Functional Characterization of *Plasmodium falciparum* TATA-box binding-like Protein (*Pf*TLP)

Lize van der Linden

Dissertation presented for the degree of

Master of Science

Department of Molecular and Cell Biology

Science Faculty

University of Cape Town

September 2019

Supervised by

Dr Thomas Oelgeschläger

Declaration

- I am presenting this dissertation in full fulfilment of the requirements for my degree.
- I know the meaning of plagiarism and declare that all of the work in the dissertation, save for that which is properly acknowledged, is my own.

Signed Signed by candidate

Acknowledgements

Thank you to the University of Cape Town (UCT) and the National Research Foundation (NRF) for financial support in the form of grant money and scholarships. The opinions and conclusions reached in this study are those of the author and is not attributed to UCT or the NRF.

Thank you to my supervisor Dr Thomas Oelgeschläger for all the help, support and guidance over the last few years.

Thank you to all my lab mates (Jasmin, Kim, Keso, Leo and Joanna) for the help, support, welcomed distractions and banter.

Lastly, thank you to all my friends from the Department of Molecular and Cell Biology for listening to me complain about my failed EMSAs and for always offering help and support.

Abstract

Plasmodium falciparum, the deadliest strain of human malaria, affected 200 million people and resulted in several hundred thousand deaths in 2017 (World Health Organization, 2018). A better understanding of the mechanisms of *P. falciparum* gene regulation can open novel avenues for the development of much needed new drugs. A key step in eukaryotic gene regulation is the process of transcription, which is largely uncharacterized in *Plasmodium*. Bioinformatic analysis identified putatuve *P. falciparum* orthologues of RNA polymerase II general transcription factors (Bing, 2014; Milton, 2017), including a TATA box-binding-like protein, *Pf*TLP. Bioinformatic analysis suggested that *Pf*TLP is a TRF2-type TBP-like protein. However, *Pf*TLP differs in several aspects from previously characterized TRF2-type proteins. These differences are thought to be *Plasmodium* specific adaptations to the parasite's intricate life cycle and AT-rich genome. This study investigates two *Plasmodium*-specific features of *Pf*TLP. Firstly, DNA binding by eukaryotic TATA-box binding protein (TBP) is mediated by four evolutionary conserved phenylalanine residues, two of which intercalate into the DNA. These residues are absent in previously characterized TRF2-type TLPs, and consistent with this, these proteins lack detectable DNA binding activity (Duttke et al., 2014). In contrast, *Pf*TLP, a TRF2-type TLP, has DNA binding activity, and all four of the DNA binding phenylalanine residues are conserved (Bing, 2014; Milton, 2017). The importance of evolutionary conserved intercalating phenylalanine residues F60 and F283 was investigated by generating mutant *Pf*TLP proteins, carrying alanine substitutions, and analysing their DNA-binding properties. The results suggest that while both phenylalanine residues are important for *Pf*TLP DNA-binding, only F60 is critical for stabilization of *Pf*TLP/DNA complexes. Secondly, *Pf*TLP possesses two low-complexity or intrinsically disordered regions (LCR1 and 2), which are absent in TLPs from model eukaryotes. These regions are located at the same positions within the two quasi-symmetrical repeats of the TLP core structure and show a non-random compositional bias towards a limited set of amino acids. A growing body of evidence supports the idea that low complexity or intrinsically disordered proteins mediate liquid-liquid phase separation (LLPS) (Alberti et al., 2019; Brangwynne et al., 2009; Elbaum-Garfinkle et al., 2015; Nott et al., 2015). Bioinformatic analysis revealed that *Pf*TLP LCRs are enriched in asparagine and lysine, and that these regions are well conserved throughout *Plasmodium* TLPs. *Pf*TLP LCRs were fused to fluorescent proteins and the fusion proteins were functionally characterized in liquid-liquid phase separating assays. The results demonstrate that *Pf*TLP LCR1 is capable of mediating LLPS, at least under certain conditions *in vitro*.

List of Abbreviations

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

A2ML - adenovirus 2 major late

ADP - adenosine diphosphate

AMP- ampicillin

ApiAP2 -Apicomplexan AP2

bp –base pairs

BRE$^{u/d}$ -upstream/downstream B-recognition element

BSA -bovine serum albumin

C-;N-terminus –carboxyl-; amino-terminus

CAM – chloramphenicol

CFP – cyan fluorescent protein

CTD – C-terminal domain

DIC - differential interference contrast microscopy

DNA -deoxyribonucleic acid

dNTP -deoxyribose nucleoside triphosphates

DPE -downstream promoter element

DREF – DNA-replication related element-binding factor

DTT -Dithiothreitol

E. coli-*Escherichia coli*

Ec - *Encephalitozoon cuniculi*

EDTA -Ethylenediaminetetraacetic acid

EMSA -electrophoretic mobility shift assay

F – phenylalanine

FUS - Fused in Sarcoma protein

GBP -*glycophorin-binding protein*

GBP-130 - glycophorin binding protein 130

GFP – green fluorescent protein

GTF -general transcription factor

HEPES -4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

hnRNPA2- Heterogeneous nuclear ribonucleoproteins A2

HRP -horseradish peroxidase

IDR – intrinsically disordered regions

Inr –initiator

IPTG -Isopropyl β-D-1-thiogalactopyranoside

ITA – immobilized template assays

KAHRP - knob associated histone rich protein

KAHRP -knob-associated histidine rich protein

L/UCST – lower/upper critical solution temperature

LB – lysogeny broth

LCR - low complexity region

LLPS – liquid-liquid phase separation

Mj - *Methanococcus jannaschii*

mRNA -messenger RNA

MSA -Multiple sequence alignments

MTE -motif ten element

MW -molecular weight

NP-40 -Nonidet™ P 40 Substitute (Sigma-Aldrich®)

NURF – nucleosome remodelling factor

OCT4 - octamer-binding transcription factor 4

OD -optical density

OD_{600} – optical density at 600 nm

ORF - open reading frame

P. falciparum*-Plasmodium falciparum*

Pab1 - Polyadenylate binding protein 1

PAGE – Poly acrylamide gel electrophoresis

Pca *- Plasmodium chabaudi adami*

PCNA – proliferating cell nuclear antigen

PCR -polymerase chain reaction

Pf *- Plasmodium falciparum*

Pf**TBP** -*Plasmodium falciparum* TATA-box binding protein

Pf**TLP** -Plasmodium falciparum TBP-like protein

Pg *- Plasmodium gaboni*

pI – isoelectric point

PIC -pre-initiation complex

Pm *- Plasmodium malariae*

PML - Promyelocytic leukemia protein

Po *- Plasmodium ovale,*

Pr *- Plasmodium reichenowi*

Prel *- Plasmodium relictum*

PrLD – prion-like domains

PTM – post-translational modifications

Pv *- Plasmodium vivax*

PVDF -polyvinylidene difluoride

Pw *- Pyrococcus woesei*

Py *- Plasmodium yoelii*

RBC -red blood cells

RNA - ribonucleic acid

RNAPII -RNA polymerase II

rRNA -ribosomal RNA

RT – room temperature

S. cerevisiae *-Saccharomyces cerevisiae*

Sc *- Saccharomyces cerevisiae*

SDS -sodium dodecylsulphate

SDS-PAGE -SDS polyacrylamide gel electrophoresis

SLiMs – small linear motifs

SOC -Super Optimal broth with Catabolite repression

SP-Sepharose FF® -sulphopropyl sepharose fast flow® (Sigma-Aldrich)

SRBs – suppressors of RNA Polymerase B mutations

Sup35 - *Saccharomyces cerevisiae* eukaryotic translation release factor 3

SWI/SNF - SWItch/Sucrose Non-Fermentable

TAE -tris base, acetic acid and EDTA

TAF -TBP-associated factor

TB – Tris base and boric acid

TBE -tris base, boric acid and EDTA

TBP -TATA-box binding protein

TDP-43 - TAR DNA-binding protein 43

TE -Tris-EDTA

TF -transcription factor

TFII (A, B, D, E, F, H) - transcription factor II A, C, D, E, F, H

TLP -TBP-like protein

TRF -TBP-related factor

tRNA -transfer RNA

TSS -transcription start site

YFP- yellow fluorescent protein

Table of Contents

Chapter 1

Introduction

1.1 *Plasmodium falciparum*

1.1.1 Impact

Human malaria is caused by five species of *Plasmodium*: *P. vivax, P. falciparum, P. ovale P. malariae*, and *P. knowlesi*. Out of these species, the majority of malaria cases are caused by *P. falciparum*, which is predominantly found in Africa. *P. falciparum* is also the most severe form of malaria, with complications such as severe anaemia, cerebral malaria, coma, pulmonary edema, renal failure, ruptured spleen, and lactic acidosis (Hermansyah et al., 2017; Idro et al., 2005; Sheehy and Reba, 1967; Trampuz et al., 2003).

World-wide there were 219 million cases of malaria in 2017, with 91 % of cases caused by *P. falciparum* in the African region (World Health Organization, 2018). Children under the age of five are the most vulnerable group, and account for roughly 61 % of the over 400, 000 malaria-related deaths. Despite massive global efforts, no significant progress was made towards the reduction of malaria cases between 2015 and 2017 (World Health Organization, 2018). The challenges preventing the reduction of malaria include: lack of funding, drug and insecticide resistance, and lack of effective vaccines. The countries with the highest malaria burden are generally poor third-world countries who rely on external funding for the anti-malaria programs and research. Multi-drug resistance, to both artemisinin and partner drugs, has been found in *P. falciparum* in the Greater Mekong Region (World Health Organization, 2018). Multi-insecticide resistant vectors have been found in all malaria-stricken regions world-wide. Efforts to produce a viable vaccine have been met with serious challenges, mostly due to the great antigenic variation seen in *Plasmodium* (Cowman et al., 2012; World Health Organization, 2018).

1.1.2 Life Cycle

Plasmodium forms part of the phylum *Apicomplexa* which includes parasitic protists that invade hosts using a specialized apical complex organelle (Aravind et al., 2003; Foth and McFadden, 2003). Most *Apicomplexa* have multiple hosts, and intricate life cycles which includes both sexual and asexual stages. *P. falciparum* infection of humans begins with the bite of an infected *Anopheles* mosquito, which injects small haploid sporozoites, present in the mosquito saliva, into the bloodstream. These infectious sporozoites travel to the liver where they invade hepatocytes, and undergo multiple asexual fission (schizogony) to produce exoerythrocytic merozoites, which are released into the bloodstream (Alano, 2007; Cowman et al., 2012; Willey et al.), where the merozoites will invade uninfected eryhtrocytes. Once inside the erythrocyte the plasmodial cell grows and develops into a trophozoite. The trophozoite's nucleus divides asexually to produce a mature schizont which has multiple nuclei, the schizont then divides to produce mononucleated merozoites, which are released when the erythrocyte bursts. The merozoites go on to infect new red blood cells (Cowman et al., 2012; Crabb et al., 1996; Willey et al.). The erythocytic stage of the infection is cyclic and highly synchronized and repeats, in case of *P. falciparum*, every 48 hours. The release of toxins, merozoites and erythrocytic debris triggers the clinical symptoms of malaria infection (Willey et al.). During the erythrocytic asexual blood stage parasites may develop, at low frequency, from the ring stage into male and female gametocytes, which are instead ingested up by the feeding *Anopheles* mosquito (Alano, 2007; Cowman et al., 2012). In the mosquito gut gametocytes mature to form male and female gametes. The gametes fuse to form a fertilized diploid zygote, which differentiates to form a motile ookinete, which develops into an oocyst in the mosquito gut. The oocyst undergoes meiosis (sporogony) to form sporozoites, which travel to the mosquito salivary glands (Alano, 2007; Aravind et al., 2003; Cowman et al., 2012; Crabb et al., 1996; Willey et al.). The infection cycle continues when the mosquito feeds on another human, injecting sporozoite carrying saliva into the new host (Fig. 1).

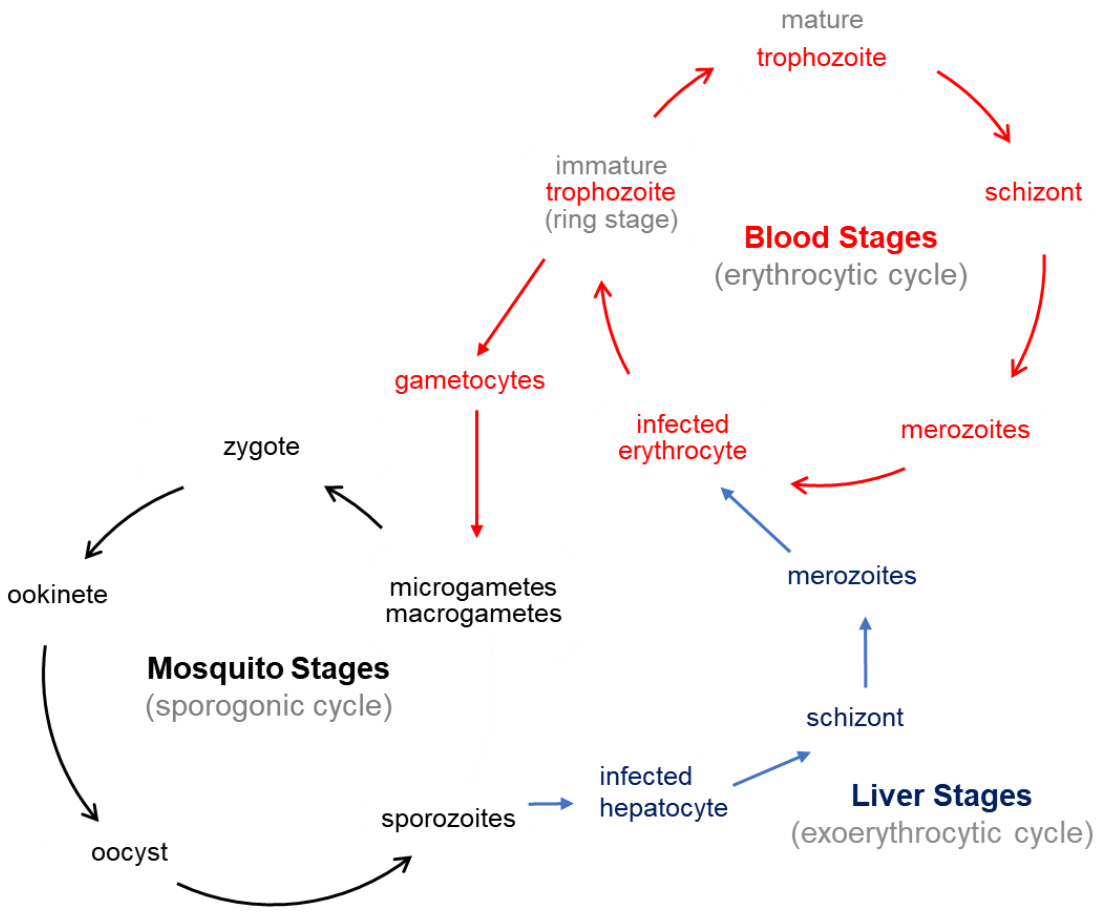


Figure 1: Schematic of the Life Cycle of *Plasmodium falciparum*

1.1.3 Gene Regulation

In 2002 the genome of *Plasmodium falciparum* was sequenced and annotated (Gardner et al., 2002). The 23-megabase nuclear genome is made of 14 chromosomes hosting predicted 5300 genes. Strikingly, the genome consists of roughly 80% A/T bases. Proteomic and transcriptomic studies established that the parasite sustains its intricate life cycle through a tightly regulated gene expression program (Bozdech et al., 2003; Florens et al., 2002; Horrocks et al., 2009). During the intra-erythrocytic cycle in the human host, over 80% of *P.falciparum's* genome is periodically regulated, in synchrony with the different developmental stages, as a continuous and highly coordinated cascade of gene expression. The repeating gene expression profile only differentiates when the parasites enters gametogenesis at low frequency (Bozdech et al., 2003). It was found that, during the gene expression program,

gene clusters of proteins involved in similar cellular functions are co-expressed (Florens et al., 2002; Horrocks et al., 2009; Le Roch, 2003). Furthermore, mRNAs from a particular cluster accumulate just prior to the cellular demands of a specific developmental stage. This observation led to the "just in time" hypothesis, which, assuming that translation immediately followed transcription, suggested that gene regulation occurred predominantly on at the level of transcription (Horrocks et al., 2009). However, later studies comparing mRNA and protein levels during various developmental stages revealed extensive post-transcriptional gene regulation, thereby opposing the "just-in-time" hypothesis (Horrocks et al., 2009; Le Roch et al., 2004).

Despite the differences in gene regulation between *Plasmodium* and model systems, there is evidence that gene regulation generally follows the principles established for model eukaryotes. Firstly, gene transcription is monocistronic and regulated by promoter regions and chromatin structure, (Lanzer et al., 1992b, 1992a). Secondly, the structure of transcribed mRNA resembles that of model eukaryotes, with a 5' cap, a polyadenylated 3' end and prototypical splice sites at intron/exon junctions. (Bischoff and Vaquero, 2010; Coleman and Duraisingh, 2008)

Recent genome-wide mapping of transcription start sites provided first insights into the dynamics of RNA polymerase II transcription initiation during the *P. falciparum* blood cycle in the human host (Adjalley et al., 2016). Transcription start sites (TSS) were found in closely spaced "clusters", over a wide area upstream of the coding region, positioned less than 1000 bp from the start codon. The paper found no evidence or signatures of a typical TATA box core promoter element directing transcription initiation 30 bp downstream (Adjalley et al., 2016). This finding, together with the observed diffuse TSS pattern might suggest two possibilities, which are not mutually exclusive. First, the RNA polymerase II pre-initiation complex (PIC) might be randomly positioned upstream of the gene coding region and initiates transcription about 30 bp downstream. Alternatively, specific core promoter regions that direct assembly of the PIC may exist but transcription is initiated at random positions at varied distances from the PIC, similar to transcription initiation in *S. cerevisiae* (Adjalley et al., 2016)..

Initially, the majority of transcription factors (TF), including general transcription factors, could not be identified in *P. falciparum* through conventional sequence homology searches (Gardner et al., 2002). This led to the assumption that there was a shortage of TFs in

P. falciparum when compared to other model eukaryotes. This suggested that a large component of gene regulation had to be post-transcriptional or epigenetic (Coulson et al., 2004). However, more *P. falciparum* TFs, including a proportion of the general transcription factors, were later identified *in silico* using a combination of profile-based searches and Hydrophobic Cluster Analysis (Callebaut et al., 2005). Authors identified putative subunits of TFIIA, TFIIB, TFIIH, TFIIE and the TFIID subunits TAF1, TAF2, TAF7, and TAF10. Interestingly, TFIID subunits (TAFs) containing histone fold domains could not be identified, suggesting the possible existence of a unique TFIID complex architecture in *P. falciparum* (Callebaut et al., 2005). Another Apicomplexan-specific class of transcription factors, ApiAP2, were also identified through sensitive sequence profile searches. The members of this class of transcription regulatory proteins are characterised by the presence of at least one AP2 (Apetala 2) intergrase DNA binding domain, found in many plants. These TFs are found to be differentially expressed during the various parasite life stages (Balaji et al., 2005). Some members of this class of transcription factors has been well characterized both *in vitro*, and in cultured *P. falciparum* cells. The ApiAP2s class of transcription factors have been extensively reviewed (Jeninga et al., 2019; Llinás et al., 2008). Despite these discoveries, *P. falciparum* still appears to contain far fewer TFs than model eukaryotes. This might be because existing TFs diverged too far from established eukaryotic TFs, and can therefore not be detected using standard sequence-based methods. Alternatively, *P. falciparum* might function with a smaller set of TFs compared to other eukaryotes.

Different to the apparent paucity of TFs in *P. falciparum*, a large number of chromatin-structuring and remodelling proteins, as well as histone proteins were identified in *P. falciparum*. Interestingly, *P. falciparum* possesses homologs for the canonical histone proteins H2A, H2B, H3 and H4, and the variant histones H2A.Z, H2Bv, H3.3 and CenH3 (Coleman and Duraisingh, 2008; Cui and Miao, 2010) but lacks the gene for histone H1, which is important for higher-order nuclear organization and compaction (Cui and Miao, 2010) in eukaryotic model organisms. A large array of histone-modification enzymes and other chromatin-associated proteins have also been identified in *Plasmodium* (Coleman and Duraisingh, 2008). Furthermore, *P. falciparum* possesses a gene for a putative DNA methyltransferase, although there is no evidence for DNA methylation. (Cui and Miao, 2010). Given the shortage of TFs and the abundance of chromatin-modifying factors, a large portion

of the gene regulation field focuses on chromatin modification and remodelling and epigenetics as methods of gene regulation (reviewed in Bischoff and Vaquero, 2010; Coleman and Duraisingh, 2008; Cui and Miao, 2010).

1.1.4 *Plasmodium falciparum* Low-Complexity Regions

Around 87% of *P. falciparum* genes contain one or more low-complexity regions (LCRs) (DePristo et al., 2006; Zilversmit et al., 2010). LCRs represent protein sequences with non-random compositional bias towards a limited number of amino acids, and are usually intrinsically disordered non-globular domains (Zilversmit et al., 2010). LCRs within protein sequences can be computationally identified using the SEG algorithm developed by Wootton and Federhen (Wootton and Federhen, 1993).

An initial bioinformatics analysis of the *P. falciparum* genome identified LCRs (*Pf*LCRs) using the SEG algorithm as well as by sequence alignment of putative *P. falciparum* proteins with eukaryotic homologs in chromosomes 2 and 3, where LCRs are identified as long stretches of amino acid insertions between globular domains (Aravind et al., 2003; Pizzi and Frontali, 2001). 90% of *Pf*LCRs were found to be hydrophilic and represent rapidly evolving nonglobular domains with a possible function in antigenic variation and host immune system evasion (Brocchieri, 2001; Pizzi and Frontali, 2001). These hydrophilic LCRs were preferentially enriched in asparagine residues (Aravind et al., 2003; Brocchieri, 2001; Pizzi and Frontali, 2001). The amino acid bias in many *P. falciparum* LCRs appears to be the result of the extremely skewed genome composition towards A/T content and likely emerged and diverged through replication slippage and recombination events (Aravind et al., 2003; DePristo et al., 2006; Xue and Forsdyke, 2003; Zilversmit et al., 2010).

Subsequent genome-wide studies identified three types of *P.falciparum* LCRs (Zilversmit et al., 2010). Firstly, heterogeneous *Pf*LCRs, encoded by highly A/T-rich genome sequences, which appear to be aperiodic and slow evolving. Secondly, A/T-rich genomic sequences that contain multiple trinucleotide repeats coding for the amino acid asparagine, known as poly N sequences. Finally, LCRs stemming from genomic sequences with high G/C content, created through frequent recombination events (Aravind et al., 2003; Zilversmit et al., 2010).

The abundance of asparagine in *Pf*LCRs strongly suggested the presence of prion-like domains in these intrinsically disordered domains. Prion-like domains (PrLDs) are amino acid

sequences enriched in glutamine (Q) and asparagine (N) with set sequential and structural characteristics that confer the ability to adopt different conformational states and the propensity to form amyloids (Pallarès et al., 2018). PrLDs were found in roughly 10% of the *P. falciparum* genome (Pallarès et al., 2018). Because cells generally select against proteins with aggregation-prone sequences, the conservation of PrLDs suggest they might have important functional roles (Pallarès et al., 2018).

1.2 Eukaryotic Transcription

Transcription initiation is regulated through various means and mechanisms. Briefly, chromatin structure is modified to create nucleosome-free promoters and enhancer DNA sequences, which will interact with transcription factors, activators, and coactivators to generate an active transcriptional state (Cramer, 2019; Haberle and Stark, 2018).

Chromatin structure needs to be manipulated so that active promoters are accessible to the transcription apparatus, and this is achieved by the removal or shifting of nucleosomes. The accessibility of a given DNA sequence is dependent on the action of chromatin remodelling factors, as well as the readout of histone modifications and DNA methylation (Cramer, 2019; Haberle and Stark, 2018). The accessible promoter sequences are situated in the vicinity of the transcription start sites (TSS) and interact with the basal transcription machinery to form the preinitiation complex (PIC), which allows transcription to start. Promoter sequences often contain specific DNA sequence elements, called core promoter elements, which are recognized by the basal, or general, transcription factors (GTFs). Some promoters lack defined sequence elements, suggesting promoter recognition by transcription factors occurs through indirect readout, whereby proteins recognize physical properties of DNA sequences, such as bendability. Alternatively, binding sites may be demarcated by nucleosome-free regions flanked by positioned nucleosomes, such as +1 nucleosome found immediately upstream of the transcription start site (Cramer, 2019; Haberle and Stark, 2018). Distal (enhancer) regulatory sequences modulate the activity of the core promoter and basal transcription factors and stimulate the assembly of the PIC at the transcription start site. These enhancer sequences contain binding sites for sequence-specific transcription regulatory proteins, such as repressors and activators. The function of activators are mediated by coactivators such as the Mediator complex, which links transcription factors with the PIC. Enhancers communicate

with target genes through proximity, which is governed by chromatin architecture (Cramer, 2019; Haberle and Stark, 2018). Some TFs are referred to as pioneer factors, based on their ability to bind to nucleosomal DNA and to recruit histone modifying and chromatin remodelling factors, which in turn remodel chromatin architecture to make promoter sequence accessible (Cramer, 2019; Haberle and Stark, 2018). Together, these means of gene regulation converge on the formation of the preinitiation complex (PIC) by the general transcription machinery (GTFs), which is the ultimate target for gene expression pathways (Haberle and Stark, 2018).

1.2.1 RNA Polymerase

The process of transcription is dependent on the enzyme RNA polymerase, which converts information from DNA into RNA. In eukaryotes four classes of RNA polymerase have been identified: RNA polymerases I, II, III, and IV. Only one RNA polymerase has been found in prokaryotes. The RNA polymerases in eukaryotes have different biochemical properties, cellular localizations, and mostly function in a nonredundant manner (Roeder and Rutter 1970; Thomas and Chiang 2006). RNA polymerase I is found in the nucleoli and synthesizes 18S and 28S rRNA. RNA polymerase II transcribes mostly mRNA and is found in the nucleoplasm together with RNA polymerase III, which transcribes 5S rRNA and tRNA. The fourth RNA polymerase was identified in plants and is in control of the synthesis of siRNA (Thomas and Chiang 2006). RNA polymerases lack sequence-specific DNA binding activity and require general transcription factors (GTFs) to recognize and bind promoter sequences.

1.2.2 RNA Polymerase II general transcription factors

General transcription factors function to recruit RNAP-II to the core promoter and to assemble with RNAPII to form a so-called preinitation complex (PIC), required for transcription initiation. They include transcription factors II A (TFIIA), -IIB, -IID, -IIE, -IIF, and TFIIH. The PIC may form in either a sequential or two -step manner. In the sequential step-wise assembly the promoter region is recognized and bound by TFIID, and the PIC is formed by step-wise association of TFIIA, TFIIB, RNAP-II, TFIIF, TFIIE and TFIIH. In the two-step assembly model a complex of TFIID and TFIIA will bind the promoter region and then recruit a pre-assembled RNAPII holoenzyme containing the remaining GTFs required for transcription initiation (Akhtar and Veenstra, 2011; Thomas and Chiang, 2006; Zehavi et al., 2015).

1.2.3 Core Promoters

Core promoters span a region of roughly 100 bp around a TSS. Core promoters are generally nucleosome free and accessible to the transcription machinery, and contain specific sequence elements. These core promoter sequence elements are recognized by components of the general RNAPII transcription machinery to recruit and position the RNAPII transcription initiation complex (PIC) and to determine the location of the transcription start site (reviewed in Haberle and Stark, 2018; Kadonaga, 2012; Smale and Kadonaga, 2003). Several core promoter elements have been identified in model eukaryotes, such as:

The TATA-box element, which is found in a small subset of genes, has the consensus sequence of TATA(A/T)A(A/T)(A/G), and is recognized and bound by the TATA-box binding protein (TBP), which forms part of the TFIID complex (Haberle and Stark, 2018). The TATA-box is usually found between nucleotide positions -31 and -24 upstream of the TSS. The binding of TBP to the TATA-box element is the initial step in the sequential PIC formation (Thomas and Chiang, 2006). Additional core promoter elements that have been identified in eukaryotes include the upstream and downstream TFIIB Recognition Element (BRE^U/BRE^D), the initiator (Inr) sequence, the downstream promoter element (DPE), the motif ten element (MTE) the downstream core element (DCE). Core promoters are complex in that they contain several core promoter sequence elements, in various combinations, and that there is a large degree of structural diversity between core promoters. Importantly, none of the core promoter elements identified to date is universally required for transcription initiation. Finally, core promoter regions of many genes appear not to contain any of the known core promoter elements, suggesting that there might be additional sequence elements yet to be discovered (Haberle and Stark, 2018; Lee et al., 2005; Thomas and Chiang, 2006; Vo ngoc et al., 2017).

1.2.4 TATA-box Binding Protein

The TATA-box binding protein (TBP) is a key component of the TFIID complex and is responsible for recognizing and binding the TATA box promoter element. TBP is well conserved from archaea to man and is the founding member of a class of paralogue eukaryotic proteins.

The structure of TBP can divided into two parts, a conserved core C-terminal domain and a variable N-terminal domain. The C-terminal core structure adopts a saddle-like shape, made up of two pseudo-symmetric halves. The core domain consists of four α-helices which connect the two halves, and ten β-strands that make contact with the DNA (Fig. 2; Kim and Burley, 1994; Kim et al., 1993; Nikolov et al., 1992, 1996, 2002). The crystal structures of human, yeast, and *Arabidopsis* TBP bound to the TATA box revealed the DNA binding mechanism of TBP. TBP unwinds the DNA as the concave part of the saddle structure binds to the minor groove, interacting with the edges of the base pairs. Through a two-step induced-fit mechanism, TBP severely bends the TATA box sequence after the initial binding event. The DNA-bending caused by TBP results from the action of four conserved phenylalanine residues. Two of these residues intercalate between bases 1 and 2 and between 7 and 8 of the TATA box sequence, causing DNA kinking (Fig. 2), while the other two residues stabilize base intercalation (Kim et al., 1993; Nikolov et al., 1992, 2002). Mutational analysis of human TBP showed that substitution of these phenylalanine residues (F197, F214, F288, and 305 in *Hs*TBP) results in loss of TBP TATA box binding activity (Klejman et al., 2005; Zhao et al., 2003). Like any other DNA-binding protein, TBP also possess non-specific DNA-binding activity (Coleman and Pugh, 1995). It has been shown that the TATA-specific DNA binding activity of TBP is not required for transcription form TATA-less promoters (Martinez et al., 1995).

The convex side of the saddle shaped C-terminal domain mediates most of the TBP interactions with protein binding partners, such as TFIIB and TFIIA, and residues involved in these interactions have been shown to be well conserved (Thomas and Chiang 2006). In contrast to the C-terminal DNA-binding domain, the unstructured N-terminal portion of TBP is not conserved in sequence or length from species to species and its function is not well understood. (Akhtar and Veenstra, 2011; Thomas and Chiang, 2006; Zehavi et al., 2015).

