Dissertation
submitted to the
Combined Faculty of Natural Sciences and Mathematics
of the Ruperto Carola University Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by
M.Sc. Kevin Marucha
born in: Kisii, Kenya on 09.09.1987
Oral examination: 07.12.2018

The suppressive role of 4EIP and PUF3 in gene expression during differentiation of *Trypanosoma brucei*

Referees:

Prof. Dr. Christine Clayton

Prof. Dr. Luise Krauth-Siegel

Acknowledgements

I thank my supervisors Prof. Dr. Christine Clayton, Prof. Dr. Luise Krauth-Siegel and Prof. Dr. Sven Diederichs for their guidance and suggestions during TAC meetings.

I thank my colleagues Dr. Bin Liu, Dr. Larissa Melo Do Nascimento, Franziska Egler, Albina Waithaka and Tania Bishola for creating a friendly working environment, for the numerous fun moments we shared and scientific discussions. Special thanks to Dr. Claudia Helbig and Ute Leibfried for technical assistance. You enabled my work run efficiently. I also thank members of Krauth-Siegel and Papavasiliou labs for their suggestions during our lab meetings.

Many thanks to Franziska Egler for translating the summary of this thesis to German. I also thank Johanna Braun for assisting me with some experiments during her lab rotations.

I thank the DAAD (Deutscher Akademischer Austauschdienst) for funding my stay in Germany, Deutsche course and other expenses. I am grateful to Dr. Henning Belle of DAAD Freundeskreis for the countless football matches, movies, theatres and music concerts. I thank HBIGS (Heidelberg Biosciences International Graduate School) for the short courses they organized, for awarding me a travel grant to present part of this work in the 29th Annual Molecular parasitology meeting and for facilitating my doctorate program.

I thank Prof. Keith Matthews for providing PAD1 antibody and Prof. Michael Rosbash for providing the Hrp48-TRIBE plasmid.

I thank my family for their understanding throughout the time I have been away from home. I thank God for the health of mind and body. *Gratias tibi Deus, gratias tibi*.

Table of Contents

Zusammenfassung

Afrikanische Trypanosomen sind begeißelte Protozoenparasiten, welche die Schlafkrankheit beim Menschen und die Nagana-Seuche bei Rindern verursachen. Während ihres Lebenszyklus verändern sie sowohl ihre Morphologie als auch ihren Metabolismus mittels ausgeprägter genregulatorischer Mechanismen. Trypanosomen sind gekennzeichnet durch die einzigarte, polycistronische Anordnung ihrer Gene und müssen daher fast ausschließlich auf post-transkriptionelle Regulationsmechanismen zurückgreifen, um die Genexpression zu regulieren, was Verarbeitung, Export, Stabilität und Translation der mRNA umfasst. Daran sind RNA-bindende Proteine wie auch Translationsfaktoren beteiligt. Trypanosomen haben sechs eIF4Es mit unterschiedlichen Affinitäten für die mRNA-Kappe und fünf eIF4Gs, welche vielerlei Möglichkeiten für die Translationsregulation darbieten. TbEIF4E1 interagiert mit keinem der eIF4Gs, sondern stattdessen mit einem 4E-interagierenden Protein, 4EIP, auf welchem der Focus des ersten Teils dieser Thesis liegt. Sowohl TbEIF4E1 als auch Tb4EIP reprimieren die Expression einer Reporter-mRNA, wenn sie artifiziell an diese gebunden sind („Tethering Assay“), jedoch braucht die Inhibition durch TbEIF4E1 Tb4EIP. Blutstrom-Formen von Trypanosoma brucei, welchen Tb4EIP fehlt, zeigen nur einen geringen Wachstumsnachteil. Bei hoher Parasitämie hören die Blutstrom-Formen auf sich zu teilen, unterdrücken die Translation und werden zu Stumpy-Formen, welche an die Differenzierung zu prozyklischen Formen angepasst sind, sobald sie von einer Tsetsefliege aufgenommen werden. Interessanterweise beeinträchtigt das Fehlen von Tb4EIP die Entwicklung zur Stumpy-Form, und dieser Effekt kann durch eine verkürzte Version von 4EIP gerettet werden, welche nicht an TbEF4E1 binden kann. Stumpy-Formen der Tb4EIP-defizienten Zellen haben ungewöhnlich hohe Protein-Syntheseraten, was darauf hindeutet, dass Tb4EIP für die Unterdrückung der Translation während der Differenzierung zu Stumpy-Formen benötigt wird.

Das RNA-bindende Protein PUF3 gehört zu den 11 T. brucei Proteinen mit einer Pumilio-Domäne. Bisher war wenig über PUF3 bekannt, außer dass es die Translation einer Reporter-mRNA unterdrücken kann und dass es zusammen mit Poly(A)-mRNA aufgereinigt werden kann. Daher liegt auf diesem der zweite Schwerpunkt meiner Thesis. Monomorphe und pleomorphe Blutstrom-Formen, in welchen die Menge an PUF3 erheblich vermindert wurde, haben einen sehr geringfügigen Wachstumsdefekt, jedoch zeigen Zellen mit einem PUF3 Gen-Knockout diesen Defekt seltsamerweise nicht. Bei Verwendung der PUF3-depletierten Zellen für Differenzierungsexperimente zu Stumpy-Formen oder prozyklischen Formen, haben diese Zellen interessanterweise eine Verzögerung bei diesen Differenzierungsprozessen gezeigt, was mit einer geringen Expression des Stumpy-Form Markerproteins PAD1 bzw. der prozyklischen Oberflächenproteine EP/GPEET einherging. Nichtsdestotrotz sind die Zellen letztendlich zu prozyklischen Formen differenziert. Interessanterweise haben sich die Blutstrom-Formen ohne PUF3 zu einem späteren Zeitpunkt angepasst und haben diesen Differenzierungsdefekt nicht gezeigt. Allerdings persistierte dieser Defekt in monomorphen Zellen, was einen Anpassungsmechanismus in pleomorphen Zellen nahelegt, welcher den zellulären Metabolismus angleicht an ein Leben ohne PUF3. Mittels der TRIBE (Targets of RNA binding proteins Identified by Editing) Methode wurden 295 mutmaßliche Ziel-mRNAs von PUF3 in Stumpy-Formen identifiziert. 79 dieser Ziel-mRNAs stammen von Genen, welche vorzugsweise in Blutstrom-Formen exprimiert werden, was mit einer Rolle für PUF3 als Repressor übereinstimmt. Diese umfassen mRNAs, welche für Zytoskelett-Proteine kodieren, sowie Proteinkinasen, RNA-bindende Proteine, Proteine mit Leucin-reichen repetitiven Sequenzen, sogenannte Expression-Site Associated Genes, Chaperone und Translationsfaktoren. Die dargestellten Ergebnisse deuten darauf hin, dass Tb4EIP und PUF3 bei der Feinabstimmung von mRNA-Mengen in Vorbereitung auf die Differenzierung wichtig sind.

Summary

African trypanosomes are flagellated protozoan parasites that cause sleeping sickness in humans and nagana in cattle. During their life cycle, they change their morphology and metabolism through robust gene regulation processes. Trypanosomes have a unique polycistronic gene arrangement and have to rely almost entirely on post-transcriptional regulation mechanisms, which include mRNA processing, export, stability and translation, in order to regulate gene expression. RNA binding proteins and translation factors are involved. Trypanosomes have six eIF4Es of varying cap-binding affinities and five eIF4Gs, suggesting numerous possibilities for translation regulation. *Tb*EIF4E1 does not interact with any of the eIF4Gs but instead interacts with 4E-interacting protein, 4EIP, which is the first focus of this thesis. Both *Tb*EIF4E1 and *Tb*4EIP repress a reporter mRNA in a tethering assay, but suppression by *Tb*EIF4E1 requires *Tb*4EIP. Bloodstream form *Trypanosoma brucei* lacking *Tb*4EIP have only a mild growth defect. At high parasitemia, bloodstream forms stop dividing, suppress translation and become stumpy forms, which are adapted to differentiate to procyclic forms when taken up by a tsetse fly. Interestingly, lack of *Tb*4EIP compromises stumpy formation, and the defect can be rescued by a truncated *Tb*4EIP that is unable to bind *Tb*EIF4E1. *Tb*4EIP knockout stumpy forms have abnormally high protein synthesis rates indicating that *Tb*4EIP is required for translation suppression during differentiation to the stumpy form.

RNA binding protein PUF3 is among 11 *T. brucei* pumilio domain containing proteins. Little was known about PUF3 save for its repression of a reporter mRNA and that it co-purifies with poly(A) mRNA. It is therefore the second focus of this thesis. PUF3-depleted monomorphic and pleomorphic bloodstream cells have a marginal growth defect but PUF3 knockout cells strangely lack this defect. Interestingly, when put to differentiate to stumpy and procyclic forms, PUF3-depleted pleomorphic bloodstream cells experience a delayed differentiation manifested by a low expression of the stumpy form marker PAD1 and procyclic surface coat proteins EP/GPEET. Nevertheless these cells eventually differentiated to viable procyclic forms. Surprisingly, pleomorphic bloodstream cells without PUF3 later seemed to adapt and lack this differentiation defect. The defect however persists in monomorphic cells, suggesting an adaptation mechanism in pleomorphic cells that equilibrates the cellular metabolism to life without PUF3. Using TRIBE (Targets of RNA binding proteins Identified by Editing) 295 putative targets of PUF3 were identified in stumpy-like cells. 79 of these targets are enriched in bloodstream forms while only 12 are enriched in procyclic forms, consistent with a role of PUF3 as a repressor during differentiation to procyclic forms. These targets include mRNAs encoding cytoskeleton proteins, protein kinases, RNA binding proteins, leucine rich repeat proteins, expression-site associated genes, chaperones and translation factors. The results here suggest that *Tb*4EIP and PUF3 fine-tune gene expression in readiness for differentiation.

1.0 Introduction

1.1 Background

Trypanosomes are flagellated protozoan parasites that belong to the class *Kinetoplastida*. Kinetoplastids have a characteristic organelle called kinetoplast at the base of the flagellum, which contains mitochondrial DNA. They are also known for compartmentalizing glycolytic enzymes in a microbody-like organelle called the glycosome, instead of in the cytosol [1], and *trans*-splicing of a 39 nucleotide long sequence called the spliced leader on to all mRNAs [2,3] . Members of *Kinetoplastida* include free-living organisms and parasites of plants and animals.

The family *Bodonidae* and *Cryptobiidae* have two flagella and consists mostly of free-living aquatic species exemplified by *Bodo* species, with a few parasites e.g. the fish parasite *Cryptobia salmositica*. The family *Trypanosomatidae* consists of uni-flagellated parasites of plants, ciliates, vertebrates and invertebrates [4]. The genus Phytomonas infects plants but causes disease in only few plants including oil palm, coconut and coffee trees [5]. Other two genera of medical and veterinary importance are *Trypanosoma* and *Leishmania*. *Trypanosoma brucei* cause human and animal African trypanosomiasis and are transmitted by *Glossina* spp (Tsetse fly), whereas *T. cruzi* cause Chagas disease (also known as American trypanosomiasis) and are transmitted by *Triatominae* bugs. *Leishmania* species cause visceral, cutaneous or muco-cutaneous leishmaniasis and are transmitted by *Phlebotomus* and *Lutzomyia* spp (Sand fly).

1.2 Diseases caused by African trypanosomes

Trypanosoma brucei rhodesiense and *T.b. gambiense* cause Human African trypanosomiasis (HAT), also known as sleeping sickness, and are responsible for 3% and 97% of HAT morbidities respectively [6]. HAT is characterized by two stages [7]. In the early stage, parasites multiply in the skin, blood and lymph causing non-specific signs and symptoms such as itchiness, headaches and joint pains. The second or late stage occurs when the parasite crosses the blood-brain barrier and manifests with more obvious symptoms like altered sleeping cycle, sensory disorders, visual impairment and motor weakness. If not treated the patient progresses into coma and death [7,8].

T.b. rhodesiense is found in eastern and southern Africa and it causes an acute form of the disease, which enters the late stage within a few weeks or months. Suramin and melarsoprol

are used to treat the early and late stages, respectively, of the *T.b. rhodesiense* HAT [8,9]. *T.b. gambiense* causes a chronic form of HAT whose symptoms manifest more slowly. The early stage of *T.b. gambiense* HAT is treated with pentamidine whereas the late stage is treated with melarsoprol or nifurtimox-eflornithine combination therapy (NECT) [8,10,11].

The burden of treatment is in the toxicity of the drugs, the large amounts required per dose and the long period of administration that requires a health centre and specialists [7,12]. Melarsoprol, for example, causes encephalopathy in 10% of patients, which is lethal in 50% of such cases [7,8,13]. Despite the challenges in diagnosis and treatment, surveillance and control programs have contributed to a decrease in new reported cases to 3,797, the lowest since 2010 [6].

Animal African trypanosomiasis (AAT) is mainly caused by *Trypanosoma congolense* and *Trypanosoma vivax*. Other species that cause nagana and similar diseases include *T.b. brucei* and *T. simiae*; they infect domestic (cattle, pigs, sheep, goat, camels and dogs) and wild animals [14,15]. *T. evansi* is also found in Asia, central and southern America and it causes Surra in camels, horses, buffalos, elephants and dogs. Its transmission is mechanical as it lacks the kinetoplast maxicircle DNA, which encodes mitochondrial proteins required for survival in the tsetse fly host [16]. *T. equiperdum* causes dourine in horses and donkeys worldwide, and is sexually transmitted [17]. AAT is treated mainly using homidium bromide and chloride, isometamidium and diminazene aceturate. These are very old drugs with high toxicity and have a huge problem of resistance [15].

Collectively, with infected and affected humans unable to work, and their infected animals unproductive in both food production and horse power, the socio-economic impact of African tryanosomiasis and tsetse flies is enormous; Food and Agriculture Organization (FAO) estimates that the combined impact of HAT and AAT in terms of gross domestic product is a loss of approximately $4.7 billion per annum (http://www.fao.org/ag/againfo/programmes/en/paat/disease.html). In the spirit of finding better treatments and disease control mechanisms, great advances have been made over a century on trypanosome cell biology and have led to discoveries like antigenic variation and *trans*-splicing which have broad applications. Trypanosomes are also genetically amenable and have become a model in studying aspects in parasitology and cell biology.

1.3 Life cycle of *T. brucei* and *in vitro* differentiation

The life cycle of *T. brucei* involves the cyclical transmission between mammals and tsetse flies and is summarized in Figure 1. In the mammal, *T. brucei* divide as long slender forms in the blood and other tissue fluids and are referred to as bloodstream forms. They have a rudimentary mitochondrion and utilize substrate-level phosphorylation for ATP production in a specialized microbody called the glycosome, which compartmentalizes the first seven glycolytic enzymes [1,18].

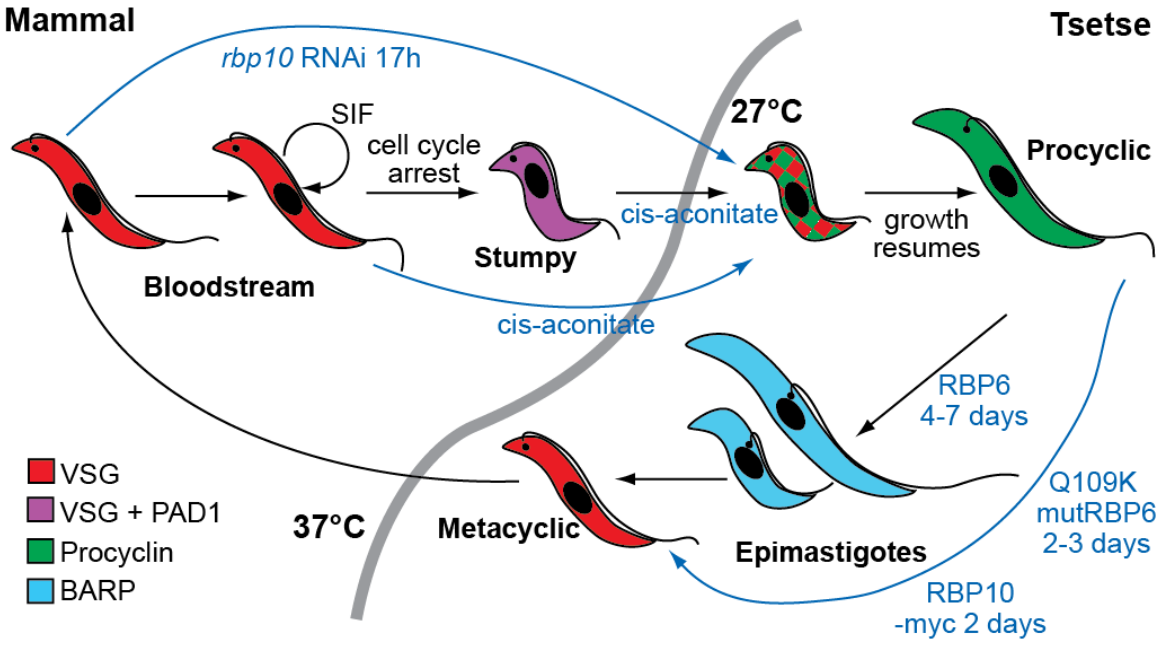


Figure 1: Life cycle of *Trypanosoma brucei*. Surface proteins are shown in different colour shades. Blue text and arrows indicate *in vitro* introduced manipulations that mimic the natural course of the life cycle. SIF – stumpy induction factor. The image was made by Christine Clayton and it includes minor modifications.

Bloodstream forms have a dense GPI-anchored coat of variant surface glycoproteins (VSGs). Its periodic replacement with a new variant enables the parasites evade the host immune machinery, a phenomenon called antigenic variation [19–21]. As the host immune catches up with the parasites, a sub-population that has switched to expressing a new VSG can survive, thus creating rising and falling waves of parasitemia. It is known that trypanosomes have up to 2000 VSG genes and gene fragments, and only one VSG is expressed at a time, guaranteed by an enigmatic allelic exclusion mechanism [21]. With this repertoire of genes and antigenic variation, the parasite is always steps ahead of the host immune and vaccination against trypanosomes is probably impossible.

At peak parasitemia, the long slender forms respond to a stumpy induction factor (SIF) in a density dependent mechanism, become arrested at G1/G0 phase, and differentiate into cells with short-stumpy morphology [22–25]. It is thought that this growth arrest ensures that cytological repositioning events that accompany differentiation are not interfered with by those accompanying cell division [25]. Stumpy forms have reduced translation but express some new proteins including some mitochondrial proteins, are tolerant to low pH and proteolysis and are thus pre-adapted to differentiate further once in the insect midgut [23,26]. In addition, they express carboxylate transporters called PAD (proteins associated with differentiation), which convey the differentiation signal cis-aconitate. PAD1 in particular is specifically expressed in stumpy forms and is thus a convenient stumpy marker [27]. PAD2 is expressed in response to cold [27], whereas PAD4, 6 and 8 seem to be expressed in epimastigotes [28]. The roles of most PAD proteins is still unclear.

To prevent premature differentiation to stumpy forms, a tyrosine phosphatase *Tb*PTP1 in slender forms phosphorylates and inhibits the DxDxDx serine threonine phosphatase, *Tb*PIP39 (PTP1 interacting protein, 39kDa). During differentiation, *Tb*PTP1 is inhibited, and *Tb*PIP39 is dephosphorylated, localized to the glycosomes and stumpy formation is initiated [29]. The targets of *Tb*PIP39 in the cascade are still unknown.

Apart from becoming stumpy in shape, another cytological rearrangement that occurs during stumpy formation is relocation of the lysosome. In slender forms, the lysosome is approximately equidistant between the kinetoplast and the nucleus. As they become stumpy, it relocates from the posterior to the anterior of the nucleus producing the relatively short distance between the nucleus and kinetoplast characteristic of stumpy forms [25,30].

In culture, slender to stumpy differentiation is achieved by growing slender cells to peak densities in viscous HMI-9 media containing 1.1% methyl cellulose [31,32]. Alternatively, slender cells can be grown in the presence of hydrolysable cAMP analogs that induce stumpy formation [33–35]. These compounds however generate unhealthy cells lacking a fully stumpy morphology [35]. Disruption of the active VSG expression sites can also induce stumpy formation [36].

During a blood meal from an infected mammal, the fly ingests the stumpy forms along with the blood. The stumpy forms develop a functional mitochondrion, the kinetoplast repositions to a position midway between the posterior edge and the nucleus, and they change their VSG coat to GPEET and EP procyclins to become procyclic forms. GPEET is highly phosphorylated while EP is highly glycosylated making both proteins appear like a smear in Western blots.

Dividing procyclic forms utilize amino acids (mainly proline) as their main carbon source, and oxidative phosphorylation for ATP production [37]. Usually not all flies get infected after a blood meal; by the fifth day, fly infections fall into either non-detectable or heavy infection, partly due to the fly's immune barriers [38]. GPEET is repressed after day six and this time coincides with the crossing of the procyclic forms into the ectoperitrophic space, implicating GPEET in a specific function while the parasites are still within the bloodmeal [39,40].

