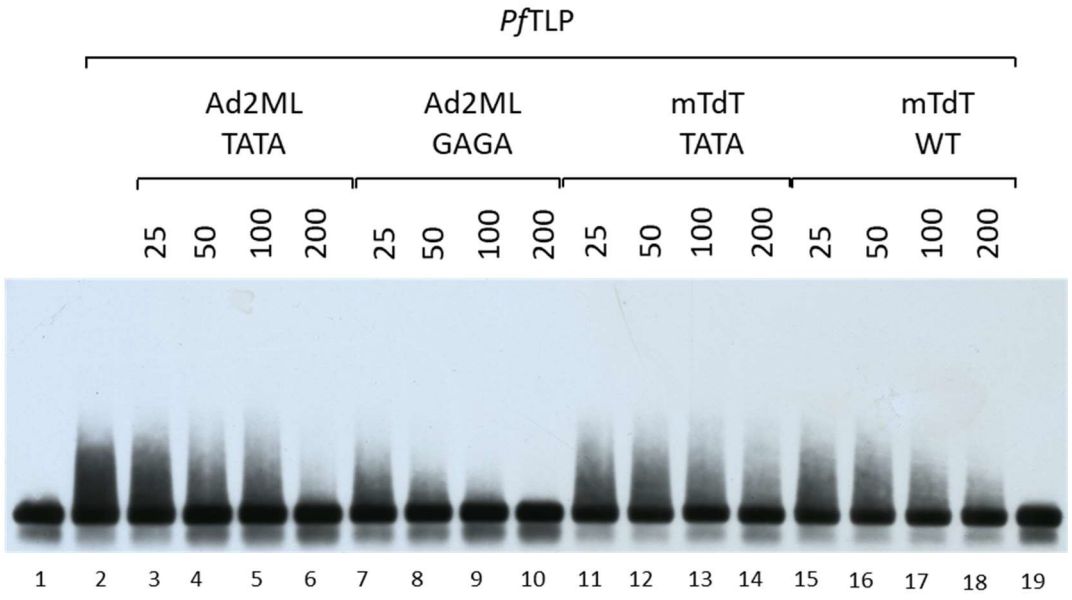


Figure S7. *Pf*TLP does not possess sequence-specific TATA box-binding activity. DNA-binding reactions (10 µl) contained 5 fmol biotinylated 282bp GBP-130 DNA probe, 10 ng poly(dG-dC), 11 nM *Pf*TLP, and 100 – 800 fold 40 bp unlabelled Ad2ML-TATA, Ad2ML-GAGA, mTdT-TATA or mTdT-WT competitor DNA as indicated. Fold excess competitor relative to the labelled GBP-130 probe was calculated in molar nucleotides. DNA binding reactions were incubated on ice for 45 min before being resolved by 1.4 % gel electrophoresis.

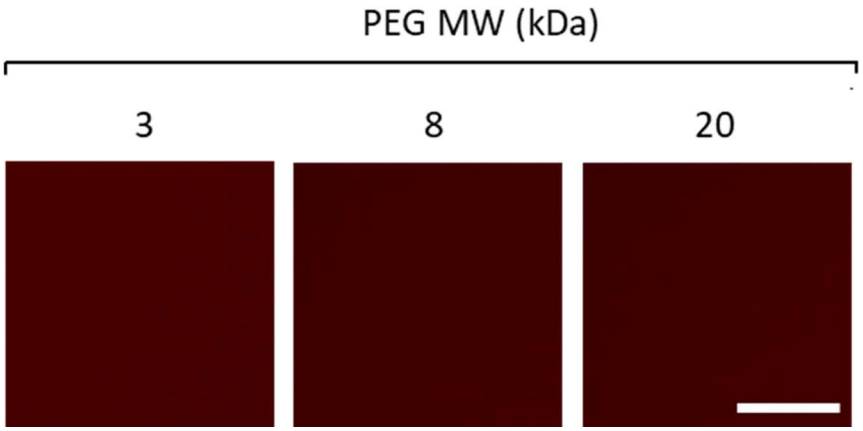


Figure S8. 6His:mCherry does not undergo *in vitro* condensation at a concentration of 50 µM in the presence of various MW PEG molecular crowders. Representative fluorescent microscopy images of condensation reactions containing 50 µM 6His:mCherry, 60 mM KCl, 5 mM magnesium acetate, and 10 % PEG molecular crowder. The molecular weight (MW) of PEG is indicated in kDa. Scale bar=10 µM.

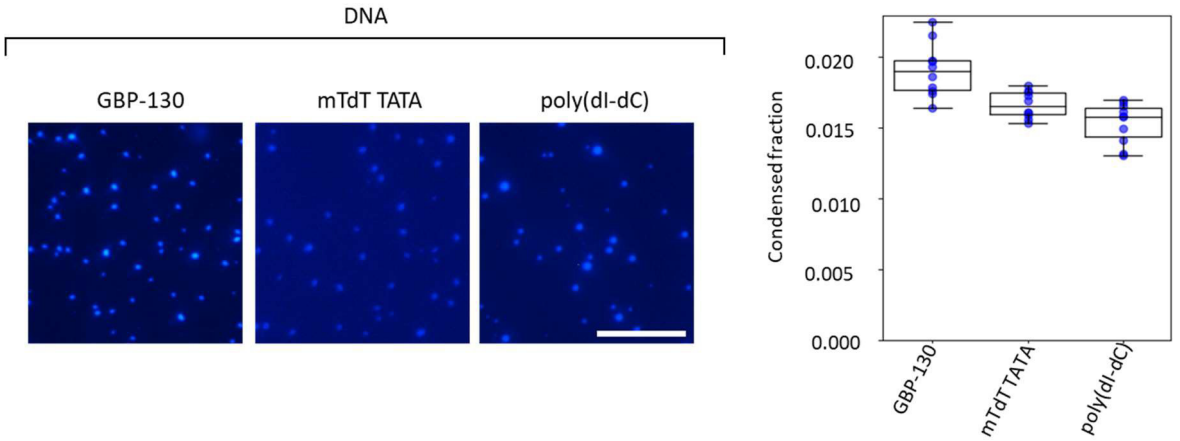


Figure S9. 282 bp GBP-130 dsDNA, 260 bp mTdT-TATA dsDNA and poly(dI-dC) form phase separated condensates in the presence of molecular crowder and magnesium ions. *Left:* Representative fluorescent microscope images. Condensation reactions contained 50 ng/µl of the indicated dsDNA, 10 % PEG 20000, 150 mM KCl, 5mM magnesium acetate and 5 µM DAPI. Scale Bar: 10 µm. *Right:* Quantification of the condensed fraction. Image analysis and quantification of condensed fraction was performed for 10 fields of view per condition.