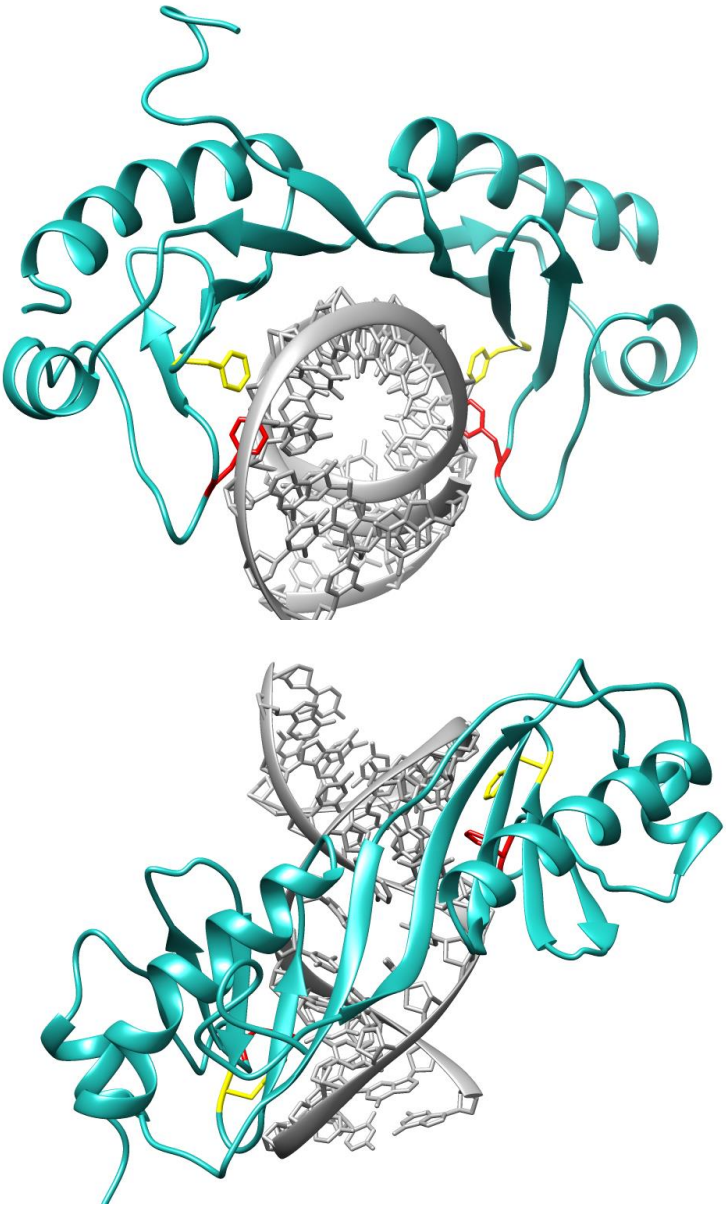


Figure 2: Crystal structure of human TBP bound to the minor groove of the TATA box promoter element. The x-ray structure model of the human TBP/TATA complex (PDB accession 1C9B) was rendered using *UCSF Chimera* (Pettersen et al., 2004). The *Hs*TBP saddle-shaped core domain is shown as a cartoon model in cyan with evolutionary conserved phenylalanine residues highlighted. Intercalating phenylalanine residues are shown in red and supporting phenylalanine residues in yellow. The TATA box DNA in the complex is shown in grey. The model in the lower panel is rotated by 90˚ about the horizontal axis towards the viewer compared to the model shown above

1.2.5 TBP-like Protein (TLP)

Metazoans possess TBP paralogues, called TBP-like proteins (TLPs) or TBP-related factors (TRFs). Some of these TLPs are well characterized in model eukaryotic organisms and are thought to be involved in cell type-specific transcription and in development and differentiation.

TRF1 was the first TLP to be identified (Holmes and Tjian, 2000), and is specifically found in insects, where it is required for RNA polymerase III transcription from most (Takada et al., 2000) but not all (Verma et al., 2013) promoters.

TFR3-type TLPs are found in vertebrates and are most closely related to TBP, sharing roughly 93% amino acid sequence homology. TRF3 has been shown to interact with both TFIIB and TFIIA, to have TATA box-binding activity, and to nucleate the RNA Polymerase II PIC (Akhtar and Veenstra, 2011; Bártfai et al., 2004; Zehavi et al., 2015). The precise function of TRF3-type TLPs has not been fully elucidated, but they appear to play species-specific roles in cell differentiation and development. In *Xenopus laevis*, TRF3 is predominantly expressed in oocytes, which appear to contain very little to no TBP, and is responsible for a large subset of vital genes during embryogenesis (Jallow et al., 2004). Consistent with this, knock-out of TRF3 in mice results in female sterility, due to the loss of transcription from oocyte-specific genes (Gazdag et al., 2009). In zebrafish, TRF3 knockout results in a phenotype consistent with mis-regulation of embryonic patterning (Bártfai et al., 2004).

To date, TRF2-type TLPs are by far the best characterized TLPs. TRF2s share about 60% homology to TBP and are found throughout multicellular organisms, in every cell type. TRF2s are predicted to adopt the same saddle-shaped core domain as TBP. While amino acid residues mediating interactions with TFIIA and TFIIB appear to be conserved, TRF2-type TLPs contain substitutions at amino acid positions critical for TBP DNA-binding activity including the critical phenylalanine residues involved in TATA box binding. Consistent with this, TRF2s appear to lack sequence-specific DNA-binding activity (Akhtar and Veenstra, 2011; Berk, 2000; Duttke et al., 2014; Thomas and Chiang, 2006; Wang et al., 2007; Zehavi et al., 2015).

TRF2 has different functions to that of TBP, and nucleate the PIC on promoters distinct from those of TBP (Akhtar and Veenstra, 2011; Berk, 2000; Zehavi et al., 2015). For example, TRF2

was shown to be involved in the regulation of TATA-less G/C-rich promoters of ribosomal proteins (Isogai et al., 2007; Wang et al., 2007). Additionally, in *Drosophila* TRF2 mediates transcription from DPE-containing promoters (Kedmi et al., 2014).

TRF2 has, in some cases, been shown to regulate proteins involved in chromatin structure. For example, it has been shown to regulate the histone H1 TATA-less promoters. Additionally, in *Drosophila,* TRF2 co-purified with the nucleosome remodelling factor (NURF) and the DNA-replication related element-binding factor (DREF), and directs core promoter recognition of the proliferating cell nuclear antigen (PCNA) gene via a TRF2-DREF complex (Hochheimer et al., 2002). Furthermore, TRF2 depletion resulted in structural defects of polytene chromosomes (Isogai et al., 2007).

Multiple studies in eukaryotic model organisms have shown that TRF2 might be important for embryogenesis. For example, in *Caenorhabditis elegans*, TRF2 is expressed at similar levels as TBP, and is responsible for the regulation of a large subset of genes involved in embryonic development, such as early patterning genes (Dantonel et al., 2000; Kaltenbach et al., 2000). Furthermore, in *Xenopus laevis,* TRF2 is important for the transcription of a subset of genes involved in embryo catabolism (Jacobi et al., 2007; Veenstra et al., 2000). In zebrafish, depletion of TRF2 in embryos resulted in downregulation of specific regulatory genes and gastrulation failure (Müller et al., 2001). Mutational studies suggested that TRF2 also plays an important role in *Drosophila* embryogenesis (Kopytova et al., 2006). In contrast, TRF2 is not required for embryonic development in mice (Akhtar and Veenstra, 2011; Zehavi et al., 2015).

TRF2 is also important for cell type specific differentiation. For example, in mice a TRF2 deficiency had no effect on viability, but the male mice were sterile, due to an arrest of late spermiogenesis (Martianov et al., 2002; Zhang et al., 2001). Furthermore, in *Drosophila,* TLP mutations led a reduction of gene expression of ecdysone-controlled genes, which are essential for proper metamorphosis (Bashirullah et al., 2007) .

1.3 Liquid-Liquid Phase Separation

1.3.1 Membraneless Organelles

Cells can organize reactions and processes using membrane bound organelles, such as the mitochondria or the nucleus. However, cells can also form membraneless organelles lacking a delimiting membrane, such as the nucleolus, stress granules, P-granules, Cajal bodies, and the centriole (Boeynaems et al., 2018). A landmark paper by Brangwynne et al (2009) showed that P-granules have liquid like properties; they are spherical droplets capable of fusing, dripping, deforming under shear stress, and show fast internal reorganization (Brangwynne et al., 2009; Gomes and Shorter, 2019). This discovery led to a large body of study that established that membraneless organelles can be described as biomolecular condensates that can be understood as liquid droplets that form through liquid-liquid phase separation (Boeynaems et al., 2018; Gomes and Shorter, 2019).

1.3.2 Liquid-Liquid Phase Separation

Liquid-liquid phase separation (LLPS) occurs when a homogenously mixed liquid solution separates into two distinct liquid phases, or regions of space with distinct concentration of molecules. LLPS is a well-known phenomenon in polymer chemistry and physics, but has only recently been studied with complex biomolecules (Boeynaems et al., 2018).

The physics underlying LLPS has been extensively reviewed (Bergeron-Sandoval and Michnick, 2018; Boeynaems et al., 2018; Martin and Mittag, 2018; Zhou et al., 2018). Briefly, LLPS is driven by the lowering of free energy of mixing, given by the equation:

$$\Delta G_m = \Delta H_m - T\Delta S_m$$

ΔS_m represents the entropy of mixing and is described by the Flory and Huggins equation of an ideal polymer (Boeynaems et al., 2018; Martin and Mittag, 2018), which is based on the volume fractions occupied by both polymer and solvent. The larger the polymer, the lower the entropic cost of confining the polymer in a dense phase. This in turn means that an increase in polymer size will decrease the concentration at which it phase separates. ΔH_m represents the enthalpic component of mixing, and is described by the mean field theory, which is based on the comparison of the interactions between solvent-solvent, polymer-

polymer and solvent-polymer. This gives rise to an interaction parameter. Taken together, these two equations can be combined into the free energy of mixing equation to form the Flory-Huggins Expression. The expression denotes that the entropy of mixing is always positive. However, when polymer-polymer interactions are favoured, energy minima or inflection points are created within the free energy curve, which represent points at which the solution is unstable and can de-mix into polymer dense and light phases (Martin and Mittag, 2018).

The Flory-Huggins expression fails to account for sequence dependent interactions of heteropolymers. It also fails to explain more complicated entropic effects. Entropy inherently favours mixing, and this is even more favourable at higher temperatures. However, for hydrophobic polymers there is a gain of entropy at higher temperatures as polymers are compacted and solvent molecules are released. These are known as lower critical solution temperature (LCST) polymers, as they phase separate at higher temperatures due to an entropic gain. On the other hand upper critical solution temperature (UCST) polymers phase separate at lower temperatures due to molecular interactions (Martin and Mittag, 2018). Phase separation can be distinguished from aggregates on the basis of thermodynamic reversibility, wherein phase separated droplets can dissolve again based on a change in signalling parameters (Zhou et al., 2018).

Phase separation can be driven by proteins or nucleic acids that are known as “scaffolds”, in which multivalency of so-called “adhesive domains” is a key feature. Multivalency can be achieved in numerous ways. Firstly, through folded protein domains that facilitate protein-protein interactions. Secondly, through folded domains that can be linearized through linkers. Finally, through intrinsically disordered low-complexity regions (IDRs, LCRs), which are hallmarked by conformational heterogeneity and a non-random compositional bias towards a limited number of amino acids. IDRs/LCRs can contain stretches of hydrophobic, polar, charged or aromatic residues that form small linear motifs (SLiMs), charge blocks, or degenerate repeats, which act as adhesive patches that facilitate interactions (Boeynaems et al., 2018; Martin and Mittag, 2018; Wheeler and Hyman, 2018). Interactions between scaffold proteins needs to be strong enough to support a dense phase, and, at the same time, be dynamic and transient enough to form a liquid and not an aggregate. These interactions include electrostatic, dipole-dipole, π - π, cation – π and hydrophobic interactions, as well as

hydrogen bonding (Boeynaems et al., 2018; Gomes and Shorter, 2019; Martin and Mittag, 2018; Zhou et al., 2018). The differences in solution properties and behaviours of dense phases depend on the primary sequence of the scaffold. RNA can also act as a scaffold, by acting as a binding partner to proteins containing an RNA-binding domain (RBD), or through intermolecular base pairing that promotes phase separation (Boeynaems et al., 2018).

1.3.3 Regulation of Liquid-Liquid Phase Separation

Many proteins have been shown to be capable of LLPS, and cells need to tightly control this process to ensure appropriate cellular functions (Alberti, 2017; Franzmann and Alberti, 2019). LLPS is highly sensitive to environmental conditions and cellular signals, because the system functions at the hypersensitive phase boundary, where small changes can induce phase transition, which in turn can result in a system-wide response at little to no energy expenditure. This makes LLPS an appealing means of cellular organization and regulation (Alberti, 2017; Franzmann and Alberti, 2019; Yoo et al., 2019). Changes in osmotic pressure, pH, temperature, salt, and protein concentration can induce or dissolve LLPS droplets (Zhou et al., 2018). One example is the yeast protein Sup35, which undergoes LLPS in response to a decrease of cellular pH under stress (Franzmann et al., 2018; Ruff et al., 2018). Another example is the poly (A)-binding protein Pab1, which binds heat shock chaperone 3' mRNA ends. In response to elevated temperatures, Pab1 releases bound RNA and phase separates. This allows the translation of heat shock chaperones in response to heat stress. In turn, these chaperones then help to dissolve the phase separated Pab1 droplets, returning the cell to its default state (Riback et al., 2017; Ruff et al., 2018).

Post-translational modifications (PTM) also play an important role in regulating the LLPS behaviour of scaffold proteins. PTM of the IDR/LCRs changes intrinsic sequence properties and intermolecular interactions potential, thereby changing the LLPS behaviour (Boeynaems et al., 2018; Zhou et al., 2018). Multiple studies have demonstrated that changes in the phosphorylation status affect LLPS of IDRs/LCRs, and that phosphorylation of even a single serine residue in a scaffold IDR/LCR can modulate LLPS (Aumiller and Keating, 2016; Guo et al., 2019). Therefore, LLPS can be dependent on the activity levels of kinases and phosphatases involved in regulating the phosphorylation level of scaffold proteins (Li et al., 2012). Another example for regulation of LLPS by PTM is the methylation of arginine residues

in the RNA-binding protein hnRNAP2, which was shown to completely disrupt LLPS (Ryan et al., 2018). SUMOylation has also been shown to be involved in LLPS regulation, where SUMOylation of promyelocytic leukemia protein (PML) is required for the formation of nuclear bodies (Shen et al., 2006). Finally, poly (ADP) ribosylation induces LLPS of proteins like Fused in Sarcoma protein (FUS) and TAR DNA-binding protein 43 (TDP-43) (Altmeyer et al., 2015; Gomes and Shorter, 2019).

RNA is a major component of many membrane-less organelles, suggesting that it likely plays a major role in regulating LLPS. Many coding and noncoding RNAs can nucleate the formation of nuclear bodies, such as Men ϵ / β noncoding RNA, which nucleates the formation of nuclear speckles (Gomes and Shorter, 2019; Shevtsov and Dundr, 2011). Some proteins can compete with droplet-forming proteins for RNA-binding partners, thereby regulating LLPS of those proteins (Zhou et al., 2018). Furthermore, RNA has been shown to both promote and limit LLPS, as in the case of Pab1 (Boeynaems et al., 2018; Riback et al., 2017).

Finally, nuclear transport systems controls both the concentration and localization of scaffold proteins and RNA which in turn regulates LLPS and pathological protein aggregation (Boeynaems et al., 2016, 2018).

1.3.4 Functions of Membraneless Organelles

The full repertoire of the cellular functions of LLPS and membrane-less organelles is only starting to be explored. Current literature has implicated LLPS in cellular sensing, protein concentration buffering, concentration, localization and sequestering of molecules, mechanical driving forces, physiochemical filtering and gene regulation.

Because LLPS events are highly sensitive to parameters such as pH, temperature, and salt concentrations, LLPS formation can be exploited by the cell to sense and respond to changes in the environment. It has already been shown that Pab1 acts as a sensor for cellular stress such as temperature (*vide supra*; Riback et al., 2017). Likewise, the *Saccharomyces cerevisiae* eukaryotic translation release factor 3 (Sup35) phase separates at low pH conditions, thereby protecting itself during cellular stress and acting as a cellular sensor (Franzmann et al., 2018).

Phase-separated droplets can also be used to buffer protein concentrations. The concentration of a particular protein can be kept constant in the dilute phase, as a rise in

concentration can be buffered by including protein in the dense droplets, thereby increasing their volume. As the concentration in the dilute phase drops, proteins are released from the dense droplets to restore the equilibrium (Alberti et al., 2019).

LLPS can also be used by the cell as a means of concentrating molecules. The multi-protein assembly miSISC concentrates molecules in order to facilitate RNA processing. miSISC uses microRNA to identify mRNA that needs to be suppressed, it then phase separates and recruits deadenylation factors into the droplet to facilitate decapping of mRNA (Sheu-Gruttadauria and Macrae, 2018). The pre-centriolar material scaffold proteins forms phase separated droplets that recruit other proteins and tubulin into the droplet where it is concentrated enough to nucleate microtubule asters (Woodruff et al., 2017). Phase separation can also be a means to suppress reactions by sequestering key molecules into droplets, preventing their entry into specific pathways (Alberti et al., 2019). LLPS Droplets can selectively recruit specific proteins and RNA species, while excluding others (Alberti et al., 2019). For example, stress granules recruit the proteosomal shuttle factor UBQLNZ to droplets, wherein the factor interacts with ubiquitinated substrates to reverse the LLPS and enable the release of client proteins from the droplets (Dao et al., 2018).

Phase separated droplets can also be used to exert mechanical force within the cell. Interactions of phase-separated proteins can result in deformation of the cellular plasma membrane, thereby facilitating endocytosis (Bergeron-Sandoval et al. 2017). There is also evidence that LLPS can also be used to construct physiochemical sieves within the cell. Nuclear pore complexes form hydrogel sieves through LLPS of LCRs, that act as barriers that selectively allows the active transport of specific molecules (Schmidt and Görlich, 2016).

Recently, LLPS has been shown to be an important component of gene regulation. It was shown that heterochromatin has liquid droplet-like properties and likely forms via LLPS (Larson and Narlikar, 2018). Another aspect of gene regulation that has been linked to phase separation is the action of super enhancers. The Mediator subunit MED1 can undergo LLPS at super enhancers, and, in doing so, recruit and concentrate the transcription apparatus (Sabari et al., 2018). The super enhancer target genes can associate with these condensates during transcription activation (Chong et al., 2018; Hnisz et al., 2017; Sabari et al., 2018; Shrinivas et al., 2019). The C-terminal domain (CTD) of RNA Polymerase II can also form liquid-like condensates with the mediator complex; this, however, is dependent on its phosphorylation

state (Chong et al., 2018; Guo et al., 2019; Gurumurthy et al., 2019). Finally, activator domains of transcription factors such as octamer-binding transcription factor 4 (OCT4) and GCN4, form phase-separated droplets with the mediator complex during gene activation. Furthermore, estrogen enhances the phase separation of droplets containing the estrogen receptor and mediator complex (Boija et al., 2018; Hahn, 2018). How these condensates recruit the basal RNA polymerase II transcription machinery to facilitate PIC assembly at target genes is not fully understood. So far, there has been no evidence implicating GTFs in LLPS (Gurumurthy et al., 2019; Sabari et al., 2018).

Aim and Objectives

Currently there is a lack of literature pertaining to the molecular mechanisms governing transcription regulation in *P. falciparum*. To bridge this gap in current knowledge our research group aims to functionally characterize *P.* falciparum GTFs in order to elucidate the molecular mechanisms underlying gene regulation at the level of transcription initiation in this globally important parasite. This includes characterization of the DNA-binding activity of *P. falciparum* GTFs, as well as the determination of specific promoter sequence elements. Recently, the research group also started investigating the possible involvement of liquid-liquid phase separation (LLPS).

The aim of this study was to functionally characterize *P. falciparum* TBP-like protein (*Pf*TLP). The work included bioinformatics analyses, characterization of the DNA-binding activity of bacterially expressed recombinant *Pf*TLP proteins, and the investigation of a potential role of *Pf*TLP low-complexity regions in driving LLPS.

Chapter 2

Materials and Methods

2.1 Bioinformatics

Multiple sequence alignments were done using the programs Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/; Sievers et al., 2014) or Muscle (https://www.ebi.ac.uk/Tools/msa/muscle/; Edgar, 2004), with default settings on the EMBL-EBI web-based platform (https://www.ebi.ac.uk/Tools/msa/; Madeira et al., 2019). Results of multiple sequence alignments were visualized using UGENE (http://ugene.net; Okonechnikov et al., 2012). Phylogenetic trees were generated using LIRMM – Phylogeny (http://phylogeny.lirmm.fr/phylo_cgi/index.cgi; Dereeper et al., 2008). Three-dimensional structure predictions were generated using Phyre 2.0 (http://www.sbg.bio.ic.ac.uk/~phyre2/; Kelley et al., 2015), and the resulting 3D structure models were visualised using UCSF Chimera 1.13 (https://www.cgl.ucsf.edu/chimera/; Pettersen et al., 2004). General protein statistics, including amino acid percentage use, pI and molecular weight, were calculated using the ProtoParam tool on the ExPasy server (https://web.expasy.org/protparam/; Gasteiger et al., 2005). Disorder predictions were carried out using the DisMeta disorder prediction server (http://www-nmr.cabm.rutgers.edu/bioinformatics/disorder/; Huang, Acton & Montelione, 2014)

2.2 Expression of Recombinant *Pf*TLP Proteins in *Escherichia coli*

2.2.1 Plasmids

The expression vectors pET11d-6His-PfTLP, -F60A, -F283A and -F60&238A were cloned by David Adebolajo, a previous laboratory member (Adebolajo, 2016). The pET11d-6His vectors are derivatives of the pET11d (Novagen®) plasmid developed for expression of recombinant proteins carrying an N-terminal six histidine tag (Hoffmann and Roeder, 1991).

2.2.2 Transformation

100 µL BL-21-CodonPlus®(DE3)-RIL competent *E. coli* cells (Agilent Technologies) were incubated on ice for 45 minutes with 100 ng of plasmid, followed by heat shock at 42 ^{o}C for 90 seconds, and cooling on ice for 2 minutes. 900 µL SOC medium (0.5% (w/v) yeast extract, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM NaCl, 2% (w/v) tryptone, 10 mM $MgSO_4$, 20 mM glucose) was added and the cells were incubated at 37^{o}C with shaking for 1 h. 100 – 200 µL were plated onto LB agar plates (100 µg/ml ampicillin (AMP), 34 µg/ml chloramphenicol (CAM)) and incubated at 37^{o}C for 16 hours.

2.2.3 Production of Glycerol Cell Stocks

Glycerol stocks of transformed bacteria were generated for long-term storage at -80 ^{o}C by mixing one volume of bacteria at logarithmic growth (OD_{600} 0.3 – 0.5) with one volume ice-cold 60% glycerol and snap-freezing in liquid nitrogen.

2.2.4 Expression of Recombinant *Pf*TLP in *E.coli*

The growth medium used in expression experiments was lysogeny broth (LB) (1% (w/v) tryptone, 0.5% (w/v) yeast extract, and 170 mM NaCl) supplemented with 100 µg/ml AMP, 34 µg/ml CAM, and 1% (w/v) glucose. Cells from a freshly transformed plate or glycerol stock were used to inoculate a starter culture, which was incubated for 16 h at 37 ^{o}C with shaking. The culture was centrifuged at ~3000 xg for 10 min, and cell pellets resuspended in fresh media to remove β-lactamase activity. Freshly resuspended cells were used to inoculate a larger culture volume. The culture was then incubated at 37 ^{o}C with shaking until the OD_{600} reaches 0.5. The cultures were then cold-shocked in an ice slurry and cooled down to 25 ^{o}C. Isopropyl β-D-1-thiogalactopyranoside (IPGT) was added to 500 µM to induce expression and cultures were further incubated for 3 h at 25 ^{o}C. Finally, the cells were harvested by centrifuging at ~3000 xg for 10 min.

Large scale expression was carried out using 3 L of culture, divided into five 600 ml cultures. Each 600 ml culture was inoculated with 20 ml of freshly resuspended overnight culture.

2.3 Purification of Recombinant *Pf*TLPs

All purification and chromatography steps were carried out at 4 °C. During pilot experiments it became clear that *Pf*TLP tends to binds to the walls of plastic reaction tubes, which led to low yield. For this reason, protein purification was carried out using LoBind (Protein) Eppendorf® tubes.

2.3.1 Production of Clear Lysate

Collected bacterial cell pellets were left overnight on ice. The next day, every 1 g of cell pellet (wet weight) was resuspended in 5 ml of sonication buffer (50 mM Tris-HCl pH 7.9 @ 4˚C, 20 mM imidazole, 500 mM NaCl, 0.1 % NP-40) containing 25 µl protease inhibitor cocktail (Sigma-Aldrich® P8849) and 1 µg/µl lysozyme from chicken egg white (Sigma-Aldrich® 62971). The cells were sonicated six times with a 10 s pulse, followed by 30 s cooling on ice. After sonication, cell lysates were incubated on ice for 1 h with 250 U of Benzonase®Nuclease (Sigma-Aldrich® E1014), followed by centrifugation at 26,000 xg for 35 min at 4 °C. The supernatant was collected as cleared lysate and the pellet as insoluble fraction.

2.3.2 Metal Affinity Chromatography

Metal-affinity purification was carried out in a Poly-Prep® chromatography column (Bio-Rad). 300 µl packed resin volume of TALON® Metal Affinity Resin (Clontech) was equilibrated with 10 ml sonication buffer. After loading the cleared lysate, unbound material was removed and the salt concentration adjusted by step-wise washing with 2 ml wash buffers 1-4 (20 mM Tris-HCl pH 7.9 @4˚C, 20 mM imidazole, 0.2 mM EDTA pH 8, 0.1 % NP-40) containing decreasing NaCl concentrations of 400 mM, 300 mM, 200 mM and 100 mM, respectively. NaCl was then exchanged for KCl, with a 5 ml wash using BC-buffer (20 mM Tris-HCl pH 7.9 @ 4°C, 0.2 mM EDTA pH 8, 0.1 % NP-40, 10 mM β-mercaptoethanol, 20 % glycerol) containing 100 mM KCl. Finally, bound protein was eluted in 10 ml BC-buffer containing 100 mM KCl and 250 mM imidazole. 300 µl fractions were collected in LoBind (Protein) Eppendorf® tubes, snap frozen in liquid nitrogen and stored at -80 °C.

2.3.3 Cation-Exchange Chromatography

650 µl of SP sepharose FF® (Sigma-Aldrich®) was packed in a Poly-Prep® Chromatography Column (Bio-Rad) and equilibrated with BC-buffer containing 100 mM KCl (BC-100). Cobalt affinity chromatography elution fractions with the highest protein concentration were loaded onto the column and unbound protein washed out with 8 column volumes of BC-100 buffer. Proteins were eluted, at a flow rate of 30 column volumes per hour, with a BC-buffer salt gradient from 100 mM to 500 mM KCl over 20 column volumes followed by a final elution step with 5 column volumes of BC-buffer containing 500 mM KCl (BC-500). 300 µl elution fractions were collected, snap frozen and stored at -80 ^{o}C. SDS PAGE analysis of elution fraction showed that recombinant *Pf*TLP proteins eluted around 300 mM KCl.

2.3.4 Normalization

Protein concentration of *Pf*TLP preparations were estimated by SDS PAGE and Coomassie staining using known amounts of BSA for comparison. Different preparations of 6:His tagged *Pf*TLP protein were normalized using SDS PAGE and Coomassie staining and immunoblotting using anti-6His primary antibody (see section 2.3.6). Bovine serum albumin (BSA) (New England BioLabs®) was added to all protein preparations to a final concentration of 100 ng/µl. Final protein preparations were aliquoted, snap-frozen in liquid nitrogen and stored at -80 ^{o}C.

2.3.5 SDS-PAGE analysis

Samples were mixed with 5x SDS-loading buffer (250 mM Tris-HCl pH 6.8, 10 % (w/v) SDS, 30 % glycerol, 0.02 % (v/v) bromophenol blue) and denatured by heating at 85 ^{o}C for 5 min. Samples were electrophoresed at constant 35 mA using 5% acrylamide (37:1) stacking gel, and 12 % acrylamide (37:1) resolving gel (Sigma-Aldrich®) using a mini-PROTEAN® system (Bio-Rad). Proteins were visualized using Bio-Safe G-25 Coomassie stain (Bio-Rad), or ProteoSilverTM Silver Stain Kit (Sigma-Aldrich®), following the manufacturer's instructions.

2.3.6 Immunoblotting

After SDS-PAGE, proteins were transferred to Millipore® Immobilon PVDF membrane using a Mini Trans-Blot® Electrophoretic Transfer Cell (Bio-Rad) at 100 V for 1 h at 4 °C. The membrane was washed three times for 5 min in TBST buffer (150 mM NaCl, 10 mM Tris pH 8,

0.05 % Tween 20), and once with water for another 5 min. After staining with Gold Solution (Sigma-Aldrich® 50755) for 1 to 3 h, the membrane was washed three times for 5 min in TBST, and blocked with 5 % (w/v) skim milk in TBST for 30 min. To detect recombinant 6His:tagged *Pf*TLP proteins, the blocked membrane was incubated overnight at 4 ^{o}C with anti-His probe rabbit polyclonal IgG (Santa Cruz Biotechnology, SC-803,) or anti-*Pf*TLP rabbit anti-serum (Oelgeschläger lab; Bing, 2014; Milton, 2017; Talvik, 2016)) in in TBST containing 5 % (w/v) skim milk, at a 1:3000 dilution. The membrane was washed three times for 5 min in TBST, and then incubated with secondary anti-rabbit IgG-peroxidase conjugate (Sigma-Aldrich®, A0545) in TBST, at a 1:3000 ratio, for 30 min at room temperature. Detection was carried out using WesternBrightTM Sirius (Advansta) kit following manufacturer's instructions.

2.4 DNA-binding Assays

2.4.1 DNA Probes

A 288 bp promoter region of the 130 kDa merozoite surface protein glycophorin-binding protein (*GBP-130*) (Hessler, 2014; Perkins, 1984) was used in DNA-binding experiments. The promoter sequence of GBP-130 was described previously (Lanzer et al., 1992a, 1992b; Ruvalcaba-Salazar et al., 2005). For ITA assays, a 5'biotinylated 422bp GBP-130 probe containing the 288 bp promoter region downstream of five GAL4 binding sites (Oelgeschläger et al., 1998) was used to increase the distance of the GBP-130 promoter region from the 5'biotin-streptavidin link used to immobilize the DNA probe on magnetic beads. 288bp and 422bp GBP-130 promoter probes were generated by PCR using plasmid pTHG5GBP-130 (Hessler, 2014). PCR amplification was carried out with 5'biotinylated forward and reverse primers to generate the 288bp GBP-130 EMSA probe and with a 5'biotinylated forward and an unmodified reverse primer to generate the 422bp GBP-130 ITA probe using the KAPA Taq PCR Kit (KAPA Biosystems) following the manufacturer's instructions. Shorter 30-40 bp DNA probes used in EMSA assays were generated by annealing complementary DNA strands at 1µM in TE buffer containing 100 mM NaCl, with 5min incubation at 95˚C followed by slow cooling to RT over 30 min. These included a 30 bp *GBP*-130 region as well as a 40 bp adenovirus 2 major late promoter (Ad2ML) probe with the typical TATA-box element (see Supplementary Table 2 for sequence details).

2.4.2 Immobilized Template Assays (ITAs)

In ITAs proteins are incubated with biotinylated DNA probes that are immobilized on magnetic streptavidin beads. The unbound proteins are washed away and the bound proteins eluted (Johnson et al., 2004) and analysed with immunoblotting.

2.4.2.1 Immobilization of probes

10 µg of single biotinylated DNA template was immobilized on 1 mg Dynabeads® M-280 Streptavidin (Invitrogen) according to the manufacturer's instruction and immobilised DNA beads were stored in 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 M NaCl, 100ng/µL BSA. To confirm the amount of immobilized DNA, DNA was isolated from magnetic beads by phenol extraction and ethanol precipitation, followed by agarose gel electrophoresis and comparison with known amounts of free DNA probe.

2.4.2.2 Binding reaction

Standard ITA binding reactions contained 1 pmol of immobilized DNA probe, or the equivalent amount of magnetic beads without DNA as control in a total volume of 40 µL. Prior to protein binding, beads were equilibrated with a 2:3 (v/v) ratio of template mix (20 mM HEPES, 5 mM DTT, 5 mM $MgCl_2$, 0.1 % NP-40) and BC-100 buffer. For protein binding, a reaction mix of 16µl template mix and 24 µl protein in BC-100 buffer was added to the beads and incubated for 45 minutes at 30 ^{o}C using an Eppendorf shaker/incubator. After protein binding, the magnetic beads were precipitated in a magnetic tube rack, and the unbound fraction was then removed. The beads were washed in 2:3 (v/v) template mix and BC-100 buffer and transferred to a fresh tube, and the supernatant removed. Finally, beads were eluted in 1x SDS loading buffer. The eluates were heated at 80 ^{o}C for two minutes and the supernatant analysed by SDS PAGE and immunoblotting.