In culture conditions, differentiation to procyclic forms is induced by treatment of stumpy cells with 6mM cis-aconitate and reduction in temperature from 37°C to 27°C [41]. Alternatively, long slender forms can directly but less efficiently be transformed to procyclic forms using the same treatment [42]. GPEET expression *in vitro* is maintained by addition of glycerol in the growing media and its mRNA is regulated by the action of mitochondrial enzymes, glycerol and hypoxia [39,43] in addition to the action of RNA binding proteins [44]. Differentiation can conveniently be followed by monitoring the expression of GPEET or EP procyclin [32,45]. By immunofluorescence, the repositioning of the kinetoplast to a midpoint position between the nucleus and the posterior edge can also be used to identify procyclic forms.

Depending on their ability to differentiate to stumpy forms, cells are described as either monomorphic or pleomorphic. Monomorphic cells are laboratory-adapted strains that are incapable of differentiating to stumpy forms. They can differentiate to procyclic forms but these cannot divide and eventually die. Monomorphics can however grow to high densities in bloodstream-form cultures and are routinely used in most experiments. An example is the Lister 427 strain. In contrast, pleomorphic forms can differentiate to stumpy forms and can complete the life cycle in tsetse flies. In culture, they are grown at low densities to avoid selecting for monomorphism. Bloodstream cultures can also be grown in soft agar or methyl cellulose containing media to promote retention of pleomorphism [46]. An example is the EATRO 1125 strain.

Procyclic forms multiply and migrate to the foregut to form epimastigotes. Epimastigotes express yet another GPI-anchored surface coat protein called BARP (brucei alanine-rich proteins) [47]. It was initially thought that all insect-stage forms except metacyclics express procyclins and that BARP was expressed in bloodstream forms hence its previous name 'bloodstream alanine-rich proteins' [48]; until transcriptomics showed 20-fold higher expression of BARP in epimastigotes, which led to its re-characterization and renaming [47]. Epimastigotes divide into two geometrically unequal diploid daughter cells, and some undergo meiosis [49]. The epimastigotes attach to the epithelium of the salivary glands and differentiate into

mammalian infective metacyclics, which are non-proliferative and express metacyclic VSGs as their surface coat. During the next blood meal, metacyclics are injected into the skin and differentiate to proliferative long slender forms in the bloodstream, and the cycle begins again.

Differentiation of procyclic cells to epimastigotes and to metacyclics has recently been demonstrated *in vitro* by induced expression of RNA binding protein 6 (RBP6) [50] or directly to metacyclics by induced expression of RBP10 [45]. However, the metacyclics obtained by over-expression of RBP6 could not progress to bloodstream forms *in vitro*, but only when a Q109K mutant of RBP6 was over-expressed [28]. In *T. cruzi*, this differentiation is achieved by nutritional stress [51] or by over-expression of *Tc*UBP1 [52].

Interestingly, the transmission stages (stumpy and metacyclics) are both non-proliferative, have VSG coats, are pre-adapted to their next host and are basically dead-ends if not transmitted: much more remains to be discovered in these vulnerable stages. Recently, some parasites with a different metabolic profiles from those dividing in the body fluids have been reported to inhabit adipose tissues of the mammal and have raised speculations concerning a new form of the parasite [53].

In sum, trypanosome pleomorphism requires a robust regulation of gene expression that responds to nutritional requirements, host immunity, parasite density and other environmental cues. This control ensures both a balance between parasite and host, and parasite transmission.

1.4 Gene regulation in *Trypanosoma brucei*

Trypanosomes and related kinetoplastids have a peculiar genome arrangement; genes are arranged in unidirectional transcription units and virtually all genes, with the exception of those encoding an RNA helicase and a poly (A) polymerase, lack introns [54]. This is reminiscent of bacteria operons and a deviation from the usual eukaryotic model. Nevertheless, polycistronic genome arrangement and *trans*-splicing have been reported in other unrelated eukaryotes like in nematodes [55], trematodes [56], euglenoids [57], chordates [58] and dinoflagellates [59]. Because these are remotely unrelated eukaryotes, it is not clear whether trans-splicing was in the last common eukaryote or it emerged separately several times. Therefore during transcription, individual mRNAs are excised from the primary transcript by *trans*-splicing of the 5'-end with a spliced leader sequence and polyadenylation of the 3'-end. Apart from the spliced leader, which has a discrete RNA polymerase II promoter sequence [60], transcription in trypanosomes

was thought to initiate from loci corresponding to specific histone modifications instead of promotors [61]. However, a recent study identified GT-rich promotors that lack specific motifs, and which initiate transcription and promote occupancy of less stable nucleosomes containing the histone variant H2A.Z [62]. Genes in a transcription unit are almost always unrelated and each mRNA will be found in different amounts [63] indicating control. Polycistronic gene arrangement makes control of individual mRNA at transcription level impossible and therefore control of expression of individual genes can only be achieved after transcription [64]; this can be done by regulating mRNA processing, nuclear export, stability and translation.

Trans-splicing and polyadenylation are coupled processes. Regulation at these steps includes alternative *trans*-splicing. Genome-wide mapping of RNAseq data indicates that about 90% of protein coding genes undergo alternative *trans*-splicing [65–68] and that half of these alternative events are developmentally regulated [66]. Alternative *trans*-splicing results in start codon skipping that prevents translation initiation, in/exclusion of targeting sequence that causes mislocalization, in/exclusion of regulatory elements that affect mRNA stability, and use of alternative open reading frames [66,68]. A mathematical model proposed that variation in transcription-processing rates could result in a competition with the co-transcription pre-mRNA-destruction rates, and thus affect mRNA steady states [69].

Another gene control mechanism is by modulating the half-life of mRNA, which in bloodstream-form trypanosomes is a median of 13 min [63]. This is mainly achieved by mRNA degradation, translation and storage, all these orchestrated by RNA binding proteins (RBPs). RBPs bind particular mRNA elements to either recruit mRNAs to degradation or translation machineries. During degradation, mRNAs are deadenylated by the CAF1/NOT1 complex [70,71] followed by decapping by an ApaH-like phosphatase (ALPH1) [72], and are then degraded in the 5'-3' direction by the XRNA exoribonuclease, and in the 3'-5' direction by the exosome [73].

1.4.1 Gene regulation by RNA binding proteins

In many scenarios, RBPs bind *cis*-elements in the non-coding regions of mRNA particularly the 3'UTR, and modulate mRNA stability [35,45,74–76]. Other proteins, including translation initiation factors, usually bind the 5'UTR or the mRNA cap. Regulation by elements in the coding sequence is also possible [77]. *Trypanosoma brucei* has over 155 RBPs, which is a figure higher than bioinformatics predictions, because it includes proteins that have no obvious RNA binding motif yet co-purify with poly(A) RNA [78]. RBPs have extensively been reviewed and the functions of many are as yet unknown [79–81] (Table 1).

The effect of a protein on gene expression can be analysed using a tethering assay. This method exploits the affinity between the RNA-binding domain of the lambda bacteriophage anti-terminator protein N (λN-peptide) and its target boxB nucleotide sequence [82]. The protein of interest is expressed as a fusion with the 22-amino acid λN peptide, and multiple 19-nucleotide boxB sequences are added to the 3'UTR of a reporter gene e.g. chloramphenicol acetyltransferase or luciferase. The protein of interest would therefore be recruited to the reporter mRNA and its effect determined by assaying the fate of the reporter. Alternatives to the λN-boxB setup involves MS2/R17 coat proteins and their binding sequences [83], or other known peptide-nucleotide interactions.

Table 1: Examples of RNA binding proteins in *Trypanosoma brucei*

Protein	**Proposed function**	**Targets**	**Effect on reporter**[84]
PABP1, PABP2	Translation[85]	All mRNA	Activator
UBP1, UBP2	Likely in cell division [86]	Not specific	ND
RBP42	Regulating energy metabolism[87]	Energy metabolism [87]	Activator
P*TB*1, P*TB*2	Splicing, mRNA stability [76,88]	Not specific	Activator
RBP6	Differentiation [50]	Unknown	ND
RBP10	Maintaining bloodstream profile [45]	Procyclin mRNA, RBPs, energy metabolism	Repressor
ALBA 1-4	Translation [89]	Unknown	ND
HnRNPF/H	Developmental regulation, splicing, mRNA stability [90]		ND
ZC3H11	Heat shock response [91–93]	Chaperone mRNA	Activator
ZC3H12	Unknown	XRNA (not verified) [94]	ND
ZC3H13	Unknown	Not specific [94]	Repressor
ZC3H20	Developmental regulation [95]	MCP12, trans-sialidase	Activator
ZC3H18	Differentiation [96]	Unknown	ND
ZFP1,ZFP2	Differentiation to procyclic		Likely activators [97]
ZFP3	Differentiation to stumpy	Stumpy enriched transcripts [98]	Activator [98]
PUF9	Cell cycle [75]	LIGKA, PNT1 and	Activator [75]

		PNT2	
PUF1	Unknown	Retrotransposon RNA [99]	ND
PUF2	Unknown	Unknown	Repressor[100]
PUF5	Unknown	RNAs in procyclic [101]	ND
PUF7, PUF10	rRNA maturation [102]	unknown	ND

ND: not effect detected

RNA recognition motif (RRM) domain proteins are those that have at least one RRM and include poly(A) binding proteins (PABP), U-rich-sequence binding proteins (UBP), polypyrimidine tract binding protein (PTB), heterogeneous nuclear ribonucleoproteins (HnRNP) and other RBPs. Zinc finger domain proteins of the CCCH family bind RNA and sometimes function in developmentally regulated processes. ALBA (acetylation lowers binding affinity) domain proteins are versatile proteins that regulate gene expression. Their binding to RNA is regulated by acetylation/deacetylation, hence their name. They are involved in chromatin modification, transcription/translation regulation, development and differentiation processes. In trypanosomes they have been suggested to regulate the initiation steps of translation as they interact with PABPs, ribosomal protein P0 and cap binding protein *Tb*EIF4E4 (Alba3) [89,103]. *T. brucei* has 11 pumilio domain proteins (PUF) which are mostly repressors of gene expression except for PUF9, and seem to have versatile functions, which may in some cases be redundant. The pumilio domain protein PUF3 will be a focus of my thesis.

1.4.2 Pumilio domain proteins

The name PUF was coined from Pumilio and Fem3 mRNA binding factor (FBF) which were the first pumilio domain containing proteins to be described respectively in *Drosophila melanogaster* and *Caenorhabditis elegans*. PUF proteins have a unique structure (Figure 2); it consists of eight PUF repeats, collectively called the PUF domain, arranged in a concave arc that binds RNA. A PUF repeat is made up of 3 short helices and the second helix in each repeat recognizes a single base. Five residues in the second helix coordinate binding: the residues at position 1 and 5 interact with an RNA base [104,105]. Each PUF domain generally recognizes eight bases in a target RNA. There may be exceptions as some PUF proteins have less than the usual eight repeats while others have more [75,105]. Artificially higher numbers of repeats can give more binding specificity for various applications in biotechnology [105].

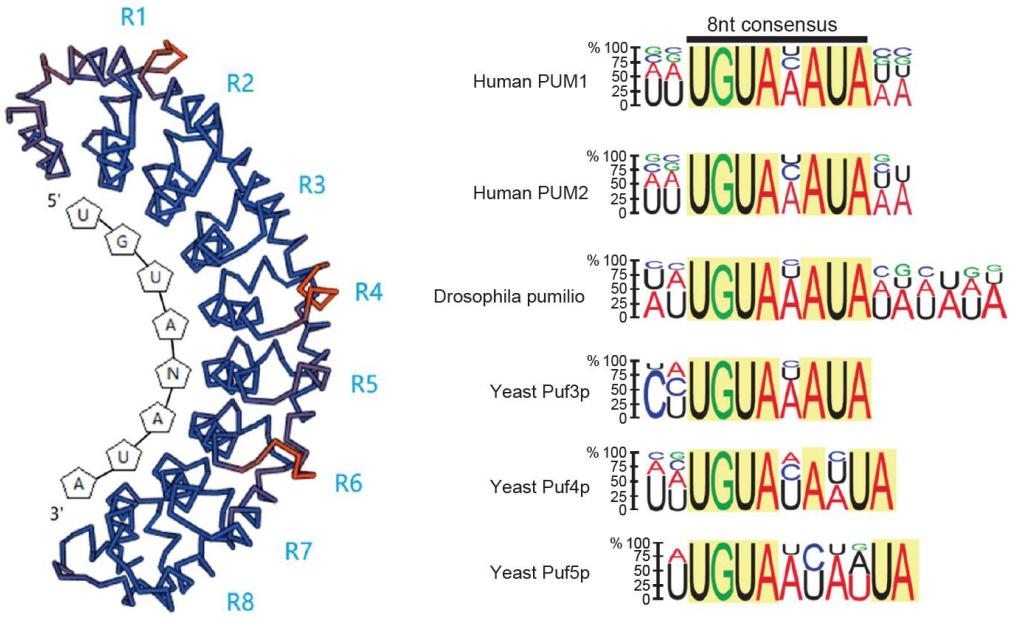


Figure 2: PUF domain structure and some consensus recognition motifs. Left panel is a stick model structure of *Arabidopsis thaliana* Pumilio 2. The eight PUF repeats (R1-R8) recognize eight bases in an RNA (UGUANAUA). The image was taken from [106]. Right panel shows consensus sequences obtained from 3'UTRs of PUF targets from human, yeast and *Drosophila*. The height of the nucleotide represents the probability that it occurs at that position, while conserved positions are shaded in yellow. The image was taken from [107]

In multicellular eukaryotes and yeast, PUF proteins are involved in cell differentiation and development functions; they bind their targets causing mRNA degradation or translation repression, though target stabilization has been reported [75,108]. Their numbers range from one in *Drosophila*, three in humans, five in yeast to 25 in *Arabidopsis* [106] and 11 in trypanosomes, attesting to their evolutionary conservation among eukaryotes. PUM and Nanos together repress the *hunchback* mRNA in *Drosophila* embryos to allow normal development of the posterior [109]. In the hermaphrodite *C. elegans*, FBF mediates the sperm-oocyte switch [110]. Human PUM2 in complex with DAZ, is implicated in maintenance of germline stem cells [111].

In trypanosomes, depletion of PUF1 has no effect on the transcriptome or growth of cells in culture [99] and neither does it associate with polysomes. It was shown to preferentially bind retroposon RNA but its function remains unknown. PUF2 also does not associate with polysomes and it represses mRNA in a tethering assay. It is essential for growth of bloodstream forms and its depletion favours a non-specific accumulation of mRNAs with short coding sequences [100]. Nonetheless, its biological function is not really understood. PUF5 is not essential for growth of either bloodstream or procyclic forms and neither is it required for differentiation of bloodstream to procyclic forms. Its over-expression seems to inhibit cell

growth in procyclic forms [101] and its function is unknown. However, its upregulation in epimastigotes suggests that it may function primarily in epimastigotes [28]. Unlike other PUF proteins, PUF7 and 10 are localized in the nucleolus and are involved in rRNA maturation [102], and together with the ribosome biogenesis protein BOP1 and NRG1 (nucleolar regulator of GPEET1), are negative regulators of GPEET mRNA [44]. PUF9 is extraordinary in that it stabilizes its target mRNAs. It stabilizes the mRNA encoding a kDNA ligase α (LIGKA) and two other targets named PUF Nine Target (PNT1 and PNT2) during S-phase via a motif UUGUACC [75].

In summary, apart from PUF7, 9 and 10, the functions of the pumilio proteins remain unclear. Despite concerns over functional redundancy [99], different phenotypes in PUF mutants, the number of PUF repeats, localization and independence of PUF7/10 (which share functions [44]) are suggestive of non-redundancy. This work will focus on PUF3 which has previously been detected among other strong repressors in a genome-wide tethering screen [84].

1.5 Control of translation

Messenger RNAs (mRNAs) in eukaryotes have a modified guanosine residue attached to the first transcribed nucleotide by a 5'-5' triphosphate linkage, followed by methylated nucleotides that make up the cap. Lower eukaryotes like yeast have cap 0 (m^7GpppN), where m is methylation, p is phosphorylation and N is any nucleotide, while higher organisms have extensively methylated caps i.e. cap 1 (m^7GpppNm) or cap2 (m^7GpppNm pN^m) [112]. During cap-dependent translation initiation, eukaryotic initiation factor eIF4E binds to the cap and recruits the scaffold protein eIF4G, which binds the DEAD-box RNA helicase eIF4A. The three initiation factors are together called eIF4F, which via eIF3 recruits the small ribosome subunit that is bound to the charged methionyl-tRNA and other factors to the cap, to form the 43S pre-initiation complex (PIC). This unwinds the 5' secondary structures and begins to scan the mRNA in the 5'-3' direction for the start codon AUG. On finding it, eIF5 and eIF5B promote hydrolysis of the GTP attached to eIF2, eIFs fall off and the large ribosomal 60S subunit is recruited to form the 80S ribosome with the methionyl-tRNA in the P-site [113]. Poly(A) binding proteins (PABP) can bind eIF4G to circularize the mRNA and create the closed-loop configuration, which enhances translation and recycling [114].

With the aid of elongation factors (eEF), the first elongation tRNA bearing an amino acid is delivered to the A-site in an eEF1A·GTP·aminoacyl-tRNA complex. The peptidyl transferase

action of the 60S subunit catalyses peptide bond formation between the amino acids on the tRNAs in the A- and P-sites, leaving behind a deacylated tRNA. Hydrolysis of the GTP promotes the mRNA to move by 3 nucleotides to position the next codon on the A-site, the peptidyl-tRNA on the P-site and the deacylated tRNA on the E-site. The next tRNA is brought into the A-site and the process repeats until a stop codon is reached. The release factor eRF1 then occupies the A-site and catalyses the hydrolysis of the ester bond linking the peptide chain to the tRNA in the P-site and terminates translation elongation [115].

Control of translation is found in all organisms particularly at the initiation step. Yeast has only one canonical (class 1) eIF4E, mammals two, *C. elegans* four, *A. thaliana* five, and *Drosophila* six. Besides these canonical ones are class 2 eIF4Es, which do not interact with eIF4G and are repressors of translation. These are exemplified by eIF4E2/4EHP in mammals, d4EHP in *Drosophila*, nCBP in *A. thaliana* and IFE-4 in *C. elegans* [115]. For example, mammalian 4EHP is recruited to a subset of mRNA by GIGYF2 and various RBPs including ZNF598, to form a repressive complex that is essential during embryonic development [116]. Another relevant RBP is tristetraprolin, which recruits 4EHP-GIGYF2 to mRNAs with AU-rich elements [117,118]. A further example is *Drosophila* 4EHP (d4EHP), which is implicated in early development: together with the RBP bicoid d4EHP represses translation of caudal mRNA [119], and with the RBPs Brat (brain tumor), Pum (pumilio) and Nos (nanos), it represses the hunchback mRNA [120]. *Arabidopsis* nCBP (novel cap-binding protein) is preferentially expressed in early stages of cell growth [121] and though classified as class 2 eIF4E [115], it interacts with the eIF4G isoform eIF(iso)4G but not with eIF4A [122]. RNAi of *IFE-4* in *C. elegans* causes defects in laying of eggs but its mechanism of action in not known [123]. The presence of multiple differentially expressed eIF4Es with only a few dedicated to general translation and some with a repressive role, is reminiscent of a control mechanism. This has potential in modulating mRNA selection for translation [124].

Regulation of translation initiation steps can also be achieved by eIF4E-binding proteins (4E-BP), which compete with eIF4G for binding eIF4E to prevent initiation [125]. Binding of eIF4G to the dorsal side of the cap binding protein eIF4E is via a canonical YXXXXLΦ motif where X is any amino acid and Φ is a hydrophobic amino acid [125]. This motif is also used by 4E-BP to bind eIF4E [126] hence binding of eIF4G and 4E-BP is mutually exclusive [127]. A second, non-canonical motif binds eIF4E on the lateral side and enhances binding by the first motif [128]. These motifs are flanked by phosphorylation sites that modulate binding to eIF4E [129,130].

4E-BPs include 4E-T, 4E-BP1-3 and Angel1 in human, Cup1, Mextli and 4E-T in *Drosophila*, Caf20 and Eap1 in yeast, and IFET-1 in *C. elegans* [125]. Mammalian 4E-BPs are phosphorylated by mTOR (mammalian target of rapamycin) kinase that prevents them from binding eIF4E, and this promotes translation of mRNAs with terminal oligopyrimidine tracts [131]. Stress conditions and inhibitors of mTOR including rapamycin and Torin 1, inhibit translation via the mTOR pathway [131,132]. *Drosophila* Cup is homologous to vertebrate 4E-T and it represses Oskar and Nanos mRNA via Bruno and Smaug RBPs respectively in early life development [133,134] in a cap-dependent mechanism. It inhibits the recruitment of the small ribosomal sub-unit to Oscar mRNA via a cap-dependent mechanism [135]. Angel is a member of the CCR4 deadenylase family and can outcompete eIF4G for eIF4E binding *in vitro* but not *in vivo* [136]. Mextli resembles eIF4G but without the eIF3A binding site, can drive translation *in vitro* and is required in embryogenesis [137]. Cap20 and Eap1 are dephosphorylated during stress response and inhibit translation [138].