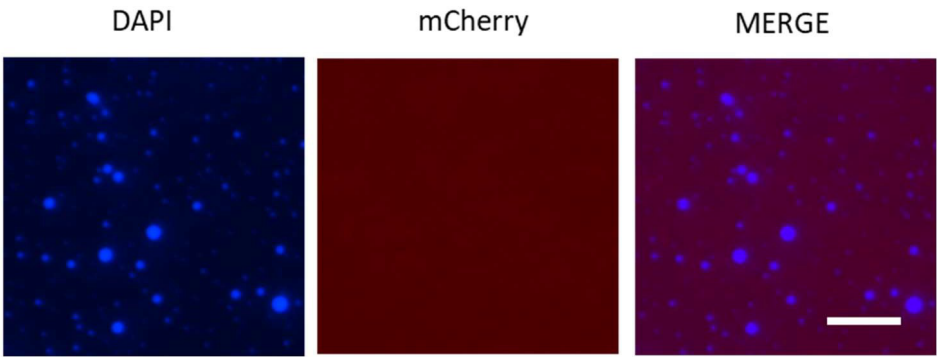


Figure S10. Poly(dI-dC) condensates do not recruit mCherry fluorescent protein.
Representative fluorescent microscope images of condensation reactions containing 200 ng/µl poly(dI-dC), 1 µM 6His:mCherry, 10 % PEG 20000, 150 mM KCl, 5 mM magnesium acetate and 50 µM DAPI. Poly(dI-dC) and 6His:mCherry were mixed thoroughly and incubated for 5 min before the addition of condensation buffer. Mixed assemblies were visualised using the indicated fluorescent channels. Scale bar = 10 µM.

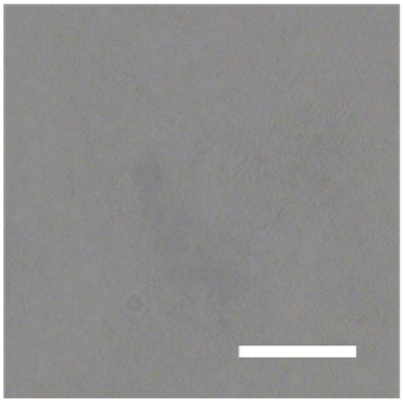


Figure S11. Total yeast RNA does not form condensates at a concentration of 200 ng/µl.
Representative brightfield microscope image showing absence of condensate formation of 200 ng/µl total yeast RNA in the presence of 10 % PEG 3000, 150 mM KCl, and 5 mM magnesium acetate. Scale bar = 10 µm.

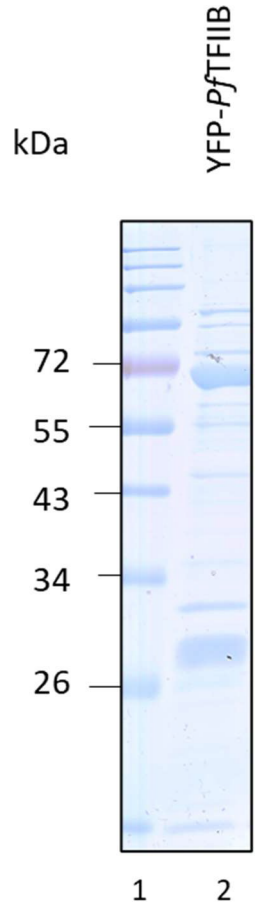


Figure S12. **SDS-PAGE analysis of bacterially expressed recombinant 6His:YFP-*Pf*TFIIB purified by IMAC.** 1 µg total protein was resolved by SDS-PAGE and visualised by Coomassie staining (lane 2). The molecular mass of protein MW standards (lane 1) is indicated. The calculated MW of 6His:YFP-*Pf*TFIIB is 71 kDa.

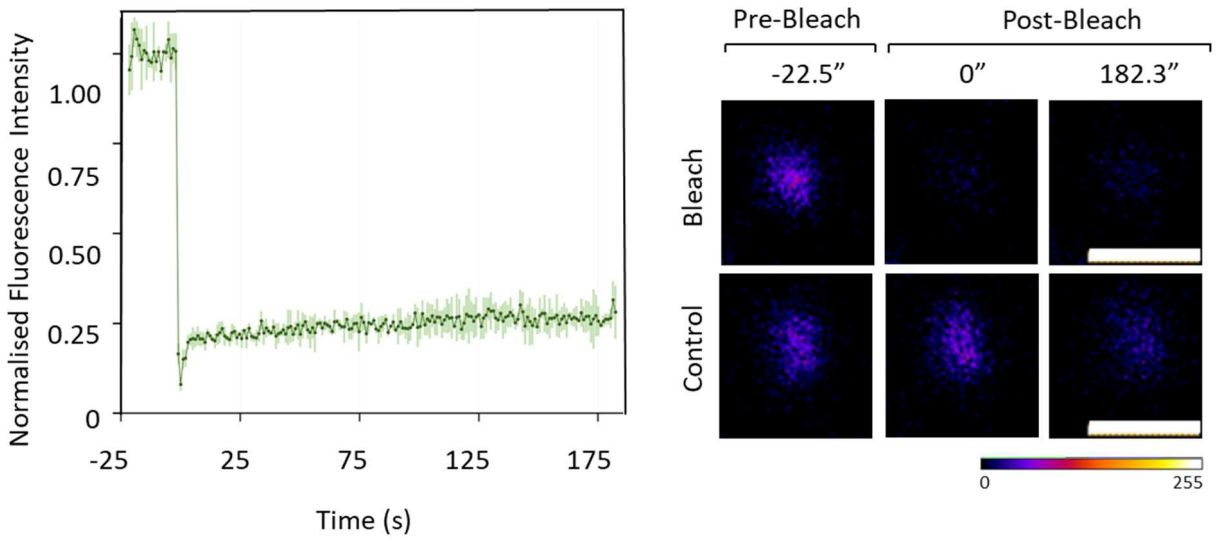


Figure S13. Fluorescence recovery after photobleaching of condensates formed with 6His:YFP-*Pf*TFIIB. Average Fluorescence recovery after photobleaching of *Pf*TFIIB condensates formed with 0.5 µM 6His:YFP-*Pf*TFIIB in the presence of 10% PEG 20000, 5 mM magnesium acetate and 150 mM KCl. Photobleaching was carried out after 20 - 40 min incubation following initiation of condensation. Left: Quantification of normalised fluorescence intensity. Error Bars represent the standard deviation (n=3). Right: Representative images of bleached (top) and control (bottom) condensates at indicated time points relative to completion of bleaching. Non-bleached control condensate in the same field of view were used for the calculation of normalised fluorescence intensity using EasyFrap (Koulouras *et al*., 2018). Images are coloured using the fire look up table in ImageJ. Scale Bar = 1 µm (white).

Table S1. Accession numbers of protein sequences from the NIH Protein Database used for bioinformatic analysis.