2.4.3 Electrophoretic Mobility Shift Assays (EMSAs)

In EMSAs biotinylated DNA probes were incubated with proteins and free DNA probe separated from protein/DNA complexes using native agarose gel electrophoresis (Zerby and Lieberman, 1997). Binding reactions were made up from 4 µl template mix (20 mM HEPES, 5 mM DTT, 5 mM Mg_2Cl, 100 ng/µl BSA, 0.5 % NP-40) supplemented with 5 fmol of biotinylated DNA probe and poly(dG-dC) competitor (as indicated) and 6 µl protein mix in

BC-100 buffer. The reaction was incubated at 30^{o}C for 45 min. EMSA loading buffer (2.5% (w/v) ficol 400) was added to each reaction, samples were loaded onto 1.4% (w/v) Mg^{2+} agarose gels in 40 mM Tris, 40 mM boric acid, 5 mM magnesium acetate and resolved for ~3 h at 100 V in 40 mM Tris, 40 mM boric acid, and 2 mM magnesium acetate at 4 ^{o}C. Alternatively, EMSAs were resolved in native agarose gels lacking magnesium acetate in the gel and running buffer and electrophoresed for 30 min to 1 h at 100 V at 4 °C. Following electrophoresis, the DNA probes and DNA/protein complexes were transferred to a Millipore® Immobilon-NY^{+} nylon membrane at room temperature (RT) for 90 min through horizontal vacuum transfer, using a modified southern/northern blot apparatus connected to a 240 mbar diaphragm pump (DA7C Charles Austen Pumps). Directly after the transfer the DNA was cross-linked using a commercial UV-cross linker, using four pulses of 120 mJ/cm^2. Detection of biotinlylated DNA was carried out using the Chemiluminescent Nucleic Acid Detection Module (Thermo-Scientific®), following manufacturer's instructions.

2.5 Generation *Pf*TLP low complexity regions fused to fluorescent protein

2.5.1 Generation of Fusion Constructs

Gene constructs for fluorescently tagged proteins were generated by PCR-driven overlap extension (Heckman and Pease, 2007) using primer sequences listed in supplementary table 1. The gene for green fluorescent protein (GFP) was cloned from a vector based on the pEGFP (Clontech) vector, kindly donated by Shakiera Sattar (University of Cape Town, Department of Molecular and Cell Biology). The ORFs for yellow and cyan fluorescent proteins (YFP, CFP) were cloned from the pEarlyGate 101 and 102 vectors (Earley et al., 2006) respectively. These vectors were kindly donated by Dr Laura Roden (University of Cape Town, Department of Molecular and Cell Biology). ORFs for fusion proteins were inserted in pET11d-6His (Hoffmann and Roeder, 1991) using *Nde*I and *Bam*HI sites (see Supplementary Fig. 1-3 for vector maps). Vector and insert sequences were verified by sequencing before being used for protein expression. A schematic of the vectors generated in this work are shown in (Fig. 28).

2.5.2 Expression and Purification of Fusion Proteins

6His-GFP-*Pf*TLP-LCR1, 6His-YFP-*Pf*TLP-LCR2, and 6His-CFP-*Pf*TLP-LCR2 were expressed in *E.coli* BL21-CodonPlus(DE3)-RIL as described in section 2.2 (Materials and Methods), except that protein expression was induced by addition of 500 µM β-D-1-thiogalactopyranoside (IPGT) when cultures reached an OD_{600} of 0.5. Induced cultures were then further incubated for 3 h at 37 ^{o}C with shaking before cells were harvested by centrifugation at ~3000 xg for 10 min. 6His-tagged LCR1 and LCR2 fused to fluorescent protein were purified by metal-affinity chromatography as described in section 2.3 (Materials and Methods). Metal-affinity eluates were collected in LoBind (Protein) Eppendorf® tubes, snap-frozen in liquid nitrogen and stored at -80 ^{o}C.

2.6 Phase Separation Assays

Glass chambers (Fig. 3) for examining phase separation of fluorescent fusion proteins were prepared based on a previously published procedure (Alberti et al., 2018). Briefly, glass slides and coverslips were silanized using Sigmacote® (Sigma-Aldrich) following manufacturer's instructions. The silanized glass slides were rinsed in water, and two strips of double-sided tape placed lengthwise on both sides of the slide (Fig. 3). The assembled slide stage was rinsed with wash buffer (25 mM HEPES pH 7.4, 150 mM KCl, 1 mM DTT) before use.

Phase separation reactions were assembled by mixing one volume protein mix in BC-100 buffer with one volume separation buffer (50 mM HEPES pH 7.4, 200 mM KCl, 1 mM DTT, 40-50% PEG 20 000). The reactions were pipetted onto the assembled slide stage, and the coverslip added (Fig. 3). Reactions were visualized with a Nikon Ti-E inverted fluorescence microscope using appropriate filters for the emission of yellow fluorescent protein (YFP; 520 – 550 nm) and cyan fluorescent protein (CFP; 460 – 500 nm). Images were rendered using Nikon NIS-Elements AR software.

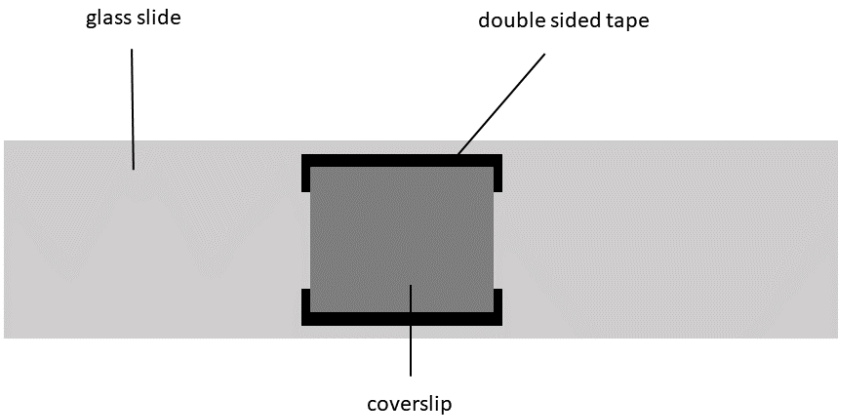


Figure 3: Schematic of the modified glass chamber used to visualize phase separation reactions by fluorescent microscopy.

Chapter 3

Results

31. Bioinformatics Analysis of *Pf*TLP

Bioinformatics analysis was performed on the *Pf*TLP amino acid sequence (See Supplementary Fig. 5 for full sequence) in order to investigate evolutionary conservation within the TBP family of proteins, as well as to gain insight into the structural and functional characteristics of *Pf*TLP.

TBP proteins consist of five classes; TBPs, and TBP-related factors (TRFs) 1-4, which expanded due to TBP gene duplication events (Akhtar and Veenstra, 2011). TRF1 was originally discovered in *Drosophila*, and is an insect-specific TLP. TRF3 is only found in vertebrates, and is the TRF most closely related to TBP with 95 % amino acid identity to the TBP core domain. The human parasite *Trypanosoma brucei* has a single TBP-family protein, identified as TRF4. TRF2s are found in metazoans, and share roughly 60% homology with TBP. It is hypothesized that TLP/TRFs evolved to function in specific gene regulation programs, where they interact with core promoter elements and transcription factors different to that of TBP (Akhtar and Veenstra, 2011; Duttke et al., 2014; Thomas and Chiang, 2006; Wang et al., 2007).

Phylogenetic tree analysis was used to evaluate the evolutionary relationship of *Pf*TLP with different TBP-family proteins. Sequences were selected from model eukaryotic species, as well as some characterized archaebacteria. As seen in Fig. 4, TBP-family proteins group roughly into thee clusters, a TBP/TRF3 cluster, a TRF2 cluster, and TRF1. *Pf*TLP groups with the TRF2 type proteins (Fig. 4), consistent with a previous analysis (Bing, 2014; Milton, 2017). The result indicated that *Pf*TLP might be a TRF2 type TLP.

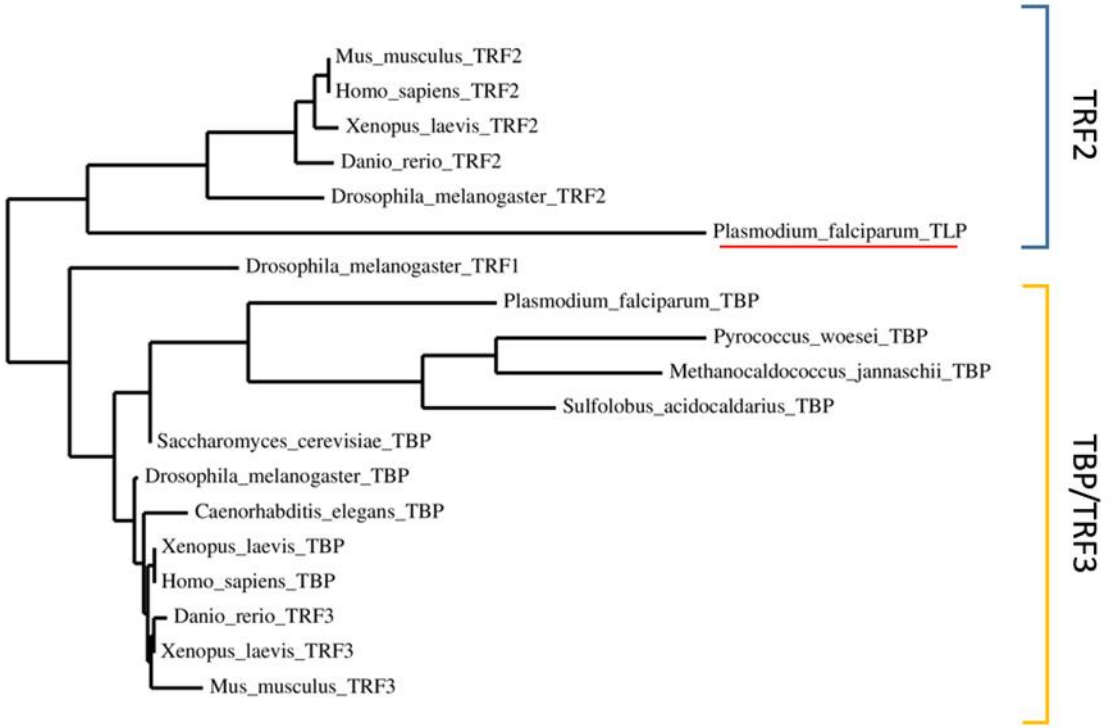


Figure 4: Phylogenetic tree analysis of TBP family proteins. A maximum-likelihood phylogenetic tree was rendered using *Phylogony.fr* (Dereeper et al., 2008), based on multiple sequence alignment results generated by MUSCLE (Edgar, 2004). Bootstrap test (100 replications) was performed. Brackets are used to indicate the groups of TBP family proteins.

Further analysis was done to see if the characteristic TBP-core domain structure was likely to be conserved in *Pf*TLP. This analysis was achieved through the web-based program *Phyre 2.0* (Kelley et al., 2015), which predicts a three dimensional protein model by combining multiple sequence alignments with secondary structure predictions into a query hidden Markov model. The query is then scanned against proteins with solved structures, and the top scoring secondary structure alignment is used to construct a crude backbone, which is then transformed into a three dimensional model after loop and side chain rendering. As seen in Fig. 5, *Pf*TLP is predicted to possess a typical saddle-shape TBP core domain consisting of eight β-strands and four α-helices. Of note are two loop structures extending out of the core domain between β-strands S3 and S4, and S3' and S4', respectively. These regions were previously identified as low-complexity regions 1 and 2 (LCR1/2), and are predicted to be intrinsically disordered (discussed further in Results section 3.4; Bing, 2014; Milton, 2017).

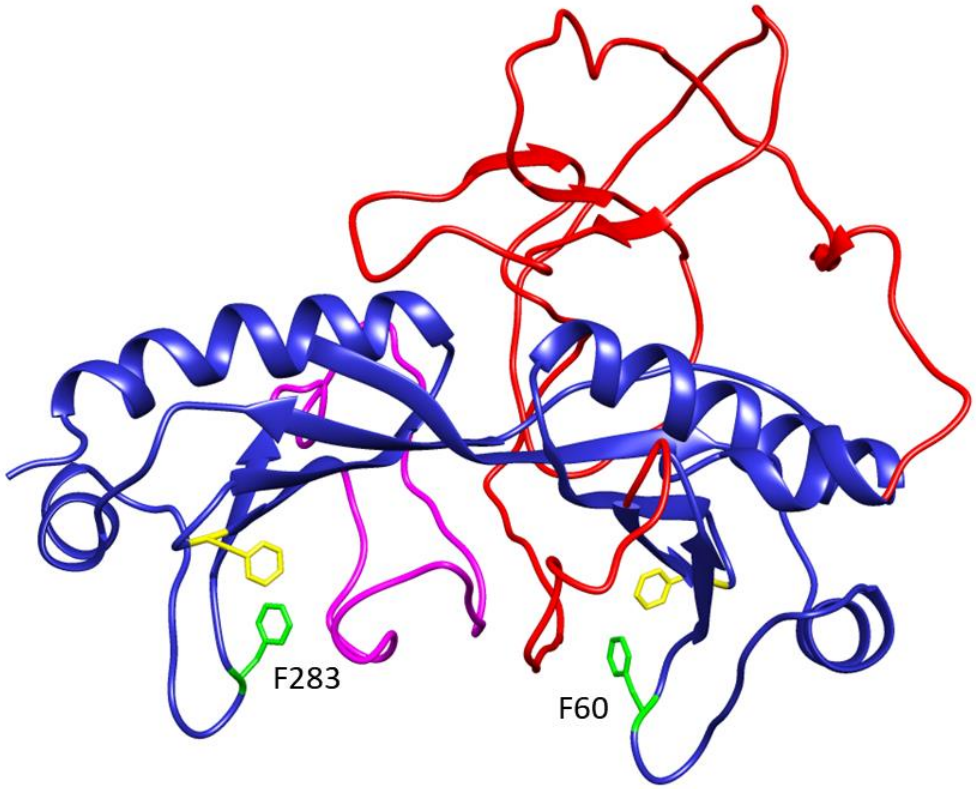


Figure 5: Predicted three dimensional structure model of *Pf*TLP. Ribbon model of the *Pf*TLP three-dimensional structure. The conserved saddle-shaped TBP core domain structure is shown in blue, low complexity regions 1 and 2 are show in red and magenta, respectively. Evolutionary conserved phenylalanine residues F60 and F283, shown to insert into the DNA during TBP binding to the TATA box, are shown in green and supporting phenylalanine residues, F209 and F333, are shown in yellow. Model generated using *UCSF Chimera 1.13* (Pettersen et al. 2004) based on *Phyre 2.0* analysis (Kelley et al., 2015).

Furthermore, the model shows that the phenylalanine residues, required for stable TATA-box binding by TBP, are conserved in *Pf*TLP, in contrast to previously characterized TRF2-type TLPs (Duttke et al., 2014).

A pair-wise sequence alignment was performed between *Pf*TLP core domain and the most closely related solved structure of *Saccharomyces cerevisiae* TBP (*Sc*TBP) (Fig. 6A). This analysis shows that the *Pf*TLP core domain has a low sequence homology with ScTBP, with only 24 % identity and 49 % similarity. The highest degree of conservation between *Pf*TLP and *Sc*TBP was observed in the critical TBP secondary structures (Fig. 6A), outlined in the schematic shown in Fig. 6B.

Collectively, this analysis suggests that *Pf*TLP is likely to adopt the same three dimensional saddle-shaped DNA-binding core domain structure as prototypical eukaryotic TBPs (*Sc*TBP), despite having a low sequence homology.

In order to further examine the conservation of the phenylalanine residues important for TBP TATA box binding in *Pf*TLP, a multiple sequence alignment was carried out with the sequences of the most closely related structures from the *Phyre 2.0* (Kelley et al., 2015) analysis. As shown in Fig. 7, this analysis revealed a high degree of sequence similarity surrounding the conserved phenylalanine residues between *Pf*TLP and eukaryotic (*Saccharomyces cerevisiae*, *Sc; Encephalitozoon cuniculi*, Ec) and archaeabacterial (*Pyrococcus woesei*, Pw, *Methanococcus jannaschi, Mji*) TBPs. This observation further supports the notion that the conserved phenylalanine residues might be important for *Pf*TLP DNA-binding activity. The role of these phenylalanine residues in *Pf*TLP DNA binding is further examined in Results section 3.3.

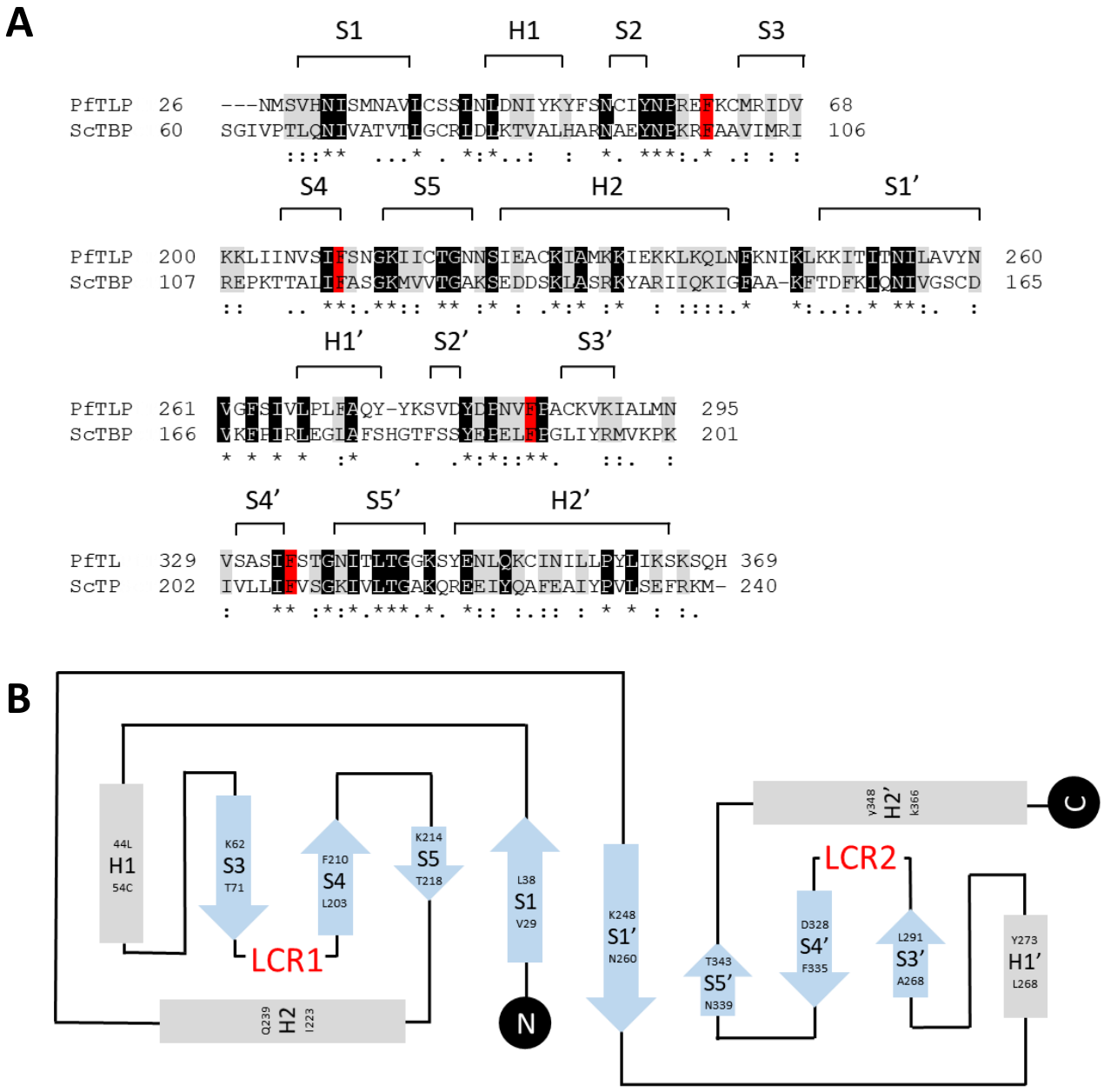


Figure 6: Predicted secondary structure architecture of *Pf*TLP. A) Pairwise alignment of *Pf*TLP core domain sequence with the core domain of *Saccharomyces cerevisiae* TBP *(Sc*TBP). Alignment generated with Clustal Omega (Sievers et al., 2014). Identical amino acid residues are highlighted in black and marked with an asterisk (*), highly similar amino acids (PAM 250 matrix score > 0.5) are highlighted in grey and marked with a colon (:), and weakly similar residues (PAM 250 matrix score < 0.5) are marked with a period (.). Conserved phenylalanine residues required for TBP binding to the TATA box are highlighted in red. Annotation of (H) α-helices and (S) β-sheets is based on the solved *Sc*TBP structure (Chasman et al., 1993). B) Schematic of the *Pf*TLP secondary structure based on the quasi symmetrical TBP core domain. (H) α-helices are shown as shaded boxes and (S) β-sheets as light blue arrows, with the identity of the flanking amino acid residues indicated. The location of the low-complexity regions (LCRs) are indicated in red. LCR1 spans the region from V70 to L203, and LCR2 from D296 to I329.

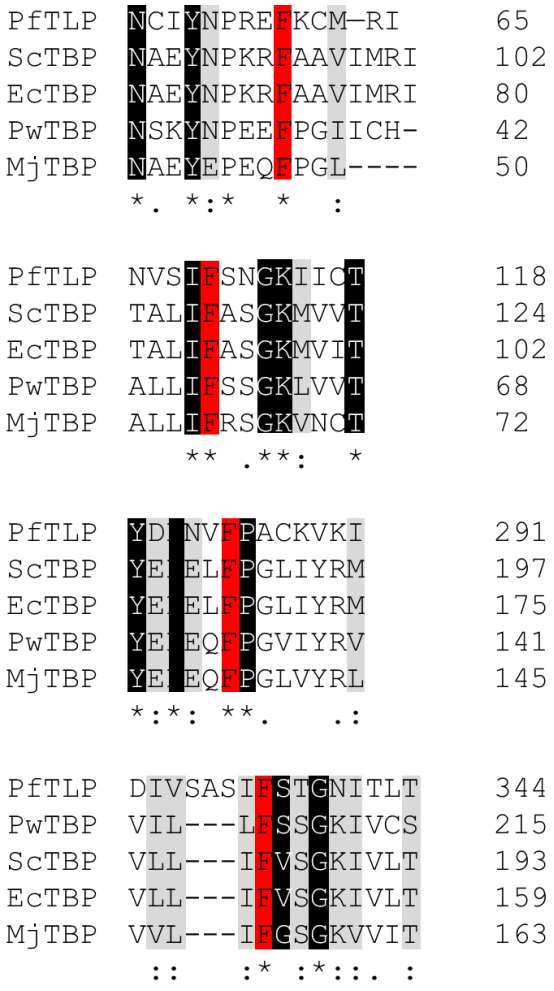


Figure 7: Multiple sequence alignments of the regions flanking the evolutionary conserved phenylalanine residues in *Pf*TLP and in eukaryotic and archaebacterial TBPs. Identical amino acid residues are highlighted in black and marked with an asterisk (*), highly similar amino acids (PAM 250 matrix score > 0.5) are highlighted in grey and marked with a colon (:), and weakly similar residues (PAM 250 matrix score < 0.5) are marked with a period (.). Conserved phenylalanine residues required for TBP binding to the TATA box are highlighted in red. *Pf* = *Plasmodium falciparum*, *Pw* = *Pyrococcus woesei*, *Sc* = *Saccharomyces cerevisiae*, *Ec* = *Encephalitozoon cuniculi*, and *Mj* = *Methanococcus jannaschii.* Alignments was carried out using Clustal Omega (Sievers et al., 2014).

3.2 Expression and Purification of *Pf*TLP

Expressing and purifying recombinant *P. falciparum* proteins comes with many caveats and limitations. The AT-rich genome of *P. falciparum* is one of the biggest obstacles in protein production. The AT-bias in the genome means that *P. falciparum* proteins not only use rare AT-rich codons, but it also leads to a major bias in genome wide amino acid usage (Yadav and Swati, 2012). This difference in amino acid usage places the bacterial host system under severe strain, leading to toxicity. Additionally, the rare codons can also lead to stalling of the translation machinery, causing truncated products (Birkholtz et al., 2008). Another major issue with expressing *P. f*alciparum proteins is the presence of low-complexity regions (Fig. 6). Proteins containing LCRs are often insoluble and aggregation prone, and therefore are often concentrated in inclusion bodies (Alberti et al., 2018; Birkholtz et al., 2008; Vedadi et al., 2007). Additionally, these proteins tend to be "sticky" and are retained on resins and certain surfaces. Special care must be taken to increase the solubility of these proteins, and this involves using low-binding reaction tubes and taking particular caution when changing salt concentrations.

A his-tagged and codon optimized expression vector, as well as an initial protocol for expression and purification for *Pf*TLP was previously established in the laboratory (Bing, 2014). Briefly, a bacterial culture of transformed *E.coli* BL21- CodonPlus(DE3)-RIL was incubated at 37 °C until the culture reached an absorbance ($0D_{600}$) of 0.5. The cells were then cold-shocked to 25 °C, and expression induced through the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG). After a further 3h incubation at 25 °C, the cells were lysed through sonication and the soluble bacterial extract collected as cleared lysate. Finally 6His:tagged recombinant *Pf*TLP (6His:*Pf*TLP) was purified from the soluble bacterial lysate through metal affinity chromatography (see Materials and Methods chapter 2).

Fig. 8 shows an SDS-PAGE analysis of samples obtained during the different steps in of the 6His:*Pf*TLP bacterial expression and purification. There was no visible band in the whole cell lysate after induction, which indicates very low levels of expression. Elution fractions contained very small amounts of a protein of the predicted molecular weight of 6His:*Pf*TLP. The estimated yield was only 6 µg soluble 6His:*Pf*TLP/l culture.

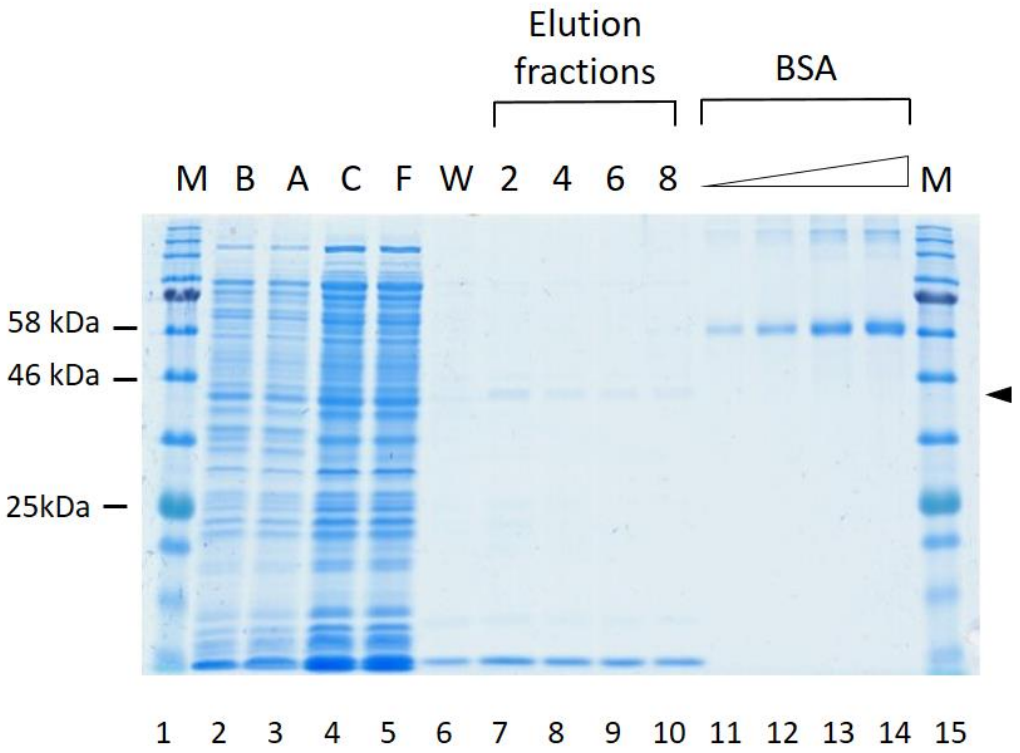


Figure 8: SDS PAGE analysis of the initial 6His:*Pf*TLP purification. Equivalent amounts of bacterial whole cell lysates before (B, lane 2), after induction with IPTG (A, lane 3), and equivalent amounts of metal affinity chromatography input (clear lysate, C, lane 4) and flow through (F, lane 5) and identical volumes of wash (W, lane 6) and elution fractions (lanes 7-8) were analyzed by 12% SDS-PAGE and visualized by Coomassie staining. 60 ng, 120 ng, 250 ng and 500 ng bovine serum albumin (BSA) protein standard was loaded for comparison (lanes 11-14). The position of the protein band corresponding to the expected size of 6His:*Pf*TLP (calculated molecular mass: 44.3 kDa) is indicated by a solid triangle. M, molecular weight standard (lanes 1,15).

SDS PAGE analysis revealed that a significant proportion of 6His:*Pf*TLP remained on the resin after elution, suggesting that *Pf*TLP aggregated on the resin during the purification process, resulting in poor yield. The previously established purification protocol involved a rapid buffer exchange from the lysis buffer, of 500 mM NaCl, to the elution buffer, of 100 mM KCl. This rapid change in salt concentration may have caused the aggregation of 6His:*Pf*TLP. To test this hypothesis and to improve 6His:*Pf*TLP yield, the salt concentration was gradually adjusted through a step-wise equilibration process, which involved sequential washes with decreasing salt concentrations (500 mM NaCl, 400 mM NaCl, etc.). As shown in Fig. 9, in the purification

using a step-wise adjustment of the salt concentration, elution fractions contained increased amounts of a protein with the expected size of 6His:*Pf*TLP (Fig. 9, lanes 6-8). Based on this the overall yield was estimated to be 18 µg of soluble 6His:*Pf*TLP/l culture. Thus gradually changing the salt concentration on the metal affinity column prevented some aggregation and increased overall protein yield.