In addition to blocking binding of eIF4G, 4E-BP can repress target mRNA independent of eIF4E as in the case of 4E-T and Cup1, or recruit decapping (Eap1) or deadenylation (Cup1) enzymes that destine the mRNA for degradation (reviewed in [125]). For example, human 4E-T is enriched in P-bodies where it is believed to represses translation of bound mRNA [139]. When tethered to a reporter mRNA, it can also represses translation in an eIF4E/P-body independent mechanism probably by its increased association with the decapping and deadenlyation machineries [140]. Tethering GIGYF2 also represses a reporter mRNA independent of 4EHP by recruiting the CCR4/NOT deadenylation complex. Repression by 4EHP requires GIGYF2 [141,142]. GIGYF2 also associates with Ago2 to repress translation via miRNA-mediated silencing [143]. *Drosophila* Cup represses Oskar mRNA independent of eIF4E by Bruno-dependent oligomerization of Oskar mRNA that form assemblies of 50S-80S “silencing particles”, which are inaccessible for translation [135].

Trypanosoma and their close relative *Leishmania* have six EIF4Es [144], five EIF4G[145], two EIF4A[146] and two poly(A) binding proteins [85] (*Leishmania* has three[147]) homologs (reviewed in [148]). In various combinations, they form different eIF4F complexes with EIF4E4-EIF4G3-EIF4A1 complex as the main eIF4F. EIF4E3 binds EIF4G4, EIF4E5 binds EIF4G1/2 and EIF4E6 binds EIF4G5, but EIF4E1 and 2 do not bind any of the eIF4Gs. Instead, EIF4E2 binds a stem loop binding protein (SLBP) – a homolog of mammalian histone mRNA-binding protein SLBP[149]. EIF4E1 binds a 4E-BP called 4E-interacting protein (4EIP) conserved only in kinetoplastids [150,151]. Works on 4EIP-EIF4E1 have been done in *Leishmania major*; *Lm*4EIP

binds *Lm*EIF4E1 in promastigotes (the insect form of the parasite) whereas in amastigotes (the form in the mammal), their interaction is abrogated and the expression of *Lm*4EIP but not *Lm*EIF4E1 is reduced [150]. Structural analysis of this interaction suggests that binding of *Lm*4EIP destabilizes the cap-binding of *Lm*EIF4E1 possibly through change in molecular conformation, thus inhibiting translation initiation [151]. However, it is not yet clear whether in the first place *Lm*EIF4E1 can initiate cap-dependent translation. An alternate hypothesis therefore is that *Lm*4EIP releases *Lm*EIF4E1 from the cap thus allowing binding of the cap by translation potent eIF4Es [151]. The exact mechanism of action is unknown.

1.5.1 *Trypanosoma brucei* 4E-interacting protein (*Tb*4EIP)

*Tb*4EIP has been studied by a number of people in our lab and their preliminary data are outlined in the following paragraphs. (Complete data was recently published in ref [152].) *Tb*4EIP, like in *Leishmania*, interacts only with *Tb*EIF4E1 and not with the other eIF4Es, at least in bloodstream forms. This interaction was abolished when the N-terminal canonical motif YXXXXLΦ was deleted. The interaction in *Leishmania* was developmentally regulated, but it is still unknown whether the interaction is sustained in the procyclic forms.

Both *Tb*EIF4E1 and *Tb*4EIP were previously identified as strong repressors in a tethering screen [78,84]. *Tb*4EIP represses mRNA stability and translation when tethered to the 3'UTR of a reporter gene. Repression was independent of *Tb*EIF4E1 interaction and the C-terminal fragment was sufficient for repression. It would be interesting to find out the reverse; whether *Tb*EIF4E1 represses gene expression in the absence of *Tb*4EIP.

Pull-down experiments illustrated that *Tb*4EIP targets mRNAs with low ribosome density and short half-lives. In addition, mass-spectrometry results showed that *Tb*4EIP consistently bound *Tb*EIF4E1 but no other obvious protein partners were found. Further, none of the yeast 2-hybrid interaction candidates, except *Tb*EIF4E1, could be confirmed in trypanosomes.

1.6 Aims of the study

The general aim was to determine the function of *Tb*4EIP and PUF3 particularly during differentiation. Specific aims were as follows:

1. To determine whether *Tb*4EIP and PUF3 are essential in *Trypanosoma brucei*
2. To determine whether *Tb*4EIP and *Tb*EIF4E1 interact in procyclic forms

3. To determine the effect of tethering *Tb*EIF4E1 in the absence of *Tb*4EIP
4. To identify protein binding partners of PUF3 and its target mRNAs
5. To determine whether *Tb*4EIP and PUF3 are required for differentiation of *Trypanosoma brucei*

2.0 Materials and methods

2.1 Culture of *Trypanosoma brucei*

All the cells used in this thesis were based on stable cell lines constitutively expressing a tetracycline repressor, which for simplicity are referred to as wild-type. Lister 427 bloodstream forms (monomorphic) were cultured in HMI-9 media supplemented with 5% PenStrep (Gibco), 1.5 mM L-cysteine, 0.2 mM β-mercaptoethanol and 10% fetal bovine serum (Heat inactivated at 55°C for 1 hr) at 37°C and 5% CO_2, and were diluted periodically as required. EATRO 1125 bloodstream forms (pleomorphic) were cultured in HMI-9 media containing 1.1 % methyl cellulose (Sigma, M0512) to maintain their pleomorphism. Procyclic forms were cultured at 27°C in MEM media supplemented with 5% PenStrep (Gibco), 7.5 mg/l hemin and 10% of heat inactivated fetal bovine serum at a density between 10^5 - 4×10^6 cells/ml.

Differentiation of bloodstream slender forms to stumpy forms was done by allowing a starting density of 5-7×10^5 cells/ml to peak at 3-4×10^6 cells/ml, and later descend to 10^6 cells/ml. This takes approximately 48 hrs. Differentiation to procyclic forms was done by adding 6mM cis-aconitate to 10^6 cells/ml slender or stumpy forms and incubating at 27°C for 24 hrs and thereafter transferring the cells to MEM.

Growth curves were performed using media without drugs whereas 100ng/µl of tetracycline was used when needed. Cell densities were determined using a Neubauer chamber.

2.2 Cell transfections

All the transfections were aimed at genomic integration by homologous recombination to obtain stable cell lines. To transfect bloodstream forms, 1.5×10^7 cells (density of 0.5-1×10^6 cells/ml) were collected at 900×g for 10 min at 4°C and the pellet re-suspended in 130 µl of ice-cold Amaxa buffer (90 mM NaH_2PO_4, 0.15 mM $CaCl_2$, 5 mM KCl, 50 mM HEPES, pH 7.3). 10 µg of linearized DNA was added to the cells in a cuvette and the cells electroporated using the free proprietary program X-001 of an Amaxa system as previously described [153]. The cells were transferred to 25ml of HMI-9 media and placed in growth conditions for 6 hours before adding the selection drug and seeding at a density not exceeding 5×10^5 cells/ml on a 24 well plate. To transfect procyclic forms, 2×10^7 cells (growing density of 1.5-3×10^6 cells/ml) were pelleted and re-suspended in ice-cold ZPFM buffer (132 mM NaCl, 8 mM KCl, 8 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 1.5 mM MgOAc.$4H_2O$, 90 µM $CaOAc_2$, pH 7.0) containing 10µg of linearized DNA in a cuvette. The cells were electroporated using the BTX system set at 1.5

volts and resistance at R2. Cells were placed in growing conditions for 6 hrs before addition of a selection drug. Cells were counted and diluted if necessary to a density of 3-5×10^6 cells/ml in MEM media, and seeded on a 24 well plate. The plate was sealed along the lid with tape to prevent evaporation of medium.

2.3 Indirect immunofluorescence microscopy

For immunofluorescence microscopy using non-chambered microscope slides, cells were smeared on a slide, air-dried and fixed using methanol for 10 min at -20°C. Alternatively, the cells were fixed in 4% paraformaldehyde at room temperature for 20 min, washed and smeared on slides. The slides were either stored at 4°C or analysed immediately. All the following steps were done at room temperature. The slides were placed in PBS for 5 min to rehydrate the cells then permeabilised using 0.2% Triton-X for 20 min. Cells were blocked with 0.5% gelatin for one hour and incubated with a primary antibody (dissolved in 0.5% gelatin) for one hour. The slides were washed three times with PBS and incubated in the dark with a secondary antibody containing a fluorescence dye for one hour. The slides were washed three times and incubated for 15 min with 1µg/ml DAPI solution. Three more washes were made and the slides air-dried before mounting with 90% glycerol-PBS. The cover slips were sealed to the slides using nail polish. Images were acquired using an Olympus IX81microscope and analysed with Olympus xcellence software or image J.

2.4 Protein detection by Western blotting

For Western blotting, 5×10^6 cells were collected at 900×g for 10 min, washed with PBS, re-suspended in Laemmli buffer (4% SDS, 20% glycerol (v/v), 120mM Tris-HCl, 0.02% (w/v) bromophenol blue) and heated at 95°C for 5 min. The cell lysates were analysed on either 8, 10 or 12% SDS-polyacrylamide gel as required. For experiments analysing PAD1, the cell lysates were not heated prior to electrophoresis and were run at 4°C as described by Dean [27]. Blotting was done by wet transfer (BIO RAD). Blots were blocked for at least one hour with 5% skimmed milk, incubated with a primary antibody for one hour, washed thrice with *TBS*-T (TBS, Tween-20) for five minutes per wash and incubated with a secondary antibody for one hour. Blots were washed thrice for 10 min per wash, reacted with ECL (perkinelmer) and the signal detected using x-ray films.

2.5 RNA preparation and Northern blotting

To obtain at least 10µg of RNA, 3×10^7 of log phase ($1.0\text{-}1.3\times10^6$ cells/ml) growing bloodstream-form cells were collected at 900×g for 10 min and RNA extracted using TriFast reagent according to the manufacturer's protocol. RNA concentrations were determined using a Nanodrop spectrophotometer and the RNA stored at -80°C or used immediately. rRNAs were depleted from whole RNA extracts before RNAseq using a cocktail of DNA oligonucleotides targeting rRNAs with an RNaseH based method as previously described [93]. The RNA was electrophoresed on a formaldehyde agarose gel, blotted on a nylon membrane by downward capillary for at least 4 hrs followed by UV cross-linking of RNA to the membrane using a Stratagene® UV cross-linker. Prior to probing, the blot was blocked using pre-hybridization buffer (5×SSC, 0.5% SDS, 5× Denhardt's solution, 100µg/ml Salmon sperm DNA) at 65°C. For probing the spliced leader, the pre-hybridization buffer was similar to the previous one except for 6×SSC and 0.05% sodium pyrophosphate. Probes made from PCR products were prepared using PrimeIt kit (Stratagene), and those from oligonucleotides using phosphonucleotide kinase (NEB) to incorporate radioactive nucleotides into the probe. The northern blot was incubated overnight with at least 2×10^6 cpm/ml of the probe at 42°C or 65°C for an oligonucleotide or a PCR-product probe respectively. The blot was washed with buffer (2 x SSC, 0.1 % SDS) twice for 10 min per wash at room temperature and once for 15 min at 65°C then exposed to a phosphorimaging film before developing the autoradiographs with a phosphorimager (Fugifilm FLA-3000).

2.6 Southern blotting

At least 5µg pf genomic DNA was digested using appropriate restriction enzymes and electrophoresed on an ethidium bromide containing agarose gel. The gel was washed in hydrolysis buffer (0.25 M HCl) for 15 min, in denaturing buffer (1.5 M NaCl, 0.4M NaOH) for 15 min and in neutralizing buffer (1.5 M NaCl, 0.5 M Tris-Cl pH7.5) for 15 min before blotting onto a nylon membrane and processing as described for the northern blotting.

2.7 Cell fractionation by digitonin

10^8 cells were collected at 900×g and washed three times in trypanosome homogenization buffer, THB (25mM Tris-Cl pH 7.8, 1mM EDTA, 150mM NaCl, 0.3M sucrose). The washes consisted of centrifugation (900×g, 5 min, 4°C) and re-suspension by gentle agitation of the

tube instead of pipetting. The cells were eventually re-suspended in 100µl of THB containing 2µg/ml leupeptin, 1mM DTT and always placed on ice. Using a stock of 0.9mg/ml of digitonin (dissolved in THB by heating at 95°C for 5min), aliquots of different digitonin concentrations were prepared and their volumes adjusted to be similar to each other and set at room temperature. 10^7 cells (10µl) were pipetted directly into the digitonin aliquots (at most, three tubes at a time) and incubated for 4 min at 25°C. Subsequently, the tubes were centrifuged (15,871×g, 1.5 min, 4°C) and both the supernatants and pellets collected separately, and dissolved in Laemmli buffer. The samples were analysed by Western blotting.

2.8 Chloramphenicol acetyltransferase (CAT) assay

2×10^7 cells were collected at 900×g for 10 min and washed three times with PBS. The pellet was re-suspended in 100µl of CAT buffer (100mM Tris-HCl pH 7.8) and freeze-thawed three times using liquid nitrogen and a 37°C heating block. The supernatants were collected by centrifugation at 10,000×g for 5 min and kept in ice. The protein concentrations were determined by Bradford assay (BioRad) according to the manufacturer's protocol. For each setup, 0.5µg of protein in 50µl of CAT buffer, 10µl of radioactive butyryl CoA (^{14}C), 2µl of chloramphenicol (stock: 40mg/ml), 200µl of CAT buffer and 4ml of scintillation cocktail were mixed in a Wheaton scintillation tube HDPE (neoLab #9-0149) and the incorporation of radioactive acetyl group on chloramphenicol was measured using program 7 of Beckman LS 6000IC scintillation counter.

2.9 Co-immunoprecipitation

10^8 cells were collected at 4°C and washed twice in ice-cold PBS. The cells were lysed in lysis buffer (50mM Tris-Cl pH 8.0, 120mM NaCl, 0.5% IGEPAL, 50ug/ml leupeptin) by passing them 15 – 20 times through a 21-gauge needle on ice. The lysate was precleared by centrifugation at 10,000×g, NaCl increased to 900mM, and then incubated with pre-washed myc/V5-antibody conjugated agarose beads for 2 hrs at 4°C. The unbound lysate was discarded and the beads washed three times with wash buffer (50mM Tris-Cl pH 8.0, 900mM NaCl, 0.5% IGEPAL) before adding Laemmli buffer and analysing by SDS-PAGE and Western blotting.

2.10 Tandem affinity purification (TAP)

6-7×10^9 cells were collected at 900×g at 4°C for 15 min in 500 ml centrifuge containers. The cell pellet was washed twice in ice cold PBS, snap frozen in liquid nitrogen and stored at -80°C or used immediately. The cells were lysed in IPP10 buffer (10mM Tris-HCL (pH 7.8), 10mM NaCl, 0.1% IGEPAL, 50µg/ml leupeptin) by passing them through a 21-gauge needle 15 – 20 times. The lysate was cleared by centrifugation at 10,000×g for 10 min and the concentration of NaCl adjusted to 150mM before loading onto a 0.8×4 cm Poly-Prep column (BioRad) containing IgG sepharose beads that had been prewashed in IPP150 buffer (10mM Tris-HCL (pH 7.8), 150mM NaCl, 0.1% IGEPAL). This was incubated at 4°C for 2 hours with gentle rocking. The beads were washed twice with 10ml IPP150 buffer and once with 10ml TEV cleavage buffer (IPP150 buffer, 1mM DTT, 0.5mM EDTA). The beads were re-suspended in 800µl TEV cleavage buffer and incubated with TEV protease at 15-20°C for 2 hours with gentle rocking. The eluate was collected and transferred to a second column containing calmodulin beads in calmodulin binding buffer (IPP150 buffer, 10 mM β-mercaptoethanol, 1 mM $MgAc_2$, 1 mM imidazole, 2 mM $CaCl_2$), and incubated for 2 hours at 4°C. The beads were washed three times with calmodulin binding buffer and the proteins were eluted using 2mM EGTA. The proteins in the eluate were concentrated using trichloroacetic acid and dissolved in Laemmli buffer. This was electrophoresed for a short time so that the sample dye ran approximately 1.5cm. The gel was then submitted to the sequencing facility at ZMBH (Heidelberg) for mass spectrometry.

2.11 RNA immunoprecipitation (RIP)

This procedure was done in a similar manner as the TAP purification with the exception that the cells were UV-irradiated before lysis and only one purification step was done. Approximately 3×10^9 cells were collected at 900 ×g at 4°C for 15 min, re-suspended on a petri dish in 30 ml of the supernatant on ice and UV- irradiated twice using a Stratagene UV crosslinker with an energy of 3 kilo joules. The cells were pelleted again in 4°C at 2000×g for 5 min and washed twice with ice cold PBS. Cells were lysed in IPP10 buffer consisting of KCl in place of NaCl and containing RNase inhibitor (RNasin, Invitrogen), by passing them through a 21-gauge needle 15-20 times. The lysate was cleared by centrifugation at 10,000×g for 10 min, KCl concentration adjusted to 150 mM and loaded onto a column containing prewashed sepharose beads similar as in the TAP purification. A small volume of the lysate equivalent to 1.5×10^7 cells was collected as the input fraction and treated with 0.2mg/ml proteinase K, 8mM

EDTA, 0.2% SDS at 45°C for 15 min before adding TriFast-FL reagent (Peqlab). The columns were incubated at 4°C for 2 hours with gentle rocking and the unbound lysate collected. This was also subjected to proteinase K treatment as described for the input fraction. Subsequently, the beads on the column were washed twice with IPP150 buffer and once with TEV cleavage buffer, re-suspended in 500µl of TEV cleavage buffer containing TEV protease and incubated at 15-20°C for 2 hours with gentle rocking. The cleaved product (bound) was treated with proteinase K and mixed with TriFast-FL reagent. The three fractions in TriFast-FL were stored at -80°C and later used for RNA extraction according to the manufacturer's protocol. RNA concentrations were determined using a NanoDrop® or a Qubit® spectrophotometer and submitted to the Next-generation sequencing facility in Bioquant (Heidelberg) for RNAseq.

2.12 RNA-sequencing data analysis

Fastq files were used as input in TrypRNAseq – an in-house pipeline for RNAseq (https://github.com/klprint/TrypRNAseq). Briefly, reads were taken through quality control, mapped on the *T. brucei* genome (TREU927), and the number of reads aligning to coding sequences (CDS) determined. The fold changes were compared between RNAi induced and un-induced samples using DESeq2 [154].

For RNA immunoprecipitation data, only genes with read counts ≥3 in all of the samples were considered in the analysis. The read counts were normalized to reads per million (RPM) and the average RPMs of the bound fraction were compared against the average RPMs of the unbound. Genes with fold change ≥2 were taken as putative targets. Motifs were predicted by DREME (Discriminative Regular Expression Motif Elicitation) [155] using 3'UTRs of the putative targets as input, while those of unbound genes served as controls. Genes lacking 3'UTR annotations were excluded from motif searches requiring 3'UTR sequence inputs.

2.13 Sequence retrieval and domain analyses

Trypanosomatid sequences were obtained from TriTrypDB (http://tritrypdb.org/tritrypdb/). Domains were detected using SMART (Simple Modular Architecture Research Tool: http://smart.embl-heidelberg.de/, [156]. Cladogram and sequence alignments were done using Simple phylogeny tool and Clustal Omega respectively, provided by EMBL-EBI services [157], and alignments were processed using Jalview [158].

Drosophila datasets previously used for analysis of Hrp48-TRIBE [159] were also used in this thesis for comparison and were downloaded from the Sequence Read Archive (SRA) database with the following ID: wild-type: SRR3177675, uninduced: SRR3177677, induced: SRR3177678. Hrp48 is a homolog of the mammalian heterogeneous nuclear ribonucleoproteins A/B family that participates in gene regulation in the circadian pathway [159].

2.14 TRIBE data analysis

2.14.1 Quality control and alignment

The analysis was similar in most aspects as originally described [159]. It involved RNA reads for three samples and DNA reads of EATRO1125 *T. brucei* strain in fastq format. All reads were first put through a stringent quality control trimming procedure to ensure very high quality reads hence high confidence in the mutations to be detected. Trimming was done using Trimmomatics [160] with the following options: HEADCROP:4 LEADING:25 TAILING:25 CROP:47 MINLEN:19 AVGQUAL:30. This allowed up to 25 bases to be removed on both 5' and 3' ends from reads with an average quality below 30 phred score. Four bases were removed from the 5' end of all reads since the quality of these first bases is usually low. Reads were cropped to 47 bases and only those with a minimum length of 19 bases were retained. The RNA reads were aligned to the genome of *T. brucei* TREU927 using BWA (Burrows-Wheeler Aligner), and samtools used to sort and index the alignments. DNA reads were instead aligned using bowtie2 [161] with the option "--sensitive". It should be noted that the TREU927 genome was used only as a reference for alignments and not for detecting TRIBE sites. Instead, DNA reads from EATRO1125 were used to avoid inclusion of single nucleotide polymorphism (SNPs). A perl script [159] was used to generate an alignment that contained only unique reads, and to convert the alignment into a table. Columns in this table represented each genomic locus while rows represented nucleotide bases. The number of rows hence represented the depth of the alignment. Tables generated for each sample were queried using MySQL version 8.0 database management system (Oracle).