Protein	Ab.	Organism	Class / Pylum	Accession Number
TBP	*Hs*	*Homo sapiens*	Mammalia	AAA36731.1
TBP	*Mm*	*Mus musculus*	Mammalia	AAH50136.1
TBP	*Gg*	*Gallus_gallus*	Aves	BAA20298.1
TBP	*Xl*	*Xenopus laevis*	Amphibia	NP_001084369.1
TBP	*Ce*	*Caenorhabditis elegans*	Nematoda	AAA03582.1
TBP	*Ov*	*Onchocerca volvulus*	Nematoda	AAA16924.1
TBP	*Dm*	*Drosophila melanogaster*	Arthropoda	AAA68629.1
TBP	*Sc*	*Saccharomyces cerevisiae*	Fungi	NP_011075.3
TBP	*Ec*	*Encephalitozoon_cuniculi*	Fungi	AGE95302.1
TBP	*Ca*	*Candida albicans*	Fungi	AAC49986.1
TBP	*Ci*	*Ciona intestinalis*	Tunicate	XP_002127120.1
TBP	*Aq*	*Amphimedon queenslandica*	Porifera	XP_003384438.1
TBP	*Hv*	*Hydra vulgaris*	Hydrozoan	XP_002160884.2
TBP1	*At*	*Arabidopsis thaliana*	Planta	AAR28027.1
TBP2	*At*	*Arabidopsis thaliana*	Planta	NP_175948.1
TBP1	*Zm*	*Zea mays*	Planta	NP_001105318.1
TBP2	*Zm*	*Zea mays*	Planta	NP_001105319.1
TBP	*Mc*	*Mesembryanthemum crystallinum*	Planta	AAA80641.1
TBP	*Pf*	*Pyrococcus_furiosus*	Archaea	WP_011012439.1
TBP	*Pw*	*Pyrococcus_woesei*	Archaea	AAA73447.1
TBP	*Tc*	*Thermococcus_celer*	Achaea	sp|Q56253.1
TBP	*Mem*	*Methanococcus_maripaludis*	Arhaea	sp|A9A840.1
TLP	*Hs*	*Homo sapiens*	Mammalia	AAD28785.1
TLP	*Mm*	*Mus musculus*	Mammalia	NP_035733.1
TLP	*Gg*	*Gallus gallus*	Aves	BAA75886.1
TLP	*Xl*	*Xenopus laevis*	Amphibia	NP_001083732.1
TLP	*Dm*	*Drosophila melanogaster*	Arthropoda	NP_996377.1
TLP	*Ce*	*Caenorhabditis elegans*	Nematoda	CAB03082.2

TLP	*Bm*	*Brugia malayi*	Nematoda	XP_001899826.1
TLP	*Ci*	*Ciona intestinalis*	Tunicate	XP_018668192.2
TRF3	*Hs*	*Homo sapiens*	Mammalia	NP_950248.1
TRF3	*Mm*	*Mus musculus*	Mammalia	NP_951014.1
TRF3	*Xl*	*Xenopus laevis*	Amphibia	NP_001080921.1
TATA_box-binding protein-like protein	*Cc*	*Crypthecodinium_cohnii*	Protozoa (Dinoflagellata (phylum) Peridiniales (order)	AAL24503.1
TBP-related_factor_ TRF1	*Eh*	*Entamoeba_histolytica*	Protozoa	ABS89251.1
TBP	*Eh*	*Entamoeba_histolytica*	Protozoa	CAA88304.1
TFIID-like_protein (TRF4)	*Lm*	*Leishmania major*	Protozoa	CAJ07235.1
TRF4	*Tb*	*Trypanasoma brucei brucei*	Protozoa	XP 828065.1
TBP1	Tv	*Trichomonas_vaginalis*	Protozoa	XP_001329427.1
TBP2	Tv	*Trichomonas_vaginalis*	Protozoa	XP_001300812.1
TBP (TATA-box binding protein)	*Bbi*	*Babesia bigemina*	Apicomplexan	XP_012767010.1
TBP2 (hypothetical protein conserved)	*Bbi*	*Babesia bigemina*	Apicomplexan	XP_012767872.1
TBP (TATA-binding protein)	*Ch*	*Cryptosporidium hominis*	Apicomplexan	XP_667132.1
TFIID-family Protein Transcription factor TFIID (or TATA-binding protein TBP) family protein	*Ch*	*Cryptosporidium hominis*	Apicomplexan	PPA65423.1/ OLQ18406.1
TBP2 (tata-box binding protein tbp2)	*Cc*	*Cyclospora cayetanensis* *CHN_HEN01*	Apicomplexan	OEH79464.1
TBP1 (TATA-box-binding protein)	*Cc*	*Cyclospora cayetanensis*	Apicomplexan	XP_026191038.1

uTBP (uncharacterised protein)	*Cc*	*Cyclospora cayetanensis*	Apicomplexan	XP_026190742.1
TBP (TATA-box binding protein, putative)	*Et*	*Eimeria tenella*	Apicomplexan	CDJ40414.1/ XP_013231164.1
TBP (putative TATA-box binding protein)	*Gn*	*Gregarina_niphandrodes*	Apicomplexan	XP_011131434.1
TFIID	*Gn*	*Gregarina_niphandrodes*	Apicomplexan	XP_011128611.1
TRP (TATA-binding protein, related)	*Nc*	*Neospora caninum*	Apicomplexan	XP_003883026.1
TBP (putative TATA box binding protein)	*Nc*	*Neospora caninum*	Apicomplexan	XP_003881260.1
TLP	*Pf*	*Plasmodium_falciparum*	Apicomplexan	PF3D7_1428800 XP_024329205.1
TBP	*Pf*	*Plasmodium_falciparum*	Apicomplexan	PF3D7_0506200 XP_001351620.1
TBP1	*Ta*	*Theileria annulata*	Apicomplexan	XP_954324.1
TBP2 transcription initiation factor TBP homologue, putative	*Ta*	*Theileria annulata*	Apicomplexan	XP_954156.1
TBP1 (TATA box-binding protein TBP1)	*Tg*	*Toxoplasma gondii ME49*	Apicomplexan	XP_002368492.1
TBP2 (TATA-box binding protein TBP2)	*Tg*	*Toxoplasma gondii ME49*	Apicomplexan	XP_002365082.1

Table S2. PCR primers used in the generation of EMSA DNA probes

Primer Name	Sequence
pTOTdTPCR-FWD	CGA GCG GAG ACT CTA GAA TTC
pTOTdTPCR-REV	AA CGC GTT GGG AGC TCT CC
pTOPCR-US-FWD	AAT TGG GCC CGA CGT CGC A
GBP-130-REV3	GAA CAT AAT TAT TTT ATA TCT TAC TAT TAT TTT TAG
GBP-130-FWD3	GTA AGA TAT AAA ATA ATT ATG TTC AAT TTA ATA AT
KAHRP-REV3	GTC GTT TTT TCT CAT TTA TTT ATT ATA ATT TAC
KAHRP-FWD3	GAG AAA AAA CGA CAA AAC ATA ACT AC

Supplementary Table 3. Primer sequences used for the generation of fluorescent protein tagged wild type *Pf*TFIIB and *Pf*TFIIB deletion variants. Primer sequences not part of the ORFs of proteins of interest are highlighted in red, restriction enzyme sites are underlined.