Some minor contaminant bands could be seen in addition to 6His:*Pf*TLP in the metal-affinity chromatography eluates (Fig. 9, lanes 6-8). Therefore, 6His:*Pf*TLP was further purified using ion-exchange chromatography. 6His:*Pf*TLP has an estimated pI of 9.28, and therefore a net

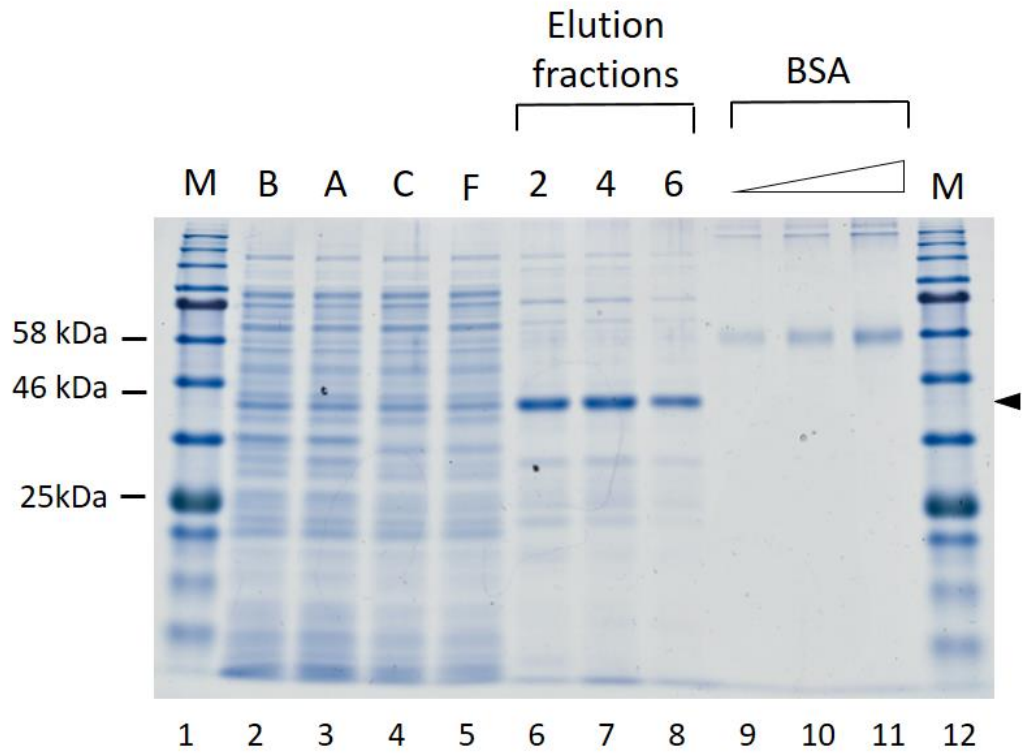


Figure 9: SDS PAGE analysis of improved 6His:*Pf*TLP purification protocol, using step-wise salt adjustment during metal affinity chromatography. Equivalent amounts of bacterial whole cell lysate before induction (B, lane 2) and after induction with IPTG (A, lane 3), and equivalent amounts of metal affinity chromatography input (clear lysate, C, lane 4), and flow through (F, lane 5) and identical volumes of the elution fractions (lanes 6-8) were analyzed by 12% SDS-PAGE and visualized by Coomassie staining. 50 ng, 100 ng , and 200 ng of bovine serum albumin (BSA) protein standard was added for comparison (lanes 9-11). The position of the protein band corresponding to the expected size of 6His:*Pf*TLP (calculated molecular mass: 44.3 kDa) is indicated by a solid triangle. M, molecular weight standard (lanes 1,12).

positive charge in our buffer systems, and could therefore be further purified using SP-Sepharose FF® cation exchange chromatography. As expected, 6His:*Pf*TLP bound to the SP-Sepharose FF® cation exchange resin. The majority of the 6His:*Pf*TLP eluted at around 300 mM KCl, and some of the contaminant bands were removed (Fig. 11 B, lanes 5-7). The final yield after SP-Sepharose purification was roughly 7 µg soluble 6His:*Pf*TLP/l culture.

The complete expression and purification protocol is shown as a schematic in Fig. 10. The same expression and purification protocol was followed to produce preparations of 6His:*Pf*TLP wild-type, and mutant proteins 6His:*Pf*TLP -F283A, -F60A, and –F60&283A. Fig. 12 shows an SDS-PAGE analysis of the purified 6His:*Pf*TLP wild-type and mutant protein preparations.

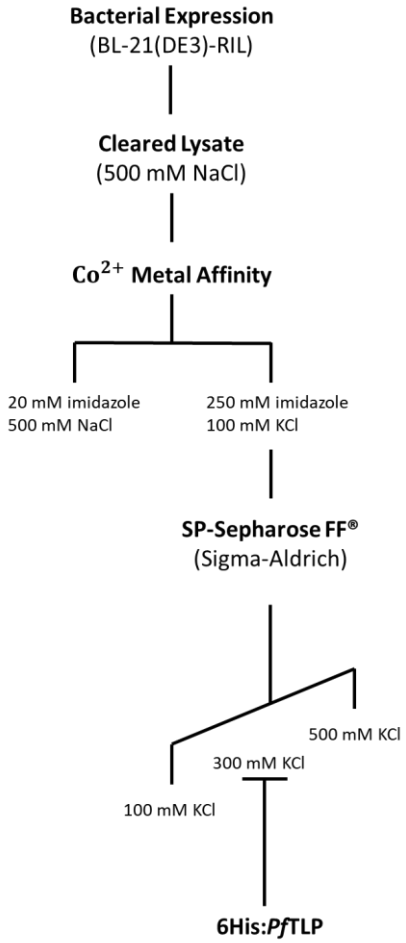


Figure 10: Purification of bacterially expressed recombinant 6His:*Pf*TLP. After binding to the Co^{2+} metal affinity column, salt is adjusted in a step-wise manner from 500 mM NaCl to 100 mM KCl (see Methods and Materials chapter 2). The metal affinity eluates are loaded onto a SP-Sepharose FF® cation exchange chromatography column and eluted with a linear gradient from 100 mM to 500 mM KCl. Majority of the 6His:*Pf*TLP elutes at a concentration of ~300 mM KCl.

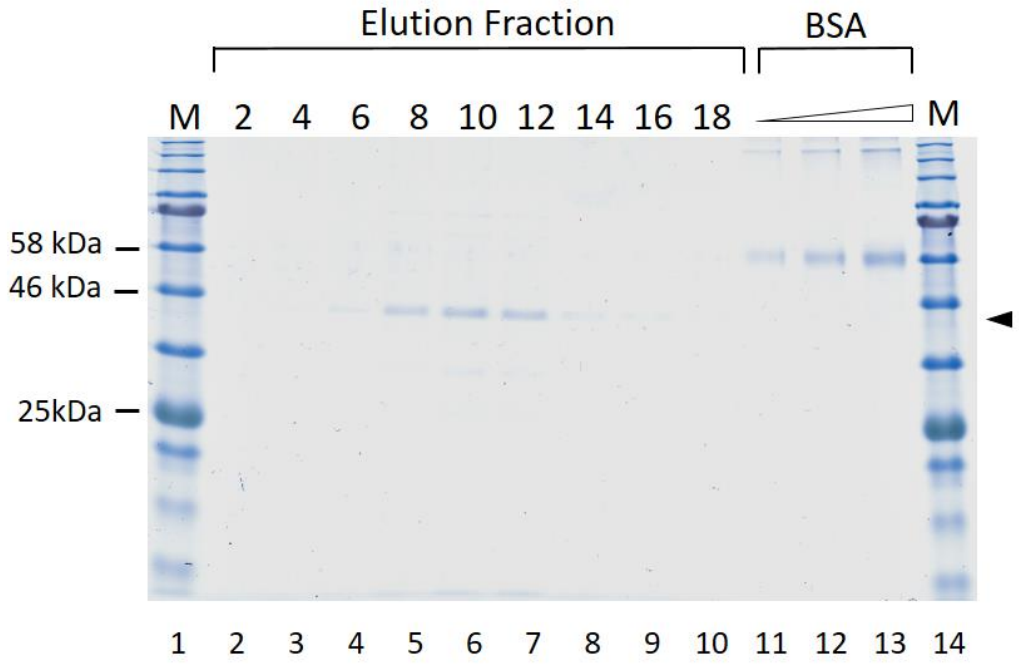


Figure 11: SDS PAGE analysis of SP-Sepharose FF® cation exchange chromatography eluate fractions. Identical volumes of elution fractions (lanes 2-10) were analysed by 12% SDS-PAGE and visualized by Coomassie staining. 50 ng, 100 ng, and 200 ng bovine serum albumin (BSA) protein standard was loaded for comparison (lanes 11-13). The position of the protein band corresponding to the expected size of 6His:*Pf*TLP (calculated molecular mass: 44.3 kDa) is indicated by a solid triangle. M, molecular weight standard (lanes 1,15).

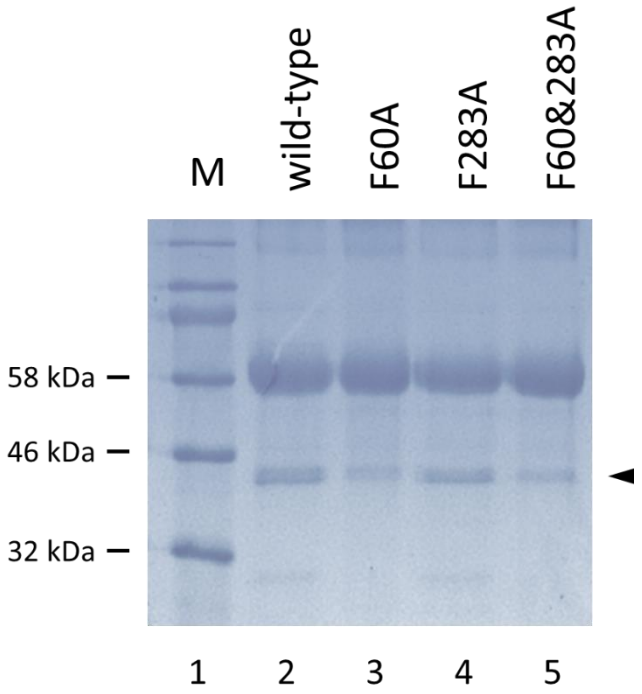


Figure 12: SDS PAGE analysis of the final 6His:*Pf*TLP wild-type and mutant protein preparations. Identical volumes of 6His:*Pf*TLP wild-type and F-to-A mutants F60A, F283A, and double mutant F60&283A (lanes 2-5) were analyzed by 12% SDS-PAGE and visualized by Coomassie stained. Note that 100 ng/µl BSA was added to the final SP-Sepharose FF® purified protein preparations. The position of the protein band corresponding to the expected size of 6His:*Pf*TLP (calculated molecular mass: 44.3 kDa) is indicated by a solid triangle. M, molecular weight standard (lane 1).

3.3 Characterization of *Pf*TLP DNA Binding Activity

*Pf*TLP was previously shown to have DNA-binding activity (Bing, 2014; Milton, 2017). This study aimed to further characterize *Pf*TLP DNA-binding activity and determine the roles of the evolutionary conserved phenylalanine residues.

The DNA-binding activity of wild-type *Pf*TLP was first investigated using an immobilized template assay (ITA). Briefly, biotinylated DNA probes are immobilized on streptavidin-coated magnetic beads and proteins are incubated with the immobilized DNA probes. The unbound proteins are washed away, and bound proteins are detected by immunoblot analysis. As a control, magnetic beads without DNA are used to ensure the proteins are binding the DNA and not the beads (Johnson et al., 2004).

As shown in Fig. 13 purified recombinantly expressed 6His:*Pf*TLP bound quantitatively to a immobilized 422 bp *P. falciparum glycophorin binding protein* (GBP-130) promoter DNA probe, whereas binding to the magnetic beads only was undetectable (Fig.13 , compare lanes 2, 4, 6 and 3, 5 and 7). This result showed that the recombinant 6His:*Pf*TLP produced in this study has DNA binding activity.

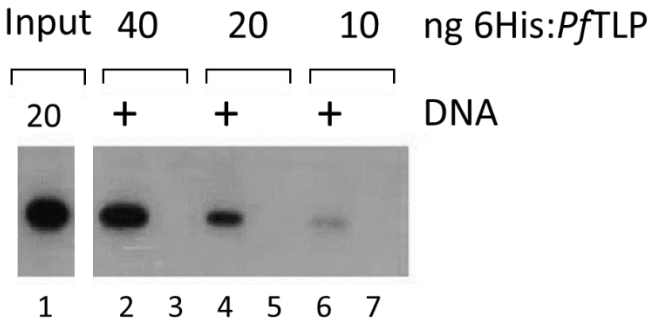


Figure 13: Binding of *Pf*TLP wild-type to an immobilized *P. falciparum* glycophorin *binding protein* (GBP-130) promoter region. 140 ng of immobilized 422 bp GBP-130 or equivalent amount of magnetic beads only (lanes 2-7) were incubated with indicated amounts of 6His:*Pf*TLP in 40 µl binding reactions. Unbound proteins were washed away with binding buffer and the bound proteins eluted with SDS loading buffer. 50 % of eluates (lanes 2-7) and 20 ng of 6His:*Pf*TLP input (lane 1) were analysed using immunoblot and detected with anti-6His (rabbit) primary antibody and anti-rabbit IgG-peroxidase secondary antibody. Protein bands were visualized on X-ray film using chemiluminescence.

The DNA-binding activity of wild-type 6His:*Pf*TLP was further investigated using the electrophoretic mobility shift assay (EMSA). In this assay the protein is incubated with labelled DNA probes and then free DNA is separated from protein/DNA complexes in native gel electrophoresis (Hellman and Fried, 2007). EMSAs are thought to be more stringent because the complexes must be stable enough to be resolved in the gel matrix during electrophoresis. Previous work in the laboratory has shown that 6His:*Pf*TLP tends to be retained in the well in EMSAs using native PAGE. This issue could be resolved by using agarose gel electrophoresis with Mg^{2+} present during electrophoresis (Bing, 2014; Milton, 2017; Zerby and Lieberman, 1997). Therefore, native agarose EMSAs were used in this study.

The EMSAs were initially performed using a 1.4% agarose gel with Mg^{2+} present in the gel and running buffer. Under these conditions 6His:*Pf*TLP DNA binding was clearly detectable, (Fig. 14A), consistent with previous observations (Bing, 2014; Milton, 2017,). However, the *Pf*TLP complexes showed only slightly reduced mobility and were not well separated from the free probe (Fig. 14A ,Bing, 2014; Milton, 2017; Talvik, 2016,). The presence of Mg^{2+} affects the EMSA assay conditions in several ways. Firstly, Mg^{2+} is required for stable complex formation of many DNA-binding proteins (Moll et al., 2002). Secondly, during electrophoresis, Mg^{2+} present in gel and running buffer migrate in the opposite direction as free DNA and DNA/protein complexes and neutralize the negative change on the DNA backbone, thereby slowing the migration of the free probe and DNA-protein complexes in the electric field. For this reason, agarose EMSA's were carried out in the absence of Mg^{2+} to improve the separation between the free DNA probe and DNA/protein complexes.

Fig. 14 panels A and B illustrates the effects of Mg^{2+} during electrophoresis on the mobility of free DNA probe and 6His:*Pf*TLP/DNA complexes in native agarose EMSA. In the absence of Mg^{2+}, the complexes were better resolved (compare Fig. 14 A and B, lanes 2 and 3). Additionally, when Mg^{2+} was omitted during electrophoresis, a larger proportion of the DNA probe was bound by 6His:*Pf*TLP (Fig. 14, compare panel A with panels B and C). This indicates that the presence of Mg^{2+} during electrophoresis reduced the stability of 6His:*Pf*TLP/DNA complexes. In the absence of Mg^{2+} ions during electrophoresis, *Pf*TLP showed similar binding activity to a 288 bp GBP-130 promoter region, and a 40 bp region of the adenovirus 2 major late promoter (Ad2ML) (See Supplementary Table 2 for sequences), containing a prototypical

TATA-box element (Fig. 14 B and C). However, in the case of the Ad2ML promoter two distinct complexes were observed, a complex with greatly reduced mobility and a second complex with only slightly reduced mobility when compared to the free DNA probe (Fig. 14 C, lanes 2-3). Together, these results suggest that *Pf*TLP can bind DNA probes of different length and sequence, and that the presence of Mg^{2+} during electrophoresis destabilizes 6His:*Pf*TLP/DNA complexes.

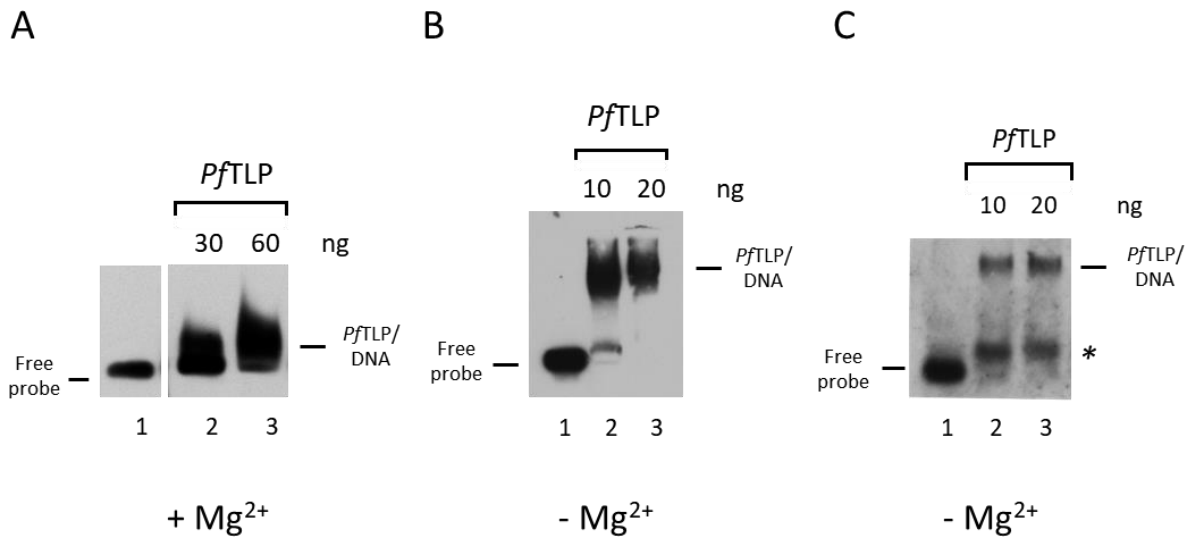


Figure 14: The effect of Mg^{2+} in 6His:*Pf*TLP agarose EMSAs. 10 µl binding reactions containing 5 fmol 288 bp GBP-130 promoter region (A ,B) or 40 bp Ad2ML promoter probes (C) and indicated amounts of 6His:*Pf*TLP were resolved in a 1.4 % agarose gel in the presence (+ Mg^{2+}) or absence (- Mg^{2+}) during electrophoresis. After electrophoresis, the biotinylated DNA was transferred to a nylon membrane and detected on X-ray film via chemiluminescence using a stabilized streptavidin-horseradish peroxidase conjugate. The positions of free DNA probe and 6His:*Pf*TLP/DNA complexes are indicated. A second Ad2ML complex with higher mobility is indicated in (C) with an asterisk.

Protein/DNA interactions are largely electrostatic and therefore generally sensitive to increased salt concentrations. It has been shown that the DNA-binding activity of eukaryotic TBP is decreased at salt concentrations above 60 mM and temperatures above 30 °C (Nakajima, 1988; Petri, 1995). Because the ideal binding conditions for 6His:*Pf*TLP DNA binding is not known, the effect of salt (KCl) concentration and temperature on DNA-binding was investigated. As seen in Fig. 15 DNA-binding of *Pf*TLP was not significantly affected by KCl concentrations up to 460 mM, and incubation temperatures up to 42 °C. However, at 60 °C no binding was detectable, presumably because *Pf*TLP denatures at this temperature. This

shows that *Pf*TLP is far less sensitive to salt when compared to typical eukaryotic TBPs that were previously characterized, and may also be less sensitive to temperature (Nakajima et al., 1988; Petri et al., 1995) Since no significant difference was seen in overall DNA-binding activity, previously established conditions for eukaryotic TBP DNA binding, with 60 mM KCl present in the binding reaction and incubation temperature of 30 °C, were used for further experiments.

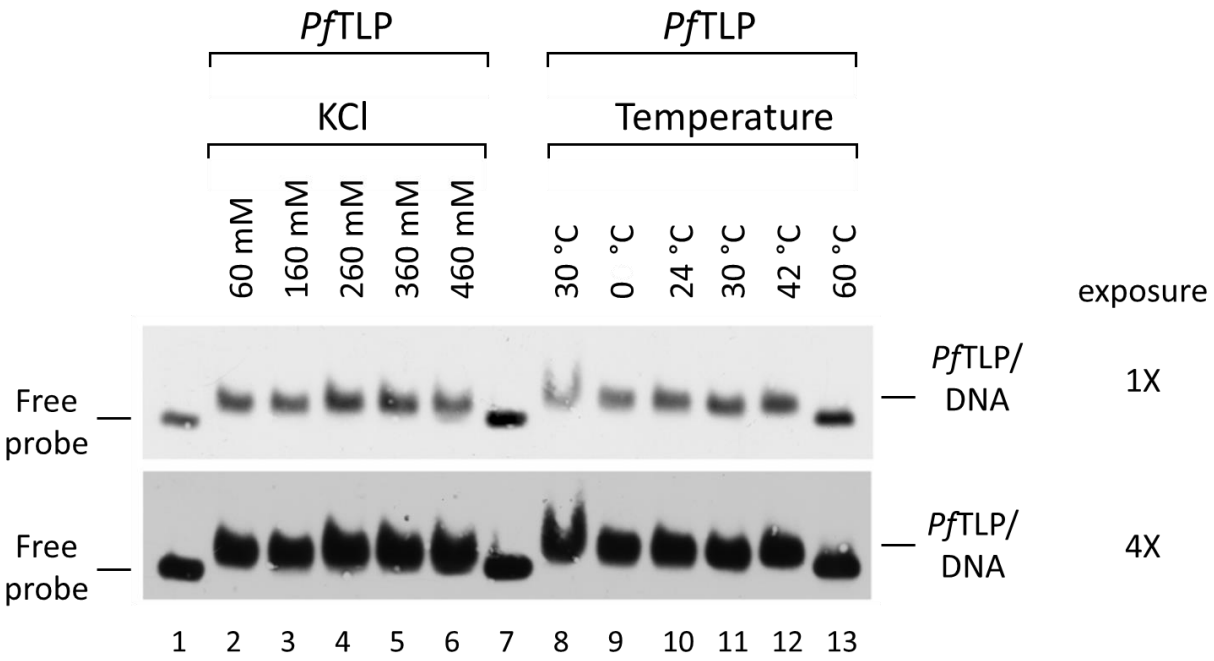


Figure 15: The effect of salt concentration and temperature on the DNA-binding activity of 6His:*Pf*TLP. 10 µl reactions containing 5 fmol 288 bp GBP-130 probe and 30 ng *Pf*TLP and were incubated for 45 min at 30 °C in the presence of indicated concentration of KCl (lanes 2-6) or at 60 mM KCl at indicated incubation temperatures (lanes 8-13). Free probe and protein/DNA complexes were resolved in 1.4% agarose with $MgCl_2$ added to the gel and running buffer. After electrophoresis the DNA was transferred to a nylon membrane and detected on X-ray film using chemiluminescence. Two different exposures of the same experiment are shown. The positions of the free probe and 6His:*Pf*TLP/DNA complexes are indicated.

3.3.1 Role of Evolutionary Conserved Phenylalanine Residues in *Pf*TLP DNA Binding

In *Pf*TLP, F60 and F283 are predicted to correspond to the two phenylalanine residues which insert into the TATA-box promoter element during DNA-binding by TBP (Fig. 6 and 7). To investigate their roles in *Pf*TLP, mutations were introduced, whereby the phenylalanine

residues were substituted with an alanine residue. Expression constructs for single mutants 6His:*Pf*TLP-F60A, -F283A, and double mutant -F60&283A were previously generated in the laboratory (Adebolajo, 2016). These mutant proteins were expressed and purified using the protocol established for wild-type 6His:*Pf*TLP (Chapter 2.2, Fig. 10).

Under standard ITA conditions, with 50 nM DNA, both the double, F60&283A, and single, F60A and F283A, mutant *Pf*TLP proteins bound the immobilized 422 bp GBP-130 probe, with similar activity as the wild-type protein (Fig. 13). This experimental condition did not yield any information about differences in binding affinity or stability in response to the F-to-A mutation. In order to create more stringent binding conditions the DNA concentration in the ITA was decreased to 10 nM and 2.5 nM. Indeed, as shown in Fig. 16, decreasing the amount of DNA in the ITA decreased the overall *Pf*TLP DNA binding activity. Moreover, decreasing the DNA concentration also revealed differences in the binding activity of wild-type and mutant *Pf*TLPs. The DNA binding of the mutant proteins carrying the F60A mutation was clearly diminished when compared to the wild-type *Pf*TLP. In contrast, the binding of the F283A mutant is indistinguishable from the wild-type protein. This suggested that the F60A mutation had a detrimental effect on *Pf*TLP DNA-binding activity.

The DNA binding of the wild-type and mutant *Pf*TLP proteins was further investigated in EMSA experiments in the absence Mg^{2+} added during electrophoresis. As shown in Fig. 17, both the wild-type and the F283A mutant bound the 40 bp Ad2ML promoter with similar activity, whereas, the binding activity of the single, F60A,mutant and the double, F60&283A, mutant was clearly diminished. This result is consistent with the ITA carried out at low DNA concentrations (Fig. 16), where the F60A mutation reduced *Pf*TLP DNA binding activity.

The effects of Mg^{2+}, during EMSA electrophoresis, on the stability of the protein/DNA complexes formed by the F-to-A mutant proteins was further investigated. As seen in Fig. 18A, in an EMSA with Mg^{2+} added during electrophoresis, the binding of *Pf*TLP wild-type and F283A to the 288 bp GBP-130 promoter was clearly detectable, whereas the binding of the single, F60A, and double, F60&283A, mutants was undetectable. This further corroborates the observation that the F60A mutation greatly diminishes DNA binding activity.

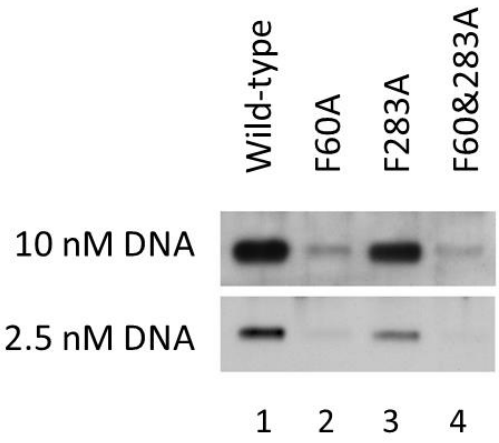


Figure 16: The effect of decreasing amounts of DNA on DNA binding of recombinant *Pf*TLP wild-type and mutant proteins in immobilized template assay (ITA). 25 ng of recombinant 6:His:*Pf*TLP wild-type or mutant proteins carrying the F60A, F283A, and F60&283A substitution were incubated with indicated concentrations of immobilized 422 bp GBP-130 in 40 µl binding reactions. Unbound proteins were washed away with binding buffer and the bound proteins eluted with SDS loading buffer. 50 % of eluates (lanes 1-4) were analysed by immunoblotting and detected with rabbit anti-6His primary antibody and anti-rabbit IgG-peroxidase secondary antibody. Protein bands were visualized on X-ray film using chemiluminescence.

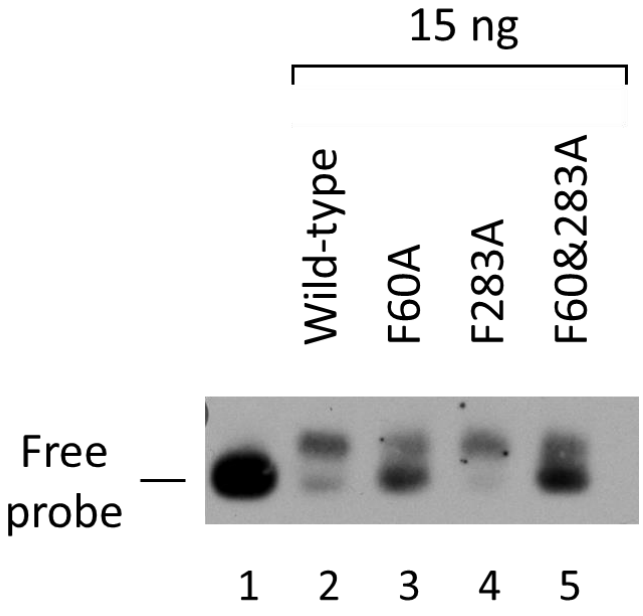


Figure 17: Effects of F-to-A mutations on *Pf*TLP DNA binding. 10 µl EMSA binding reactions containing 5 fmol of 40 bp Ad2ML promoter with 15 ng of 6His:*Pf*TLP wild-type (lane 2), F60A (lane 3), F283A (lane 4), and F60&283A (lane 5) were resolved in a 1.4 % agarose gel in the absence of Mg^{2+}. After electrophoresis, DNA was transferred to a nylon membrane and detected via chemiluminescence on X-ray film. The positions of free DNA probe is indicated.

In contrast, in EMSAs without Mg^{2+} present during electrophoresis, the binding activity of all mutant *Pf*TLP proteins to the 288 bp GBP-130 promoter were diminished when compared to that of wild-type *Pf*TLP (Fig. 18B, compare lanes 2, 5, 8, and 11). Furthermore, under these conditions, F60A and F283A point mutations diminished *PfT*LP binding activity to a similar extent and in a cumulative fashion, with the binding activity of the F60&283A double mutant further reduced compared to the F60A and F283A single mutant *Pf*TLP proteins. This observation suggests that both F60 and F283 are important for *Pf*TLP DNA binding. However the observation that the presence of Mg^{2+} during electrophoresis destabilized the *Pf*TLP/DNA complex for the mutants carrying the F60A mutation suggests that, while both F60 and F283 contribute to *Pf*TLP DNA binding activity, the stability of the *Pf*TLP/DNA complexes is critically dependent on F60, but not on F283.

Interestingly, in EMSA experiments carried out in the absence of Mg^{2+} during electrophoresis, the mobility of the wild-type *Pf*TLP/DNA complex appeared reduced when compared to that of the mutant *Pf*TLP proteins (Fig. 18B). For this reason, an additional experiment was conducted to investigate the differential complex mobility by comparing the *Pf*TLP/DNA complexes formed in the presence of saturating amounts of wild-type and mutant *Pf*TLPs side-by-side. The result of the experiment (Fig. 19) confirms that, in an EMSA without Mg^{2+} during electrophoresis, the wild-type *Pf*TLP/DNA complex has a lower mobility when compared to the mobility of the single F60A and F283A, and double F60&283A mutants. This result might indicate different conformations of nucleoprotein complexes formed with wild-type and mutant *Pf*TLPs.

A competition experiment was conducted to further investigate the contribution of the F283 residue on the stability of the *Pf*TLP/DNA complexes. Reactions containing saturating amounts of *Pf*TLP wild-type and F283A, which results in the quantitative binding of the 288 bp GBP-130 promoter DNA probe, was challenged by a non-specific competitor DNA. Poly(dG-dC) was chosen as an ideal non-specific competitor, given the A/T-rich genome of *P. falciparum*. The experiment was conducted under more stringent conditions, with the addition of Mg^{2+} to the gel and running buffer. As shown in Fig. 20, binding of both wild-type *Pf*TLP and the single F283A mutant to the 288 bp GBP-130 promoter DNA probe decreased with increasing amounts of poly(dG-dC) competitor. However, the presence of 25ng poly(dG-dC) (~ 77 pmol nucleotides) did not significantly affect wild-type *Pf*TLP DNA

complexes, but was enough to recover free probe in the presence of the *Pf*TLP F283A mutant. (Fig. 20, compare lanes 1, 5 and 6). Considering that the standard binding reaction only contains 5 fmol of 288 bp GBP-130 probe (~2.9 pmol, 82.5% A/T), that means a ~27 fold excess of competitor DNA was required to compete for binding of the F283A mutant, whereas ~54 fold excess was required to compete for the binding of the wild-type *Pf*TLP. This result suggests that *Pf*TLP has a preference for A/T rich DNA sequences, which is diminished by the F283A mutation.

Together these results show that wild-type *Pf*TLP has DNA-binding activity, and forms complexes with various DNA probes. *Pf*TLP/DNA complexes are disrupted in an EMSA experiment when Mg^{2+} is present during electrophoresis. The DNA-binding activity of wild-type *Pf*TLP was shown not to be sensitive to high concentrations of salt, which may indicate that the binding is mostly mediated by hydrophobic interactions. The *Pf*TLP/DNA complexes are also stable under higher temperatures, which differs from characterized eukaryotic TBPs. The F-to-A mutations in the evolutionary conserved phenylalanine residues revealed that the phenylalanine residues are not required for general DNA binding activity, but affect DNA complex stability, presumably at A/T-rich target sequences. Overall, the F60A mutation affected the stability of the *Pf*TLP/DNA complexes to a much greater extent than the F283A mutation. Of note, is that the DNA complexes formed with F-to-A mutant *Pf*TLP proteins showed increased mobility in agarose EMSAs when compared to the wild-type *Pf*TLP, this indicating that these mutations affect the overall conformations of the *Pf*TLP/DNA complexes.