2.14.2 Detection of edited sites

All A>G mutations were defined as loci with A>80% and G=0% in genomic DNA table and G>0% in RNA table. For genes in the reverse strand, T>C mutations were considered. These first results were referred to as raw results. Endogenous edit sites defined as loci with <10%

A>G editing and alignment depth below 20, were removed from each raw result. All edited sites in control cell line were subtracted from the experimental cell lines to give the actual edited sites. Annovar [162] was used to annotate edited sites while Microsoft Excel was used for analysis. Sequence motifs were predicted using DREME (Discriminative Regular Expression Motif Elicitation) [155] with sequences upstream/downstream of the edited site as the input.

2.15 List of plasmids and primers

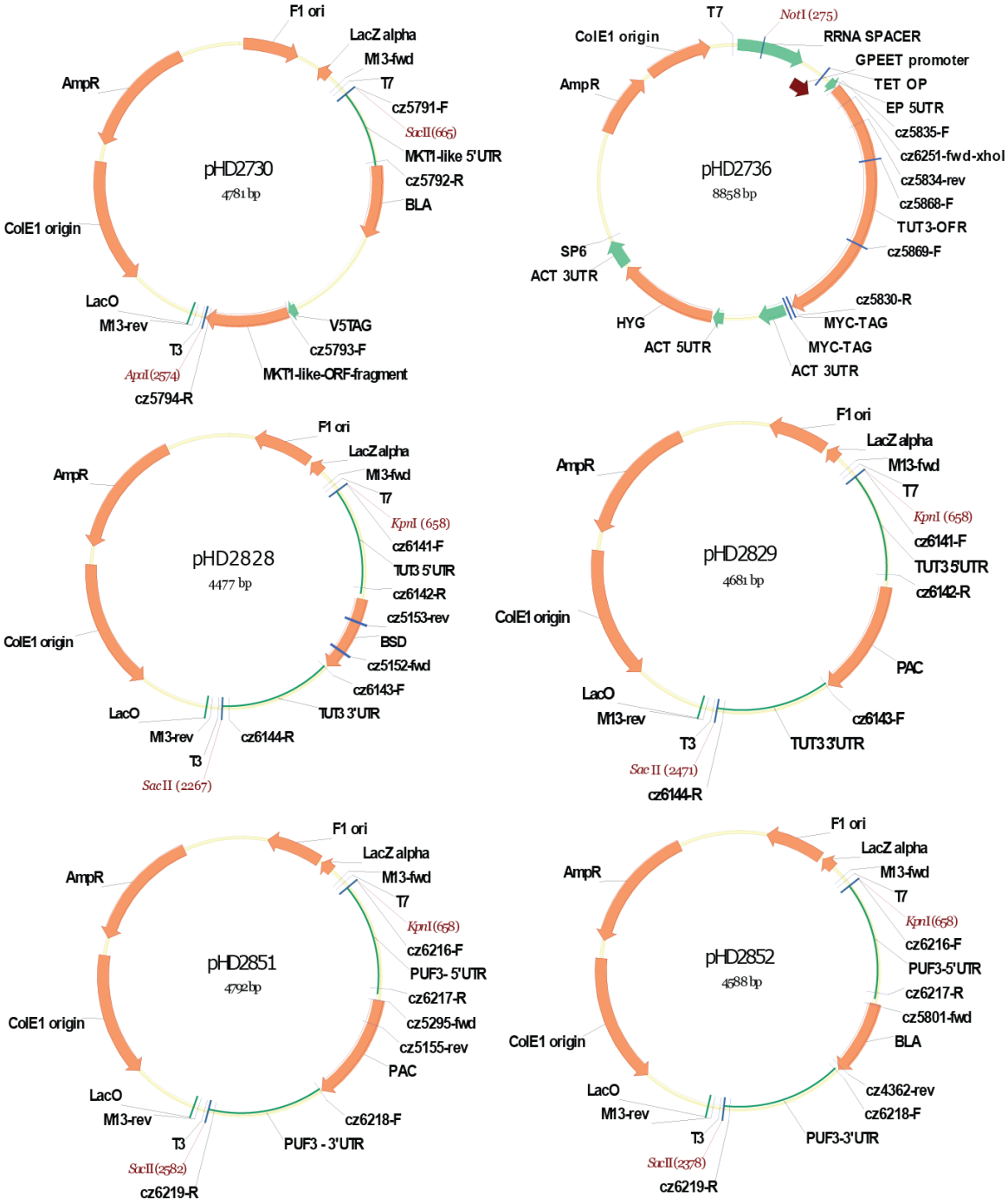
ID	Sequence	Plasmid	Purpose
cz5789	taacaagcttcatgaggaaagctggagca	pHD2729	MKT1-like-myc
cz5790	tattggatccgtgaggaatcggttcttg	pHD2729	MKT1-like-myc
cz5791	tataccgcggaatcatctttagcgtttc	pHD2730	V5-MKT1-like
cz5792	tatatctagattgtgcttcccccgtctg	pHD2730	V5-MKT1-like
cz5793	atatctcgagatgaggaaagctggagca	pHD2730	V5-MKT1-like
cz5794	atatgggccctatttgctccacctccca	pHD2730	V5-MKT1-like
cz5830	tatagttaacgtttttctgacagttgtacgcct	pHD2736	TUT3-myc
cz5831	tataccgcggcttttctcattttccagg	pHD2737	V5-TUT3
cz5832	tatatctagaatcctggtaaatgggcta	pHD2737	V5-TUT3
cz5834	tatagggcccagagagagcgcggcatggtgcac	pHD2737	V5-TUT3
cz5835	tatagggcccatgatatttttttttaaaaaacgataccacgtgttcttg	pHD2736	TUT3-myc
cz5841	tatactcgagatgatatttttttttaaaaaacgataccacgtgttcttg	pHD2737	V5-TUT3
cz5866	ggcccccggctggtattggtggtggcgg		sequencing pHD2729
cz5867	aatagcagcgtaaattatctagacaggg		sequencing pHD2729
cz5868	agctggagccccttgtgcggacccgcg		sequencing pHD2736
cz5869	taacctcggccgtcacctatcccctgac		sequencing pHD2736
cz6141	tataggtaccccacttctggcgctgctactgtttac	pHD2828 and pHD2829	TUT3 KO
cz6142	tatactcgaggcaagctaaggtacttctcgacctttgc	pHD2828 and pHD2829	TUT3 KO
cz6143	tatagaattcttgatgttggtacttcttcgctgaagtgg	pHD2828 and pHD2829	TUT3 KO
cz6144	tataccgcggcgaatttgtacatggtgactaaacattaaccg	pHD2828 and pHD2829	TUT3 KO
cz6194	catttcaacaacaacttccgttgacctttgttgtgcg		screen TUT3 KO
cz6216	tataggtaccgtggttagcggaagtagcagtagtgcaag	pHD2851	PUF3 KO

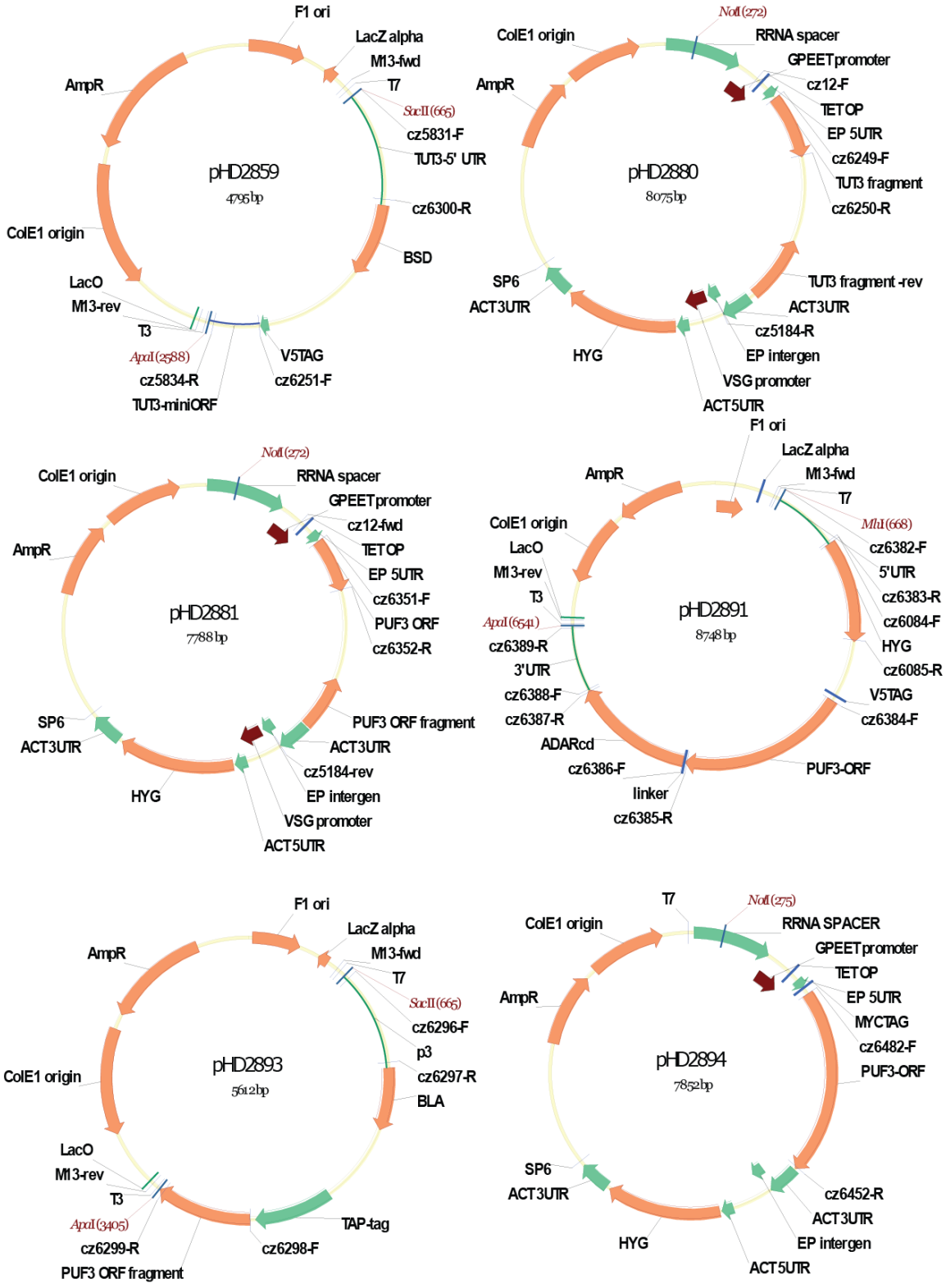
		and pHD2852	
cz6217	tatactcgagcctatatacccaaaatgataccttaagtg	pHD2851 and pHD2852	PUF3 KO
cz6218	tatagaattcactcccatggccaagtgttttagtggtgg	pHD2851 and pHD2852	PUF3 KO
cz6219	tataccgcggcgtggaaacacttgtggtaaccccgttc	pHD2851 and pHD2852	PUF3 KO
cz6220	gcagtggaacttttgcggaccgtgcaagcgg		screen PUF3 KO
cz6249	tataagatctgcatgcggtgaacggcctagtgatgt	pHD2880	TUT3 RNAi_stemloop
cz6250	tatagaattcgtcgacatcatccccctctttgcttt	pHD2880	TUT3 RNAi_stemloop
cz6251	tatcctcgagatgttcacatcggtgcctgtagctcc	pHD2859	V5-TUT3
cz6296	tatagagctcgtggttagcggaagtagcagtagtgcaag	pHD2893	TAP-PUF3
cz6297	tatacatatgcctatatacccaaaatgataccttaagtg	pHD2893	TAP-PUF3
cz6298	tataaagcttccatgtggactgtgcaggaggacgagtac	pHD2893	TAP-PUF3
cz6299	tatagggccccctttgccagctcgtaaacgtttccc	pHD2893	TAP-PUF3
cz6300	tatatctagagcaagctaaggtacttctcgacctttgc	pHD2859	V5-TUT3
cz6328	ttcttggttcaaaagctg		PUF3 probe
cz6329	tagcgtgattaccattcg		PUF3 probe
cz6330	acgcgggagggcctttcg		TUT3 probe
cz6331	tggacttttgcgctgcacag		TUT3 probe
cz6351	tataagatctgcatgccggacgataatgagcgaaat	pHD2881	PUF3 RNAi_stemloop
cz6352	tatagaattcgtcgaccctcagcaatttgcaaacag	pHD2881	PUF3 RNAi_stemloop
cz6382	taatccgcggacgcgtgggagaatataaagattc	pHD2891	TRIBE_C-terminal
cz6383	tatatctagacctatatacccaaaatgat	pHD2891	TRIBE_C-terminal
cz6384	tatagtcgacgcatgcatgtggactgtgcaggag	pHD2891	V5-PUF3-ADAR
cz6385	tatgaattcgccggagagcggttggtt	pHD2891	V5-PUF3-ADAR
cz6386	attgaattctgcaatatttcctacagtccaatggtggtgccacag	pHD2891	TRIBE_C-terminal
cz6387	tataagatcttcattcggcaagaccgaac	pHD2891	TRIBE_C-terminal
cz6388	aatctcgagtccgaattctccagatctactcccatggccaagtgt	pHD2891	TRIBE_C-terminal
cz6389	atatgggcccgggagtcgtggaaacact	pHD2891	TRIBE_C-terminal
cz6451	tatagggcccatgtggactgtgcaggaggac	pHD2895 and pHD2904	PUF3 and PUF3-myc
cz6452	atatgttaactcagccggagagcggttgg	pHD2895 and pHD2894	PUF3 and myc-PUF3
cz6453	atatgttaacgccggagagcggttgg	pHD2904	PUF3-myc
cz6482	atagggcccTTatgtggactgtgcaggagg	pHD2894	myc-PUF3

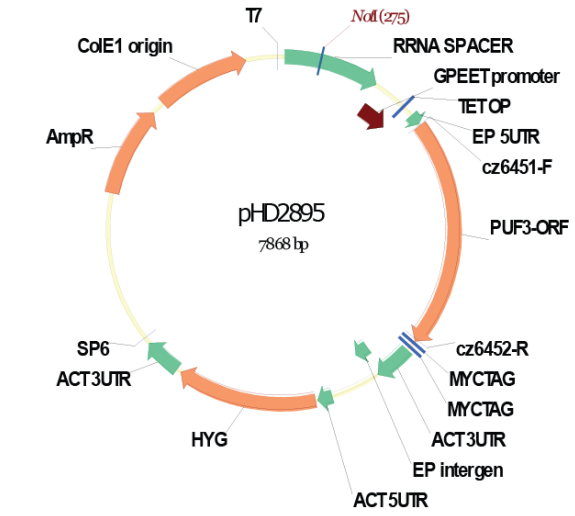
cz6728	tataagatctgcatgcgctgaatcaagcgcg	pHD2952	*Tb*EIF4E1 RNAi_stemloop
cz6729	tatagaattcgtcgacttacagtagtgaagcatcgc	pHD2952	*Tb*EIF4E1 RNAi_stemloop
cz6749	tatagggcccatgatggctgaatcaag	pHD2953	Tethering *Tb*EIF4E1
cz6750	tatagttaacggccttgctagcgccatg	pHD2953	Tethering *Tb*EIF4E1

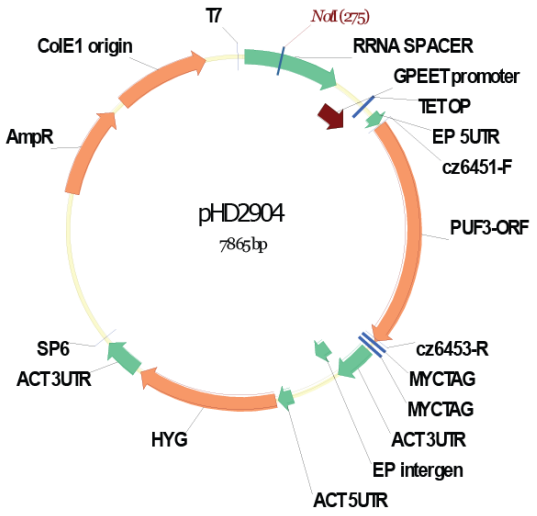
2.16 Plasmids used and their maps

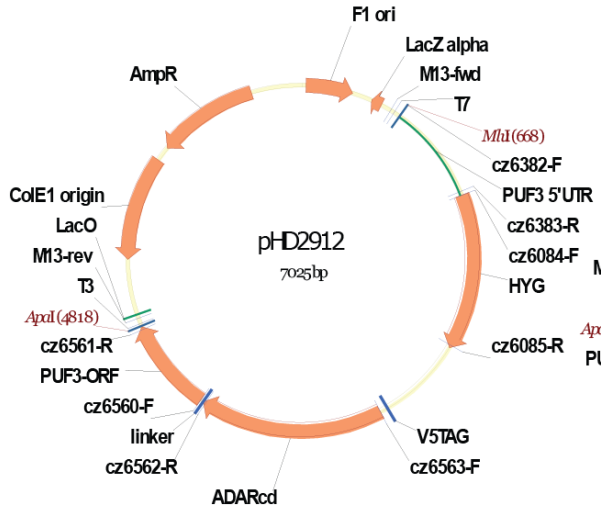
Enzymes used to linearize the DNA before transfection are written in red. Primers used for cloning genes/gene fragments are prefixed with the letters “cz”.

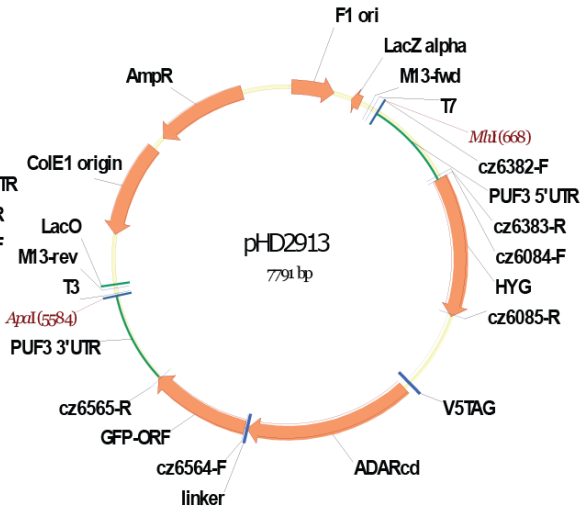


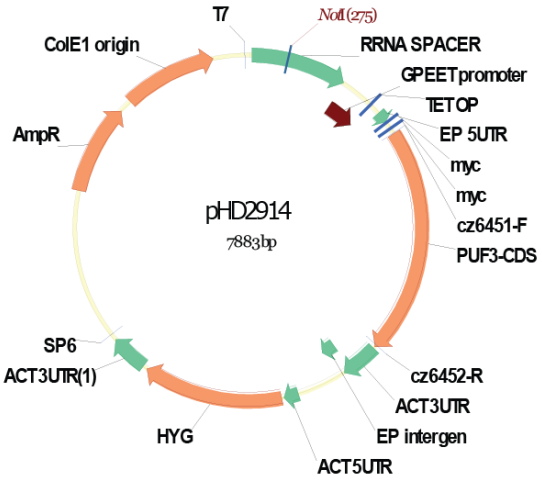

pHD2859
4795 bp
F1 ori
LacZ alpha
M13-fwd
T7
SacII (665)
cz5831-F
TUT3-5' UTR
cz6300-R
BSD
V5TAG
cz6251-F
ApaI (2588)
cz5834-R
TUT3-miniORF
LacO
M13-rev
T3
AmpR
ColE1 origin
pHD2880
8075 bp
NotI (272)
RRNA spacer
GPEET promoter
cz12-F
TETOP
EP 5UTR
cz6249-F
TUT3 fragment
cz6250-R
TUT3 fragment -rev
ACT3UTR
cz5184-R
EP intergen
VSG promoter
ACT5UTR
HYG
SP6
ACT3UTR
ColE1 origin
AmpR
pHD2881
7788 bp
NotI (272)
RRNA spacer
GPEET promoter
cz12-fwd
TETOP
EP 5UTR
cz6351-F
PUF3 ORF
cz6352-R
PUF3 ORF fragment
ACT3UTR
cz5184-rev
EP intergen
VSG promoter
ACT5UTR
HYG
SP6
ACT3UTR
pHD2891
8748 bp
F1 ori
LacZ alpha
M13-fwd
T7
MluI (668)
cz6382-F
5'UTR
cz6383-R
cz6084-F
HYG
cz6085-R
V5TAG
cz6384-F
PUF3-ORF
linker
cz6385-R
cz6386-F
ADARcd
cz6387-R
cz6388-F
3'UTR
cz6389-R
ApaI (6541)
LacO
M13-rev
T3
AmpR
ColE1 origin
pHD2893
5612 bp
F1 ori
LacZ alpha
M13-fwd
T7
SacII (665)
cz6296-F
p3
cz6297-R
BLA
TAP-tag
cz6298-F
PUF3 ORF fragment
cz6299-R
ApaI (3405)
LacO
M13-rev
T3
AmpR
ColE1 origin
pHD2894
7852 bp
T7
NotI (275)
RRNA SPACER
GPEET promoter
TETOP
EP 5UTR
MYCTAG
cz6482-F
PUF3-ORF
cz6452-R
ACT3UTR
EP intergen
ACT5UTR
HYG
SP6
ACT3UTR
AmpR
ColE1 origin