Primer Name	Sequence	RE site
mCherry-FWD	GGCAGC**CAT**ATGGTGAGCAAGGGCGAGGA	NdeI
mCherryREV-IIBFL	ACTTTTTATTTGGAGAAATGGATGACATCTTGTACAGC TCGTCCATGCC	
mCherry-PfIIBFL-FWD	GGCATGGACGAGCTGTACAAGATGTCATCCATTTCTCC AAATAAAAAGT”	
PfIIB-REV	GCAGCCAGATCTCTATTAGTTTTTTTTCCTTTTGTCTTCA CTAAG	BglII
mCherry-PfIIB140-367FWD	GGCATGGACGAGCTGTACAAGAATGATCAAACATTAA TTTCCGCATTTAATA	
mCherry-IIB140REV	TATTAAATGCGGAAATTAATGTTTGATCATTCTTGTAC AGCTCGTCCATGCC	
PfIIB-FWD	GGCAGCCATATGTCATCCATTTCTCCAAATAAAAAG	NdeI

PfIIB1-140mcherryRev	TCCTCGCCCTTGCTCACCATTTTATTTATTTGTGTCATCA TATTTAAATGTTG	
PfIIB1-140mcherryFWD	CAACATTTAAATATGATGACACAAATAAATAAAATGGT GAGCAAGGGCGAGGA	
PfIIB-38-140 FWD	GGCAGCCATATGAATGATCAAACATTAATTTCCGCATT TAATATC	NdeI
mCherry-REV	GCAGCCGGATCCCTATTA CTTGTACAGCTCGTCCATGCC	BamHI
YFP-FWD	GCGGCAGCCATATGGTGAGCAAGGGCGAG	NdeI
YFP-REV	GCAGCCGGATCCCTATTACTTGTACAGCTCGTCCATG	BamHI
YFPTFIIB-FWD	GGCAGC**CAT**ATGGTGAGCAAGGGCGAGGA	NdeI
YFPTFIIB-REV	GGCAGCCATATGCTTGTACAGCTCGTCCATGCC	NdeI

Table S4. PROCHECK results for *Pf*TBP and *Pf*TLP tertiary structure predictions. Values reflect the structural evaluation by Ramachandran plot produced by PROCHECK. Asterisks indicate values for experimentally determined structures (Buendía-Orozco, Guerrero and Pastor, 2005). Note that for both *Pf*TBP and *Pf*TLP, residues within disallowed regions are located in regions predicted to be intrinsically disordered.

Structure	Residues in most favoured regions (%)	Residues in additional allowed regions (%)	Residues in generously allowed regions (%)	Residues in disallowed regions (%)
1C9B	89.3*	10.7*	0.0*	0.0 *
1VOL	78.1*	18.8*	1.1*	2.0*
1D3U	91.3*	8.1*	0.6*	0.0*
1AIS	88.2*	10.4*	0.9*	0.6*
*Pf*TLP	76.3	16.0	4.9	2.9
*Pf*TBP-C	84.7	12.6	2.8	0
*Pf*TBP-FL	78.1	15.4	2.9	3.5