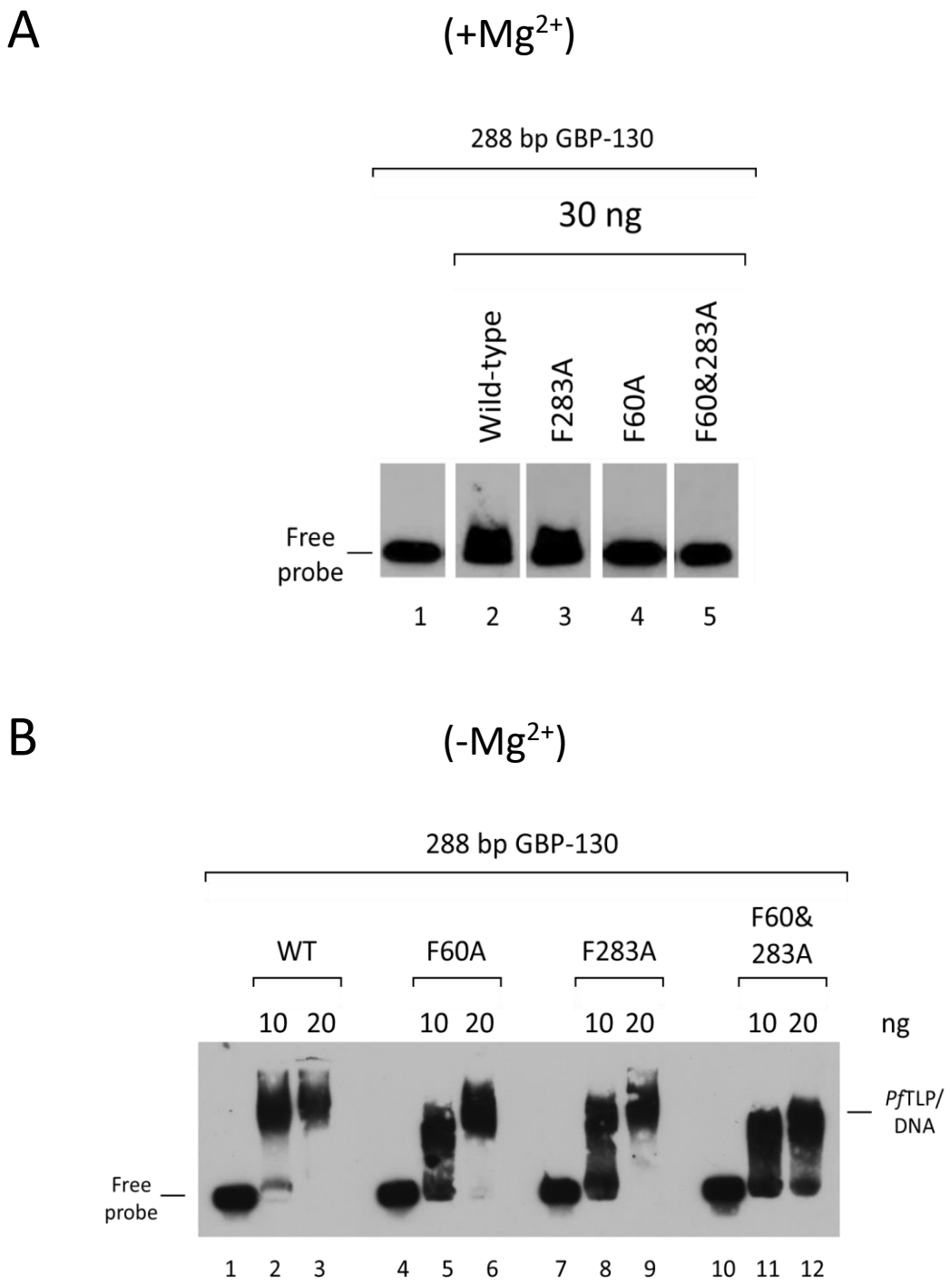


Figure 18: Effects of F-to-A mutations on *Pf*TLP in agarose EMSAs in the presence and absence of Mg^{2+} during electrophoresis. (A) 10 µl EMSA binding reactions containing 5 fmol of 288 bp GBP-130 promoter DNA probe and 30 ng of 6His:*Pf*TLP wild-type (lane 2), and F60A (lane 3), F283A (lane 4) and F60&283A mutant variants (lane 5) were resolved in a 1.4% agarose gel with Mg^{2+} present during electrophoresis. (B) EMSA conditions identical to (A) were electrophoresed in 1.4 % agarose gel without Mg^{2+} present during electrophoresis. After electrophoresis, DNA was transferred to a nylon membrane and detected via chemiluminescence on X-ray film. The positions of free DNA and *Pf*TLP/DNA complexes are indicated.

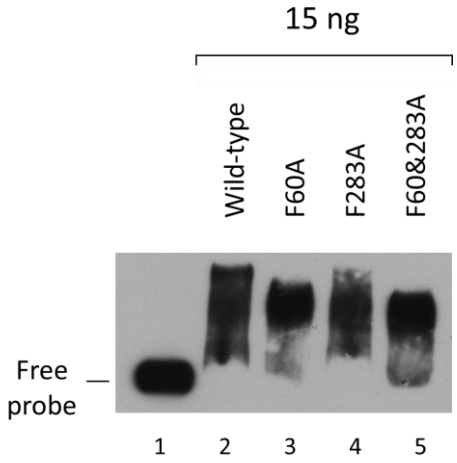


Figure 19: Differential mobility of DNA complexes formed with recombinant *Pf*TLP wild-type and F-to-A mutant proteins. 10 µl EMSA binding reactions contained 5 fmol of 288 bp GBP-130 promoter DNA probe and indicated amounts of 6His:*Pf*TLP wild-type (lane 2), F60A (lane 3), F283A (lane 4), and F60&283A (lane 5). Protein/DNA complexes were resolved in a 1.4% agarose gel in the absence of Mg^{2+}. After electrophoresis, DNA was transferred to a nylon membrane and detected via chemiluminescence on X-ray film. The position of free DNA probe is indicated.

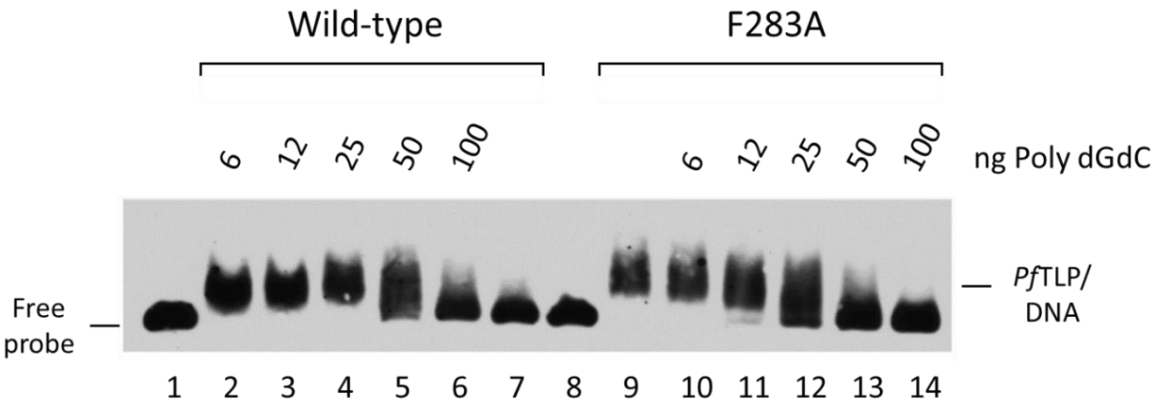


Figure 20: The effect of non-specific competitor DNA on DNA-binding of recombinant *Pf*TLP wild-type and F283A. 10 µl EMSA binding reactions contained 5 fmol 288 bp GBP-130 promoter DNA probe and 60 ng of 6His:*Pf*TLP wild-type (lanes 2-7) and F283A mutant (lanes 9-14) with indicated amounts of the competitor DNA poly(dG-dC). Binding reactions were resolved in a 1.4% agarose gel in the presence Mg^{2+}. After electrophoresis, DNA was transferred to a nylon membrane and detected via chemiluminescence on X-ray film. The positions of free DNA probe and 6His:*Pf*TLP/DNA complexes are indicated.

3.4 Characterization of *Pf*TLP Low Complexity Regions

3.4.1 Bioinformatics

Bioinformatics analysis revealed that *Pf*TLP contains two low complexity regions, LCR1 and 2, which are located between β-strands S3 and S4, and S3' and S4' within the two intra-molecular repeats of the TBP core structure (Fig. 6B, Bing, 2014; Milton, 2017) .

LCRs often coincide with intrinsically disordered regions (IDR). Several different programs and algorithms have been developed to identify and characterize low-complexity (LCR) and intrinsically disordered regions (IRDs). A powerful tool to predict disorder based on amino acid sequence is *DisMeta* (Huang et al., 2014). *Dismeta* is a meta-server that collates and curates the outputs from various disorder prediction programs. The output from *DisMeta* is a consensus disorder prediction for every amino acid residue position, based on the predictions from eight different sequence-based disorder prediction programs. The server also includes the SEG algorithm, the original low-complexity prediction program designed by Wootton and Federhen at the National Center for Biotechnology Information (NCBI) (Wootton and Federhen, 1993). As seen in Fig. 21, the output from *DisMeta* identified both LCR1 and LCR2 as regions with a high probability of disorder. These regions were also categorized as low-complexity through the SEG algorithm (data not shown).

Once identified as LCR/IDR, the sequence composition, bias, and patterning of the *Pf*TLP LCRs was investigated. The relationship between protein sequence composition, propensity to form disordered regions, and phase separating behaviour was previously analysed by sorting the amino acid residues of a given sequence into set categories, specifically highlighting polar, hydrophobic, charged, and aromatic residues and comparing their ratios (Martin and Mittag, 2018). For example, glycine is generally overrepresented in these IDRs and has been shown to play a role in the phase separating behaviour of many IDRs. The process of categorizing amino acid residues was applied to both *Pf*TLP LCR1 and LCR2, in order to investigate the amino acid bias in these two regions. As seen in Fig. 22 both LCRs contain a large proportion of charged residues, as well as randomly distributed aromatic residues (Fig. 22). Both *Pf*TLP LCR1 and 2 are enriched in polar residues such as asparagine and serine residues (Fig. 22). In addition, *Pf*TLP LCR1 contains repeating units of asparagine and serine residues in various

combinations (Fig. 22, A). Interestingly, glycine is underrepresented in both sequences (Fig. 22). An abundance of charged and aromatic residues was correlated earlier with a high propensity for phase separation (Martin & Mittag, 2018). The lack of glycine in *Pf*TLP LCR1 and LCR2 may be explained by the fact that the codon for glycine is GC-rich (GGN). Thus, in the context of the AT-rich *P. falciparum* genome, glycine may have been substituted with the AT-rich codon of asparagine (AAT, AAC).

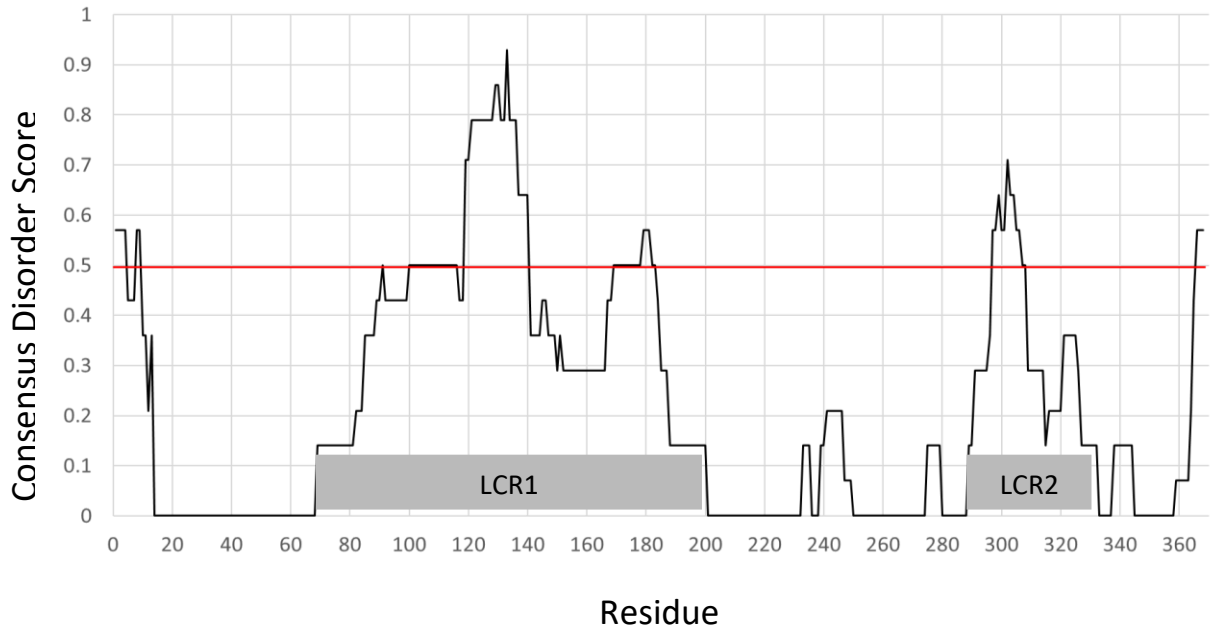


Figure 21: Disorder prediction profile of *Pf*TLP. Output from *DisMeta* provides a disorder prediction for every amino acids residue (Huang et al., 2014), based on the scores from eight different sequence-based disorder prediction programs. The standard threshold for a high probability of disorder is identified at 0.5 with a red line. The positions of *Pf*TLP LCR1 and LCR2 are indicated.

In order to establish whether the observed bias in amino acid use towards asparagine and serine residues was specific to *Pf*TLP LCRs, or simply a result of the general AT rich codon bias in the *P. falciparum* genome, the amino acid use in the *Pf*TLP LCRs was compared to that of the structured core domain of *Pf*TLP and the entire *P. falciparum* proteome (Fig. 23, Yadav & Swati, 2012). As seen in Fig. 23, the *P. falciparum* proteome is enriched in asparagine (14%) and lysine (12%) residues compared to other amino acids. Both asparagine and lysine residues have AT-rich codons, and therefore this enrichment might be a result of the overall genome AT bias. The amino acid composition of *Pf*TLP core domain roughly reflects that of the *P. falciparum* proteome, with a few exceptions. The core domain is slightly enriched in branched aliphatic amino acids, such as isoleucine (10.7%) and valine (6.8%), whereas

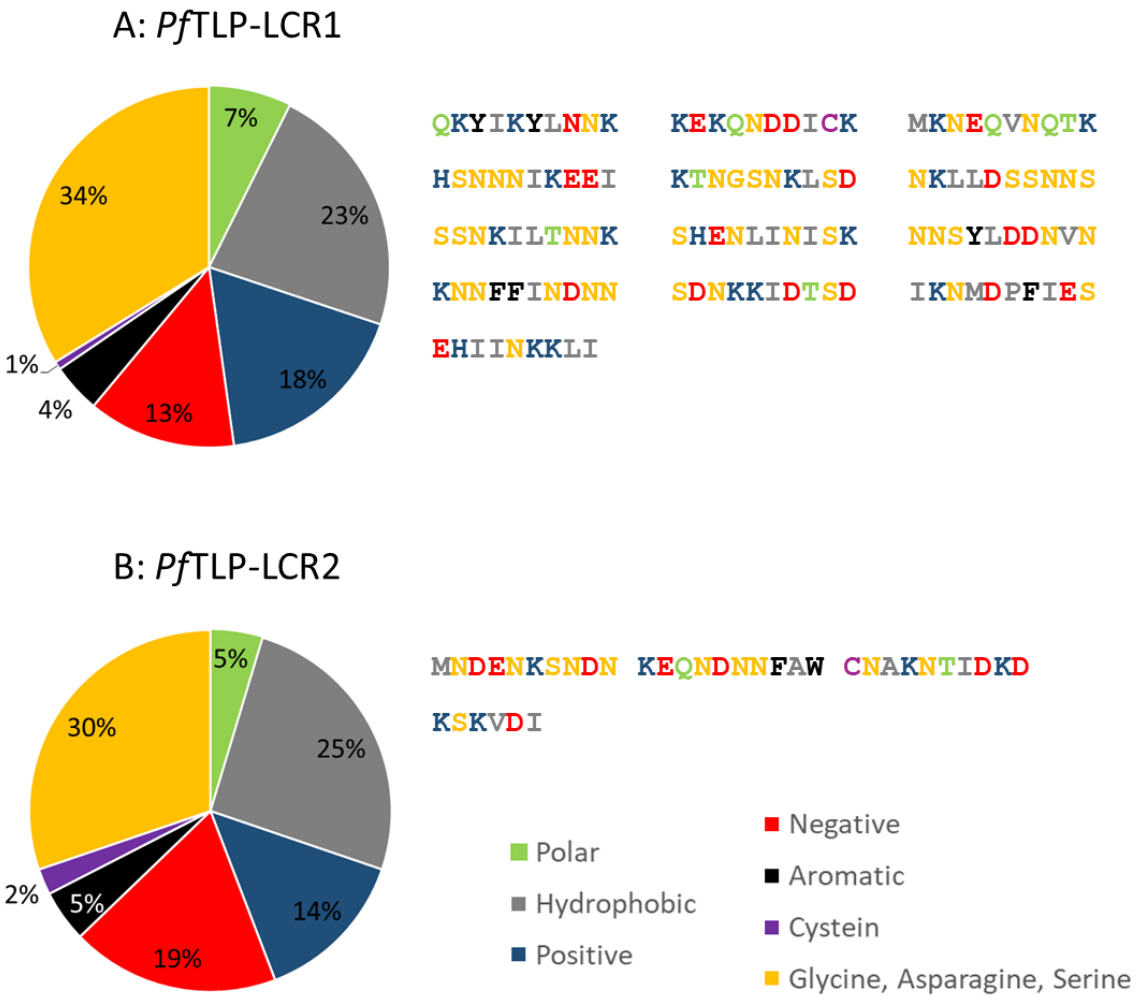


Figure 22: Amino acid composition of *Pf*TLP LCRs. The amino acid sequence of *Pf*TLP LCR1 (A) and LCR2 (B), are shown with amino acid highlighted according to the colour key. Pie charts represent the percentage of each category present in the respective full length LCR sequence. Data generated with ProtoParam (Gasteiger et al., 2005).

negatively charged amino acids like aspartic and glutamic acid are underrepresented (2.4% and 3.4% respectively) (Fig. 23). In the DNA-binding domain of eukaryotic TBP branched hydrophobic amino acids contact base pairs inside the minor groove of the DNA. Furthermore, the DNA binding surface of TBP is also characterized by a net positive surface charge in order to facilitate interaction with the negatively charged DNA backbone (Kim and Burley, 1994; Kim et al., 1993; Nikolov et al., 2002). Thus, the observed enrichment in branched aliphatic amino acid residues and underrepresentation of negatively charged amino acid residues is consistent with the result of the Phyre2 analysis which predicts that *Pf*TLP adopts a TBP-like structured core domain. *Pf*TLP LCR1 and LCR2 are both enriched for asparagine (23.3% and 25%), lysine (16.3% and 16.7%) and serine (11.6% and 11.1%) residues when compared to the *Pf*TLP core domain and *P. falciparum* proteome. Interestingly LCR2 is also specifically enriched in aspartic acid (16.7%) (Fig. 23), and this contributes to the overall negative charge of LCR2. Taken together, this analysis shows that the amino acid content of *Pf*TLP LCR1 and LCR2 is different to the *Pf*TLP structured core domain and the entire *P. falciparum* proteome as LCR1 and LCR2 are specifically enriched in certain amino acid residues, and that this bias is not merely a result of the *P. falciparum* genome-wide codon bias.

To address the question of whether these LCRs are a general feature of TLPs in *Plasmodium,* multiple sequence alignments were performed using putative TLP orthologues from various *Plasmodium* species, which were identified using a BLAST search. Multiple sequence alignment of putative *Pf*TLP orthologues revealed the presence of LCRs in the same positions as *Pf*TLP LCR1 and LCR2 (full alignment available in Supplementary Fig. 4). *DisMeta* analysis confirmed that the LCR regions in *Plasmodium* species *Pf*TLP orthologues are also predicted to be highly disordered (Fig. 24). These observations suggests that the presence of two LCRs with a propensity to form disordered regions, located between β-strands S3/S4 and S3'/S4' within the structured *Pf*TLP core domain, is a general feature of *Plasmodium* TLPs.

Next, the sequence conservation of the LCR sequences in the TLP orthologues from different *Plasmodium* species was investigated. LCR1 sequences could not be meaningfully aligned by multiple sequence alignment, suggesting that there is very low sequence conservation in the LCR1 sequences from different *Plasmodium* species (data not shown). An exception to this was the *Laverania* subgenus, which includes the closely related primate-specific *P. falciparum*,

P. reichenowi, and *P. gaboni*, that showed a high degree of sequence homology (Loy et al., 2017). In contrast, there is a high level of conservation between the LCR2 sequences from various *Plasmodium* species, with a central (F/Y)AW motif conserved in all orthologues analysed (Fig. 25). As with LCR1, there is a higher degree of homology between LCR2s from the subgenus *Laverania.* (Fig. 25, Loy et al., 2017). Overall evolutionary conservation of sequence for *Plasmodium* TLP LCR1s could not be detected, but there is a high level of sequence conservation for the LCR2 regions from *Plasmodium* TLP orthologues. This might suggest that the LCR2 region serves a specific function that is conserved within *Plasmodium*.

Previous studies have shown that the tendency for intrinsic disorder and the ability to phase separate is not mediated by specific amino acid sequence, but rather by regions of a particular amino acid composition (Brangwynne, Tompa & Pappu, 2015; Martin & Mittag, 2018; Turoverov et al., 2019). Therefore, further analysis was done to compare the amino acid compositions of the LCR1 regions from the various *Plasmodium* orthologues. As shown in Fig. 26 these regions have considerable variation in the amino acid compositions between different species. This variation may be a consequence of the difference in GC content of the various *Plasmodium* genomes, which in turn may bias amino acid codon usage. To further illustrate the variation in amino acid distribution, the amino acid composition of the TLP LCR1 regions from *P. falciparum*, with a genome GC content of 19% (Carlton et al., 2008), was compared to that of *P. vivax*, with a genome GC content of 57% (Carlton et al., 2008). As seen in Fig. 27, *Pf*TLP LCR1 is enriched in asparagine (23.3%), lysine (16.3%), serine (10.9%), isoleucine (10.1%), and aspartic acid (8.5%) residues. In contrast, *Pv*TLP LCR1 is enriched in proline (9.9%), threonine (21.8%), glycine (15.8%), glutamic acid (16.8%), and glutamine (10.9%) residues. It is interesting to note that *Pf*TLP LCR1 is specifically enriched in the phosphor acceptor serine, whereas *Pv*TLP LCR1 is specifically enriched in the phosphor acceptor threonine. Similarly, the major charge carrying residue in *Pf*TLP LCR1 is lysine, whereas in *Pv*TLP LCR1 it is glutamic acid. Interestingly, glycine is overrepresented in *Pv*TLP LCR1, which is a hallmark generally associated with LCRs with a high propensity to phase separate (Martin & Mittag, 2018). Together, these results indicate that there is little conservation in sequence or amino acid composition within the LCR1 regions of the various *Plasmodium* TLP orthologues. However, certain features, for example the enrichment of

potential phosphor acceptor residues, appear to be conserved. This may point to a potential role of phosphorylation in the regulation of *Pf*TLP function through these regions.

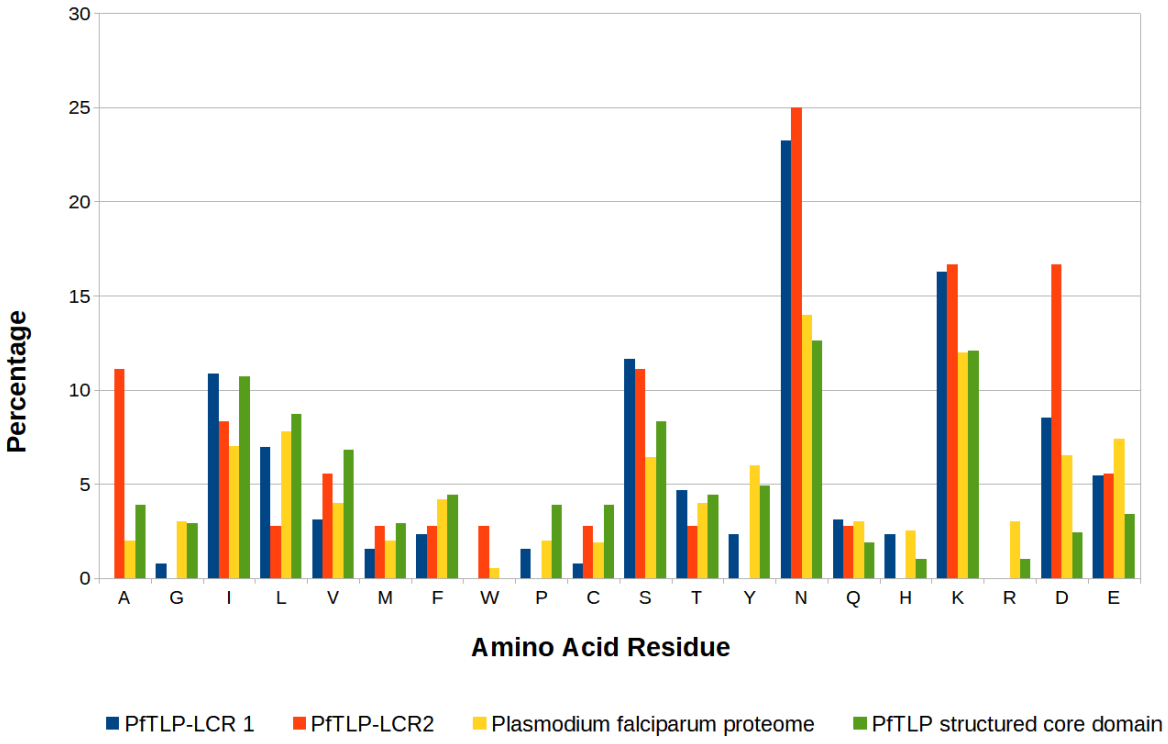


Figure 23: Amino acid usage in *Pf*TLP and the *P. falciparum* proteome. Percentage amino acid residue use in *Pf*TLP-LCR1 (blue) and -LCR2 (orange), *Pf*TLP structured domain (green) and the entire *P. falciparum* proteome (yellow) are shown. Proteome data were adapted from (Yadav and Swati, 2012)**.** *Pf*TLP data generated with ProtoParam (Gasteiger et al., 2005).

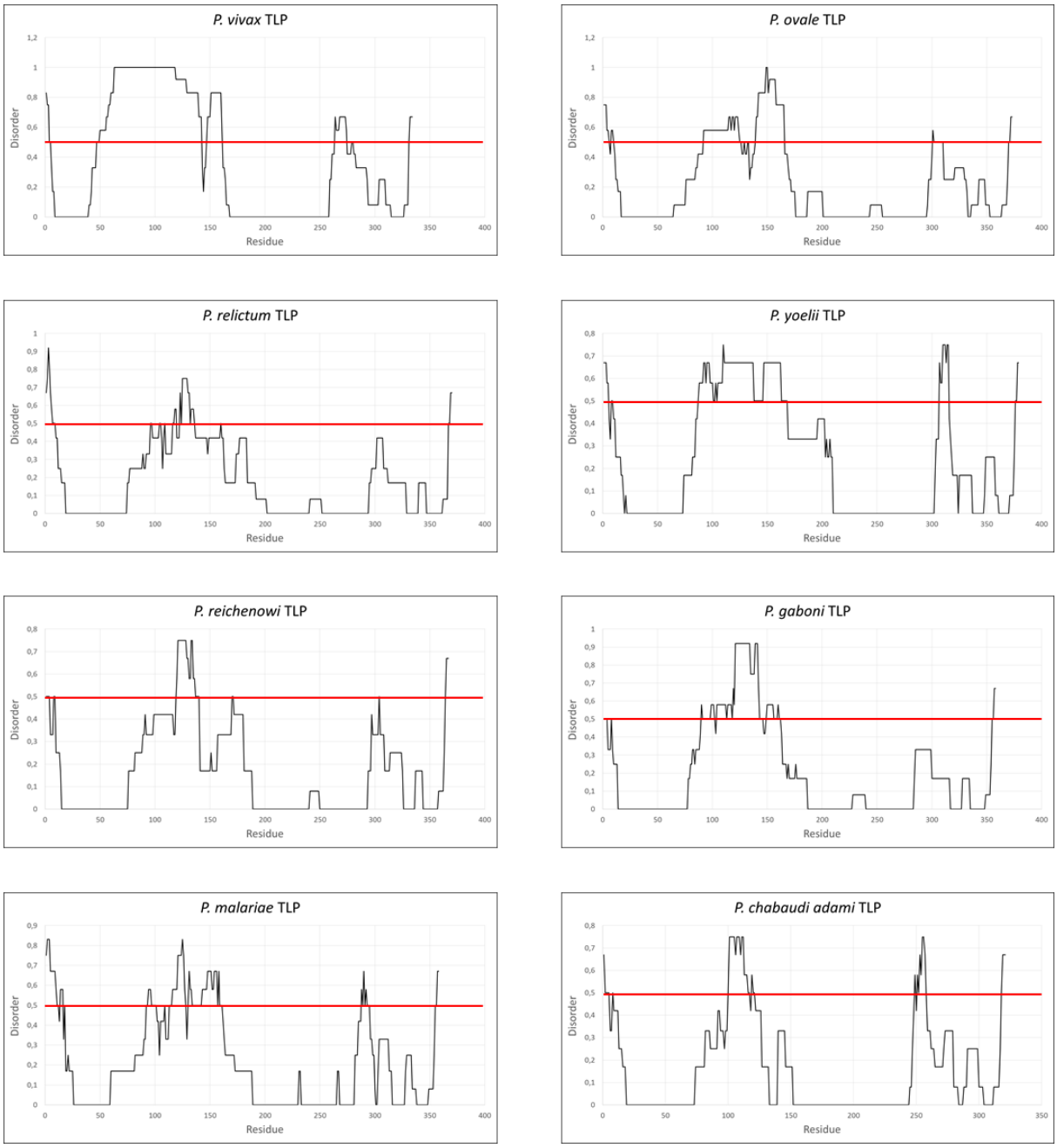


Figure 24: Consensus disorder prediction for TLP orthologues from various *Plasmodium* species. The output from *DisMeta* (Huang et al., 2014) shows the consensus disorder prediction for every amino acid position, based on the scores from eight different sequence-based disorder prediction programs for every *Plasmodium* TLP orthologue sequence. The standard threshold for disorder is identified at 0.5 with a red line. Two peaks of heightened disorder can be observed in all the *Plasmodium* TLPs corresponding to LCR1 and LCR2 in *Pf*TLP.