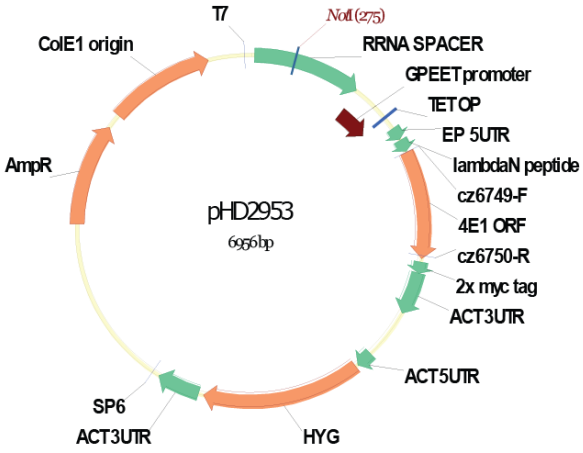

pHD2895
7868 bp
T7
NotI (275)
RRNA SPACER
GPEET promoter
TETOP
EP 5UTR
cz6451-F
PUF3-ORF
cz6452-R
MYCTAG
MYCTAG
ACT3UTR
EP intergen
ACT5UTR
HYG
SP6
ACT3UTR
AmpR
ColE1 origin


pHD2904
7865bp
T7
NotI (275)
RRNA SPACER
GPEET promoter
TETOP
EP 5UTR
cz6451-F
PUF3-ORF
cz6453-R
MYCTAG
MYCTAG
ACT3UTR
EP intergen
ACT5UTR
HYG
SP6
ACT3UTR
AmpR
ColE1 origin


pHD2912
7025bp
F1 ori
LacZ alpha
M13-fwd
T7
MluI (668)
cz6382-F
PUF3 5'UTR
cz6383-R
cz6084-F
HYG
cz6085-R
V5TAG
cz6563-F
ADARcd
cz6562-R
linker
cz6560-F
PUF3-ORF
cz6561-R
ApaI (4818)
T3
M13-rev
LacO
ColE1 origin
AmpR


pHD2913
7791 bp
F1 ori
LacZ alpha
M13-fwd
T7
MluI (668)
cz6382-F
PUF3 5'UTR
cz6383-R
cz6084-F
HYG
cz6085-R
V5TAG
ADARcd
linker
cz6564-F
GFP-ORF
cz6565-R
PUF3 3'UTR
ApaI (5584)
T3
M13-rev
LacO
ColE1 origin
AmpR


pHD2914
7883bp
T7
NotI (275)
RRNA SPACER
GPEET promoter
TETOP
EP 5UTR
myc
myc
cz6451-F
PUF3-CDS
cz6452-R
ACT3UTR
EP intergen
ACT5UTR
HYG
SP6
ACT3UTR(1)
AmpR
ColE1 origin


pHD2953
6956bp
T7
NotI (275)
RRNA SPACER
GPEET promoter
TETOP
EP 5UTR
lambdaN peptide
cz6749-F
4E1 ORF
cz6750-R
2x myc tag
ACT3UTR
ACT5UTR
HYG
SP6
ACT3UTR
AmpR
ColE1 origin

3.0 Results

PART 1 – EIF4E1 interacting protein *Tb*4EIP

3.1 *Tb*4EIP is located in the cytosol

Pleomorphic bloodstream cells lacking endogenous *Tb4EIP* (generated by Johana Braun) and ectopically expressing a tetracycline inducible *Tb*4EIP with a myc tag on the C-terminal end, were used to determine the cellular location of *Tb*4EIP by immunofluorescence microscopy. Procyclic cells were examined 3 days after differentiating the bloodstream cells (detailed in section 3.7) in the presence/absence of tetracycline. Cells were stained with anti-myc antibody (1:500, mouse) followed by anti-mouse secondary antibody (1:700) conjugated with Cy3 fluorophore. *Tb*4EIP signal appeared with higher intensity from the cytosol though the signal in bloodstream forms was weaker relative to procyclic forms (Figure 3). Some signal was observed in control cells possibly because of marginal expression in the absence of tetracycline. Using TrypTag [163], an online resource that contains localization data of many proteins in *T. brucei*, *Tb*4EIP also appears in the cytosol but in granules. Since *Tb*4EIP is known to congregate in stress granules [164] it is possible that the cells used in the imaging were already stressed due to incubation in PBS during the live cell imaging. The results and those from TrypTag suggest that most *Tb*4EIP is in the cytosol.

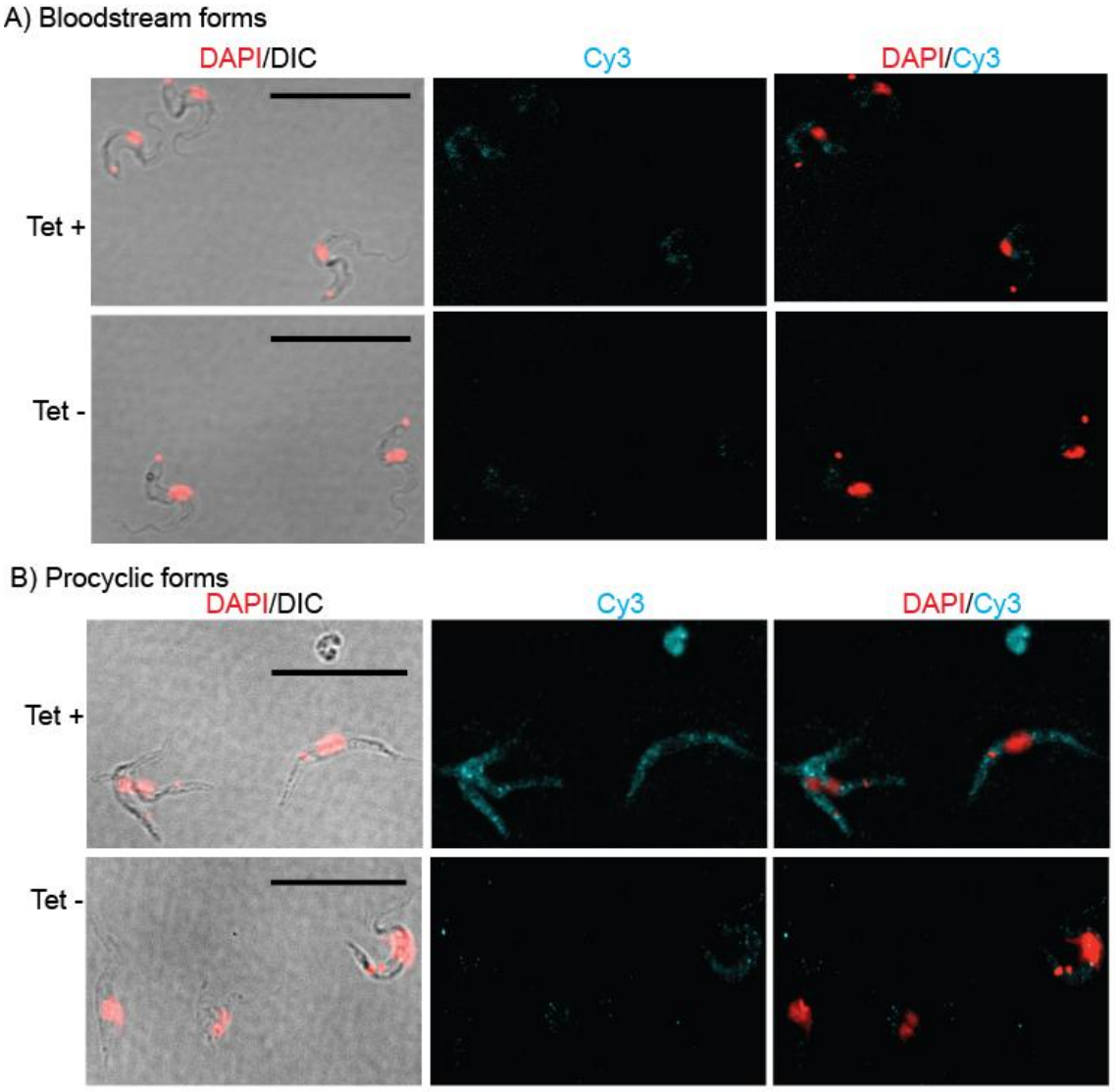


Figure 3: Cellular localization of *Tb*4EIP in bloodstream and procyclic forms. (**A**) Location of *Tb*4EIP in bloodstream cells. Bloodstream cells induced/non-induced (Tet+/-) with tetracycline to express 4EIP-myc for 24 hrs, were stained for microscopy. DAPI represents the nucleus and kinetoplast DNA staining, Cy3 represents *Tb*4EIP-myc and DIC is the differential interference contrast. The Cy3 signal was very faint and barely visible (**B**) Location of *Tb*4EIP in procyclic cells. Three day old procyclic cell growing in the presence/absence of tetracycline were processed similar as in (A). Bar =20µM

3.2 Truncated *Tb*4EIP does not interact with *Tb*EIF4E1 *in vivo*

Preliminary data showed that *Tb*4EIP interacts with *Tb*EIF4E1 and not with any of the other five eIF4Es. The interaction had been demonstrated in a yeast-two hybrid (Y2H) assay and in bloodstream forms by co-immunoprecipitation (co-IP). *Tb*4EIP-*Tb*EIF4E1 interaction was however lost when *Tb*4EIP lacking the canonical eIF4E binding motif (YXXXXLΦ) was used in

Y2H assay, showing that the motif is required for binding in yeast. Structural analysis of *Leishmania* protein versions showed a second helical (non-canonical) motif that increases the *Lm*4EIP-*Lm*EIF4E1 binding affinity 9-fold [151]. To check this result for trypanosome cell lysates, and whether the second motif is sufficient for binding *in vivo*, I used co-IP. The truncated *Tb*4EIP (*Tb*ΔN4EIP) was not able to pull down *Tb*EIF4E1 whereas the full-length protein was co-precipitated (Figure 4A). This confirmed that the N-terminal canonical motif is required for interaction *in vivo* even when the second motif is present.

In *Leishmania*, *Lm*4EIP-*Lm*EIF4E1 interaction was present in the insect stage promastigotes and not detected in vertebrate stage amastigotes [150]. To test whether the *Tb*4EIP-*Tb*EIF4E1 interaction is also developmentally regulated in trypanosomes, bloodstream forms were differentiated to procyclic forms and maintained for 4 weeks as procyclic forms before performing a co-IP. In contrast to *Leishmania* [150], *Tb*4EIP-*Tb*EIF4E1 interaction was detectable for both parasite forms (Figure 4B). Consistently, *Tb*ΔN4EIP in procyclic forms was also unable to pull down *Tb*EIF4E1 confirming that the canonical N-terminal motif is essential for binding in both parasite forms.

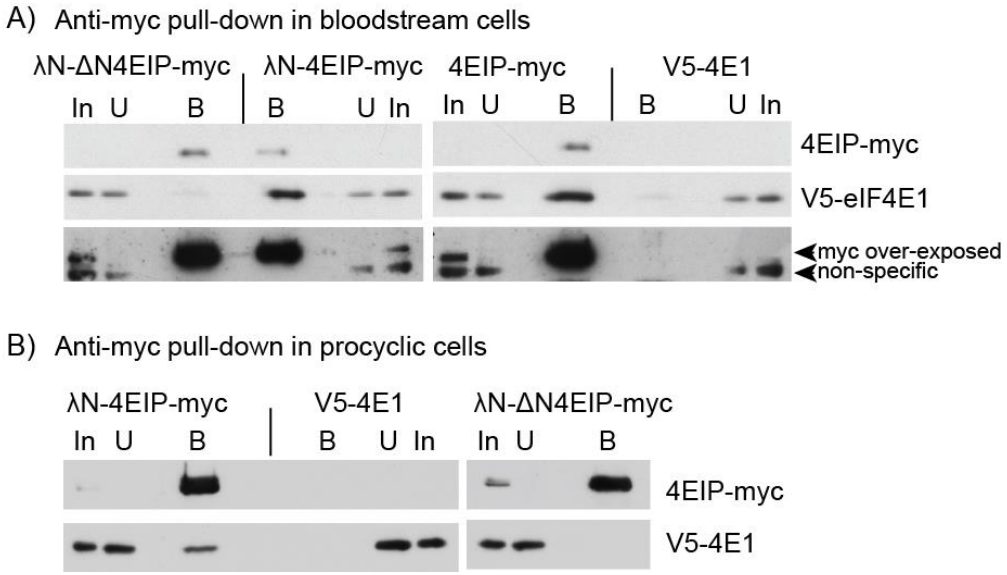


Figure 4: Interaction of truncated *Tb*4EIP with *Tb*EIF4E1 in bloodstream and procyclic forms. (**A**) Western blot showing presence of *Tb*EIF4E1 in *Tb*4EIP precipitates. Cells co-expressing a V5-tagged *Tb*EIF4E1 (V5-4E1) and myc tagged *Tb*4EIP (4EIP-myc) were induced for 24 hrs to express *Tb*4EIP. λN is a short peptide that can bind boxB mRNA sequences, but is not relevant here. ΔN4EIP is the truncated *Tb*4EIP lacking the canonical motif YXXXXLΦ required for binding to *Tb*EIF4E1. A cell line lacking *Tb*4EIP-myc was used as a negative control while those expressing full length *Tb*4EIP were used as positive controls. 'In' is input (~5×10^6 cells), 'U' is unbound (~5×10^6 cells) and 'B' is bound (~1×10^8 cells). The lower panel is an over-exposed version of the upper panel that shows myc expression in the inputs. A non-specific

band appears below 4EIP-myc. (**B**) Western blot showing a similar experiment as (A) but using procyclic cells.

Interactions in yeast were therefore similar to the interactions detected in trypanosomes by the co-IP. The results collectively show that *Tb*4EIP-*Tb*EIF4E1 interaction depends on the canonical YXXXXLΦ motif and that the interaction is not developmentally regulated, at least in bloodstream and procyclic forms.

3.3 *Tb*EIF4E1 requires *Tb*4EIP for gene suppression

Both *Tb*4EIP and *Tb*EIF4E1 are strong repressors of expression when tethered to a reporter gene [78]. *Tb*4EIP lacking the canonical *Tb*EIF4E1 binding motif and which does not interact with *Tb*EIF4E1 remained repressive, implying that repression by *Tb*4EIP does not require binding of *Tb*EIF4E1, though it may be required *in vivo* to enhance binding or stabilize *Tb*4EIP. The C-terminal end of *Tb*4EIP was also found to be sufficient for repression[152]. To test the reverse hypothesis i.e. whether repression by *Tb*EIF4E1 requires *Tb*4EIP, I tethered *Tb*EIF4E1 to *CAT* (chloramphenicol acetyl transferase) mRNA in cells lacking *Tb*4EIP (see knockout experiment in section 3.4). Interestingly, *Tb*EIF4E1 does not repress the CAT reporter expression both at mRNA and at protein level (Figure 5).

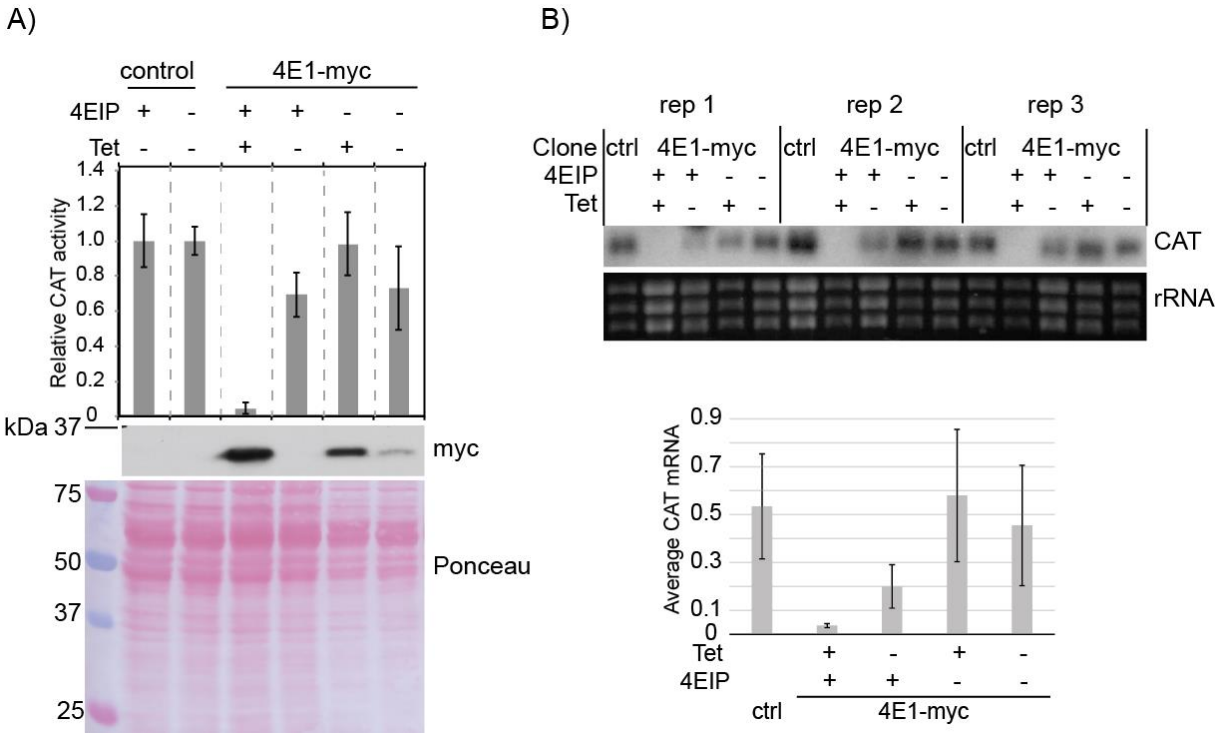


Figure 5: Gene repression by *Tb*EIF4E1 requires *Tb*4EIP. (**A**) CAT (chloramphenicol acetyltransferase) activity in the presence/absence of *Tb*4EIP and Western blot showing expression of *Tb*EIF4E1-myc. *Tb*EIF4E1-myc tagged with a λN peptide on the N-terminal end (pHD2953) was transfected into cells lacking 4EIP and constitutively expressing a CAT gene. *Tb*EIF4E1 was tethered to the 3'UTR of the CAT gene which contains a boxB sequence capable of binding λN peptide of *Tb*EIF4E1-myc, by inducing with tetracycline for 24 hrs. Cell lines lacking *Tb*EIF4E1 with the λN peptide served as controls. Cell lysates were used for measuring CAT activity, for Western blotting and for northern blotting. (**B**) A northern blot and the quantification of the CAT mRNA relative to the rRNA.

Lysates from the tethering cell line lacking *Tb*4EIP had similar CAT activity to those from control cell lines despite leaky expression of 4E1-myc in the absence of tetracycline. A similar repressive trend was observed at the mRNA level; *Tb*EIF4E1 was only repressive in the presence of *Tb*4EIP. Therefore tethered *Tb*EIF4E1 requires *Tb*4EIP to repress expression and it is not a repressor by itself. A similar scenario has been reported in *Drosophila* Cup [135], and mammalian 4E-T [140] and GRB10-interacting GYF GIGFY2 [141,142]. Mammalian eIF4E homologous protein 4EHP is repressive only when interacting with GIGYF2. It is therefore thought that 4EHP prevents translation by competing with eIF4E for the mRNA cap and repress expression by binding to the repressor GIGYF2 [142]. This could be a possible mechanism also for trypanosome *Tb*4EIP. The results therefore demonstrate that the repression activity of *Tb*4EIP-*Tb*EIF4E1 complex is inherent to *Tb*4EIP.

3.4 *Tb*4EIP binds Terminal uridylyl transferase 3

Repression activity of *Tb*4EIP could be mediated by recruitment of the deadenylation complex and/or the degradation machinery. To determine the proteins that are recruited by 4EIP, immunoprecipitation assays were done by my colleagues using a myc-tagged version of 4EIP followed by mass spectrometry of the precipitate. In addition to *Tb*EIF4E1, TUT3 (terminal uridylyl transferase 3) and MKT1-like were co-purified.

First, I sought to validate *Tb*4EIP-TUT3 and *Tb*4EIP-MKT1-like interactions using a co-immunoprecipitation assay (Figure 6). TUT3 was myc-tagged on the C-terminus (pHD2736) and MKT1-like V5-tagged on the N-terminus (pHD2730). By pulling down *Tb*4EIP, it was possible to detect TUT3 in the precipitate (Figure 6A) and likewise, pulling down TUT3 co-purified *Tb*4EIP (Figure 6B). Further, *Tb*4EIP-TUT3 interaction was confirmed in the presence of RNase A to abolish possible RNA dependent interactions (Figure 6A, right panel). However *Tb*4EIP did not co-purify with MKT1-like (Figure 6C). The results indicated that *Tb*4EIP binds TUT3 *in vivo* in an RNA independent fashion but does not bind MKT1-like. Sections 3.8 – 3.10 show preliminary studies conducted on TUT3.