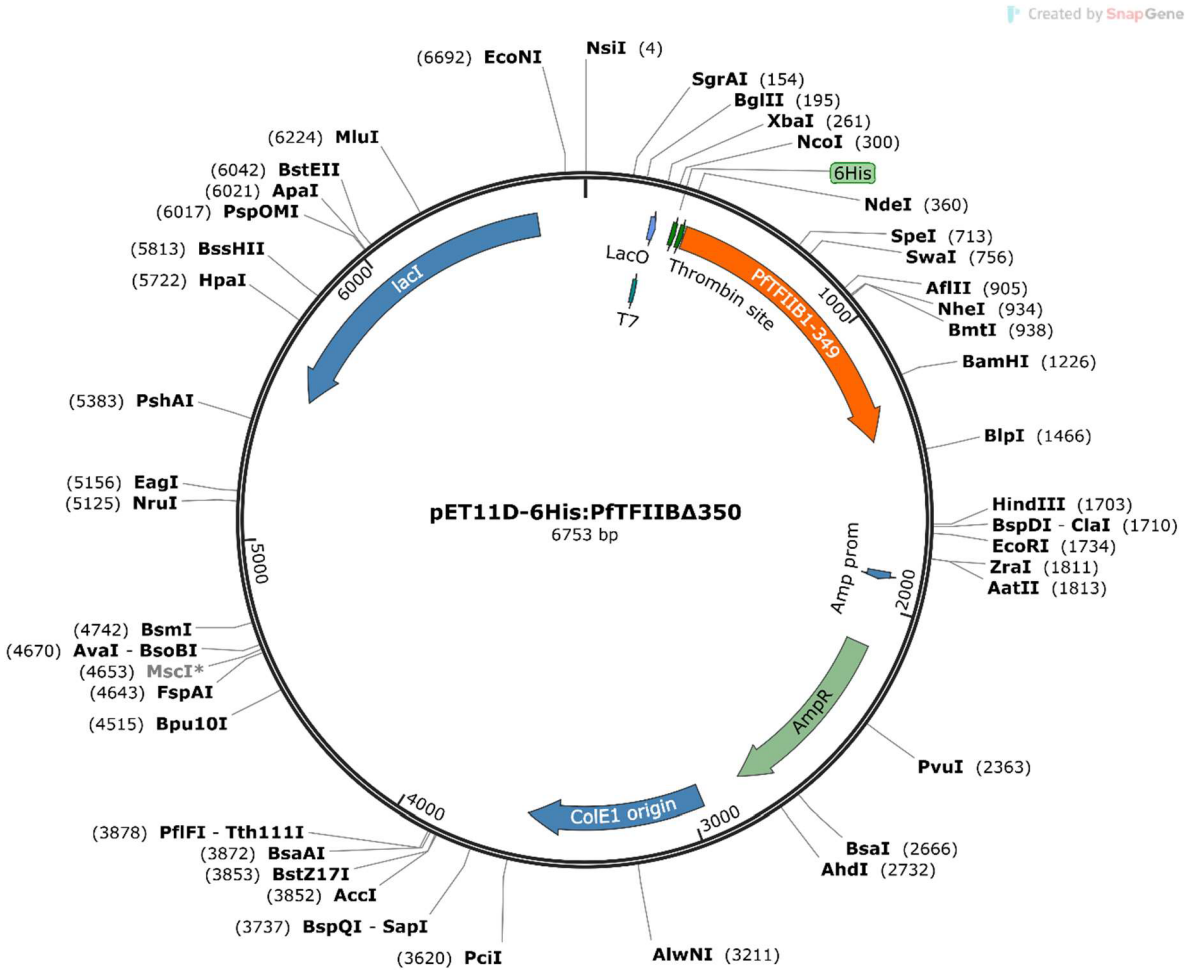


Fig 14. Plasmid map of pET11D-6His:*Pf*TFIIBΔ350

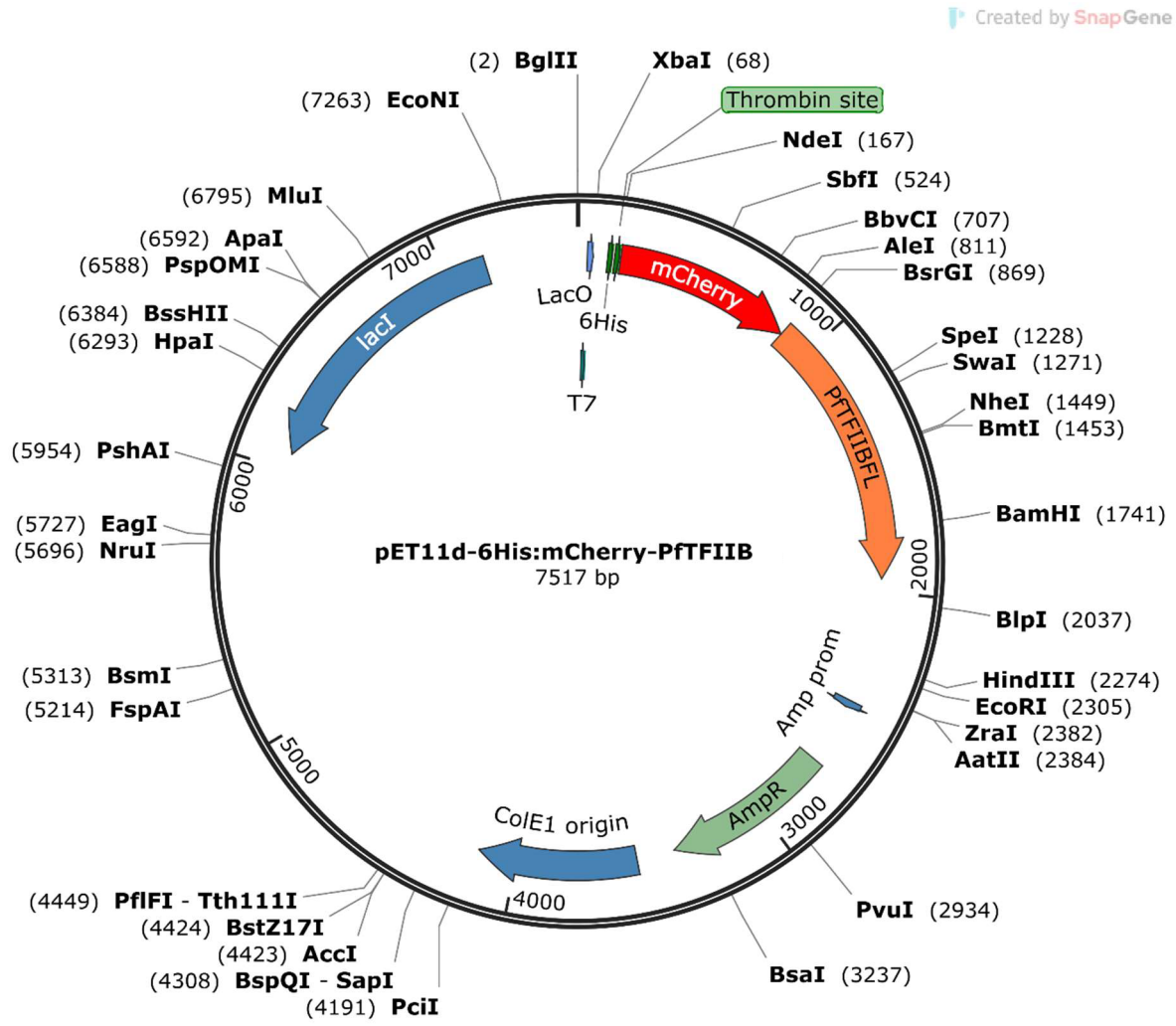


Fig S15. Plasmid map of pET11d-6His:mCherry-*Pf*TFIIB