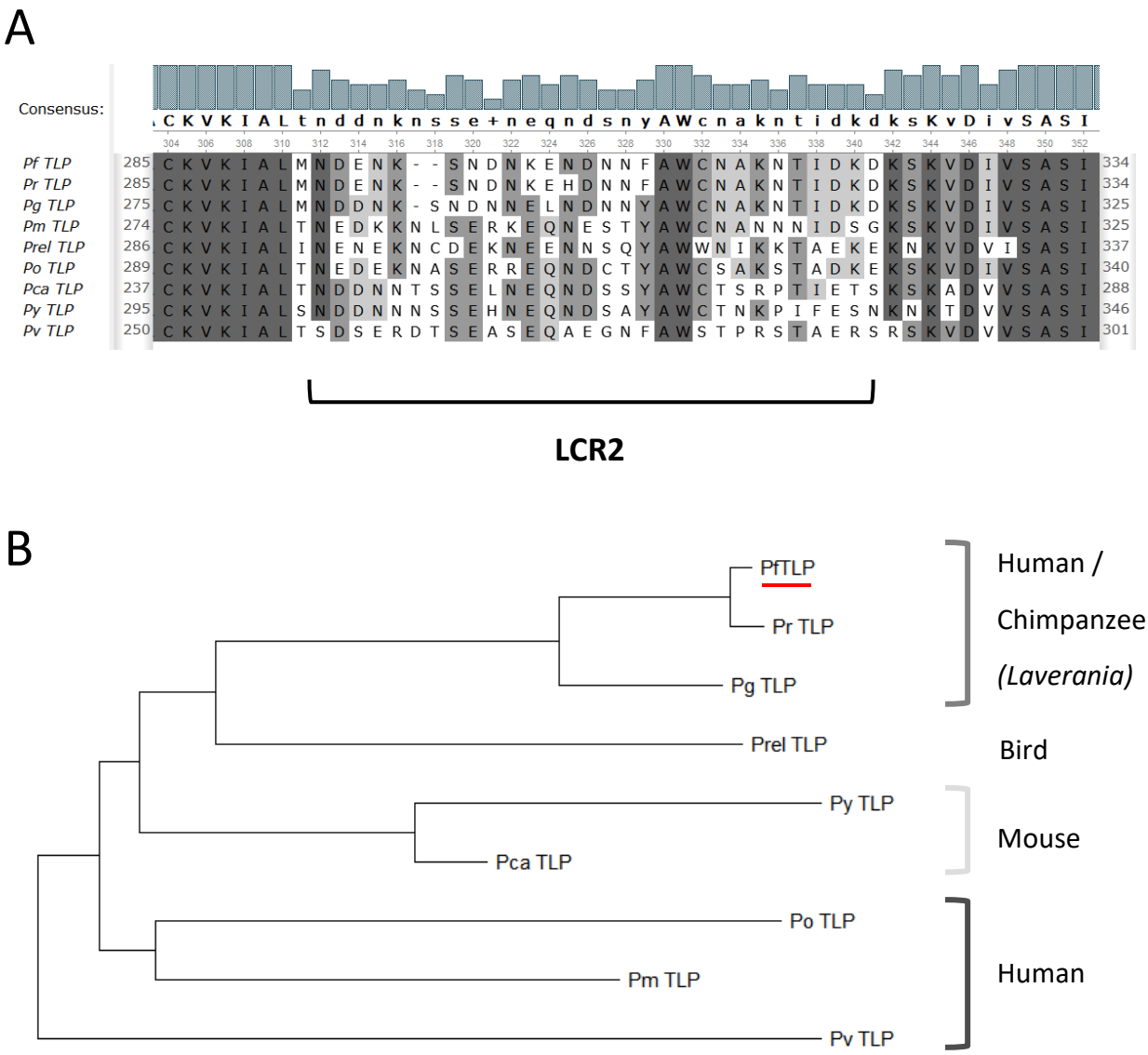


Figure 25: Multiple sequence alignment and phylogenetic tree analysis of TLP2 regions in various *Plasmodium* species. A) Multiple sequence alignment of the TLP LCR2 regions of various *Plasmodium* species. Alignment done with MUSCLE (Edgar, 2004) and visualized in UGENE (Okonechnikov et al., 2012). Light grey highlighting represents residues with a > 50% identity, medium grey residues with > 80% identity, and dark grey residues with 100% identity. B) A maximum-likelihood phylogenetic tree of the TLPs from various *Plasmodium* species was rendered through UGENE (Okonechnikov et al., 2012), based on the multiple sequence alignment results, generated by MUSCLE (Edgar, 2004). Bootstrap test (100 replications) was performed. Brackets are used to indicate the various hosts of the different *Plasmodium* species, with the primate-specific subgenus of *Laverania* indicated. *Pg = P. gaboni, Pr = P. reichenowi, Pm = P.malariae, Po = P.ovale, Prel = P. relictum, Pv = P.vivax, Py = P. yoelii, Pca = P. chabaudi adami, and Pf = P. falciparum*.

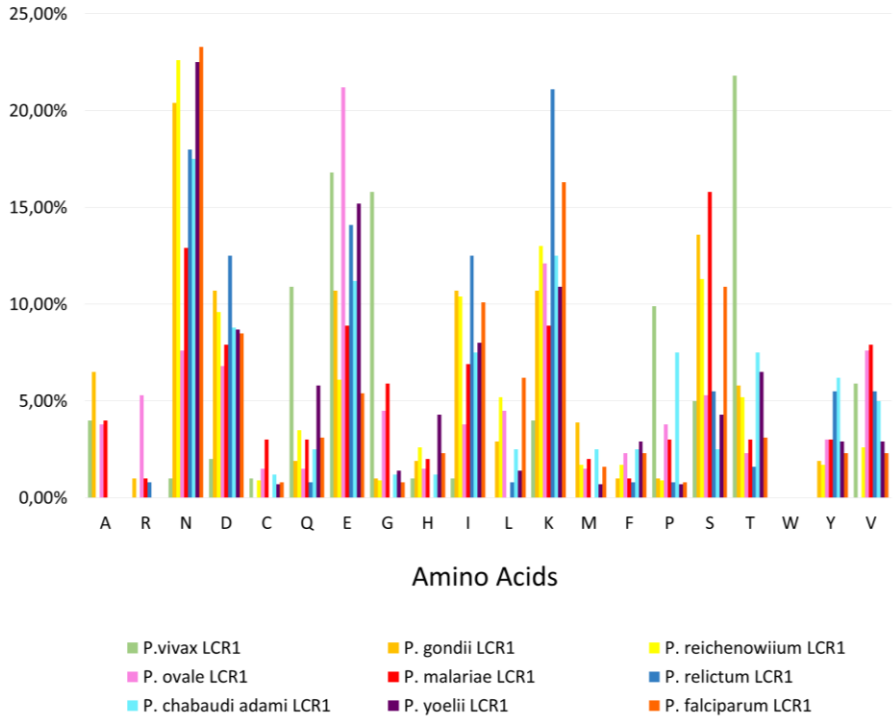


Figure 26: Percentage amino acid use in the LCR1 region of various *Plasmodium* TLP orthologue proteins. Data generated with ProtoParam (Gasteiger et al., 2005).

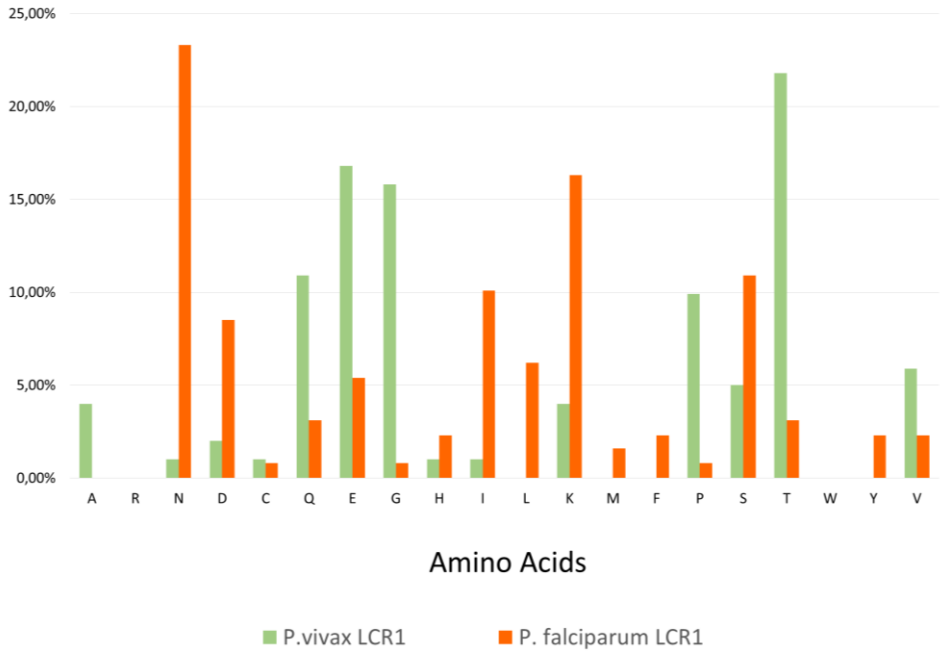


Figure 27: Percentage amino acid use in *P. falciparum* TLP-LCR1 (orange) compared to *P. vivax* TLP-LCR1 (green). Data generated with ProtoParam (Gasteiger et al., 2005).

3.4.2 Generation of LCR fusion Proteins

In order to investigate the potential of the *Pf*TLP LCRs to drive liquid-liquid phase separation (LLPS), recombinant 6His tagged versions of LCRs fused to fluorescent protein were generated (Fig. 28A), bacterially expressed and purified from bacterial lysates by metal-affinity chromatography (Materials and Methods, Chapter 2.5). An SDS PAGE analysis of purified LCR fluorescent fusion proteins is shown in Fig. 28B.

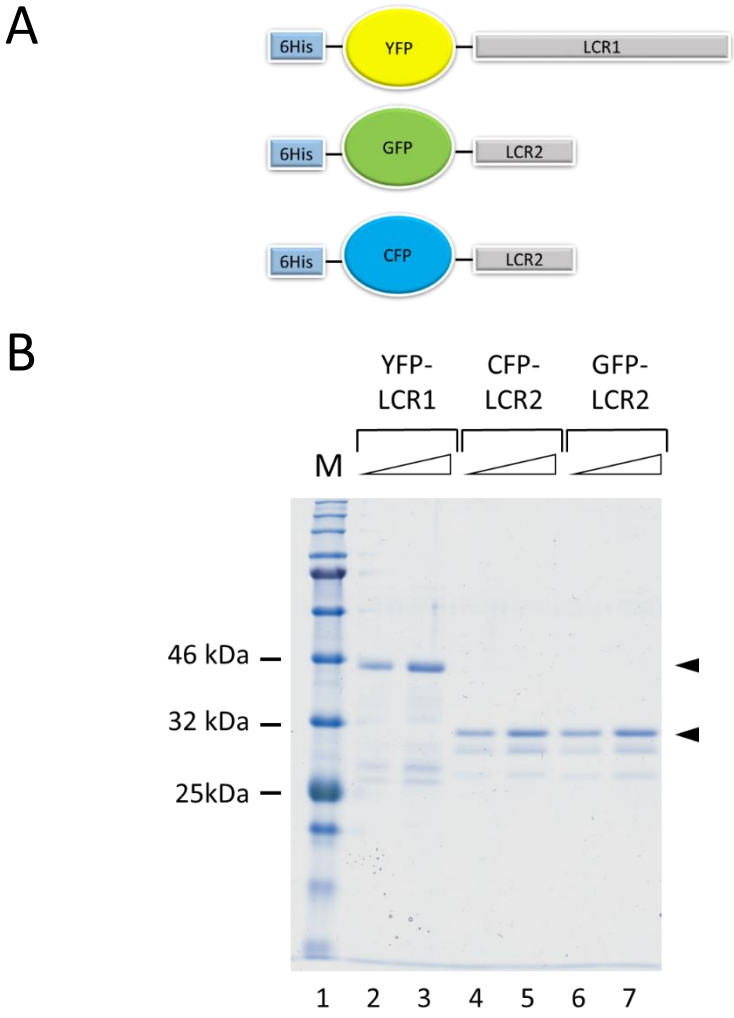


Figure 28: Recombinant bacterially expressed 6His:tagged *Pf*TLP LCR-fluorescent fusion proteins. (A) Schematic of the *Pf*TLP LCR fusion proteins constructs. N-terminal 6His tagged green, yellow or cyan fluorescent protein (GFP, YFP, and CFP) was fused to the N- terminus of either *Pf*TLP LCR1 or *Pf*TLP LCR2. (B) SDS PAGE analysis of purified recombinant 6His-tagged TLP LCRs fused to fluorescent proteins. 125 ng and 250 ng of purified recombinantly expressed 6His:YFP-*Pf*TLP-LCR1 (lanes 2-3), 6His:CFP-*Pf*TLP-LCR2 (lanes 4-5), and 6His:GFP-*Pf*TLP-LCR2 (lanes 6-7) were analysed by 12% SDS PAGE and visualized by Coomassie staining. The positions of selected molecular weight standards are indicated. The calculated molecular mass of LCR1 and LCR2 fusion proteins are 38 kDa and 33 kDa, respectively. M, molecular weight marker, YFP; yellow fluorescent protein, CFP; cyan fluorescent protein, GFP; green fluorescent protein.

3.4.3 Phase Separation Assays

In order to test if *Pf*TLP-LCR1 and -LCR2 are able to drive LLPS the YFP-*Pf*TLP-LCR1 and CFP-*Pf*TLP-LCR2 fusion proteins were used in phase separating assays, based on published methodology. Briefly, proteins are mixed into a phase separation buffer, which has a set pH, salt concentration, and percentage crowding agent such as polyethylene glycol (PEG) or dextran (Alberti et al., 2018; Wang et al., 2018). The reactions are then examined on modified glass slides by fluorescent microscopy (see Materials and Methods, chapter 2.6). In the phase separation assays used in this study 14 µM protein was incubated with 7-25% PEG and 150 mM KCl at a pH of 7.4.

In the initial assays YFP-*Pf*TLP-LCR1 or CFP-*Pf*TLP-LCR2 were assayed separately. At a protein concentration of 14 µM protein, and in the presence of 7-15 % PEG no phase separation was observed, and both proteins stayed in solution. However, in the presence of 20-25 % PEG YFP-*Pf*TLP-LCR1 formed spherical droplets with increased fluorescence, that were also visible in phase contrast or differential interference contrast (DIC) microscopy (Fig. 29, panel A). In contrast, LLPS droplet formation could not be observed with 14 µM CFP-*Pf*TLP-LCR2 in the presence of 25% PEG, (Fig. 29, panel A), and the protein remained uniformly distributed in solution. Thus *Pf*TLP-LCR1, but not *Pf*TLP-LCR2, is able to undergo LLPS under these conditions.

Proteins able to drive LLPS and the formation of membraneless organelles on their own are referred to as scaffold proteins, whereas proteins that are selectively recruited to these phase separated droplets are called client proteins (Bergeron-Sandoval and Michnick, 2018; Ditlev et al., 2018; Posey et al., 2018). The recruitment of these client proteins to phase separated droplets might be mediated through modular protein domains or LCRs (Ditlev et al., 2018). In order to test if CFP-*Pf*TLP-LCR2 could be recruited into LLPS droplets formed by YFP-*Pf*TLP-LCR1, YFP-*Pf*TLP- LCR1 and CFP- PfTLP-LCR2 were combined at a concentration of 14 µM in the presence of 25 % PEG. Fluorescent microscopy revealed formation of droplets containing YFP but lacking CFP (Fig. 29, panel B). CFP-*Pf*TLP-LCR2 remained uniformly in solution. Interestingly, the droplet size of YFP- *Pf*TLP-LCR1 increased significantly with time (Fig. 29, panel C). However, even after prolonged incubation times recruitment of CFP-*Pf*TLP-LCR2 into the droplets formed by YFP-*Pf*TLP-LCR1 was still undetectable. Taken together,

these results provide first evidence that *Pf*TLP-LCR1 can undergo LLPs. However, a role of *Pf*TLP-LCR2 in LLPS could not be observed under these conditions.

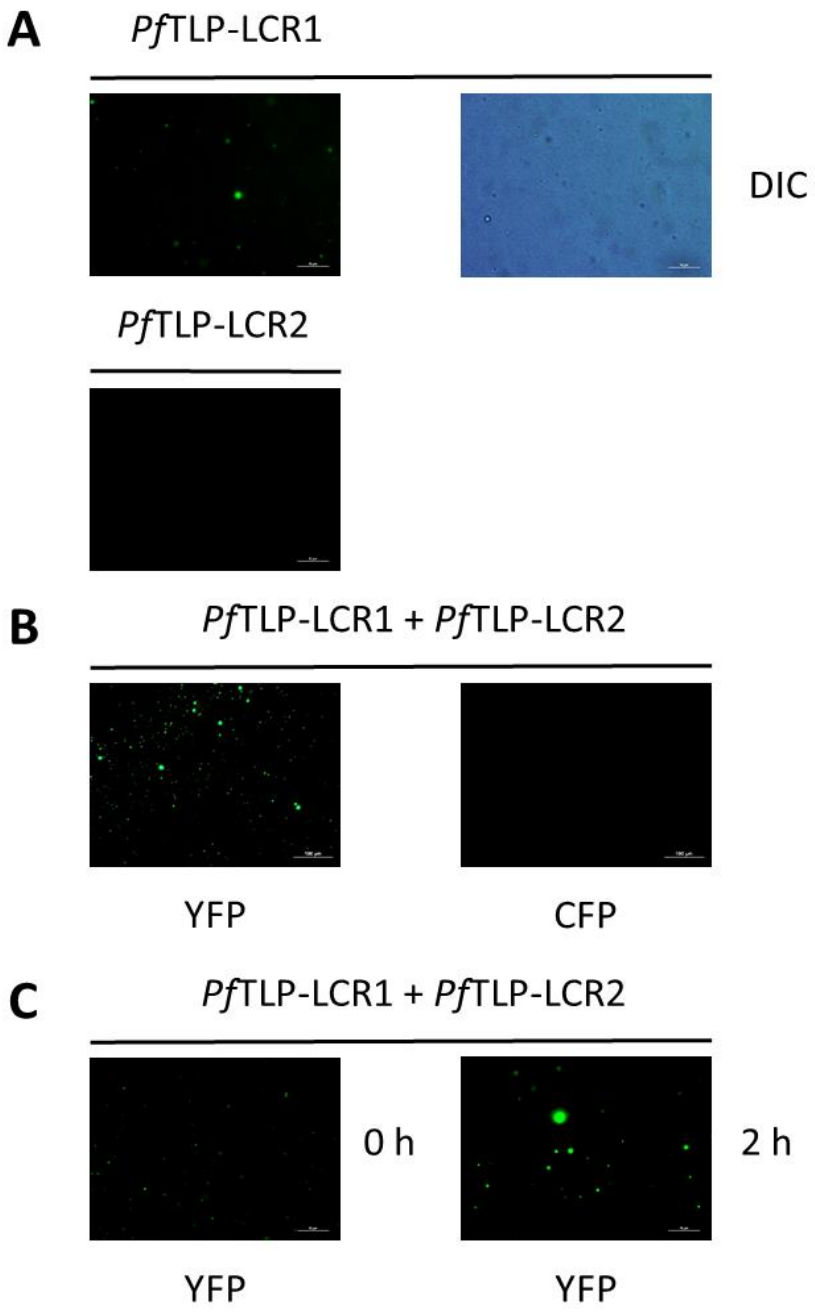


Figure 29: *Pf*TLP LCR1 mediates liquid-liquid phase separation. Microscopy images of phase separation experiments. (A) 14 µM YFP-*Pf*TLP-LCR1 or CFP-*Pf*TLP-LCR2 or (B, C) 14 µM YFP-*Pf*TLP-LCR1 and 14 µM CFP-*Pf*TLP-LCR2 were incubated in the presence of 25 % PEG and 150 mM KCl (pH 7.4). Reactions were visualized at immediately (A, B) or after 2 h incubation at room temperature (C), fluorescent microscopy, using indicated filters for the detection of YFP or CFP, or by differential interference contrast microscopy (DIC, panel A). Panels represent the same reactions, but not the same images. Image exposure adjusted for the detection of phase separated droplets. Scale bar represents 10 µm.

Chapter 4

Discussion

4.1 Bioinformatics analysis of *Pf*TLP

Phylogenetic tree analysis grouped *Pf*TLP with the TRF2-type TLPs. However, *Pf*TLP has four key phenylalanine residues, vital for TBP DNA binding, conserved, whereas typical TRF2 type TLPs lack one or more of these phenylalanine residues. Another difference between *Pf*TLP and typical TRF2-type TLPs is the presence of two low complexity regions that extend out of the core domain structure. These regions are predicted to be intrinsically disordered and mediate LLPS.

A previous study on the evolution of the TRF2-type TLPs postulated that the gene duplication event, which led to the emergence of TRF2-type TLPs, happened prior to the development of the bilateria, but after the evolutionary split between the bilaterian and nonblilaterian animals (Duttke et al., 2014). This is consistent with the observation that typical TRF2-type TLPs, lacking the phenylalanine residues required for TBP DNA binding, are only found in metazoans (Dantonel et al., 1999). It is hypothesized that the emergence of the TRF2-type TLPs allowed for an expansion of gene regulatory programs that facilitated the development of complex body plans (Dantonel et al., 1999; Duttke et al., 2014). The common ancestor of TRF2-type TLPs appears to have emerged in a lineage that branched off from the protists. Furthermore, while there are protist TBP/TLPs that lack the evolutionary conserved phenylalanine residues, these may have resulted from convergent evolution.

Distinct TBP and TBP-like genes have been found in a range of eukaryotic organisms, including plants and protists, suggesting that different organisms evolved unique TBP/TBP-like proteins (Dantonel et al., 1999; Guillebault et al., 2002; Ruan et al., 2004). For example the only TBP protein found in the unicellular eukaryotic organism *Crypthecodinium cohnii* (Guillebault et al., 2002), was found to be structurally closely related to eukaryotic TBPs, with the signature core domain repeats and the interactions sites for TFIIA and TFIIB conserved. However, this TBP lacked the four phenylalanine residues, required for TBP sequence-specific DNA binding,

and also showed only very weak TATA box affinity. Thus, functionally this *C. cohnii* TBP be considered a *C. cohnii* specific TBP-like protein (Guillebault et al., 2002). Similarly, the human parasite *Trypanosoma brucei* lacks a gene for a *bona fide* TBP, but possesses a TBP-like factor, called TRF4. Interestingly, *P. falciparum* TBP, possesses only three of the four evolutionary conserved phenylalanine residues required for TBP DNA binding, with one of the intercalating phenylalanine residues substituted by isoleucine. Furthermore, while *Pf*TBP appears to bind preferentially AT-rich sequences, it lacks TATA box-specific DNA-binding activity (Milton, 2017).

Taken together, these observations suggest *that P. falciparum* may have independently evolved two unique TBP-like proteins with unusual properties, as a means to facilitate the complex life cycle of the parasite. This might indicate that, although phylogenetic analysis based on amino acid sequence homology places *Pf*TLP in the TRF2-type TLP clade, it might in actual fact be a *Plasmodium*-specific TBP-like protein, not functionally related to TRF2-type TLPs identified in metazoans.

4.2 *Pf*TLP DNA binding

Characterization of *Pf*TLP DNA binding activity led to the following observations. First, *Pf*TLP binds stably to DNA sequences even at elevated temperatures of up to 42 °C and at elevated salt concentrations up to 460 mM KCl (Chapter 3.3, Fig. 15).

Eukaryotic TBP binds to the TATA box promoter element through the minor groove of the DNA via an induced fit mechanism (Kim et al., 1993; Nikolov et al., 1992, 1996), slightly unwinding the DNA helix to expose and interact with the base pairs (Kim and Burley, 1994). The TBP/DNA complex is clamped in place through the actions of four phenylalanine residues. Two of these residues intercalate between the first and last base pairs of the TATA box element. These phenylalanine residues will interact with the nucleobases through van der Waals and $\pi - \pi$ interactions to compensate for the loss of base stacking. The intercalation by the phenylalanine residues causes kinks in the DNA and the two additional phenylalanine residues support the kinks by interacting with the ribose backbone through van der Waals interactions. The two kinks compress the DNA towards the major groove, causing a dramatic bend in the DNA of roughly 90° (Chapter 1.2.4, Fig. 2, Kim, Nikolov & Burley, 1993; Kim &

Burley, 1994). Previous studies have shown that substitutions of intercalating phenylalanine residues results in a drastic loss of TBP in vitro DNA-binding activity, suggesting that these residues are essential for TBP DNA binding (Kamenova et al., 2014; Klejman et al., 2005; Zhao et al., 2003). In *Pf*TLP, conserved phenylalanine residues appear not to be required for DNA binding activity, but appear to play an important role in *Pf*TLP/DNA complex stability, with F60 playing a much more important role than F283. Furthermore, the F283 residue appears to contribute to the preferential binding of *Pf*TLP to A/T-rich sequences (chapter, Fig. 20).

4.2.1 *Pf*TLP DNA binding conditions

Previous studies have determined the ideal binding conditions for prototypical TBPs isolated from human, *S. cerevisiae*, and *Drosophila* cells. These studies have revealed that human TBP is completely inactivated by heat treatment at 47 °C for 15 min , however, when complexed with DNA TBP is heat stable up to 70 °C (Coleman and Pugh, 1995; Nakajima et al., 1988). The ideal incubation temperature for TBP DNA binding experiments was determined to be 30 °C, with TBP DNA binding activity diminished in temperatures above or below 30 °C (Petri et al., 1995). In the case of *Pf*TLP no difference in DNA binding activity was seen at incubation temperatures from 0 °C to 42 °C, however at 60 °C, *Pf*TLP appears to denature (Fig. 15). Although the ideal *Pf*TLP DNA binding conditions were not determined in this work, it can be concluded that the standard 30 °C incubation temperature established for TBP DNA binding experiments is still suitable for studying *Pf*TLP DNA binding.

TBP DNA binding involves both hydrophobic as well as electrostatic interactions. TBP binding to the minor groove of the DNA involves hydrophobic interactions with the edges of base pairs, this process is entropically driven due to the release of water molecules during complex formation (Privalov et al., 2007, 2011). In addition there are electrostatic interactions between the overall positively charged DNA binding surface and the negatively charged DNA backbone. The contribution of electrostatic interactions to TBP DNA binding activity is illustrated by the observation that salt concentrations above 60 mM KCl in the binding reaction drastically diminishes TBP DNA binding activity (Petri et al., 1995). In contrast, *Pf*TLP DNA binding was unaffected by high salt concentrations, and stable binding was still detectable at 460 mM KCl. This observation suggests that, compared to TBP, *Pf*TLP DNA binding may be mediated through mostly hydrophobic interactions, and to a lesser extent by

electrostatic interactions. This interpretation is supported by the presence of hydrophobic amino acid residues in the predicted *Pf*TLP DNA binding surface (Chapter, Fig. 6A and 23). Of note, DNA binding by TBP from the hyperthermophilic archaeon *Pyrococcus woesei* at high salt concentrations is also primarily mediated through hydrophobic interactions (O'Brien et al., 1998).

4.2.2 The effects of the phenylalanine residues on *Pf*TLP DNA binding

In *Pf*TLP the core domain architecture of model eukaryotic TBP appears conserved, and is enriched in hydrophobic residues (Chapters 1.2.4). Furthermore, the phenylalanine residues predicted to intercalate into the DNA during TBP sequence-specific DNA binding are conserved in *Pf*TLP. These observations support the idea that *Pf*TLP binds DNA in a similar manner to eukaryotic TBP, through a hydrophobic DNA binding surface and the action of the conserved phenylalanine residues, and imply that *Pf*TLP may also have sequence-specific DNA binding activity. In this study, the two phenylalanine residues that were predicted to intercalate into the DNA during *Pf*TLP sequence specific binding, F60 and F283A, were mutated to alanine residues to produce two single *Pf*TLP mutants, F60A and F283A, and a double mutant, F60&283A.

In DNA binding experiments using EMSAs without Mg^{2+} present during electrophoresis the double mutant F60&283A showed lower DNA binding activity than the two single point mutants, suggesting that both F60 and F283 contribute to some extent to overall *Pf*TLP DNA binding activity. In ITA experiments, the DNA-binding activity of *Pf*TLP carrying the F60A substitution was strongly diminished, whereas the F283A mutation had only little effect. Furthermore, in EMSA experiments with Mg^{2+} present during electrophoresis DNA binding activity of the F60A mutant was undetectable, suggesting that F60 is critically important for *Pf*TLP/DNA complex stability (Chapter 3.3.1). These results contrast earlier observations made for human and yeast TBP, where both phenylalanine residues were found to be equally essential for DNA binding (Kamenova et al., 2014; Klejman et al., 2005; Zhao et al., 2003). Interestingly, a comparison of the crystal structures of *At*TBP2 and *Hs*TBP bound to the Ad2ML TATA box sequence demonstrated differences in the extent of phenylalanine residue intercalation, which in turn affected the trajectory of the DNA exiting the respective DNA/protein complexes (Nikolov et al., 1996). This observation raises two important points.

First, the two phenylalanine residues in a single TBP molecule may not intercalate to the same extent. Second, the degree of base intercalation by conserved phenylalanine residues differs between TBPs from different species, resulting in different degrees of overall DNA bending and, consequently, in a different trajectory of the DNA emanating from the TBP/TATA complex. It seems therefore possible that the extent to which F60 and F283 intercalate into DNA upon DNA binding by *Pf*TLP differs considerably or that only the F60 residue intercalates into the DNA at all. This may suggest that the degree of DNA bending and thus the overall topology of the *Pf*TLP/DNA complex may be dramatically different to that seen in prototypical TBP/DNA complexes.

4.2.3 DNA binding modes and sequence specificity

The results of this study suggest two modes of *Pf*TLP DNA binding. First, a mode that is independent of the conserved phenylalanine residues, based on the observation that the double mutant, F60&283A, retained DNA binding activity. Second, a mode of DNA binding that is depended the conserved phenylalanine residues, based on the observation that F-to-A substitutions resulted in reduced DNA binding activity. For DNA-binding proteins three basic modes of DNA binding may be considered – specific DNA binding, nonspecific DNA binding and unspecific DNA binding. Specific DNA binding refers to the recognition of and binding to a specific DNA sequence element, with significantly higher affinity compared to unrelated DNA sequences. Nonspecific DNA binding refers to binding of a sequence-specific DNA-binding protein to a sequence unrelated to it is recognition sequence. Lastly, unspecific DNA binding occurs when a protein shows no sequence preference (Murphy & Churchill, 2000). Eukaryotic TBP possesses both sequence-specific and nonspecific DNA binding activity, both of which are mediated through the hydrophobic DNA binding interface of TBP (Coleman & Pugh, 1995). Mutational analysis suggest eukaryotic TBP sequence specific DNA binding depends on the four evolutionary conserved phenylalanine residues (Zhao et al., 2003; Klejman et al., 2005; Kamenova, Warfield & Hahn, 2014). In this study, binding to an A/T-rich GBP-130 DNA probe by the *Pf*TLP F283A mutant was shown to more sensitive to challenge by an excess poly(d-G-dC) competitor DNA than wildtype *Pf*TLP. This result suggests binding selectivity of *Pf*TLP for A/T-rich DNA mediated by (a) conserved phenylalanine residue(s). We have so far not been able to demonstrate TATA sequence-specific DNA binding by *Pf*TLP or

*Pf*TBP (Jasmin Knopp and Thomas Oelgeschläger, unpublished) and the question whether or not *Pf*TLP has preference for a specific A/T sequence remains to be answered.

4.2.4 Outlook

The sequence specificity of *Pf*TLP may be further investigated using modified EMSA and ITA competition experiments, or by using a systematic evolution of ligands by exponential enrichment (SELEX) approach (Milton, 2017). If sequence-specificity can be established it would also be interesting to compare wildtype *Pf*TLP with the mutants carrying F-to-A substitutions in order to further elucidate the role of phenylalanine residues in *Pf*TLP sequence-specific DNA binding. Recent work on *Pf*TBP in the laboratory established A/T preference for *Pf*TBP DNA binding but a specific A/T sequence motif could not be determined using a SELEX approach (Milton, 2017). Further work is required to determine whether *Pf*TBP and *Pf*TLP represent two unique *Plasmodium*-specific TBP/TBP-like proteins with distinct DNA-binding properties.

While results of this work suggests a role of conserved phenylalanine residues in *Pf*TLP DNA binding, it remains unclear whether *Pf*TLP DNA binding also results in DNA bending. Provided, sequence specificity can be established for *Pf*TLP, the degree of DNA bending by *Pf*TLP and the effect of F-to-A mutations could be investigated using circular permutation assays (Bernués et al., 1996; Horikoshi et al., 1992; Starr et al., 1995).

Finally, the important question of the effect of *Pf*TFIIA and *Pf*TFIIB interactions with *Pf*TLP need to be further investigated (Milton, 2017; Talvik, 2016). Purified recombinant *Pf*TFIIA and *Pf*TFIIB have been produced in the laboratory (Bing, 2014; Milton, 2017; Talvik, 2016), that could be used to investigate whether *Pf*TFIIA and/or *Pf*TFIIB interactions affect *Pf*TLP DNA-binding activity and, potentially, compensate for the effect of F-to-A substitutions in *Pf*TLP.