Subsequent tandem affinity purifications of *Tb*4EIP done by my colleague Monica Terrao could not detect TUT3 or MKT1-like on mass spectrometry. This was not surprising considering the tiny amount of protein co-purified using either *Tb*4EIP or TUT3. Such weak interactions may not survive a tandem purification.

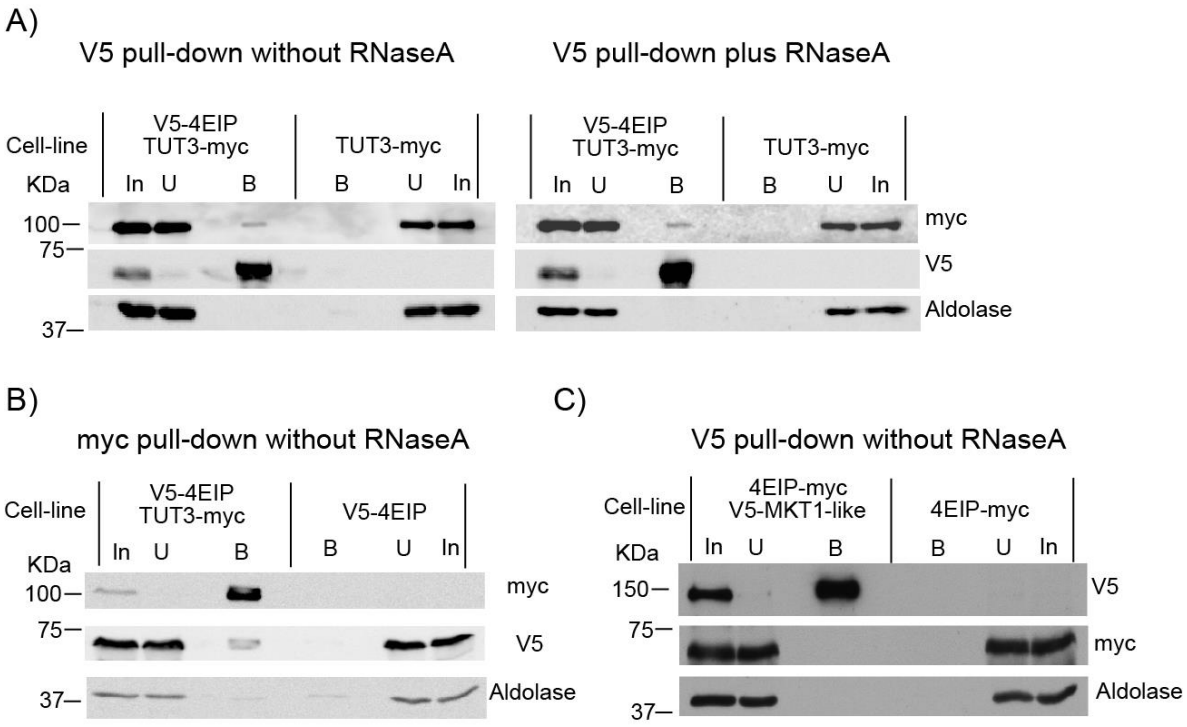


Figure 6: *Tb*4EIP interacts with TUT3 but not with MKT1-like. Western blots showing presence of *Tb*4EIP, MKT1-like and TUT3 in respective pull-downs. Cells expressing V5/myc tagged versions of *Tb*4EIP, TUT3 or MKT1-like were subjected to a co-immunoprecipitation assay. Pull-downs were done using anti-V5 or anti-myc agarose beads using cells lacking V5 or myc accordingly as controls. Also pull-downs were done in the presence or absence of RNase A to control for RNA dependent interactions. 'In' is input (~5×10^6 cells), 'U' is unbound (~5×10^6 cells) and 'B' is bound (~1×10^8 cells). (**A**) shows test for *Tb*4EIP-TUT3 interaction using anti-V5 agarose beads in the presence and absence of RNase A. (**B**) is the anti-myc pull-down similar to (A). (**C**) shows test for *Tb*4EIP-MKT1-like interaction using anti-V5 agarose beads in the absence of RNase A. Aldolase was used as a loading control.

3.5 *Tb*4EIP is not essential for survival of bloodstream and procyclic forms

By interacting with a translation initiation factor, it is conceivable that *Tb*4EIP plays a role in translation regulation and therefore its role is presumably indispensable. Previously, depletion of *Tb4EIP* by RNAi was attempted but some protein expression persisted with a marginal growth defect [152]. A complete deletion was therefore attempted in both bloodstream (done by a student on rotation, Johanna Braun) and procyclic forms by sequential replacement of both alleles of *Tb4EIP* by homologous recombination with drug resistance cassettes (pHD2797 and pHD2898). Genomic DNA was extracted from clones growing on blasticidin and puromycin and a polymerase chain reaction (PCR) done to detect *Tb4EIP* DNA (Figure 7).

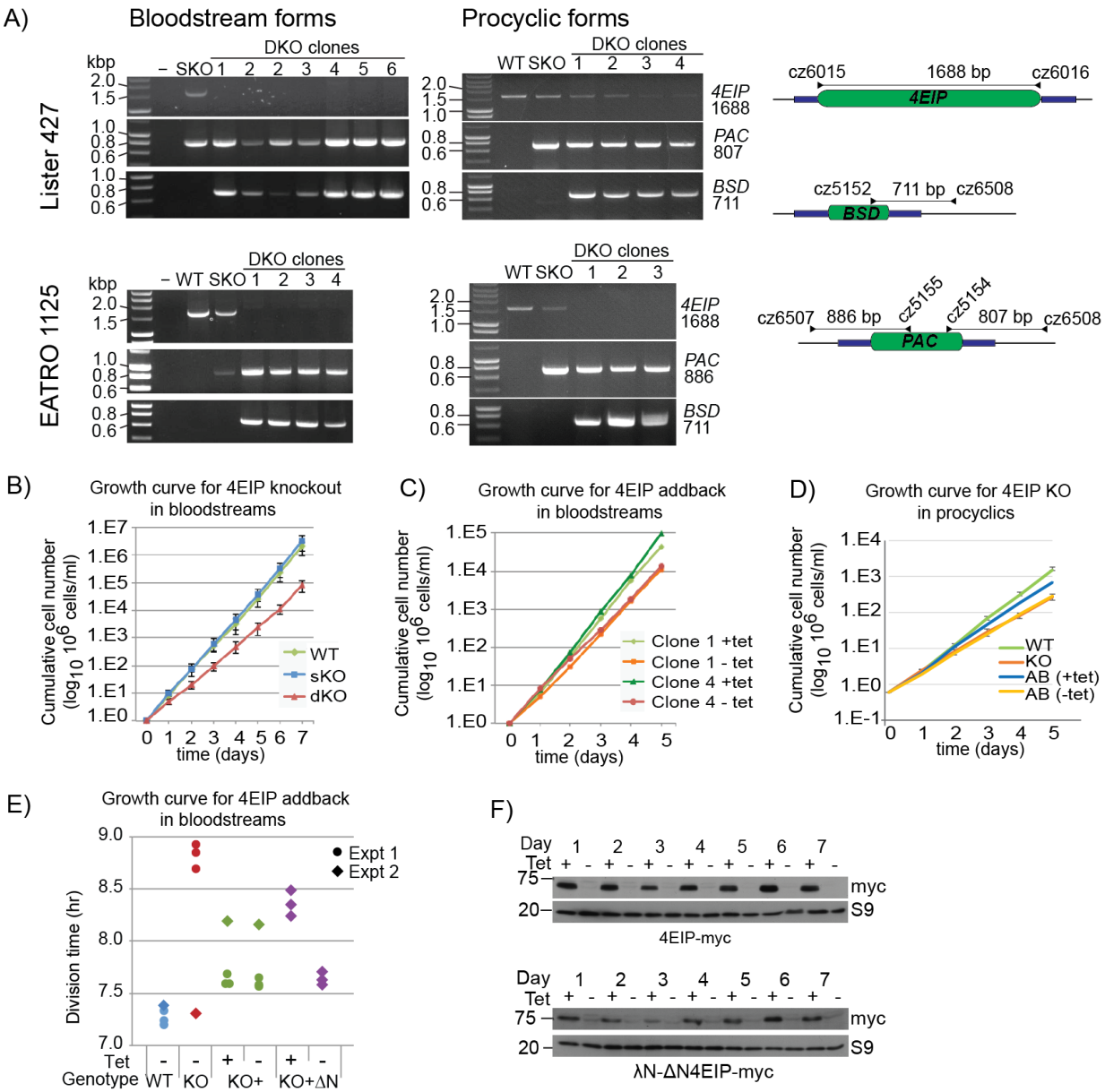


Figure 7: 4EIP is not essential for survival of *T. brucei*. (**A**) PCR to detect *4EIP* and drug resistance genes. DNA was extracted from candidate double knockout (DKO) clones, wild-type (WT) or single knockout (SKO) clones and used as template in a PCR assay targeting *4EIP*, puromycin N-acetyltransferase (*PAC*) and blasticidin-S deaminase (*BSD*). Primers are prefixed with "cz" and the region they target is indicated by an arrowhead. Open reading frames of genes are shaded green while fragments of untranslated regions used for cloning are shaded blue. (**B**) Growth curves of DKO clones in comparison with WT and SKO. Error bars represent standard deviation of three independent experiments. This was done by Johanna Braun. (**C**) Growth curves of DKO clones inducibly expressing *Tb*4EIP from an ectopic locus (4EIP add-back, AB). (**D**) Growth curves of WT, double knockout (KO) and AB procyclic cells. Error bars represent standard deviation. The add-back represents one experiment. (**E**) Doubling times of similar cell lines in B and C as well as an AB expressing a truncated *Tb*4EIP lacking the N-terminal canonical motif for binding *Tb*EIF4E1. Two

experiments were done approximately 8 weeks apart. Graph was drawn by Christine Clayton. (**F**) Western blots showing inducible expression of 4EIP-myc and ΔN4EIP-myc in the DKO.

The double knockout (DKO) clones tested had no detectable *Tb4EIP* DNA and instead had both puromycin N-acetyltransferase (*PAC*) and blasticidin-S deaminase (*BSD*) in the *Tb4EIP* locus (Figure 7A). Lister 427 procyclic forms however had faint bands of *Tb4EIP* despite having both BSD and PAC correctly integrated. This could possibly be explained by the presence of a small population that is growing on both selection drugs but are not DKO. The Lister 427 cells were not used for any further analyses. Knockout was nevertheless successful in the pleomorphic EATRO1125 cells indicating that *Tb*4EIP is not required for survival of bloodstream and procyclic forms.

To determine whether the DKO cells had a growth defect, cells densities of EATRO1125 bloodstream cells were monitored daily in media without selecting drugs (Figure 7B). *Tb4EIP* knockout mutants had a slight but reproducible growth defect (Figure 7B) which could be rescued by re-expression of *Tb*4EIP from an ectopic locus (Figure 7C and D). While generating the add-back clones (with full-length and truncated *Tb*4EIP) which took approximately 8 weeks, the DKO clones which were meanwhile maintained in culture had strangely recovered from their growth defect (Figure 7E, Experiment 2). Expression of the ectopic copy of *Tb4EIP* also caused an increase in division times compared to the non-induced counterparts. This could be attributed to an adaptation mechanism to life without *Tb*4EIP and over-expression would therefore cause a growth defect associated with reverting to life with *Tb*4EIP. It is also conceivable that excess of *Tb*4EIP could cause non-specific targeting hence affect cell growth. Collectively, the results indicate that *Tb*4EIP is not required for survival of bloodstream and procyclic forms of the parasite but is probably required for normal growth in both bloodstream and procyclic forms.

3.6 4EIP is required for translation repression during stumpy formation

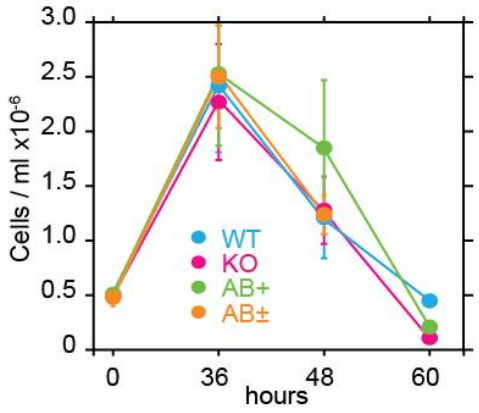
Bloodstream forms divide logarithmically as slender forms and stop when high density is attained; they become arrested at G1/G0 phase and transform to stumpy forms which is the transmissible stage pre-adapted for differentiation. In the midgut of the insect the stumpy forms differentiate further to procyclic forms. Differentiation to stumpy forms is marked by global suppression of gene expression with synthesis of select proteins including mitochondrial enzymes and proteins associated with differentiation (PAD) [23]. Since tethered *Tb*4EIP represses expression of a reporter (see Figure 5), I tested whether *Tb*4EIP plays a

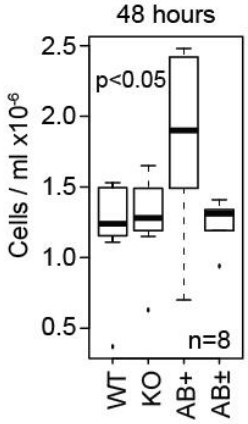
role in translation repression during stumpy formation. Cells were grown to high densities in methyl-cellulose-containing media and monitored for incorporation of methionine at different time-points during stumpy formation.

At maximum density ~2.5×10^6 cells/ml which corresponds to the onset of stumpy formation, the WT cells had reduced incorporation of methionine, which is the normal suppression of translation during stumpy formation. Interestingly, the *Tb4EIP* KO cells continued to incorporate significantly more methionine even at 48 hrs when their density corresponded to that of stumpy forms. Further, they did not express the stumpy marker PAD1. *Tb4EIP* add-back (AB) reverted the increased methionine incorporation showing that the defective incorporation was caused by absence of *Tb*4EIP. The effect of the un-induced AB cells did not exactly reproduce the effect of KO cells probably because of leaky expression, and were therefore referred to as AB± (Figure 8B).

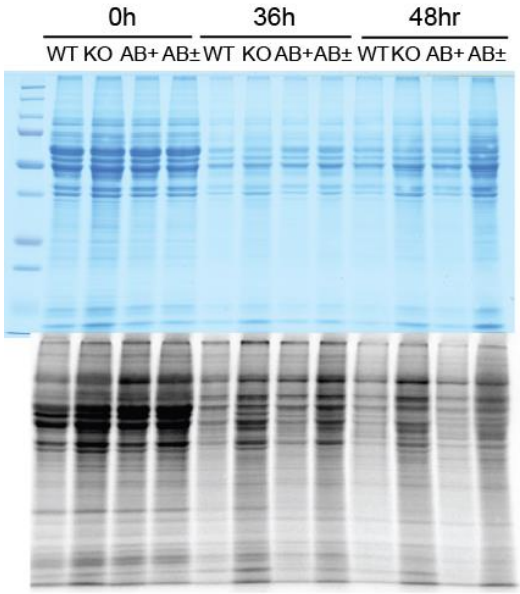
In addition, the densities of cells declined at different rates. *Tb4EIP* KO cell densities declined similar to WT cells but the complemented KO (add-back, AB) cells significantly remained at higher densities for longer (Figure 8A). This suggests that excess *Tb*4EIP promotes cell survival at high densities.

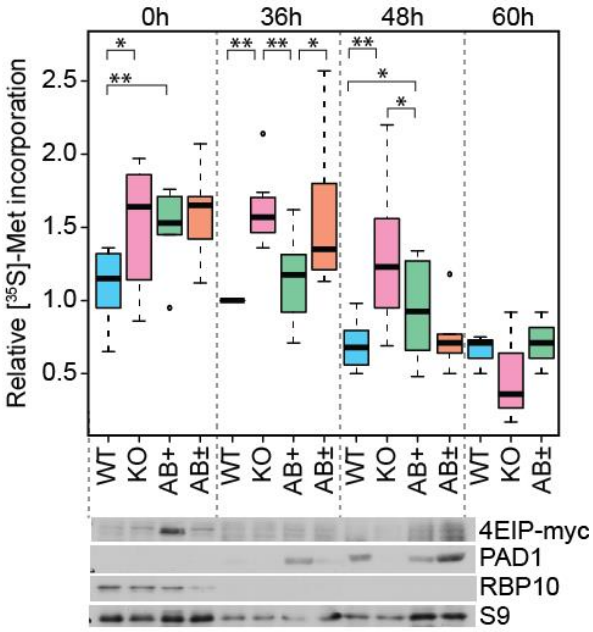
A) Cell densities





B) Methionine - ^{35}S incorporation





C) mRNA content

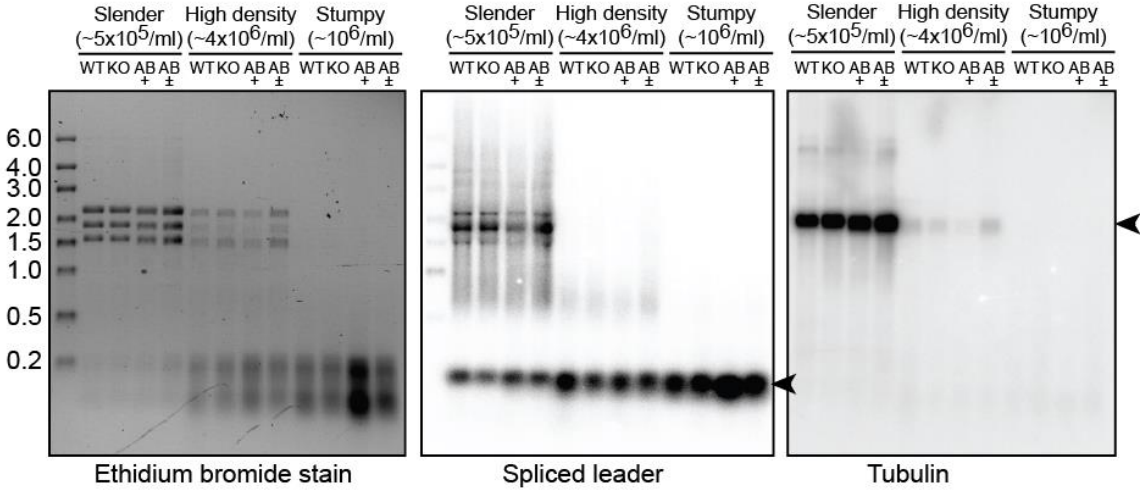


Figure 8: *Tb*4EIP is required for translation repression during stumpy formation. (**A**) Densities of cells during stumpy formation. The 0hr time point represent five independent experiments, the 36 and 48 hr time points represent eight independent experiments and the 60 hr time point represent three experiments. Means and standard deviations are plotted. The box plot on the right depicts cell densities at the 48 hr time point. Boxes represent quartiles, the middle lines are medians and the whiskers are minimum and maximum values. (**B**) A representative coomassie stained protein gel and its autoradiograph of cell lysates from (A). The graph on the right represents relative quantities of incorporated methionine. Cells were collected at each time point shown in (A) and pulsed with methionine [^{35}S] for 20 min following a one hour methionine starvation. Cell lysates were analysed by SDS-PAGE and autoradiography. ** represent student test p-value P<0.01 and * represent P<0.05. The graph was made by Christine Clayton. The expression of AB 4EIP-myc, PAD1 and RBP10 are shown in the Western blot below the graph. The ribosomal protein S9 and Ponceau stain served as loading (**C**) Northern blot showing quantity of mRNA during stumpy formation. Cells were collected at each time point shown in (A) for RNA extraction. Equal amount of RNA were used per sample and analysed by Northern blotting, probed for spliced leader (SL) and tubulin that are shown by arrow-heads.

The inability to repress translation could be as a result of higher mRNA content in the KO cells. To test this hypothesis, total RNA was obtained from cells during stumpy formation. All trypanosome mRNAs have a SL (spliced leader) at the 5' end [3]. Messenger RNA content can therefore be estimated using a probe against the SL on a northern blot. Total RNA was extracted from cells at different densities during stumpy formation and analysed on a northern blot. However, sufficient intact RNA could not be obtained especially from cells after high density (Figure 8C). This could probably be caused by high concentration of RNases from dying cells in the decline phase during stumpy formation. Nevertheless, there was no detectable difference in the SL signal between samples at 0 hr and at 36 hr. The signal of tubulin mRNA which is abundant in cells, was also minimal or completely undetectable in the dense cells. The SL RNA (Figure 8C) was however detected in increasing amounts as the cells differentiated to stumpy. Interestingly, there was an accumulation of small RNAs demonstrated by the ethidium bromide stain, but the reason for this is unknown. Technical challenges could not be ruled out because of lengthy sample collection from the semi-solid media (see materials and methods, 2.1). The northern blot was therefore not interpretable at the moment.