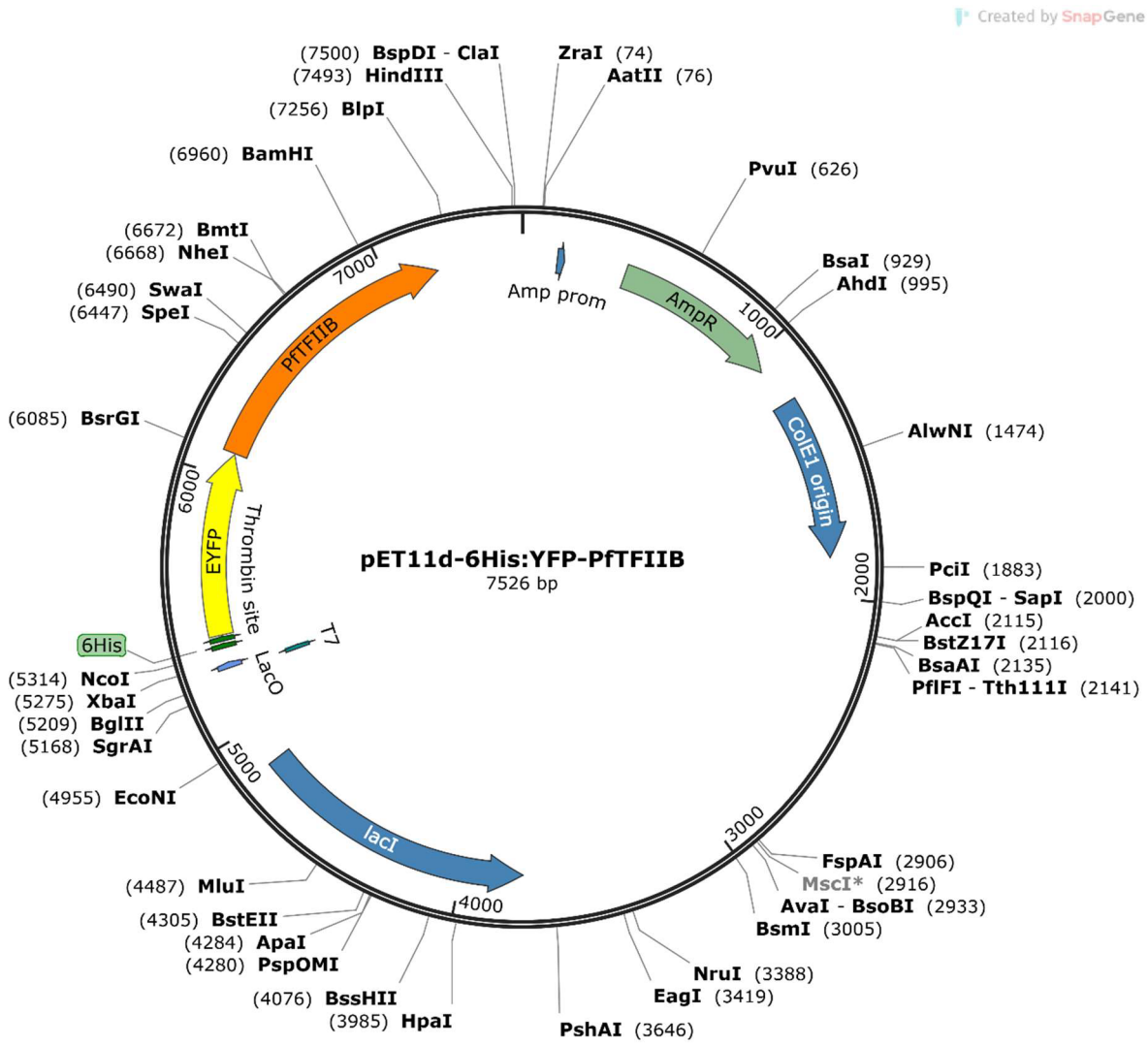


Fig S16. Plasmid map of pET11d-6His:YFP-*Pf*TFIIB

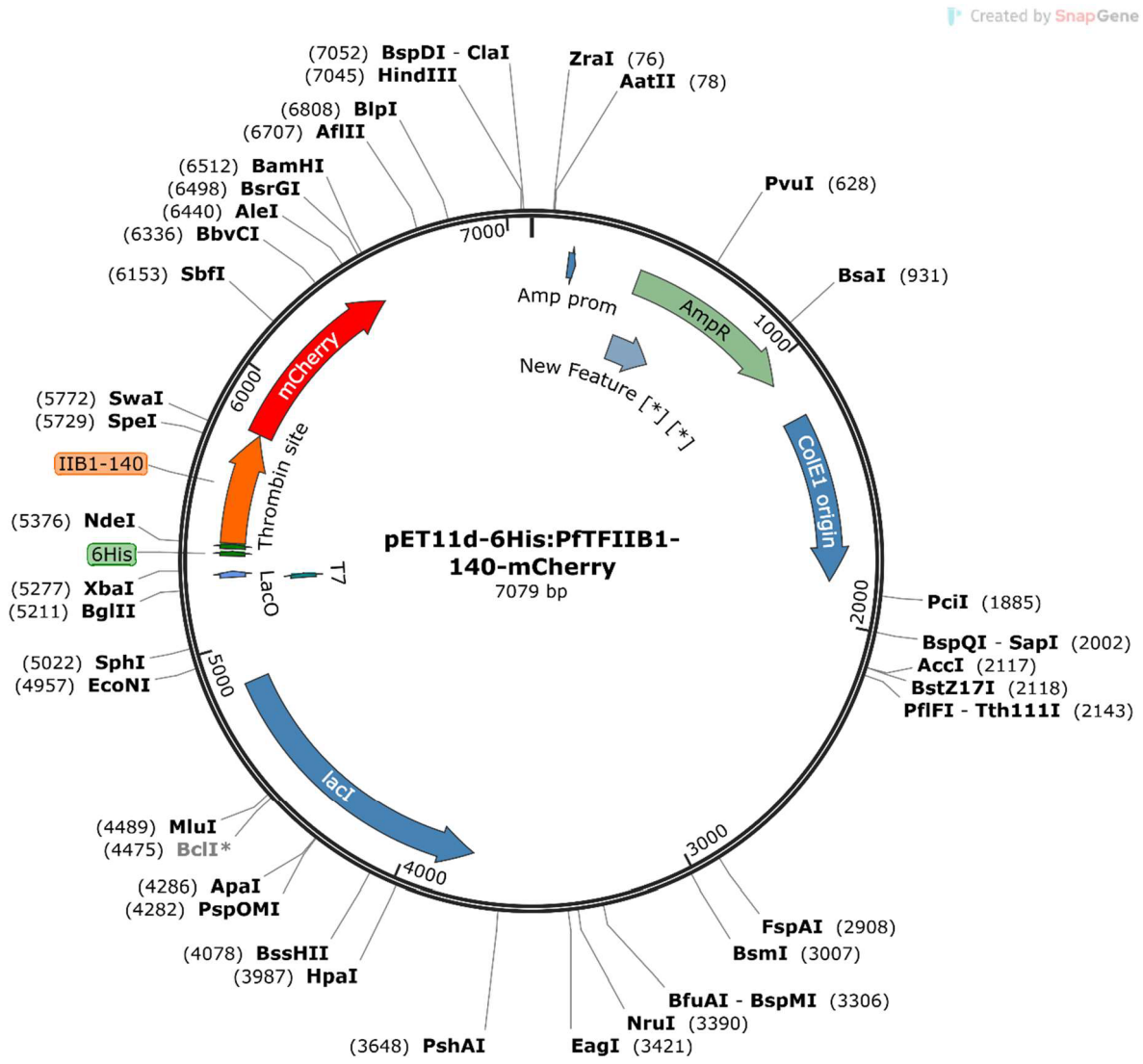


Fig S17. Plasmid map of pET11d-6His:*Pf*TFIIB1-140-mCherry