4.3 Low Complexity Regions

4.3.1 *Pf*TLP LCRs amino acid enrichment is consistent with literature

Recent studies have identified an increasing number of proteins, involved in essentially all key cellular processes, that are able to drive LLPS, of which a large proportion have regions of low complexity (LCR), which in many instances gives rise to intrinsic disorder (Ditlev et al., 2018).

The relationship between low complexity and the propensity to drive LLPS has been discussed for many proteins (reviewed in Martin and Mittag, 2018). Briefly, due to the multitude of LCR functions these regions come in a variety of "flavours" or archetypes. For example, polar LCRs that mediate LLPS have been shown to be particularly enriched in polar amino acids like serine, glutamine, and, asparagine. These regions tend to be further enriched in glycine residues, which provide a large degree of conformational flexibility. Polar LCRs that mediate LLPS tend to have upper critical solution temperatures (UCST), and phase separate at lower temperatures, through a process mediated by multivalent interactions. The phase separating activity of many polar LCRs appears to be mediated by the following amino acid sequence features. Firstly, these regions tend to contain randomly dispersed aromatic residues, which are able to engage in π – π and π – cation interactions (Martin and Mittag, 2018). Secondly, these regions are often punctuated with charged amino acids, which add to random electrostatic interactions (Martin and Mittag, 2018).

The analysis of the amino acid enrichment and patterning of the LCR1 and LCR2 regions of *Pf*TLP reflected some of these observations. Firstly, both *Pf*TLP LCR1 and LCR2 regions are particularly enriched in polar amino acids, such as asparagine and serine. Particular enrichment in asparagine may be due to the genome wide A/T bias (Aravind et al., 2003; Pizzi and Frontali, 2001). Secondly, *Pf*TLP LCR1 and LCR2 regions are also punctuated with aromatic and charged amino acid residues. One key difference between the reported properties of polar LCRs and *Pf*TLP LCR1 and LCR2 is the lack of glycine. It seems possible that the underrepresentation of glycine in *P. falciparum* is due to its G/C codon. Taken together these observations supported a possible involvement of *Pf*TLP LCR1 and LCR2 in LLPS, and suggests that these regions, due to their polar nature might be classified as UCST LCRs.

4.3.2 LCR conservation in *Plasmodium* genus

Multiple sequence alignments and *DisMeta* analysis demonstrated the existence of LCR1 and LCR2 regions, of similar length as in *Pf*TLP, between the core domain structures of β-strands S3/S4 and S3'/S4' in all *Plasmodium* TLP orthologues (Fig. 6B). This analysis also revealed a high level of identity between the TLP amino acid sequences from *P. falciparum*, *P. reichenowi*, and *P. gaboni* LCRs. This was not an unexpected result since these species are part of the *Plasmodium* subgenus of *Laverania*, which are evolutionary very closely related

and therefore share over 90% amino acid sequence identity. The subgenus *Laverania* specifically infects primate hosts, such as humans, chimpanzees, and gorillas in sub-Saharan Africa. In contrast, TLP orthologues outside of the subgenus *Laverania* showed very little sequence conservation in LCR1 regions, and do not appear to be enriched in the same amino acid residues either. This is not surprising as LCR/IDRs generally show high rates of evolution (Turoverov et al., 2019). Importantly, *Plasmodium* species outside *Laverania* and therefore more distantly related to *P. falciparum* have different genome nucleotide biases, which impact the enrichment of specific amino acid residues within LCRs. This might indicate that *Pf*TLP LCR1 regions have evolved different functions within the various species. However, LCRs that mediate LLPS have been shown to be able to accommodate high levels of sequence divergence without loss of LLPS function (Turoverov et al., 2019). In contrast to LCR1, LCR2 sequences were found to have a significant degree of sequence similarity between the various *Plasmodium* TLP orthologues. This might indicate a specific evolutionary conserved function of TLP LCR2, perhaps outside of LLPS, within the various *Plasmodium* species.

4.4 Liquid-Liquid Phase Separation Assays

4.4.1 Observation of LLPS

This study provides first evidence for the possible role of *Pf*TLP LCR1 region in driving LLPS, and hence forms the basis for future research into this important new aspect of gene regulation. The results from phase separation assays demonstrated that *Pf*TLP-LCR1 is capable if driving LLPS *in vitro*. *Pf*TLP LCR1 might be considered to have an upper critical solution temperature (UCST), given that it phase separates at room temperature.

4.4.2 Relevance

A number of recent studies have highlighted the importance of LLPS in the regulation of RNA polymerase II transcription. However, these studies have focused on RNA polymerase II, the mediator complex coactivator and the DNA binding transcription factor activator domains (Boija et al., 2018; Chong et al., 2018; Guo et al., 2019; Gurumurthy et al., 2019; Larson et al., 2011; Sabari et al., 2018; Shrinivas et al., 2019). Studies at super-enhancers suggest that LLPS involving the mediator complex and activation domains serve to concentrate the transcription machinery at target genes during gene activation (Boija et al., 2018; Chong et al., 2018;

Gurumurthy et al., 2019; Sabari et al., 2018; Shrinivas et al., 2019). It was further shown that the phosphorylation status of the C-terminal domain of RNA polymerase II (RNAPII) determines the association of RNAPII with either the mediator complex at super-enhancers (hyperphosphorylated) or membraneless splicing organelles (hypophosphorylated) (Cho et al., 2018; Guo et al, 2019.). Up to now, there is no literature on the possible involvement of the general transcription machinery in LLPS. This study provides first evidence for the possible involvement of a general transcription factor in LLPS.

4.4.3 Limitations and Outlook

More work needs to be done to further characterize the LLPS properties of *Pf*TLP and to investigate the relevance of the *in vitro* observations.

Firstly, the conditions under which *Pf*TLP LCR1 drives LLPS need to be determined. This can be done through systematic analysis of *Pf*TLP LCR1 LLPS at different concentrations of protein, salt, and crowding agents (PEG/Dextran) as well as under different temperatures and incubation times. Secondly, the precise liquid-like properties of the observed spherical protein assemblies needs to be examined by observing fusion events between droplets, and by fluorescence recovery after photobleaching (FRAP) experiments, which demonstrate rapid diffusion of *Pf*TLP LCR1 between droplets and the surrounding solution (Alberti et al., 2019). Thirdly, the role of the full length wild-type *Pf*TLP in LLPS has to be determined, by testing its ability to drive LLPS, or to be recruited into membraneless organelles.

Fourth, the amino acid determinants mediating the LLPS activity of *Pf*TLP LCR1 needs to be determined. For example, repeating units of phenylalanine and arginine with glycine (FG/RG) were shown to be critical for the phase separating behaviour of the DDX4 protein (Nott et al., 2015). Furthermore, alternating the positions of charged residues within the DDX4 protein also leads to a change in phase separating behaviour (Nott et al., 2015). Additionally, studies of the fused in sarcoma (FUS) RNA binding protein found that mutating repeating units of tyrosine to serine resulted in a change in hydrogel formation and cellular localization (Kato et al., 2012). Together, this indicates that certain amino acid patterns within LCRs have a higher propensity to drive LLPS than others. Similarly, the role of the repeating asparagine and serine residues in *Pf*TLP-LCR1 LLPS can be investigated by generating deletion and substitution mutant proteins and testing their LLPS properties.

Additionally, the phase separation assays were so far conducted in a one (LCR1) or two component (LCR1 + LCR2) system. However, cellular membraneless organelles are complex assemblies that contain multiple protein and nucleic acid partners. It would therefore be interesting to test whether *Pf*TLP can form LLPS assemblies with other *P. falciparum* basal transcription factors. The majority of human and *P. falciparum* (putative) basal transcription factors, including TFIIA, TFIIB, TFIIE, TFIIF and TFIIH appear to contain LCRs with the potential to phase separate (Leonidas Karamanof and Thomas Oelgeschläger, unpublished). It would be interesting to see how these transcription factors interact, and if they can form higher-order phase separated assemblies. If such higher order assemblies are identified, it would be important to see which proteins act as scaffold proteins, and recruit additional proteins into phase separated droplets. The ability of membraneless organelles to recruit and concentrate molecules and proteins involved in similar cellular functions is well documented and can be experimentally corroborated *in vitro* and *in vivo* through recruitment assays. (Alberti et al., 2019; Dao et al., 2018; Sheu-Gruttadauria and Macrae, 2018; Woodruff et al., 2017). It would be interesting to test if the *P. falciparum* and human proteins have different LLPS properties, as this might open new avenues for anti-malarial drug development.

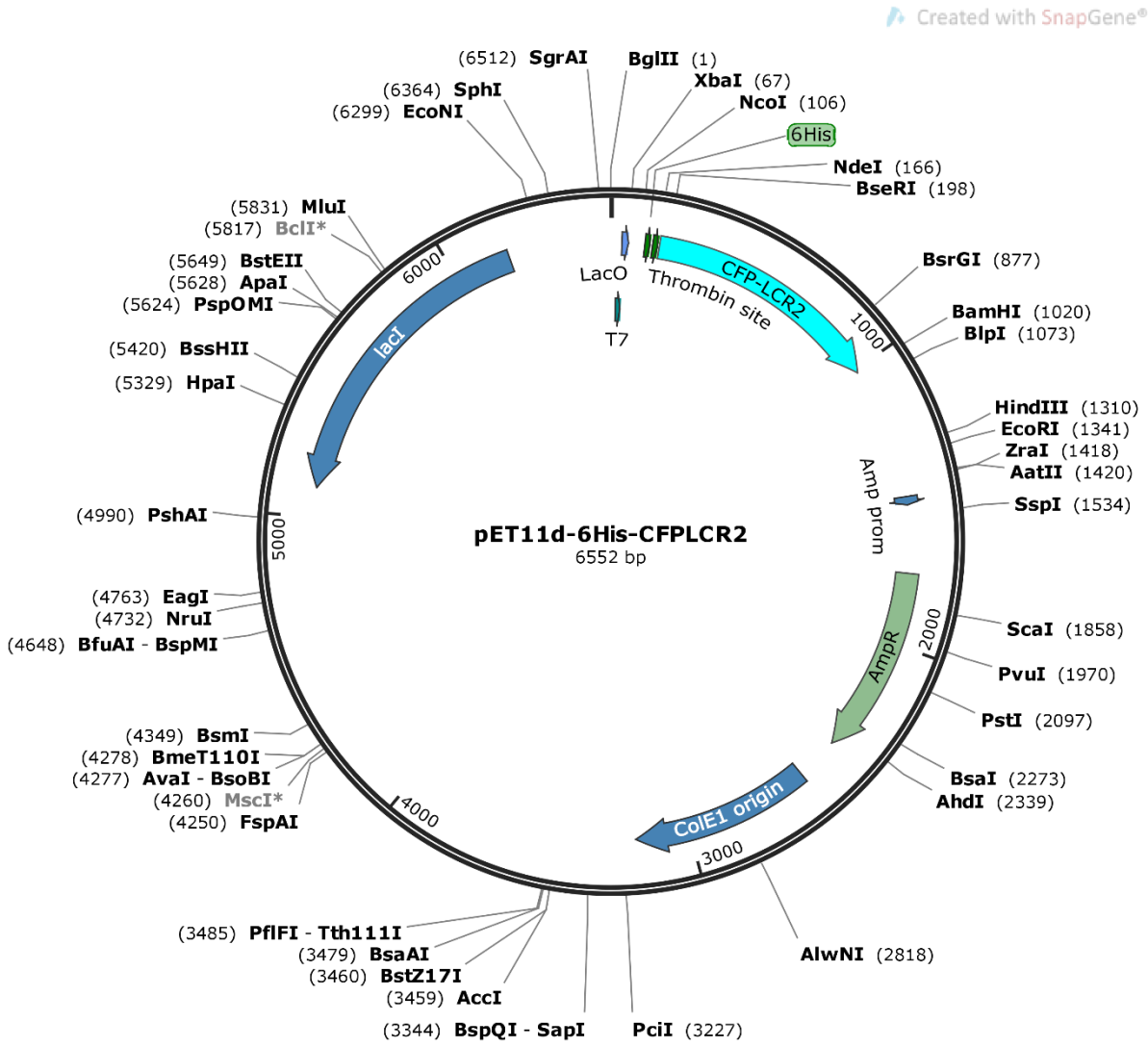
Finally, the role of *Pf*TLP LCR2 in LLPS could not be determined in this study, and this is likely due to the limited set of conditions that were tested. A wider range of experimental conditions needs to be employed to determine a possible involvement of LCR2 in *Pf*TLP LLPS.

Conclusion

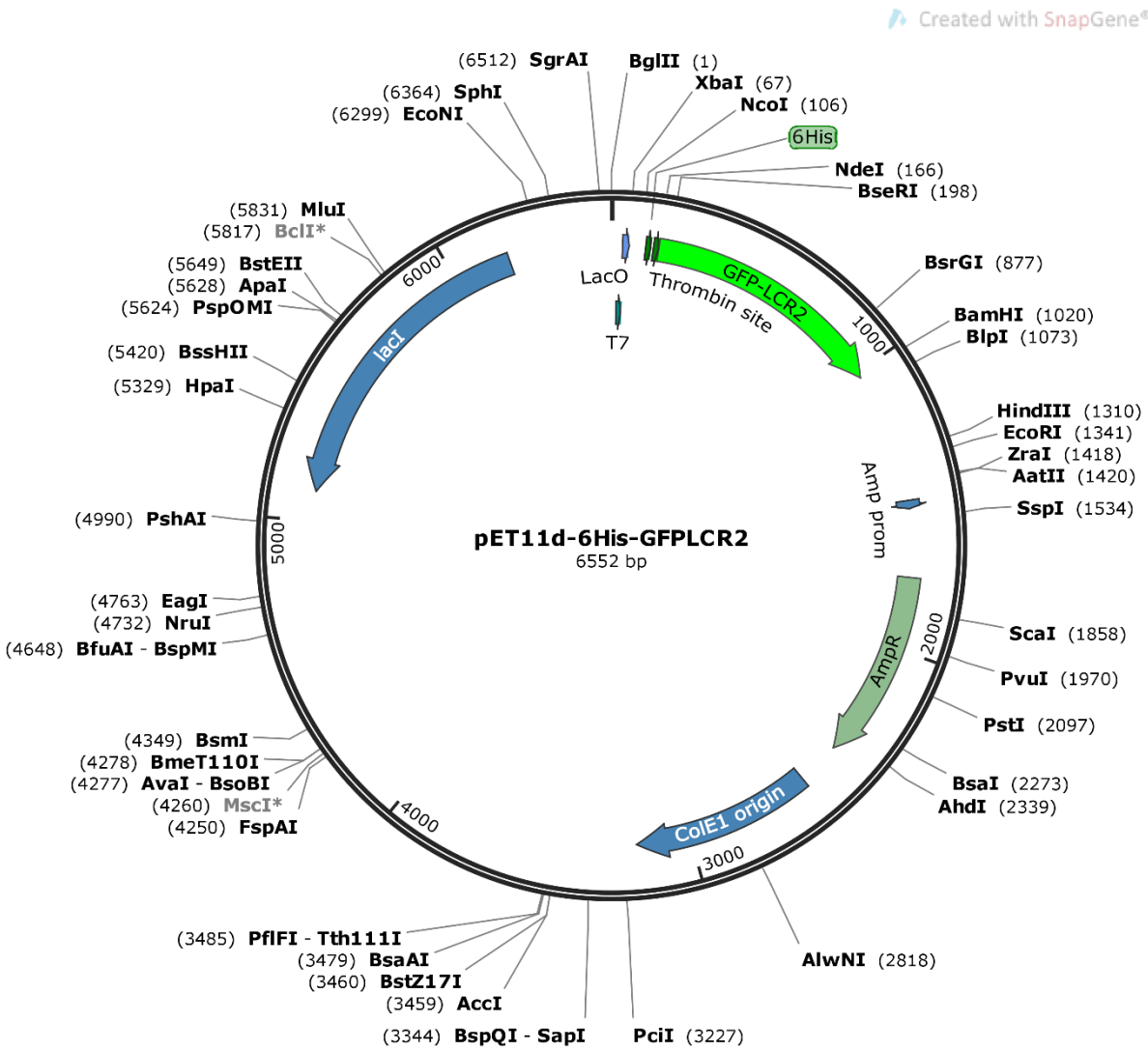
This study provided the initial characterization of a *Plasmodium*-specific general transcription factor, *Pf*TLP. It was shown that *Pf*TLP possesses DNA binding activity, which might be mediated through hydrophobic interactions. DNA binding was not dependent on the action of the two evolutionary conserved phenylalanine residues, F60 and F283, predicted to be important for TBP DNA binding. However, these two phenylalanine residues, especially F60, are important for *Pf*TLP/DNA complex stability. Additionally, it was shown that *Pf*TLP possesses two low complexity regions that extend out from the core domain structure, which are conserved in the *Plasmodium* genus. These LCRs have the hallmarks of polar LCRs that drive LLPS, and it was shown that *Pf*TLP LCR1 is able to phase separate *in vitro*. Taken together,

this study provides first insight into the possible functions of *Pf*TLP, and contributes to a better understanding of transcription initiation and gene expression in this globally important parasite

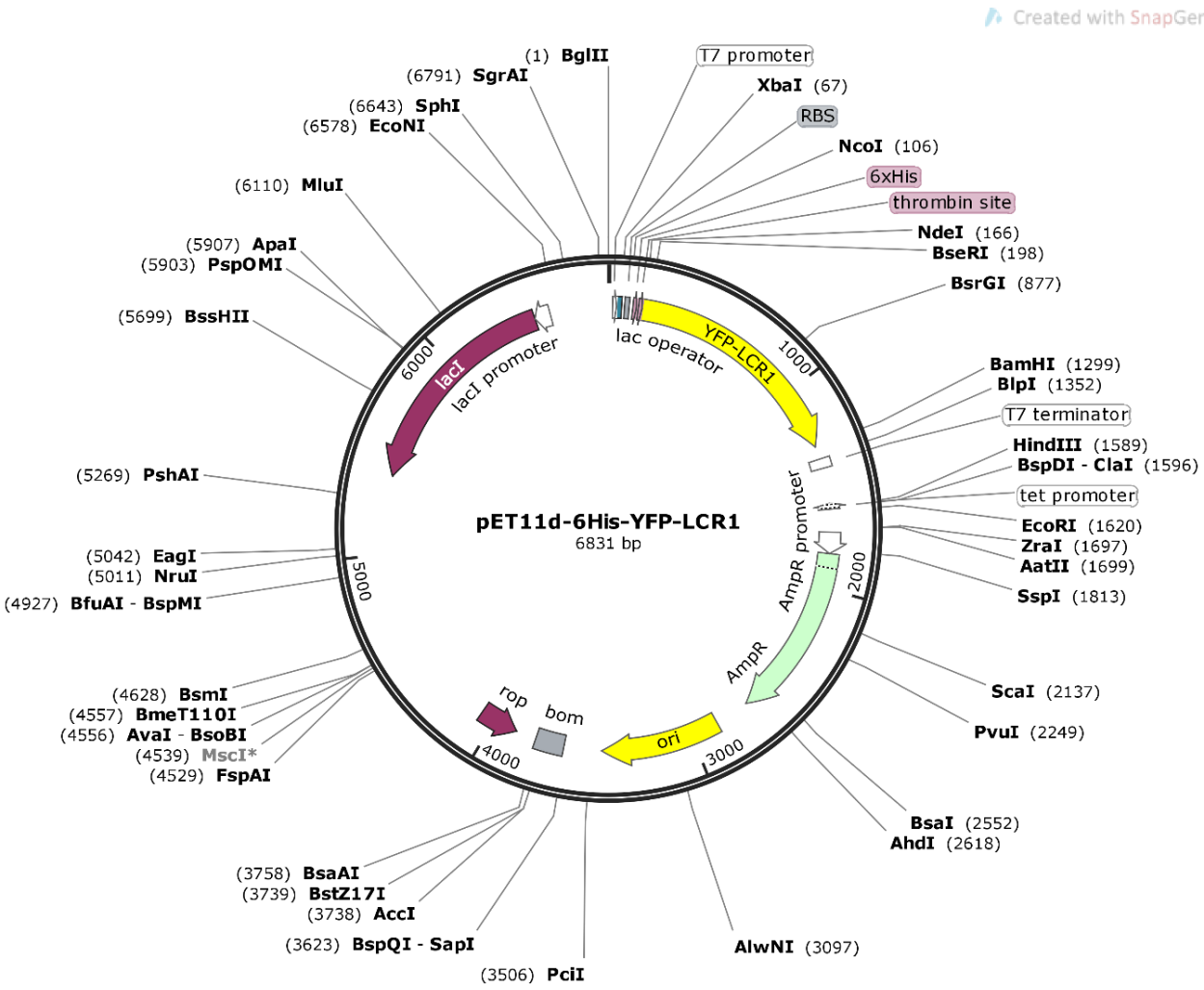
Supplementary Material



Supplementary Figure 1: Plasmid map of pET11d-6His-CFP-LCR2



Supplementary Figure 2: Plasmid map of pET11d-6His-GFP-LCR2



Supplementary Figure 3: Plasmid map of pET11d-6His-YFP-LCR1

Supplementary Table 1: Primers used for the generation of fluorescently labeled LCRs.

Primer	**Sequence (5'-3')**
GFP-LCR1-FWD	tctcggcatggacgagctgtacaagCCGGTTACCCTGAGTACGG
GFP-LCR1-REV	gcaccgtactcagggtaaccggCTTGTACAGCTCGTCCATGC
GFP-LCR2-FWD	cggcatggacgagctgtacaagATTGCCCTGATGAACGATGAAAAC
GFP-LCR2-REV	tgttttcatcgttcatcagggcaatCTTGTACAGCTCGTCCATGC
LCR1-REV	tagcagccggatccttatcaGATCAGTTTTTTGTTGATGATATGTTC
LCR2-REV	tagcagccggatccttatcaACTGGCAGAGACGATATCAACTT
eGFP-FWD	gcggcagccatATGGTGAGCAAGGGCGAGG
pEarlyGate FWD	gcggcagccatatgGTGAGCAAGGGCGAGGAG

CLUSTAL O(1.2.4) multiple sequence alignment of Plasmodium TLPs (page 1 of 2)

```
Pv          ---------------------------MSVHNISMNATLCSSLNLDSLYRHFANCIYNPR      33
Pg          --MYPPCKKKKLNNNEVTNIFLKNENKMSVHNISMNAVLCSSLNLDNIYKYFSNCIYNPR      58
PfTLP       --MYPPCKKKKLNNNEVTNIFLKNENNMSVHNISMNAVLCSSLNLDNIYKYFSNCIYNPR      58
Pr          --MYPPCKKKKLNNNEVTNIFLKNENNMSVHNISMNAVLCSSLNLDNIYKYFSNCIYNPR      58
Po          --MSQPVKKSKLSYTESKNIFVNNENSMSVHNISMNAVLCSTLNLDSIYRYFANCIYNPR      58
Pm          MKMSQPLKRSKLNNNESQNIFVKNENSMSVHNISMNAILCSSLNLDSIYKYFSNCIYNPR      60
Prel        --MCNPSKRTKLNNNEINNIFIKNEHNMSVHNISMNAILCSSLNLDSIYKYFSNCIYNPR      58
Pca         --MEHALKKIKLNNNESKNIYINNANNMSVHNISMNAILCSSLNLDSIYKYFSNCVYNPR      58
Py          --MEHVSKKIKLDNNESKNIYINNSNNMSVHNISMNAILCTSLNLDNIYKHFSNCVYNPR      58
                                     ********** **::****.:*::*:**:****

Pv          EFKCMRIDVPVSTRSISKYVRFVQQKGGRQTGETQTGETQTGETQ-TGETQT--GETQTG      90
Pg          EFKCMRIDVPVSLNTVHKYIKYLNNKKERQNDDISKMKNEQENEAIHSNNYIKEENITTE      118
PfTLP       EFKCMRIDVPVTLSTVQKYIKYLNNKKEKQNDDICKMKNEQVNQTKHSNNNIKEEIKTNG      118
Pr          EFKCMRIDVPVTLSTVQKYIKYLNNKKEKQNDDICKMKNEQVNQTKHSNNNIKEEIKTTG      118
Po          EFKCMRIDVPVSLKTVNKYVNYLREKRERGKRERVMEEMEREELEKCSREQLKKEEKDEA      118
Pm          EFKCMRIDVPVSLKTVHKYVNYVKNKRKKEQQGIEVEINKQEDSSYKGVDNNCKEVSNNC      120
Prel        EFKCMRIDVPVSLRTVNKYVNYINNKKERDDKYKLKDEK-YKQIKEKNEKKF-DDIKNEK      116
Pca         EFKCMRIDVPVSLSTVNKYVNHIKNKEKLKTE--------------KNQNEVCMDVENDS      104
Py          EFKCMRVDVPVSLNTVNKYIDYIKNKEKIQTEQTEQTEQ-TNLVTKNNKNEICMPVENNS      117
            ******:****:  :: **: .:.:*                      .

Pv          EKPTGEEQ-----TGEKQTGEKPTGEEQKG-SEGVIPPSDGT----SVVTP-SAPPPDVA      139
Pg          SNKLNDNK----LSDS--NN---------SSSGKILTDNKSD-ENSINISKNKSNIDD--      160
PfTLP       SNKLSDNK----LLDS-SNN---------SSSNKILTNNKSH-ENLINISKNNSYLDD--      161
Pr          SNKLSNNK----LLDS-SNN---------SSSDKILTNNKSH-ENLINISKNNSYLDD--      161
Po          KEKKEDEIDVAEIVTEQCNNE--VPEEGKVSPPEKVNKDTG-IH-ESD-IPNNEIPAAE-      172
Pm          DMPSGGNTS-FPIDDGKCSMP--DS--VHVSS-----------TANKASS----------      154
Prel        VKK-DDDI-----ENEKEESE--EK--KEIVSEKIIDNNKNYIKDNKNISNDNYIPFDET      166
Pca         NKPLPTQT-----NTT----------------------------NFNINIEPNNYYPPEIK      132
Py          NKDTSTQL-----FNKKSNNN--DN--IEHNNDNIEHNNDNIEHNNDNIEHNNYGSSGIK      168
                  :

Pv          TT--CQVVE--------AT-------TSQGGGNPPEEHTEANEKVIINVSIFANGKIICT      182
Pg          NNSEIK-----------KNTDIIYDMET--MDPFNESEHTMNKKLIINVSIFSNGKIICT      207
PfTLP       NVNKNNFFI--NDNNSDNKKIDTSDIKN--MDPFIESEHIINKKLIINVSIFSNGKIICT      217
Pr          NVNKNNFVI--NDNNSDNKKIDTYDIQT--MDPFIESEHIINKKLIINVSIFSNGKIICT      217
Po          -----------SATLSDDFFLKGGYEKLVSNDFSEHYNDNVGEKLVINVSIFSNGKIICT      221
Pm          ---SNAEVSARNDHISGESNYINSVI----EK-INENYKDVTEKLVINVSIFSNGKIICT      206
Prel        EN--INSII--NEDISDTKNDKSNDIVEYS----IENNNVVNKKLIINVSIFSNGKIICT      218
Pca         YIDGIPM----------------------FEDKNDEDD--YTSEKLVINVSIFSNGKIICT      169
Py          HNDEIQNFEEKNDEV-QNFEEKNDEIQKFEEKNDEDDHYTTSEKLVINVSIFSNGKIICT      227
                                           .      :*::******:*******

Pv          GNNSIEACKIAMKKIERKLKQLNFKNISIKNVTITNILAVYNVGFSIVLPLFAQYYKSVD      242
Pg          GNNSIEACKVAMKKIEKKLKQLNFKNIKLKKITITNILAVYNVGFSIVLPLFAQYYKSVD      267
PfTLP       GNNSIEACKIAMKKIEKKLKQLNFKNIKLKKITITNILAVYNVGFSIVLPLFAQYYKSVD      277
Pr          GNNSIEACKIAMKKIEKKLKQLNFKNIKLKKITITNILAVYNVGFSIVLPLFAQYYKSVD      277
Po          GNNSIEACKIAMKKVEKKLKFLNFKNIKLKKITITNILAVYNVGFSIVLPLFAQFYKNVD      281
Pm          GNNSIEACKIAMKKVEKKLKQLNFKNIKLKKITIANILAVYNVGFSIVLPLFAQYYKSVD      266
Prel        GNNSIEACKIAMKKVEKKLKQLNFKNIKLKKITITNILAVYNIGFSIVLPLFAQYYKSVD      278
Pca         GNNSIEACKIAMKKVEKKLKQLNFKNIKLKKIAITNILAVYNVGFSIVLPLFAQYYKSVD      229
Py          GNNSIEACKIAMKKVEKKLKQLNFKNIKLKKIAITNILAVYNVGFSIVLPLFAQYYKSVD      287
            *********:****:*:*** ******.:*:::*:*******:***********:**.**
```

```
CLUSTAL O(1.2.4) multiple sequence alignment of Plasmodium TLPs (page 2 of 2)

Pv          YDPNVFPACKVKIALTSDSERDTSEASEQAEGNFAWSTPRSTAERSRSKVDVVSASIFST     302
Pg          YDPNVFPACKVKIALMNDDNKS-NDNNELNDNNYAWCNAKNTIDKDKSKVDIVSASIFST     326
PfTLP       YDPNVFPACKVKIALMNDENKS-NDNK-ENDNNFAWCNAKNTIDKDKSKVDIVSASIFST     335
Pr          YDPNVFPACKVKIALMNDENKS-NDNK-EHDNNFAWCNAKNTIDKDKSKVDIVSASIFST     335
Po          YDPNVFPACKVKIALTNEDEKNASERREQNDCTYAWCSAKSTADKEKSKVDIVSASIFST     341
Pm          YDPNVFPACKVKIALTNEDKKNLSERKEQNESTYAWCNANNNIDSGKSKVDIVSASIFST     326
Prel        YDPNVFPACKVKIALINENEKNCDEKNEENNSQYAWWNIKKTAEKEKNKVDVISASIFST     338
Pca         YDPNVFPACKVKIALTNDDNNTSSELNEQNDSSYAWCTSRPTIETSKSKADVVSASIFST     289
Py          YDPNVFPACKVKIALSNDDNNNSSEHNEQNDSAYAWCTNKPIFESNKNKTDVVSASIFST     347
            *************** .:.:.  .:     :  :** . .   :  :.*.*::*******

Pv          GNITLTGGKSYQNLQRCIDILYPYLIKSKSQH        334
Pg          GNITLTGGKSYENLQKCINILLPYLIKSKSQH        358
PfTLP       GNITLTGGKSYENLQKCINILLPYLIKSKSQH        367
Pr          GNITLTGGKSYENLQKCINILLPYLIKSKSQH        367
Po          GNITLTGGKSYENLQRCIDILLPYLIKSKSQH        373
Pm          GNITLTGGKSYENLQRCIDILLPYLLKSKSQH        358
Prel        GNITLTGGKSYENLQKCIDILLPYLIKSKSQH        370
Pca         GNITLTGGKSYDNLKKCIDILLPYLIKSRSQH        321
Py          GNITLTGGKSYDNLKKCIDILLPYLIKSKSQH        379
            ***********:**::**:** ***:**:***

Pv = Plasmodium vivax
Pg = Plasmodium gondii
Pr = plasmodium reichenowiium
Po = plasmodium ovale
Pm = plasmodium malariae
Prel = plasmodium relictum
Pca = plasmodium chabaudi adami
Py = plasmodium yoelii
```

Supplementary Figure 4 : Multiple sequence alignment of Plasmodium TLP orthologues. Light grey letters indicated amino acid residues in LCR1 and LCR2 of *Pf*TLP. Conserved phenylalanine residues, corresponding to those required for TBP sequence-specific DNA binding are shown in red.

Supplementary Table 2: DNA probes used in this study. Uppercase letters refer to PCR primer sequences, italics refer to GAL4 binding sites.