3.7 *Tb*4EIP is required for differentiation to procyclic forms

Tb4EIP KO cells do not express the stumpy marker PAD1 when put to differentiate. This is despite them growing to high densities, which is a trigger for stumpy formation - a mechanism that involves stumpy induction factor (SIF) [22,165]. In addition, *Tb4EIP* KO cells are unable to

repress translation at high densities. These cells could therefore be defective in stumpy formation and likely unable to differentiate to procyclic forms. To test this, the cells were treated with 10µM of the cAMP analog, 8-pCPT-2'-O-Me-cAMP, for 36 hrs or grown to high density in semi-solid media (see section 3.6) to make stumpy forms, and then treated with 6mM cis-aconitate (CA) which induces differentiation of stumpy forms to the insect stage procyclic forms. EP procyclin is the surface protein in procyclic cells and is therefore a convenient marker for monitoring differentiation to procyclic forms. WT cells expressed EP already at 6 hrs after inducing differentiation whereas in KO cells, only small amounts of EP were detectable at 24 hrs (Figure 9A).

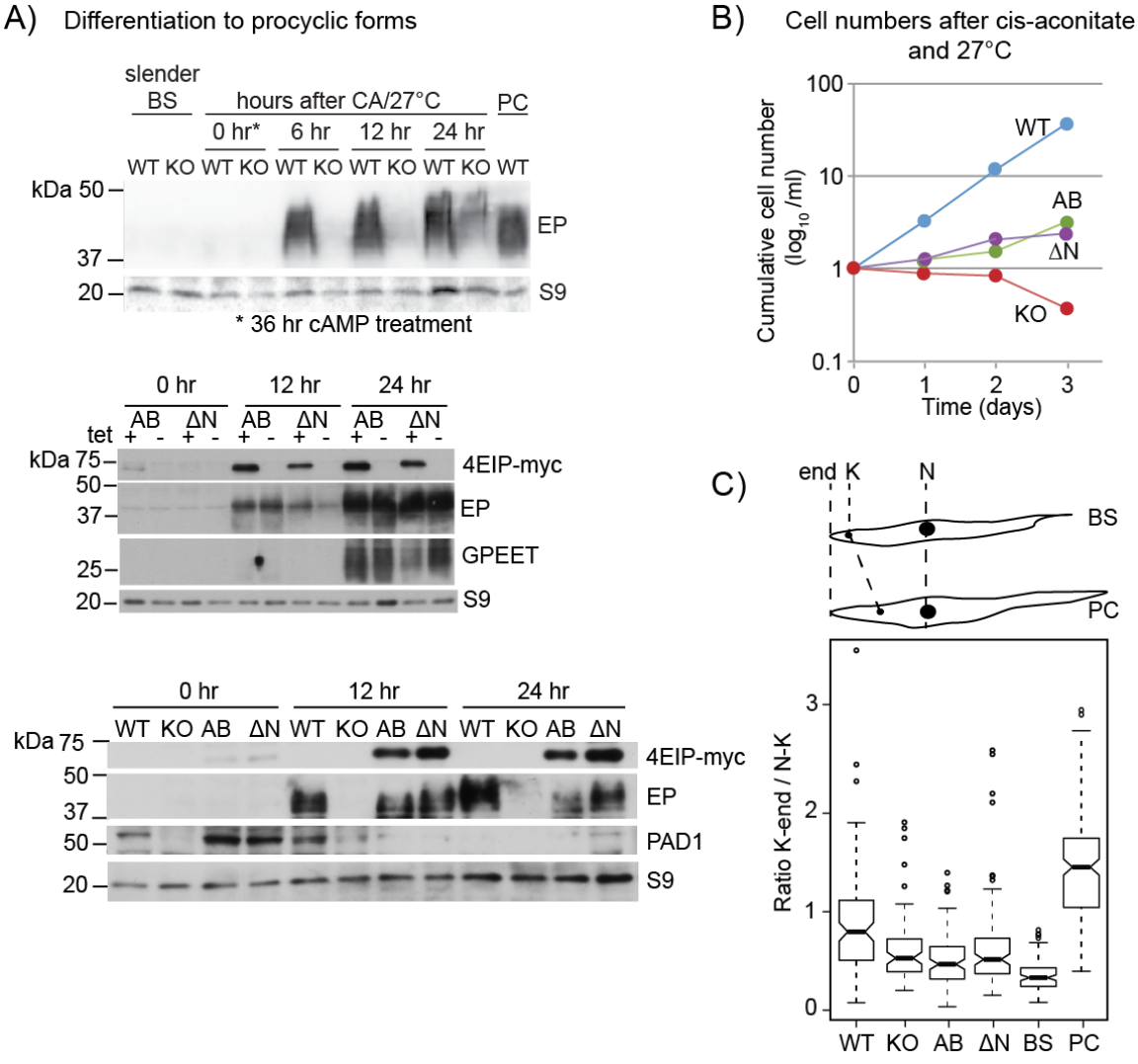


Figure 9: Differentiation of stumpy to procyclic forms. (**A**) Western blots showing presence of EP during differentiation. Slender cells were grown to high density or treated with 10µM cAMP analog (8-pCPT-2'-O-Me-cAMP) for 36 hrs (indicated by *) to make stumpy forms (0 hr) and treated with 6mM cis aconitate for 24 hr to differentiate to procyclic forms. Cells were collected

at various time points during differentiation and analysed by Western blotting. Cells used are WT; wild type, KO; 4EIP knockout, AB; KO with inducible expression of 4EIP-myc, ΔN; KO with inducible expression of ΔN4EIP-myc. Procyclins EP/GPEET were used to monitor differentiation, PAD1 indicate stumpy forms while S9 indicate loading. (**B**) Growth curve of cells from (A) 24 hrs into differentiation to procyclic cells. (**C**) Cells at 24 hrs after cis-aconitate were stained with DAPI and the kinetoplast repositioning quantified in comparison to bloodstream (BS) and procyclic (PC) forms. The cell metrics and box-plots were done by Christine Clayton because the slides needed to be read blind. The first Western blot was done by a rotation student, Johanna Braun.

This defect could be rescued by ectopic expression of *Tb4EIP*. In addition a truncated *Tb*4EIP (ΔN4EIP) that does not bind *Tb*EIF4E1 could similarly rescue the defect in *Tb4EIP* KO indicating that *Tb*4EIP-*Tb*EIF4E1 interaction is not essential for normal differentiation (Figure 9A, middle and lower panel).

Further, cell numbers were monitored 24 hr after differentiation (Figure 9B). WT cells grew exponentially as procyclic forms while the KO complement cells (AB and ΔN) grew much slower. The KO cells however did not grow and eventually died between 2-3 days after differentiation. This is consistent with the KO cells inability to form viable stumpy forms and as a result do not differentiate further to procyclic forms, thus highlighting the need for *Tb*4EIP in stumpy formation.

During differentiation to procyclic forms, the kinetoplast DNA migrates to a position such that the distance between the nucleus and kinetoplast is approximately equal to the one from the posterior edge to the kinetoplast, a phenomenon called kinetoplast repositioning [166]. Bloodstream cells therefore have a k-end/K-N ratio less than one while procyclic cells have ratios higher than one. The K-end/K-N ratio for all the newly differentiated cells (3 days old) were higher than that of bloodstream cells indicating successful repositioning of the kinetoplast and a sign of procyclic life (Figure 9C). The ratios were however lower than in late procyclic forms a phenomenon already seen elsewhere in early procyclic forms [45]. There was however no difference between the KO and AB/ΔN cells, though the WT cells had a slightly higher ratio.

Together, the results indicate that *Tb*4EIP is required for differentiation at least to repress gene expression during stumpy formation. The interaction with *Tb*EIF4E1 is not required for differentiation and possibly for translation repression since *Tb*ΔN4EIP does not interact with *Tb*EIF4E1 and is still capable of repressing expression of a reporter and of rescuing differentiation defect of *Tb4EIP* KO cells. *Tb*4EIP is therefore required for normal differentiation to stumpy and procyclic forms.

3.8 TUT3 is not essential for survival of bloodstream trypanosomes

Terminal uridylyl transferases (TUT) add uridine residues at the end of mRNA. In mammals TUT are implicated in mRNA degradation. TUT4 and TUT7 uridylate mRNA with short poly(A) tail and the LSM1-7 complex binds the short U-tail and facilitate decapping by DCP1/2 complex, or the short U-tail is recognised by DIS3L2 or the exosome that degrades the mRNA in the 3'-5' direction or in the 5'-3' by XRN1 [167]. In trypanosomes, uridylation has been detected in GPEET mRNA in early procyclic forms where GPEET is gradually repressed and replaced by EP [168]. This could implicate TUT3 in uridylation of GPEET mRNA for degradation in early procyclic. I have earlier demonstrated that TUT3 and *Tb*4EIP are binding partners and therefore are likely to participate in similar pathways. According to a poly (A)-mRNA-proteome screen, TUT3 does not associate with mRNA [78]. It is neither a repressor nor an activator of gene expression [84] suggesting its need for *Tb*4EIP form mRNA binding.

To obtain a first impression of its role, I attempted RNAi of *TUT3* in bloodstream and procyclic cells. A single allele of *TUT3* was V5 tagged (plasmid pHD 2859) and a plasmid for RNAi (plasmid pHD 2880) expressed ectopically under the control of tetracycline repressor. RNAi was induced every 24 hrs and cell densities determined. Some V5 tagged protein was still detectable after 5 days of RNAi in both bloodstream and procyclic forms thus depletion was not entirely successful (Figure 10A). Further, the cells had no detectable growth defect in both parasite forms compared to WT cells probably because the remaining protein after RNAi was sufficient for normal growth.

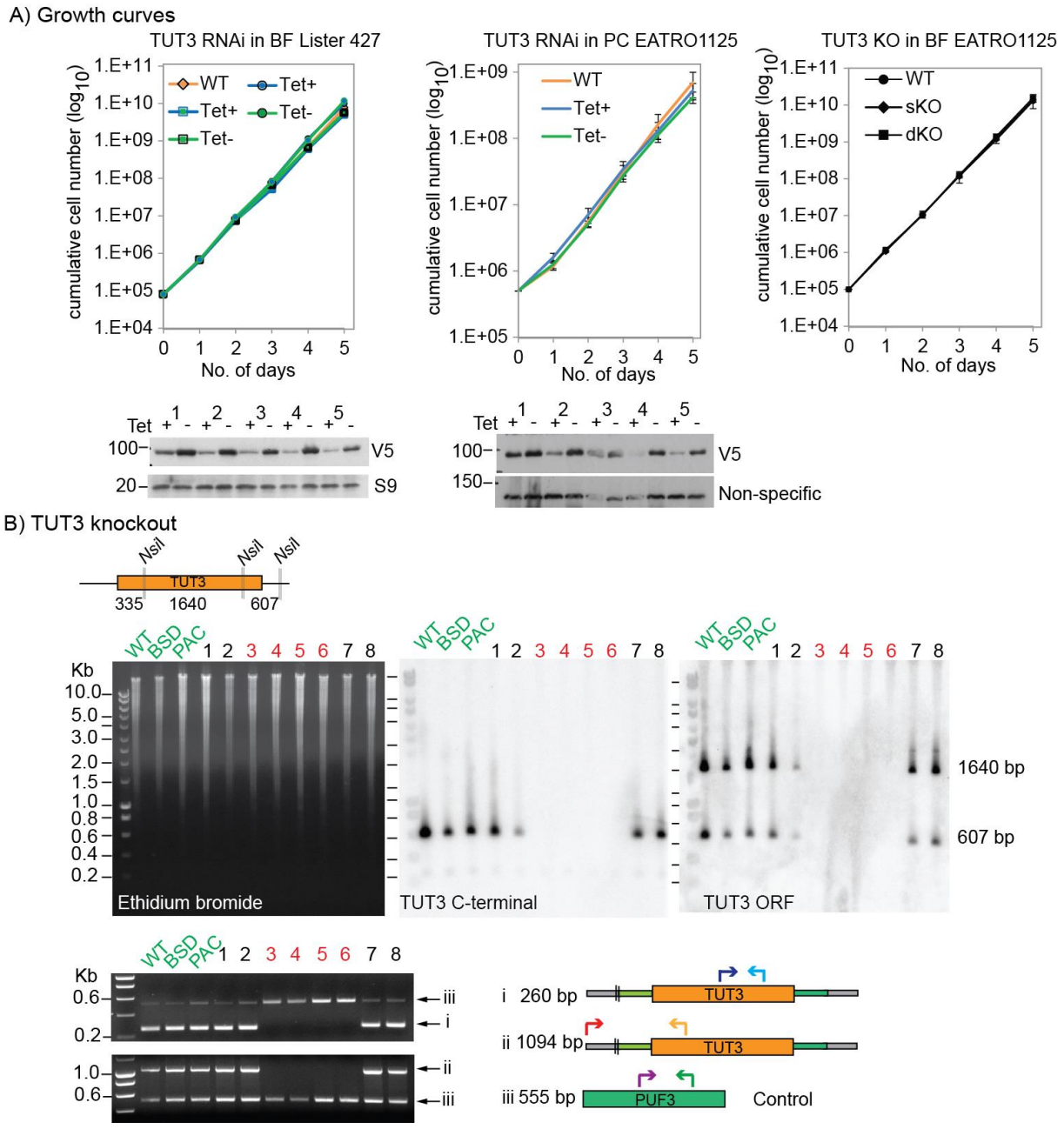


Figure 10: TUT3 is not essential for growth and survival. (**A**) Growth curve of cells depleted or lacking *TUT3* and Western blot showing expression of TUT3. For RNAi, cells were induced with tetracycline every 24 hrs and their cell density determined. Cells were collected at each time-point for Western blot to monitor expression of V5-TUT3. Ribosome subunit S9 or a non-specific band were used as a loading control. Left panel represents growth curve for two monomorphic bloodstream (BF) clones marked with squares and circles respectively while the WT is marked with diamond shape. Middle panel represents average growth curve of three pleomorphic procyclic (PC) clones induced (+) or non-induced (-) for RNAi. Right panel represents growth curve for three *TUT3* knockout clones (dKO) in comparison to WT and single knockout (sKO). (**B**) Southern blotting and PCR to detect presence of *TUT3*. Wild-type (WT), sKO containing blasticidin-S deaminase (BSD), sKO containing puromycin acetyltransferase (PAC) and candidate dKO cells were used. Genomic DNA obtained from

these cells was digested with *NsiI* which cleaves TUT3 gene into three as shown in the upper panel. The digested DNA was separated in an ethidium bromide stained gel and transferred to a nylon membrane for southern blotting. A probe targeting the C-terminal end and another targeting the whole TUT3 coding sequence were used as indicated in the middle panel. The DNA used for southern blotting was used for PCR using the primers prefixed with “cz” as indicated in the cartoon. Primers to amplify *PUF3* were included in the reactions as a loading control. True KO clones are labelled in red.

I therefore attempted a gene knockout by replacing both *TUT3* alleles with genes encoding drug resistance in pleomorphic bloodstream cells (Figure 10B). DNA was extracted from clones that were growing on both drugs corresponding to the two drug resistance genes used for the knockout, and analysed by PCR and southern blotting. The southern blot was probed using a fragment and the whole coding sequence of *TUT3*. Half of the KO clones tested were positive *TUT3* KO cells as they gave no signal in both the PCR and southern blot.

To find out whether the knockout cells had any growth disadvantage, cell densities of three knockout clones were monitored for five days and compared to the WT. The *TUT3* KO cells grew similar to WT cells (Figure 10A, right panel). These results collectively demonstrate that TUT3 is not essential for growth and survival of *T. brucei* as bloodstream or procyclic forms.

3.9 TUT3 is not required for differentiation to procyclic forms

Since *Tb*4EIP is essential for normal differentiation, its binding partner TUT3 is presumably required in differentiation. I therefore attempted to differentiate the cells. *TUT3* knockdown and knockout bloodstream cells were induced for differentiation by treatment with 6mM cis-aconitate and reduction of temperature to 27°C, then monitored for the expression of the surface coat protein, EP procyclin. Depletion of TUT3 in pleomorphic forms was more efficient than in monomorphic strains as observed at the initial hours of differentiation (compare Figure 10A, left panel and Figure 11, middle panel). The depletion efficiency decreased as the cells became procyclic. Growth kinetics during differentiation remained similar to WT cells (Figure 11, left panel). However, contrary to *Tb*4EIP, there was no defect in the expression of EP procyclin during differentiation.

TUT3 is therefore not required for differentiation like its binding partner *Tb*4EIP. Preliminary results from our lab suggest that *Tb*EIF4E1 is also not required during differentiation, hitherto positioning *Tb*4EIP at the centre of the pathway among its binding partners.

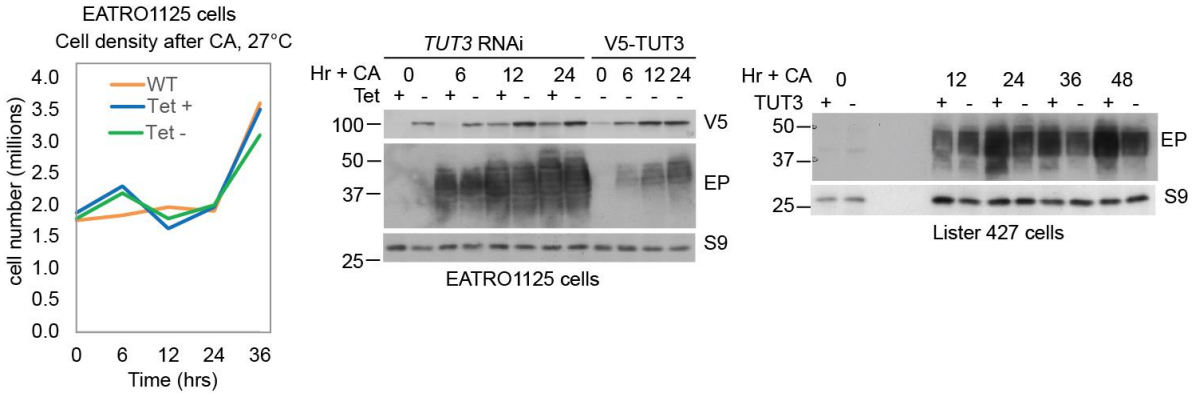


Figure 11: Differentiation of bloodstream to procyclic forms depleted/lacking *TUT3*. Left panel is a growth curve of pleomorphic cells (EATRO 1125) induced (Tet+) or uninduced (Tet-) for *TUT3* depletion by RNAi and treated with 6mM cis-aconitate and 27°C to induce differentiation. In the middle panel, expression of EP was monitored by Western blot during differentiation. Cells lacking the RNAi construct were included as control. S9 represents loading. The right panel represents expression of EP in differentiating monomorphic cells (Lister 427) with (+) or without (-) TUT3 i.e. TUT3 KO. These were put to differentiate as in the middle panel.

3.10 TUT3 is located in the cytosol

Trypanosoma brucei has five terminal uridylyl transferases mostly localized in the mitochondria and involved in RNA editing [169]. Two of these TUT3 and 4 are however thought to be cytosolic [170]. To determine the cell localization of TUT3, cells expressing TUT3 with a C-terminal myc tag (plasmid pHD 2736) were analysed by immunofluorescence microscopy. Tryparedoxin peroxidase (TxNPx) and aldolase were used as cytoplasmic and glycosomal controls respectively. TUT3 signal appeared to co-localise with both aldolase and TxNPx signals (Figure 12A and B).

To make a distinction, cell fractionation using digitonin was done whereby serial concentrations of digitonin were used to rupture cell membranes so that proteins in intact organelles remain in the pellet after centrifugation. The migration of TUT3 from the pellet to the supernatant relative to aldolase was monitored using a Western blot (Figure 12C). TUT3 was detectable in the supernatant at lower digitonin concentrations similar to the cytosolic control TxNPx. Meanwhile, aldolase migrates to the supernatant at much higher concentrations. This implies that TUT3 and aldolase are in different locations in the cell and TUT3 is not in the glycosomes. Collectively the immunofluorescence and the cell fractionation assays demonstrate that TUT3 is cytosolic.