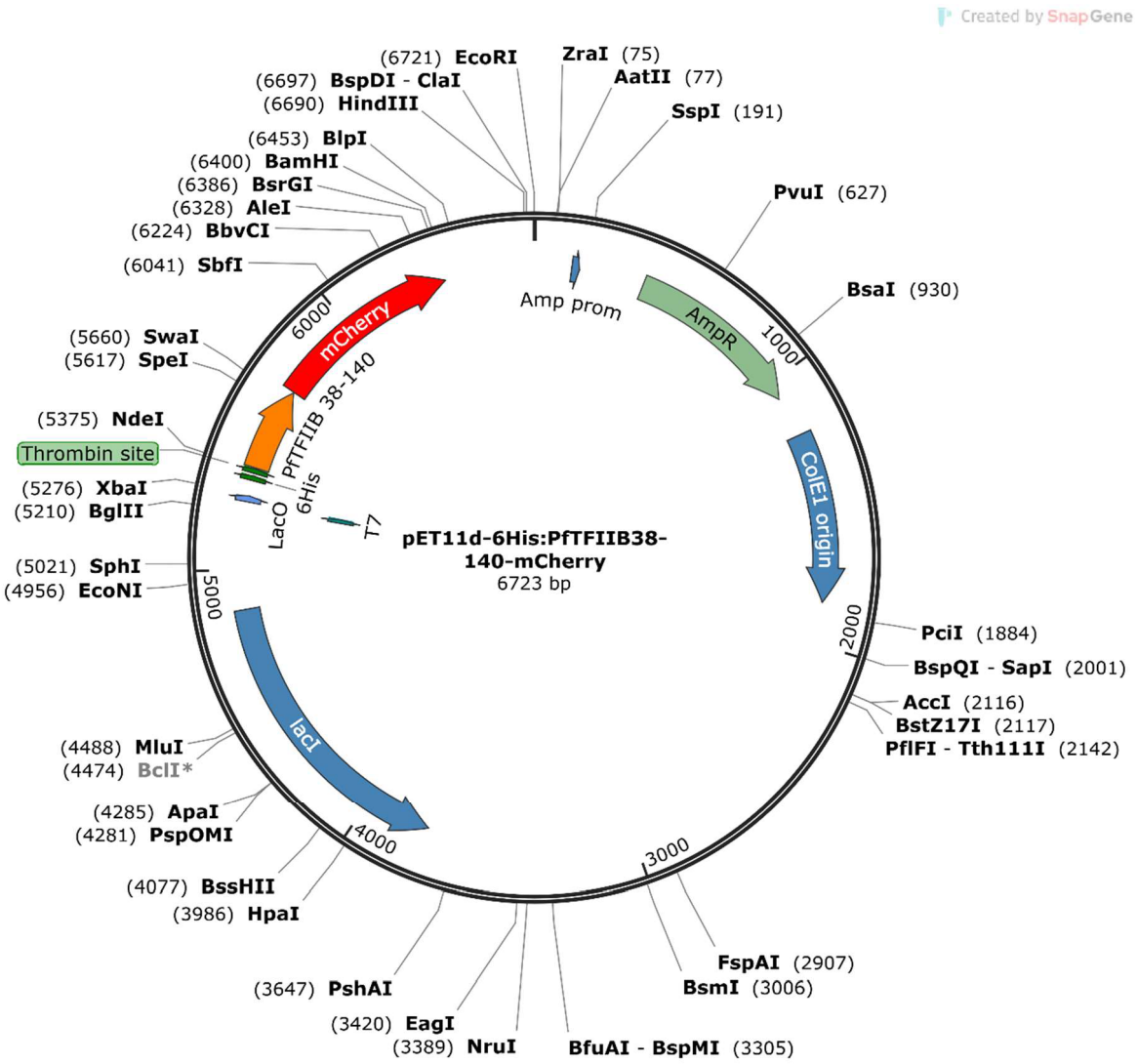


Fig S18. Plasmid map of pET11d-6His:*Pf*TFIIB38-140-mCherry

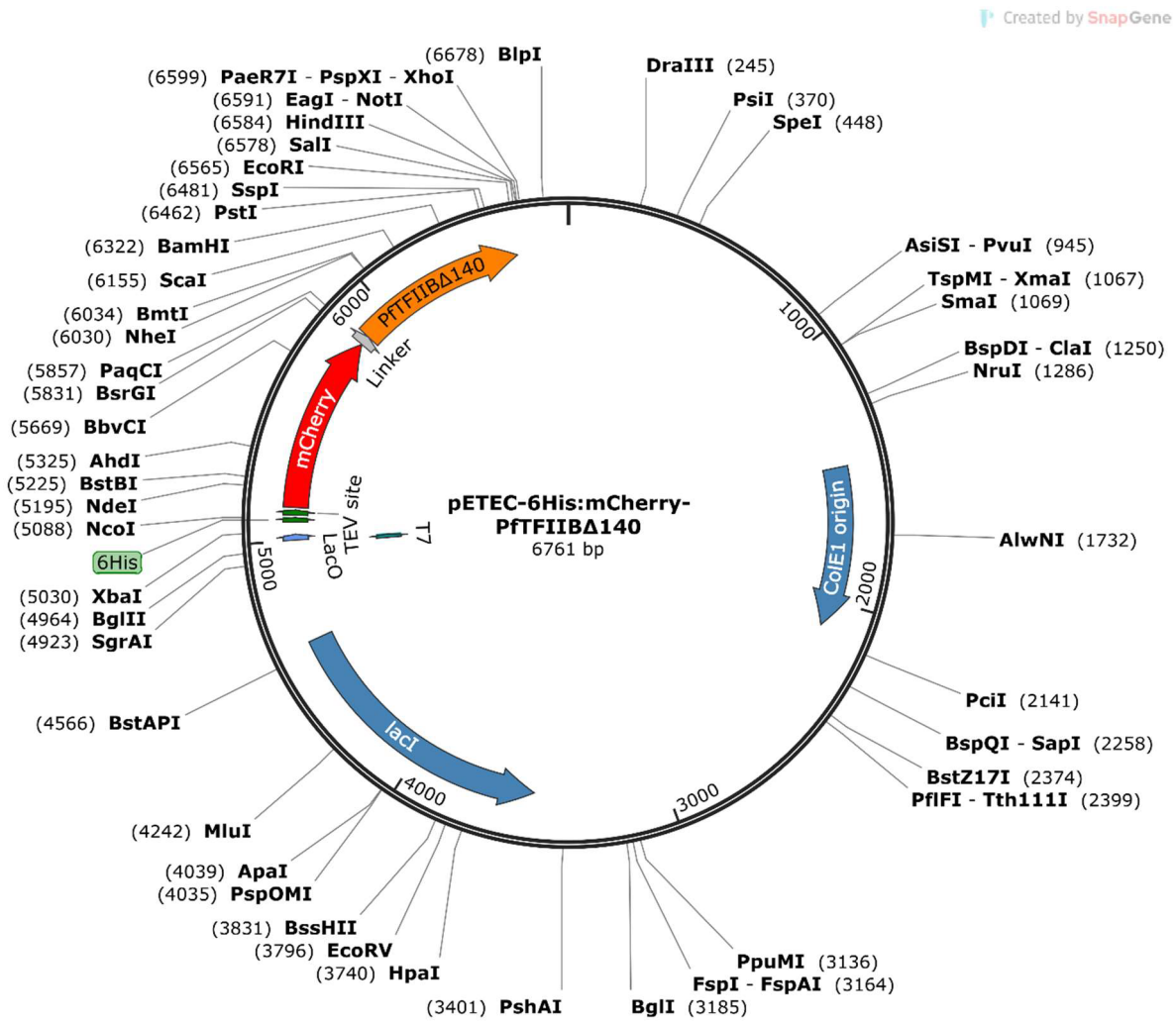


Fig S19. Plasmid map of pET11d-6His:mCherry-*Pf*TFIIBΔ140