Probe	Length	Experiment	Sequence	Biotinylation
GBP-130	422 bp	ITA	5'AATTGGGCCCGACGTCGCA*tgctcctctagagcttgcatgcctgcaggtcggagactgtcctccgagcggagtactgtcctccgagcggagtactgtcctccgagcggagtactgtcctccgagcggagactctagaattc*tttgaagtacactcaaaataagttatataccatatgtttttaacatatattatatatatatatatatatataatataatataattataattatttttaatattattaaattgaacataattattttatatcttactattatttttagaaaatttattatatatacatgcaatcataaataatgttttccctgaacctttttttcaatgaaataagttaacacaccattcctttGGATCCGGAGAGCTCCCAACGCGTT3'	5'
GBP-130	288 bp	EMSA	5'CGAGCGGAGACTCTAGaattctttgaagtacactcaaaataagttatataccatatgtttttaacatatattatatatatatatatatatatatatataatataatataattataattatttttaatattattaaattgaacataattattttatatcttactattatttttagaaaatttattatatatacatgcaatcataaataatgttttccctgaacctttttttcaatgaaataagttaacacaccattcctttttGGATCCGGAGAGCTCCCAACGCGT3'	3' and 5'
GBP-130	30 bp	EMSA	(+) 5'- GAAAATTTATTATATATACATGCAATCATA (-) 5'- TATGATTGCATGTATATATAATAAATTTTC	3' and 5'
Ad2ML	40 bp	EMSA	(+)5'- GTTCCTGAAGGGGGGCTATAAAAGGGGGTGGGGGCGCGTT (-)5'- AACGCGCCCCCACCCCCTTTTATAGCCCCCCTTCAGGAAC	3' and 5'

*Pf*TLP Predicted Protein Sequence

```
1     MYPPCKKKLNNNEVTNIFLKNENNMSVHNISMNAVLCSSL   41

42    NLDNIYKYFSNCIYNPREFKCMRIDVPVTLSTVQKYIKYLN   82

83    NKKEKQNDDICKMKNEQVNQTKHSNNNIKEEIKTNGSNKLS   123

124   DNKLLDSSNNSSSNKILTNNKSHENLINISKNNSYLDDNVN   164

165   KNNFFINDNNSDNKKIDTSDIKNMDPFIESEHIINKKLIIN   205

206   VSIFSNGKIICTGNNSIEACKIAMKKIEKKLKQLNFKNIKL   246

247   KKITITNILAVYNVGFSIVLPLFAQYYKSVDYDPNVFPACK   287

288   VKIALMNDENKSNDNKENDNNFAWCNAKNTIDKDKSKVDIV   328

329   SASIFSTGNITLTGGKSYENLQKCINILLPYLIKSKSQH     367
```

Supplementary Figure 5: Protein Sequence of *Pf*TLP, extracted from *PlasmoDB* accession number PF 3D7_1428800.1

References

Adjalley, S.H., Chabbert, C.D., Klaus, B., Pelechano, V., Steinmetz, L.M., and De, : Larsms@embl (2016). Landscape and dynamics of transcription initiation in the malaria parasite Plasmodium falciparum Graphical abstract HHS Public Access. Cell Rep. March *15*, 2463–2475.

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

Alano, P. (2007). Plasmodium falciparum gametocytes: Still many secrets of a hidden life. Mol. Microbiol. *66*, 291–302.

Alberti, S. (2017). The wisdom of crowds: regulating cell function through condensed states of living matter. J. Cell Sci. *130*, 2789–2796.

Alberti, S., Saha, S., Woodruff, J.B., Franzmann, T.M., Wang, J., and Hyman, A.A. (2018). A User's Guide for Phase Separation Assays with Purified Proteins. J. Mol. Biol. *430*, 4806–4820.

Alberti, S., Gladfelter, A., and Mittag, T. (2019). Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. Cell *176*, 419–434.

Altmeyer, M., Neelsen, K.J., Teloni, F., Pozdnyakova, I., Pellegrino, S., Grøfte, M., Druedahl Rask, M.-B., Streicher, W., Jungmichel, S., Lund Nielsen, M., et al. (2015). Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). Nat. Commun. *6*.

Aravind, L., Iyer, L.M., Wellems, T.E., and Miller, L.H. (2003). Plasmodium Biology. Cell *115*, 771–785.

Aumiller, W.M., and Keating, C.D. (2016). Phosphorylation-mediated RNA/peptide complex coacervation as a model for intracellular liquid organelles. Nat. Chem. *8*, 129–137.

Balaji, S., Babu, M.M., Iyer, L.M., and Aravind, L. (2005). Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. Nucleic Acids Res. *33*, 3994–4006.

Bártfai, R., Balduf, C., Hilton, T., Rathmann, Y., Hadzhiev, Y., Tora, L., Orbán, L., and Müller, F. (2004). TBP2, a Vertebrate-Specific Member of the TBP Family, Is Required in Embryonic Development of Zebrafish. Curr. Biol. *14*, 593–598.

Bashirullah, A., Lam, G., Yin, V.P., and Thummel, C.S. (2007). dTrf2 is required for transcriptional and developmental responses to ecdysone duringDrosophila metamorphosis. Dev. Dyn. *236*, 3173–3179.

Bergeron-Sandoval, L.-P., and Michnick, S.W. (2018). Mechanics, Structure and Function of Biopolymer Condensates. J. Mol. Biol. *430*, 4754–4761.

Bergeron-Sandoval, L.-P., Khadivi Heris, H., Chang, C., Keller, S.L., François, P., Hendricks, A.G., Ehrlicher, A.J., and Michnick, S.W. Endocytosis caused by liquid-liquid phase separation of proteins.

Berk, A.J. (2000). TBP-like Factors Come into Focus. Cell *103*, 5–8.

Bernués, J., Carrera, P., and Azorín, F. (1996). TBP binds the transcriptionally inactive TA5 sequence but the resulting complex is not efficiently recognised by TFIIB and TFIIA. Nucleic Acids Res. *24*, 2950–2958.

Bing, S. (2014). Expression and initial characterisation of the Plasmodium falciparum general transcription factors TFIIB and TLP. University of Cape Town.

Birkholtz, L.-M., Blatch, G., Coetzer, T.L., Hoppe, H.C., Human, E., Morris, E.J., Ngcete, Z., Oldfield, L., Roth, R., Shonhai, A., et al. (2008). Heterologous expression of plasmodial proteins for structural studies and functional annotation. Malar. J. *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*, 34.

Boeynaems, S., Bogaert, E., Van Damme, P., Ludo, ·, and Bosch, V. Den (2016). Inside out: the role of nucleocytoplasmic transport in ALS and FTLD. Acta Neuropathol. *132*, 159–173.

Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., Polymenidou, M., Rousseau, F., Van, L., Bosch, D., Tompa, P., and Fuxreiter, M. (2018). Protein Phase Separation: A New Phase in Cell Biology. Trends Cell Biol. *28*, 420–435.

Boija, A., Klein, I.A., Sabari, B.R., Agnese, A.D., Coffey, E.L., Zamudio, A. V, Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell *175*, 1842–1855.

Bozdech, Z., Llinás, M., Pulliam, B.L., Wong, E.D., Zhu, J., and DeRisi, J.L. (2003). The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. PLoS Biol. *1*, 85–100.

Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., and Hyman, A.A. (2009). Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. Science (80-.). *324*, 1729–1732.

Brangwynne, C.P., Tompa, P., and Pappu, R.V. (2015). Polymer physics of intracellular phase transitions. Nat. Phys. *11*, 899–904.

Brocchieri, L. (2001). Low-Complexity Regions in Plasmodium Proteins: In Search of a Function. Genome Res. *11*, 195–197.

Callebaut, I., Prat, K., Meurice, E., Mornon, J.-P., Tomavo, S., Sinden, R., Lanzer, M., Bruin, D. de, Ravetch, J., Lanzer, M., 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*, 100.

Carlton, J.M., Adams, J.H., Silva, J.C., Bidwell, S.L., Lorenzi, H., Caler, E., Crabtree, J., Angiuoli, S. V., Merino, E.F., Amedeo, P., et al. (2008). Comparative genomics of the neglected human malaria parasite Plasmodium vivax. Nature *455*, 757–763.

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

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018a). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, 412–415.

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018b). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science

361, 412–415.

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., et al. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science *361*, eaar2555.

Coleman, B.I., and Duraisingh, M.T. (2008). Transcriptional control and gene silencing in Plasmodium falciparum. Cell. Microbiol. *10*, 1935–11946.

Coleman, R.A., and Pugh, B.F. (1995). Evidence for functional binding and stable sliding of the TATA binding protein on nonspecific DNA. J. Biol. Chem. *270*, 13850–13859.

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

Cowman, A.F., Berry, D., and Baum, J. (2012). The cellular and molecular basis for malaria parasite invasion of the human red blood cell. J. Cell Biol. *198*.

Crabb, B.S., Cowman, A.F., Nossal, G.J. V, and Hall, W. (1996). Characterization of promoters and stable transfection by homologous and nonhomologous recombination in Plasmodiumfalciparum. Med. Sci. *93*, 7289–7294.

Cramer, P. (2019). Organization and regulation of gene transcription. Nature *573*, 45–54.

Cui, L., and Miao, J. (2010). Chromatin-Mediated Epigenetic Regulation in the Malaria Parasite Plasmodium falciparum. Eukaryot. Cell *9*, 1138–1149.

Dantonel, J.-C., Quintin, S., Lakatos, L., Labouesse, M., and Tora, L. (2000). TBP-like Factor Is Required for Embryonic RNA Polymerase II Transcription in C. elegans. Mol. Cell *6*, 715–722.

Dantonel, J.C., Wurtz, J.M., Poch, O., Moras, D., and Tora, L. (1999). The TBP-like factor: An alternative transcription factor in Metazoa? Trends Biochem. Sci. *24*, 335–339.

Dao, T.P., Kolaitis, R.-M., Kim, H.J., O'donovan, K., Martyniak, B., Colicino, E., Hehnly, H., Taylor, J.P., and Castañeda, C.A. (2018). Ubiquitin modulates liquid-liquid phase separation of UBQLN2 via disruption of multivalent interactions HHS Public Access. Mol Cell *69*, 965–978.

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*, 19–

30.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., et al. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. *36*, W465.

Ditlev, J.A., Case, L.B., and Rosen, M.K. (2018). Who's In and Who's Out—Compositional Control of Biomolecular Condensates. J. Mol. Biol. *430*, 4666–4684.

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

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C.S. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. Plant J. *45*, 616–629.

Edgar, R.C. (2004a). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Elbaum-Garfinkle, S., Kim, Y., Szczepaniak, K., Chen, C.C.-H., Eckmann, C.R., Myong, S., and Brangwynne, C.P. (2015). The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. Proc. Natl. Acad. Sci. *112*, 7189–7194.

Florens, L., Washburn, M.P., Dale Raine, J., Anthony, R.M., Graingerk, M., David Haynes, J., Kathleen Moch, J., Muster, N., Sacci, J.B., Tabb, D.L., et al. (2002). A proteomic view of the Plasmodium falciparum life cycle. Nature *419*, 520–526.

Foth, B.J., and McFadden, G.I. (2003). The apicoplast: A plastid in Plasmodium falciparum and other apicomplexan parasites. In International Review of Cytology, (Academic Press), pp. 57–110.

Franzmann, T.M., and Alberti, S. (2019). Protein Phase Separation as a Stress Survival Strategy. Cold Spring Harb. Perspect. Biol. *11*.

Franzmann, T.M., Jahnel, M., Pozniakovsky, A., Mahamid, J., Holehouse, A.S., Nüske, E., Richter, D., Baumeister, W., Grill, S.W., Pappu, R. V., et al. (2018). Phase separation of a yeast prion protein promotes cellular fitness. Science (80-.). *359*, eaao5654.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A.,

Nelson, K.E., Bowman, S., et al. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. Nature *419*, 498–511.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook, (Totowa, NJ: Humana Press), pp. 571–607.

Gazdag, E., Santenard, A., Ziegler-Birling, C., Altobelli, G., Poch, O., Tora, L., and Torres-Padilla, M.-E. (2009). TBP2 is essential for germ cell development by regulating transcription and chromatin condensation in the oocyte. Genes Dev. *23*, 2210–2223.

Gomes, E., and Shorter, J. (2019). The molecular language of membraneless organelles. J. Biol. Chem. *294*, 7115–7127.

Guillebault, D., Sasorith, S., Derelle, E., Wurtz, J.-M., Lozano, J.-C., Bingham, S., Tora, L., and Moreau, H. (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. J. Biol. Chem. *277*, 40881–40886.

Guo, Y.E., Manteiga, J.C., Sabari, B.R., Hannett, N.M., Krishna, S., Henninger, J., Dall'Agnese, A., Coffey, E.L., Li, C.H., Abraham, B.J., et al. (2019). RNA Polymerase II phosphorylation regulates a switch between transcriptional and splicing condensates. Nature *572*, 543–548.

Gurumurthy, A., Shen, Y., Gunn, E.M., and Bungert, J. (2019). Phase Separation and Transcription Regulation: Are Super-Enhancers and Locus Control Regions Primary Sites of Transcription Complex Assembly? BioEssays *41*, 1800164.

Haberle, V., and Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. Nat. Rev. Mol. Cell Biol. *19*, 621–637.

Hahn, S. (2018). Previews Phase Separation , Protein Disorder , and Enhancer Function. 1723–1725.

Heckman, K.L., and Pease, L.R. (2007). Gene splicing and mutagenesis by PCR-driven overlap extension. Nat. Protoc. *2*, 924–932.

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

Hermansyah, B., Fitri, L.E., Sardjono, T.W., Endharti, A.T., Arifin, S., Budiarti, N., Candradikusuma, D., Sulistyaningsih, E., and Berens-Riha, N. (2017). Clinical features of severe malaria: Protective effect of mixed plasmodial malaria. Asian Pac. J. Trop. Biomed. *7*, 4–9.

Hessler, T. (2014). No Title. Univeristy of Cape Town.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. Cell *169*, 13–23.

Hochheimer, A., Zhou, S., Zheng, S., and Holmes, M.C. (2002). TRF2 associates with DREF and directs promoter-selective gene expression in Drosophila. *420*, 1–7.

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 Res. *19*, 6337–6338.

Holmes, M.C., and Tjian, R. (2000). Promoter-selective properties of the TBP-related factor TRF1. Science *288*, 867–870.

Horikoshi, M., Bertuccioli, C., Takada, R., Wang, J., Yamamoto, T., and Roeder, R.G. (1992). Transcription factor TFIID induces DNA bending upon binding to the TATA element. Proc. Natl. Acad. Sci. U. S. A. *89*, 1060–1064.

Horrocks, P., Wong, E., Russell, K., and Emes, R.D. (2009). Control of gene expression in Plasmodium falciparum - Ten years on. Mol. Biochem. Parasitol. *164*, 9–25.

Huang, Y.J., Acton, T.B., and Montelione, G.T. (2014). DisMeta-a Meta Server for Construct Design and Optimization. Methods Mol. Biol.

Idro, R., Jenkins, N.E., and Newton, C.R. (2005). Pathogenesis, clinical features, and neurological outcome of cerebral malaria. Lancet Neurol. *4*, 827–840.

Isogai, Y., Keles, S., Prestel, M., Hochheimer, A., and Tjian, R. (2007). Transcription of histone gene cluster by differential core-promoter factors. Genes Dev. *21*, 2936–2949.

Jacobi, U.G., Akkers, R.C., Pierson, E.S., Weeks, D.L., Dagle, J.M., and Veenstra, G.J.C. (2007). TBP paralogs accommodate metazoan- and vertebrate-specific developmental gene regulation. EMBO J. *26*, 3900–3909.

Jallow, Z., Jacobi, U.G., Weeks, D.L., Dawid, I.B., and Veenstra, G.J.C. (2004). Specialized and redundant roles of TBP and a vertebrate-specific TBP paralog in embryonic gene regulation in Xenopus. Proc. Natl. Acad. Sci. *101*, 13525–13530.

Jeninga, M., Quinn, J., and Petter, M. (2019). ApiAP2 Transcription Factors in Apicomplexan Parasites. Pathogens *8*, 47.

Johnson, K.M., Wang, J., Smallwood, A., and Carey, M. (2004). The Immobilized Template Assay for Measuring Cooperativity in Eukaryotic Transcription Complex Assembly. In Methods in Enzymology, pp. 207–219.

Kadonaga, J.T. (2012). Perspectives on the RNA polymerase II core promoter. Wiley Interdiscip. Rev. Dev. Biol. *1*, 40–51.

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

Kamenova, I., Warfield, L., and Hahn, S. (2014). Mutations on the DNA Binding Surface of TBP Discriminate between Yeast TATA and TATA-Less Gene Transcription. Number 15 Mol. Cell. Biol. *34*, 2929–2943.

Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J., et al. (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell *149*, 753–767.

Kedmi, A., Zehavi, Y., Glick, Y., Orenstein, Y., Ideses, D., Wachtel, C., Doniger, T., Waldman Ben-Asher, H., Muster, N., Thompson, J., et al. (2014). *Drosophila* TRF2 is a preferential core promoter regulator. Genes Dev. *28*, 2163–2174.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. *10*, 845–858.

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

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*, 520–527.

Klejman, M.P., Zhao, X., A van Schaik, F.M., Herr, W., and Th Marc Timmers, H. (2005). Mutational analysis of BTAF1-TBP interaction: BTAF1 can rescue DNA-binding defective TBP mutants. Nucleic Acids Res *33*.

Kopytova, D. V, Krasnov, A.N., Kopantceva, M.R., Nabirochkina, E.N., Nikolenko, J. V, Maksimenko, O., Kurshakova, M.M., Lebedeva, L. a, Yerokhin, M.M., Simonova, O.B., et al. (2006). Two isoforms of Drosophila TRF2 are involved in embryonic development, premeiotic chromatin condensation, and proper differentiation of germ cells of both sexes. Mol. Cell. Biol. *26*, 7492–7505.

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

Lanzer, M., de Bruin, D., and Ravetch, J. V (1992b). A sequence element associated with the Plasmodium falciparum KAHRP gene is the site of developmentally regulated protein-DNA interactions. Nucleic Acids Res. *20*, 3051–3056.

Lanzer, M., de Bruin, D., and Ravetch, J. V (1992c). Transcription mapping of a 100 kb locus of Plasmodium falciparum identifies an intergenic region in which transcription terminates and reinitiates.; Transcription mapping of a 100 kb locus of Plasmodium falciparum identifies an intergenic region in which. EMBO J. *1*, 1–949.

Lanzer, M., De Bruin, D., and Ravetch, J. V (1992d). A sequence element associated with the Plasmodium falciparum KAHRP gene is the site of developmentally regulated protein-DNA interactions. Nucleic Acids Res. *20*, 3051–3056.

Larson, A.G., and Narlikar, G.J. (2018). The Role of Phase Separation in Heterochromatin Formation, Function, and Regulation. Biochemistry *57*, 2540–2548.

Larson, D.R., Zenklusen, D., Wu, B., Chao, J.A., Singer, R.H., Muresan, L., Dugast-Darzacq, C., Hajj, B., Dahan, M., and Darzacq, X. (2011). Real-Time Observation of Transcription Initiation and Elongation on an Endogenous Yeast Gene. Science (80-.). *332*, 475–478.

Lee, D.-H., Gershenzon, N., Gupta, M., Ioshikhes, I.P., Reinberg, D., and Lewis, B.A. (2005). Functional Characterization of Core Promoter Elements: the Downstream Core Element Is Recognized by TAF1. Mol. Cell. Biol. *25*, 9674.

Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J. V., King, D.S., Banani, S.F., et al. (2012). Phase transitions in the assembly of multivalent signalling proteins. Nature *483*, 336–340.

Llinás, M., Deitsch, K.W., and Voss, T.S. (2008). Plasmodium gene regulation: far more to factor in. Trends Parasitol. *24*, 551–556.

Loy, D.E., Liu, W., Li, Y., Learn, G.H., Plenderleith, L.J., Sundararaman, S.A., Sharp, P.M., and Hahn, B.H. (2017). Out of Africa: origins and evolution of the human malaria parasites Plasmodium falciparum and Plasmodium vivax. Int. J. Parasitol. *47*, 87–97.

Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. *47*, W636–W641.

Martianov, I., Brancorsini, S., Gansmuller, A., Parvinen, M., Davidson, I., and Sassone-Corsi, P. (2002). Distinct functions of TBP and TLF/TRF2 during spermatogenesis: requirement of TLF for heterochromatic chromocenter formation in haploid round spermatids. Development *129*, 945–955.

Martin, E.W., and Mittag, T. (2018). Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. Biochemistry *57*, 2478–2487.

Martinez, E., Zhou, Q., L'Etoile, N.D., Oelgeschläger, T., Berk, A.J., and Roeder, R.G. (1995). Core promoter-specific function of a mutant transcription factor TFIID defective in TATA-box binding. Proc. Natl. Acad. Sci. U. S. A. *92*, 11864–11868.

Milton, R. (2017). Regulation of transcription in Plasmodium falciparum,the causative agent of severe malaria: initial characterisation of PfTBP and PfTFIIA. University of Cape Town.

Moll, J.R., Acharya, A., Gal, J., Mir, A.A., Vinson, C., and Gal, J. (2002). Magnesium is required for specific DNA binding of the CREB B-ZIP domain. Nucleic Acids Res. *30*, 1240–1246.

Müller, F., Lakatos, L., Dantonel, J., Strähle, U., and Tora, L. (2001). TBP is not universally required for zygotic RNA polymerase II transcription in zebrafish. Curr. Biol. *11*, 282–287.

Nakajima, N., Horikoshi, M., and Roeder, R.G. (1988). Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity, and TATA box-promoter

interactions of TFIID. Mol. Cell. Biol. *8*, 4028–4040.

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

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

Nikolov, D.B., Chen, H., Halay, E.D., Hoffman, A., Roeder, R.G., and Burley, S.K. (2002). Crystal structure of a human TATA box-binding protein/TATA element complex. Proc. Natl. Acad. Sci. *93*, 4862–4867.

Nott, T.J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-Jones, D.P., Pawson, T., Forman-Kay, J.D., et al. (2015). Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. Mol. Cell *57*, 936–947.

O'Brien, R., DeDecker, B., Fleming, K.G., Sigler, P.B., and Ladbury, J.E. (1998). The effects of salt on the TATA binding protein-DNA interaction from a hyperthermophilic archaeon. J. Mol. Biol. *279*, 117–125.

Oelgeschläger, T., Tao, Y., Kang, Y.K., and Roeder, R.G. (1998). Transcription activation via enhanced preinitiation complex assembly in a human cell-free system lacking TAFIIs. Mol. Cell *1*, 925–931.

Okonechnikov, K., Golosova, O., Fursov, M., and UGENE team (2012). Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics *28*, 1166–1167.

Oluwafikayo Adebolajo, D. (2016). Functional Characterization of the Plasmodium falciparum TBP-like Protein (PfTLP) DNA-binding Activity. University of Cape Town.

Pallarès, I., de Groot, N.S., Iglesias, V., Sant'Anna, R., Biosca, A., Fernàndez-Busquets, X., and Ventura, S. (2018). Discovering Putative Prion-Like Proteins in Plasmodium falciparum: A Computational and Experimental Analysis. Front. Microbiol. *9*, 1737.

Perkins, M.E. (1984). Surface proteins of Plasmodium falciparum merozoites binding to the

erythrocyte receptor, glycophorin. J.Exp.Med. *160*, 788–798.

Petri, V., Hsieh, M., and Brenowitz, M. (1995). Thermodynamic and Kinetic Characterization of the Binding of the TATA Binding Protein to the Adenovirus E4 Promoter. Biochemistry *34*, 9977–9984.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

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

Posey, A.E., Holehouse, A.S., and Pappu, R. V. (2018). Phase Separation of Intrinsically Disordered Proteins. Methods Enzymol. *611*, 1–30.

Privalov, P.L., Dragan, A.I., Crane-Robinson, C., Breslauer, K.J., Remeta, D.P., and Minetti, C.A.S.A. (2007). What Drives Proteins into the Major or Minor Grooves of DNA? J. Mol. Biol. *365*, 1–9.

Privalov, P.L., Dragan, A.I., and Crane-Robinson, C. (2011). Interpreting protein/DNA interactions: Distinguishing specific from non-specific and electrostatic from non-electrostatic components. Nucleic Acids Res. *39*, 2483–2491.

Riback, J.A., Katanski, C.D., Kear-Scott, J.L., Pilipenko, E. V, Rojek, A.E., Sosnick, T.R., Allan Drummond, D., Methodology, D.A.D.;, Investigation, D.A.D.;, and Software, D.A.D.; (2017). Stress-triggered phase separation is an adaptive, evolutionarily tuned response Author contributions Conceptualization HHS Public Access. Cell *168*, 1028–1040.

Le Roch, K.G. (2003). Discovery of Gene Function by Expression Profiling of the Malaria Parasite Life Cycle. Science (80-.). *301*, 1503–1508.

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

Roeder, R.G., and Rutter, W.J. (1970). Specific Nucleolar and Nucleoplasmic RNA Polymerases. Proc. Natl. Acad. Sci. U. S. A. *65*, 675–682.

Ruan, J.-P., Arhin, G.K., Ullu, E., and Tschudi, C. (2004). Functional characterization of a Trypanosoma brucei TATA-binding protein-related factor points to a universal regulator of transcription in trypanosomes. Mol. Cell. Biol. *24*, 9610–9618.

Ruff, K.M., Roberts, S., Chilkoti, A., and Pappu, R. V. (2018). Advances in Understanding Stimulus-Responsive Phase Behavior of Intrinsically Disordered Protein Polymers. J. Mol. Biol. *430*, 4619–4635.

Ruvalcaba-Salazar, O.K., Ramírez-Estudillo, M. del C., Montiel-Condado, D., Recillas-Targa, F., Vargas, M., and Hernández-Rivas, R. (2005). Recombinant and native Plasmodium falciparum TATA-binding-protein binds to a specific TATA box element in promoter regions. Mol. Biochem. Parasitol. *140*, 183–196.

Ryan, V.H., Dignon, G.L., Zerze, G.H., Chabata, C. V., Silva, R., Conicella, A.E., Amaya, J., Burke, K.A., Mittal, J., and Fawzi, N.L. (2018). Mechanistic View of hnRNPA2 Low-Complexity Domain Structure, Interactions, and Phase Separation Altered by Mutation and Arginine Methylation. Mol. Cell *69*, 465-479.e7.

Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A. V, Manteiga, J.C., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science *361*.

Schmidt, H.B., and Görlich, D. (2016). Transport Selectivity of Nuclear Pores, Phase Separation, and Membraneless Organelles. Trends Biochem. Sci. *41*, 46–61.

Sheehy, T.W., and Reba, R.C. (1967). Complications of falciparum malaria and their treatment. Ann. Intern. Med. *66*, 807–809.

Shen, T.H., Lin, H.-K., Scaglioni, P.P., Yung, T.M., and Pandolfi, P.P. (2006). The Mechanisms of PML-Nuclear Body Formation. Mol. Cell *24*, 331–339.

Sheu-Gruttadauria, J., and Macrae, I.J. (2018). Phase transitions in the assembly and function of human miRISC In Brief HHS Public Access. Cell *173*, 946–957.

Shevtsov, S.P., and Dundr, M. (2011). Nucleation of nuclear bodies by RNA. Nat. Cell Biol. *13*, 167–173.

Shrinivas, K., Sabari, B.R., Coffey, E.L., Klein, I.A., Boija, A., Zamudio, A. V., Schuijers, J.,

Hannett, N.M., Sharp, P.A., Young, R.A., et al. (2019). Enhancer Features that Drive Formation of Transcriptional Condensates. Mol. Cell *75*, 549-561.e7.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., et al. (2014). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539–539.

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

Starr, B.D., Hoopes, B.C., and Hawley, D.K. (1995). DNA Bending is an Important Component of Site-specific Recognition by the TATA Binding Protein. J. Mol. Biol. *250*, 434–446.

Takada, S., Lis, J.T., Zhou, S., and Tjian, R. (2000). A TRF1:BRF complex directs Drosophila RNA polymerase III transcription. Cell *101*, 459–469.

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

Thomas, M.C., and Chiang, C.-M. (2006). The General Transcription Machinery and General Cofactors. Crit. Rev. Biochem. Mol. Biol. *41*, 105–178.

Trampuz, A., Jereb, M., Muzlovic, I., and Prabhu, R.M. (2003). Clinical review: Severe malaria. Crit. Care *7*, 315–323.

Turoverov, K.K., Kuznetsova, I.M., Fonin, A. V., Darling, A.L., Zaslavsky, B.Y., and Uversky, V.N. (2019). Stochasticity of Biological Soft Matter: Emerging Concepts in Intrinsically Disordered Proteins and Biological Phase Separation. Trends Biochem. Sci. *44*, 716–728.

Vedadi, M., Lew, J., Artz, J., Amani, M., Zhao, Y., Dong, A., Wasney, G.A., Gao, M., Hills, T., Brokx, S., et al. (2007). Genome-scale protein expression and structural biology of Plasmodium falciparum and related Apicomplexan organisms. Mol. Biochem. Parasitol. *151*, 100–110.

Veenstra, G.J.C., Weeks, D.L., and Wolffe, A.P. (2000). Distinct Roles for TBP and TBP-Like Factor in Early Embryonic Gene Transcription in Xenopus. Science (80-.). *290*, 2312–2315.

Verma, N., Hung, K.-H., Kang, J.J., Barakat, N.H., and Stumph, W.E. (2013). Differential Utilization of TATA Box-binding Protein (TBP) and TBP-related Factor 1 (TRF1) at Different

Classes of RNA Polymerase III Promoters. J. Biol. Chem. *288*, 27564–27570.

Vo ngoc, L., Wang, Y.-L., Kassavetis, G.A., and Kadonaga, J.T. (2017). The punctilious RNA polymerase II core promoter. Genes Dev. *31*, 1289–1301.

Wang, Y.-L., Duttke, S.H.C., Chen, K., Johnston, J., Kassavetis, G.A., Zeitlinger, J., and Kadonaga, J.T. (2007). TRF2, but not TBP, mediates the transcription of ribosomal protein genes. Genes Dev. *28*, 1550–1555.

Wang, Z., Zhang, G., and Zhang, H. (2018). Protocol for analyzing protein liquid–liquid phase separation. Biophys. Reports 1–9.

Wheeler, R.J., and Hyman, A.A. (2018). Controlling compartmentalization by non-membrane-bound organelles. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *373*.

Willey, J.M., Sherwood, L., and Woolverton, C.J. Prescott's microbiology.

Woodruff, J.B., Ferreira Gomes, B., Widlund, P.O., Mahamid, J., Honigmann, A., and Hyman, A.A. (2017). The Centrosome Is a Selective Condensate that Nucleates Microtubules by Concentrating Tubulin. Cell *169*, 1066-1077.e10.

Wootton, J.C., and Federhen, S. (1993). Statistics of local complexity in amino acid sequences and sequence databases. Comput. Chem. *17*, 149–163.

World Health Organization (2018). World Malaria Report 2018 (Geneva).

Xue, H.Y., and Forsdyke, D.R. (2003). Low-complexity segments in Plasmodium falciparum proteins are primarily nucleic acid level adaptations. Mol. Biochem. Parasitol. *128*, 21–32.

Yadav, M.K., and Swati, D. (2012). Comparative genome analysis of six malarial parasites using codon usage bias based tools. *8*, 1230–1239.

Yoo, H., Triandafillou, C., and Drummond, D.A. (2019). Cellular sensing by phase separation: Using the process, not just the products. J. Biol. Chem. *294*, 7151–7159.

Zehavi, Y., Kedmi, A., Ideses, D., and Juven-gershon, T. (2015). TRF2 : TRansForming the view of general transcription factors. Transcription *6*.

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

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

Zhao, X., Schramm, L., Hernandez, N., and Herr, W. (2003). A Shared Surface of TBP Directs RNA Polymerase II and III Transcription via Association with Different TFIIB Family Members. Mol. Cell *11*, 151–161.

Zhou, H.-X., Nguemaha, V., Mazarakos, K., and Qin, S. (2018). Why Do Disordered and Structured Proteins Behave Differently in Phase Separation? Trends Biochem. Sci. *43*, 499–516.

Zilversmit, M.M., Volkman, S.K., DePristo, M.A., Wirth, D.F., Awadalla, P., and Hartl, D.L. (2010). Low-complexity regions in Plasmodium falciparum: missing links in the evolution of an extreme genome. Mol. Biol. Evol. *27*, 2198–2209.