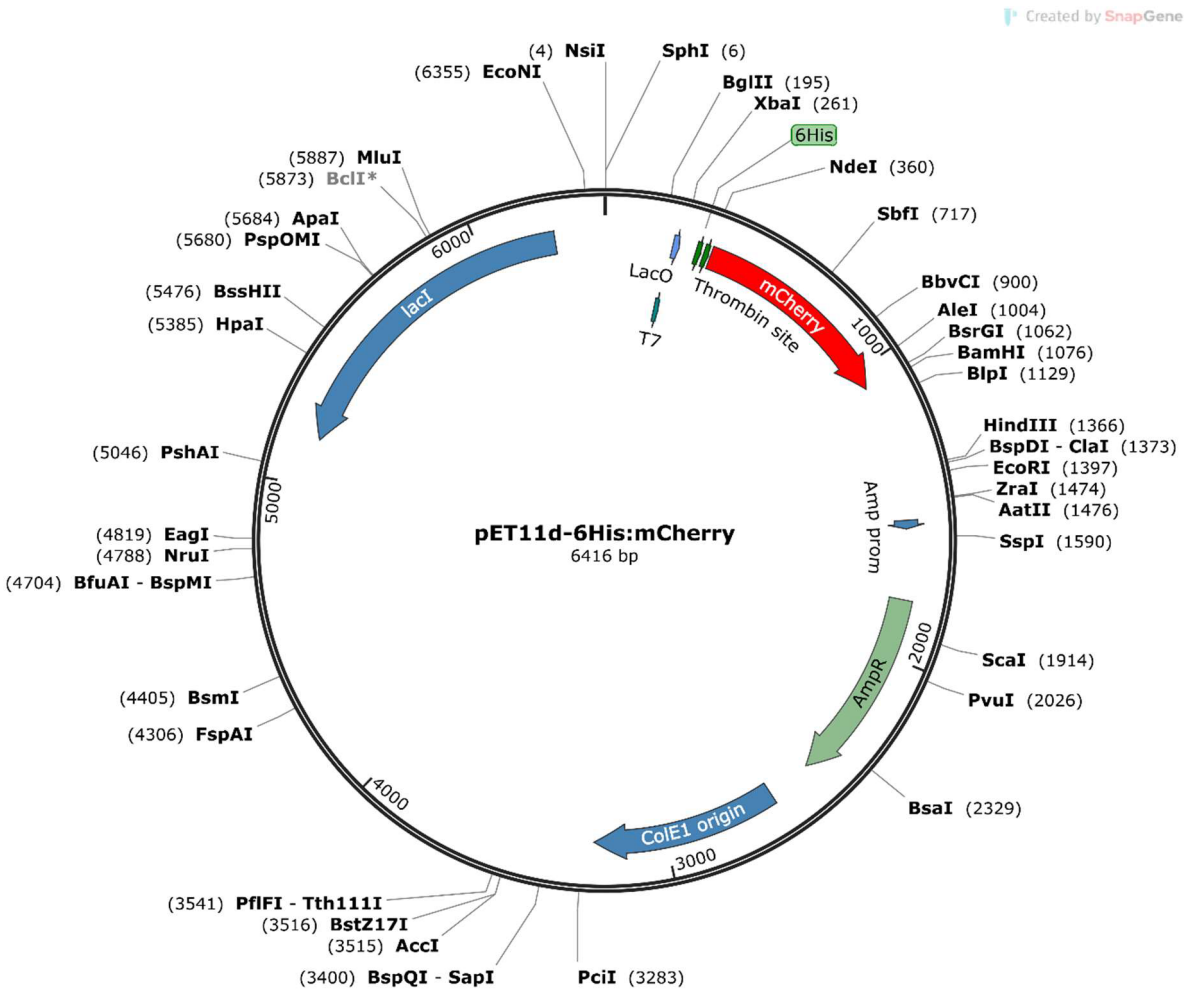


Fig S20. Plasmid map of pET11d-6His:mCherry

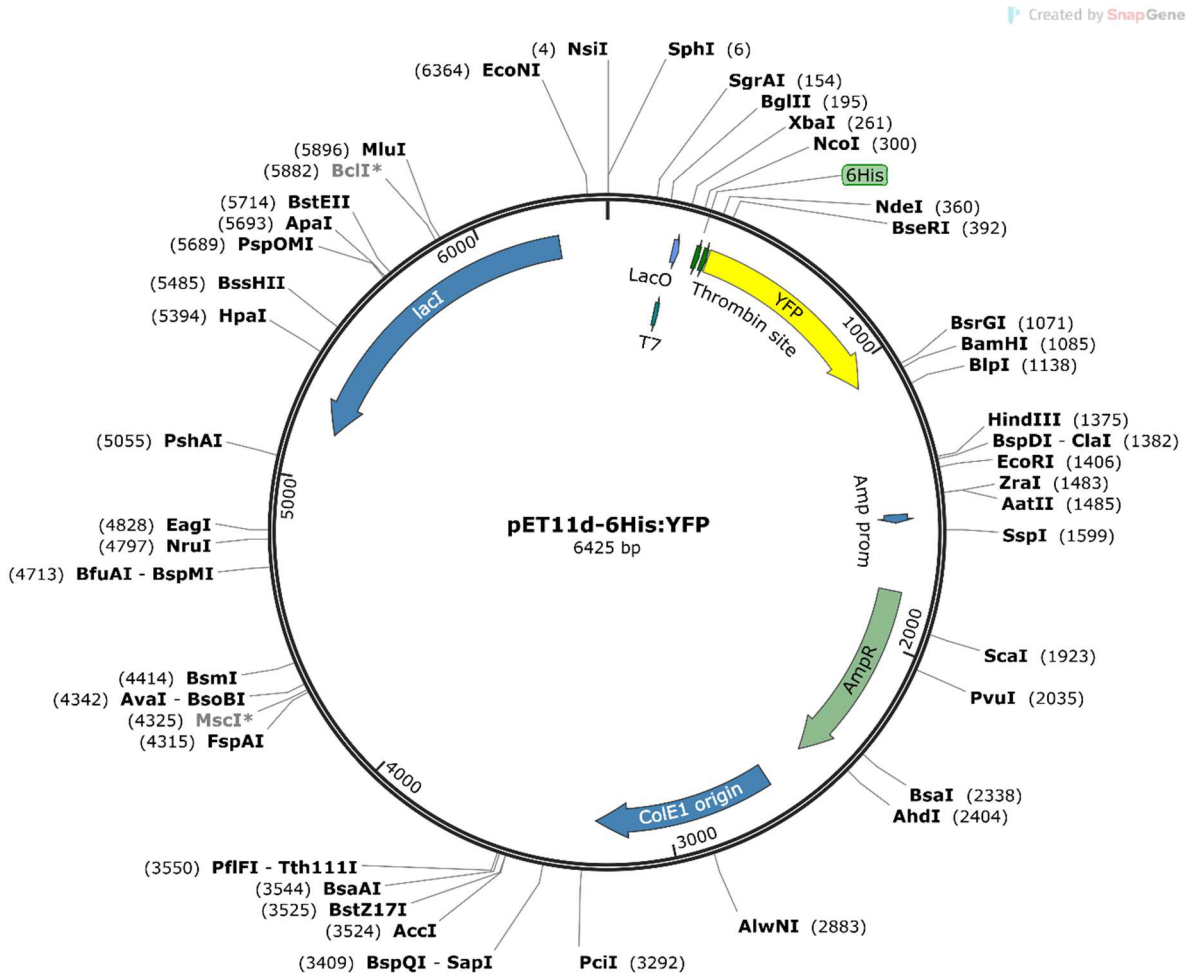


Fig S21. Plasmid map of pET11d-6His:YFP