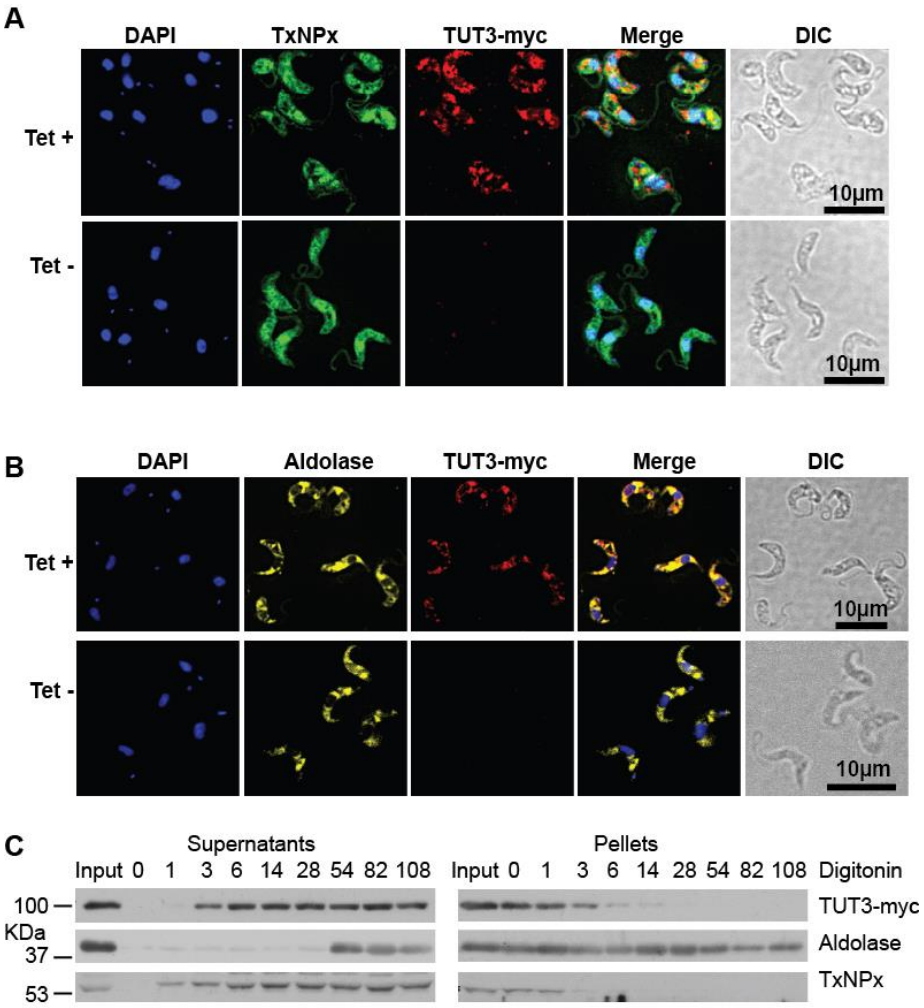


Figure 12: Localisation of TUT3. (**A**) Immunofluorescence images showing staining for TUT3 and other proteins. Bloodstream cells expressing an inducible myc tagged TUT3 from an ectopic loci (pHD2736) were stained with anti-TxNPx (rabbit-1:1000) and anti-myc (mouse-1:500) followed by corresponding secondary antibodies with fluorophore (anti-rabbit Alexa488-1:700, anti-mouse Cy3-1:700). (**B**) Cells were treated as in (A) but using anti-aldolase (rabbit-1:1000). DIC is the differential interference contrast. DAPI was used for staining DNA. The colours used are not necessarily the true colour of the fluorophore. (**C**) Western blots showing migration of proteins from pellets to supernatants. Cells similar to those in (A) were lysed using increasing concentrations of digitonin and organelle rupture was detected by the migration of the organelle's proteins to the supernatant. Aldolase and tryparedoxin peroxidase were used as glycosome and cytosolic markers respectively. The amounts of digitonin (µg of digitonin/ mg of protein) are indicated on each lane.

In summary, *Tb*4EIP and TUT3 interact, are cytosolic and are not essential for survival of growing bloodstream or procyclic forms of both monomorphic and pleomorphic cells. Nonetheless, cells lacking either of the two show different phenotypes during differentiation with *Tb*4EIP being essential during differentiation. TUT3 could play a peripheral function upstream in the pathway of *Tb*4EIP.

PART 2 – Pumilio domain protein PUF3

3.11 PUF3 domain architecture and orthology

Pumilio domain protein 3 (PUF3) orthologs are conserved across kinetoplastids and are almost always syntenic (TriTrypDB). To compare the domain architecture of PUF3, orthologs were obtained from trypanosomes (*Trypanosoma brucei*, *T. vivax*, *T. congolense*, *T. theileri* and *T. cruzi*), relative trypanosomatids (*Leishmania major*, *Blechomonas ayalai* and *Endotrypanum monterogeii*) and from a free living kinetoplastid *Bodo saltans*.

Orthologs were virtually identical in the region with the PUF repeats (Figure 13A). The N-terminal and the distal C-terminal ends were highly variable and this could imply different interactions and functions. A striking difference was observed between the free-living *Bodo saltans* and the parasitic trypanosomatids: the parasites have three major deletions flanking the PUF repeats region (Figure 13), a feature that could characterize loss or change of gene function and evolution to a dependent lifestyle similar to genome erosion in endosymbionts [171]. *T. congolense* has a unique deletion in one of its PUF repeats and has only seven left.

A cladogram using PUF3 protein sequences clusters the parasitic trypanosomatids into three (Figure 13B). The first clade consists of *Bl. ayalai*, *L. major* and *E. monterogeii*, which shares a deletion from amino acid 277-295 (Figure 13A). The second consists of *T. cruzi* and *T. theileri* while the third consists of the African trypanosomes - *T. vivax*, *T. brucei* and *T. congolense*.

To get an overview of how *T. brucei* PUF proteins compare with respect to possible redundancy, their domain structures were predicted using SMART [156]. The 11 PUF proteins of *T. brucei* have different domain architectures (Figure 13C). With the exception of PUF3, all the others have in addition, low complexity and coiled coil motifs. Apart from PUF4 and 6 which look similar, the others have unique domain arrangements. Further, only PUF2, 3, 4, and 6 have the canonical 8 pumilio repeats. These dissimilarities suggest differences in function and a remote chance of functional redundancy. PUF4 and 6 could however share functions judging by their domain similarities.

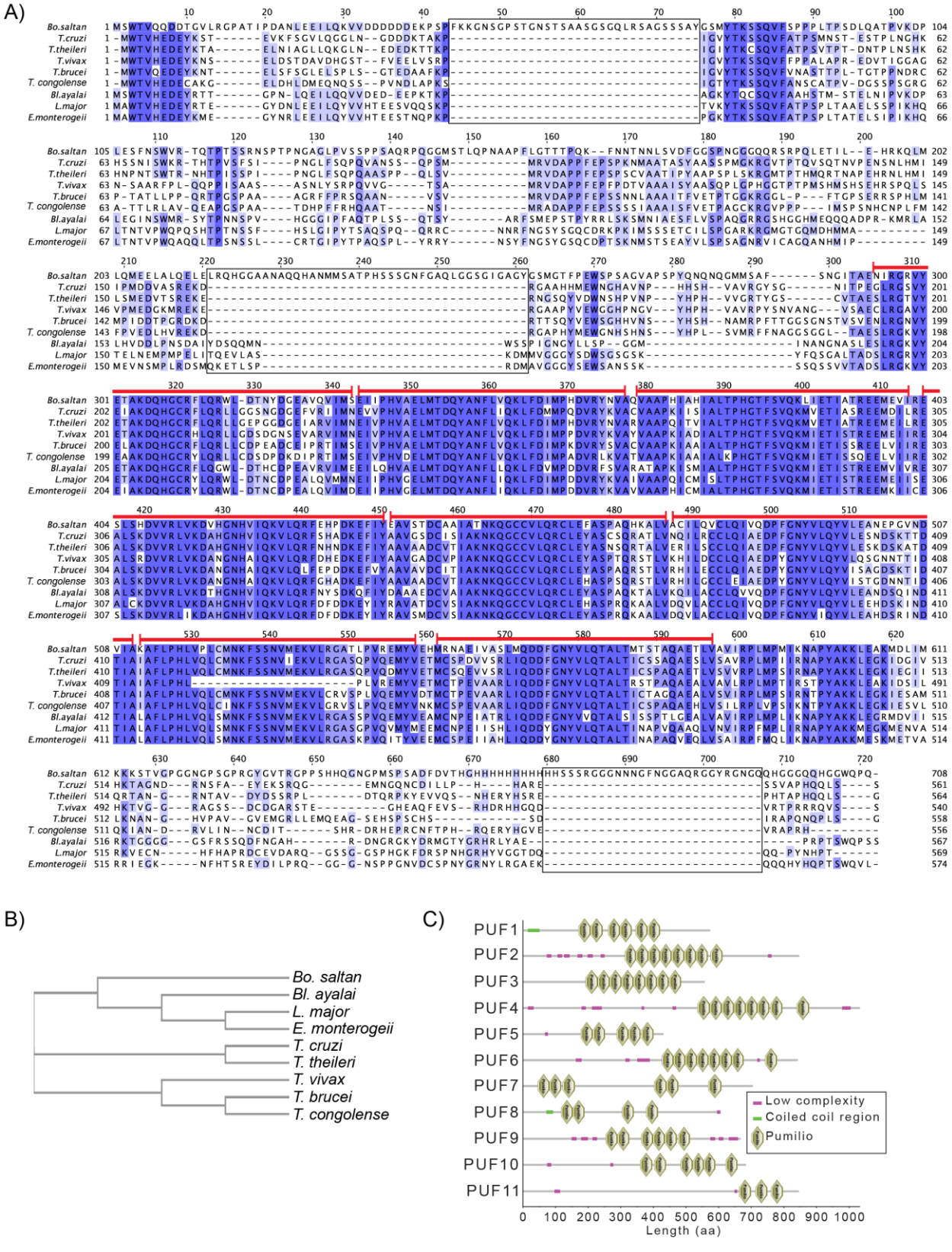


Figure 13: Domain architecture and orthology of select kinetoplastid PUF3s. (**A**) Protein sequence alignment of PUF3. PUF3 protein sequences from *Trypanosoma brucei*, *T. vivax*, *T.*

congolense, *T. theileri, T. cruzi*, *Leishmania major*, *Blechomonas ayalai, Endotrypanum monterogeii* and from a free living kinetoplastid *Bodo saltans* were aligned using Clustal Omega (EMBL-EBI) with default settings. The shading highlights similarities. Red bars indicate the position of the PUF repeats relative to *T. brucei*. (**B**) A cladogram based on the same sequences in (A), drawn using Simple Phylogeny Tool (EMBL-EBI). (**C**) Domain organization of the 11 *T. brucei* PUF proteins. The domain organization was predicted using SMART tool. The horizontal lines represent the protein lengths drawn to scale.

These observations suggest that PUF3 is unique in domain architecture from other PUF proteins in *T. brucei* and is likely to have a unique function. In addition, PUF3 appears to be highly conserved among the African trypanosomes and is likely to have a conserved function.

3.12 PUF3 is localized in the cytosol

Many proteins regulating gene expression at post-transcriptional level are localized in the cytosol. Most of the PUF proteins are cytosolic; exceptions are PUF7 and PUF10, which localise in the nucleolus and regulate GPEET mRNA and rRNA maturation [44,102]. To determine the cellular location of PUF3, bloodstream cells with tagged versions of PUF3 (V5-PUF3: pHD2127 and PUF3-myc: pHD2904) were used in immunofluorescence microscopy and cell fractionation assay (Figure 14).

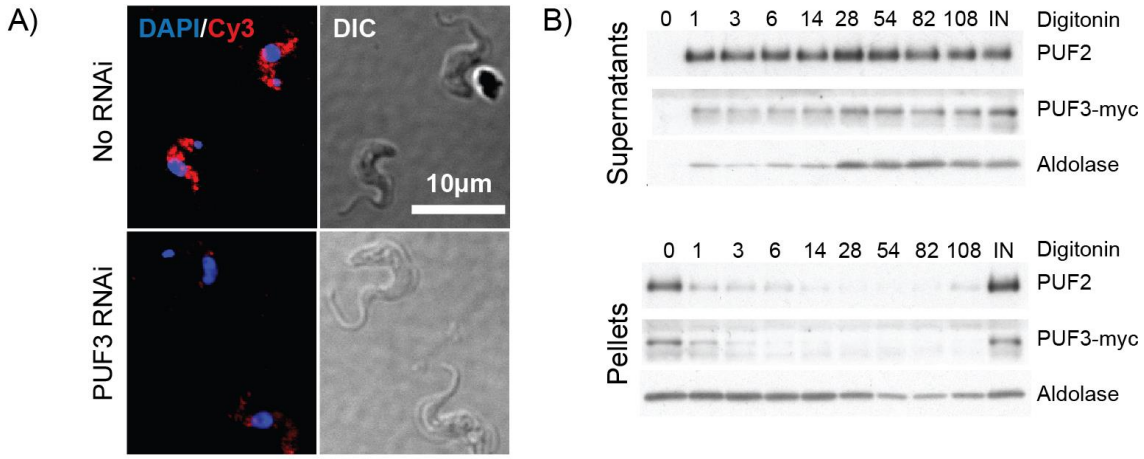


Figure 14: PUF3 is localized in the cytoplasm. (**A**) Cell images showing the location of PUF3. One allele of PUF3 was V5-tagged on the N-terminus (pHD2127) and an RNAi construct (pHD2881) expressed ectopically. Cells were induced for RNAi for 24 hrs and processed for immunofluorescence microscopy. Anti-V5 antibody (1:500, mouse) and secondary antibody with Cy3 fluorophore (1:700, anti-mouse) were used. DNA was stained with DAPI. DIC is the differential interference contrast. (**B**) Cells expressing an inducible copy of PUF3 with a C-terminal myc tag were induced for 24 hrs with tetracycline and lysed using increasing concentrations of digitonin. Organelle rupture was detected by the migration of the organellar proteins to the supernatant. PUF2 and aldolase were used as cytosolic and glycosomal

markers respectively. The amounts of digitonin (µg of digitonin/ mg of protein) are indicated on each lane.

PUF3 appeared to be spread throughout the cytosol with no signal from the nucleus. A cell fractionation assay using digitonin was used to test whether PUF3 could be in the glycosomes. Migration of PUF3 from the pellet to the supernatant relative to aldolase and the cytosolic PUF2 was monitored during increased concentrations of digitonin. PUF3 migrates to the supernatant already at 1µg of digitonin/ mg protein similar to PUF2 whereas aldolase persists in the pellet and migrates at 28µg of digitonin/ mg protein treatment. Combined, these results indicate that PUF3 is cytosolic and is not detectable in the nucleus or glycosome.

3.13 PUF3 is not essential for survival of trypanosomes

To determine whether PUF3 is essential for cell growth, *PUF3* was depleted by RNA interference (RNAi) in bloodstream forms. (This initial work on PUF3 was done in the monomorphic Lister 427 strain.) Previous transcriptome and ribosome profiling data [172] suggested that PUF3 is more expressed in bloodstream forms than in procyclic forms while a genome-wide RNAi screen suggested that PUF3 is likely to be essential in bloodstream forms [173]. RNAi was therefore attempted in bloodstream forms.

One of the *PUF3* alleles was tagged so as to express a fusion of V5-PUF3 (pHD 2127, made by Claudia Helbig) to enable detection of PUF3 protein. This cell line was transfected with a plasmid pHD2881 for *PUF3* RNAi. The effectiveness of the knockdown was determined by detecting V5-PUF3 on a Western blot.

RNAi was induced by addition of 100ng/ml of tetracycline to the growing culture while monitoring cell density every 24 hrs. Tetracycline was added daily while splitting the cells as required. In two clones tested, cells depleted of PUF3 grew similar to wild-type (WT) cells but with a marginal defect (Figure 15A) despite sufficient knockdown; no PUF3 could be detected after 24 hr of RNAi induction in both clones (Figure 15B). Though the amount of PUF3 detected is theoretically half the total (i.e. only one allele was tagged), the RNAi targets products from both PUF3 alleles. However, it is possible that small undetectable amounts of PUF3 are sufficient for the cell growth and therefore complete knockout of PUF3 was attempted.

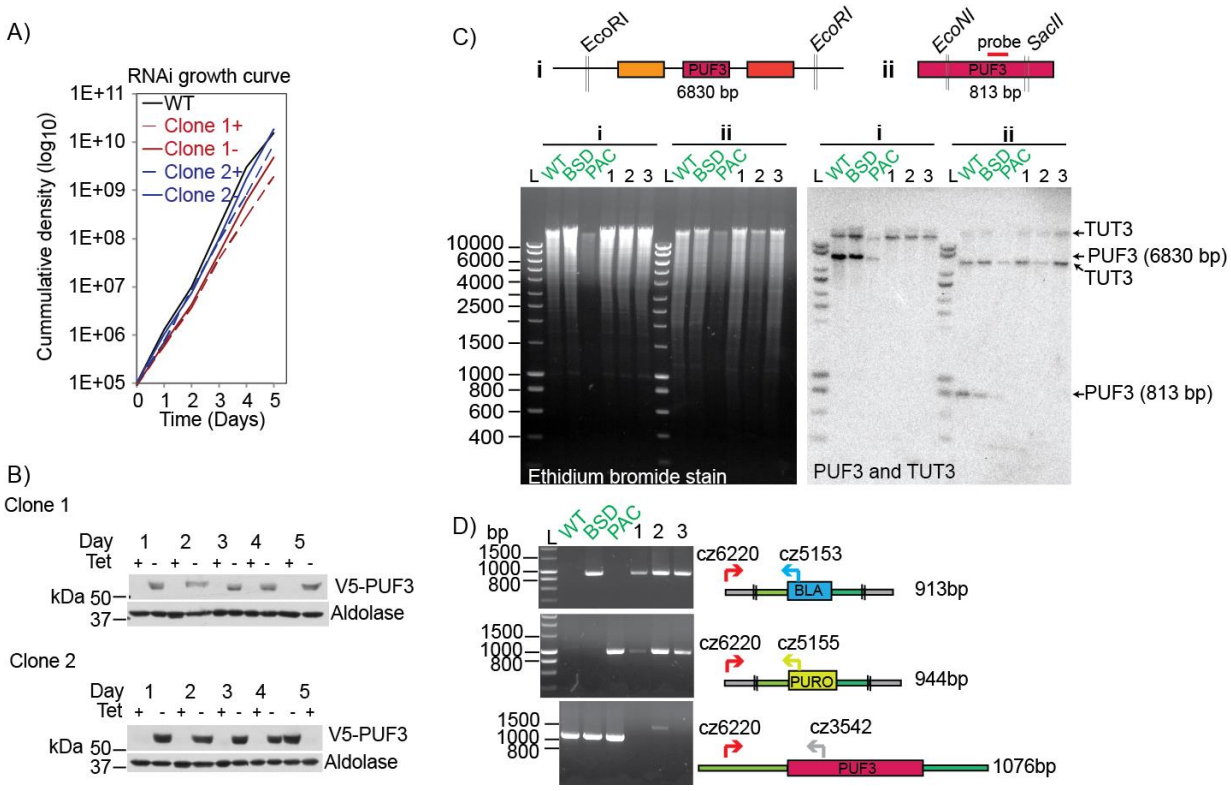


Figure 15: PUF3 is not essential for survival of monomorphic bloodstream forms. (**A**) Growth curve of two *PUF3* RNAi clones induced (+)/ uninduced (-) with tetracycline. (**B**) Western blots showing the expression of a V5-tagged PUF3 in the cell lines in (A). (**C**) The upper panel represents the design of a restriction digest. Coding sequences of genes are represented by rectangular blocks and vertical double-strokes show restriction sites. Genomic DNA of cells selected for double knockout (transfected with plasmid pHD2851 and pHD2852) were digested with *EcoRI* (experiment i) or *EcoNI* and *SacII* (experiment ii). The digested DNA for wild-type (WT), single knockouts (*BSD* and *PAC*) and candidate knockout clones were analysed by southern blotting using probes targeting *PUF3* (shown in upper panel) and *TUT3* as a loading control. 'L' is the molecular marker. (**D**) DNA from (C) were also used in a polymerase chain reaction using the primers shown (prefixed with "cz"). Double stroke lines represent the ends of the cloning fragment used for transfections.

Wild-type bloodstream forms were sequentially transfected with linearized DNA from pHD 2851 and pHD 2852 which encode puromycin N-acetyltransferase (PAC) and blasticidin S deaminase (BSD), both flanked with segments of *PUF3* UTRs for *in locus* integration of the drug resistance cassettes while replacing *PUF3* coding sequences. Knockout clones were selected using both puromycin and blasticidin. PCR using primers originating within and without the transfection DNA confirmed the absence of *PUF3* and the presence of the drug resistance cassettes in place of *PUF3* (Figure 15D). Mutations during gene replacement can occur at primer binding sites and result in a false negative PCR of *PUF3*. A southern blot was therefore done to detect *PUF3* in its locus (experiment i) and elsewhere in the genome

(experiment ii). A radioactive probe detecting a *PUF3* unique sequence was used to detect *PUF3* on an autoradiograph (Figure 15C). Both Southern blotting and PCR confirmed the absence of *PUF3* in the knockout (KO) clones confirming successful deletion of *PUF3* from the genome and that PUF3 is not required for survival of the monomorphic bloodstream forms.

A proteomic screen in differentiating trypanosomes [32] showed that PUF3 is 1.5 fold up-regulated during differentiation and a genome-wide RNAi screen [173] suggested that bloodstream cells depleted of *PUF3* were disadvantaged during differentiation. To get a more realistic effect of PUF3 requirements in differentiation, RNAi and knockout of *PUF3* were attempted in a pleomorphic *T. brucei* strain (EATRO 1125), which are capable of differentiating to growing procyclic cells. This was done in a similar manner as in the monomorphic cells described earlier. Knockdown of *PUF3* conferred a marginal growth defect reproducible as from the third day of depletion (Figure 16A). The un-induced cells also showed a slight growth defect that could result from poor regulation of RNAi in the absence of tetracycline (Figure 16B). However, a parental cell line lacking the RNAi plasmid did not show a different V5-PUF3 signal from the un-induced RNAi cells therefore excluding leaky regulation (see lane marked with # in Figure 16B). This comparison was however done only once. Other fitness costs due to genetic manipulation cannot be completely ruled out.

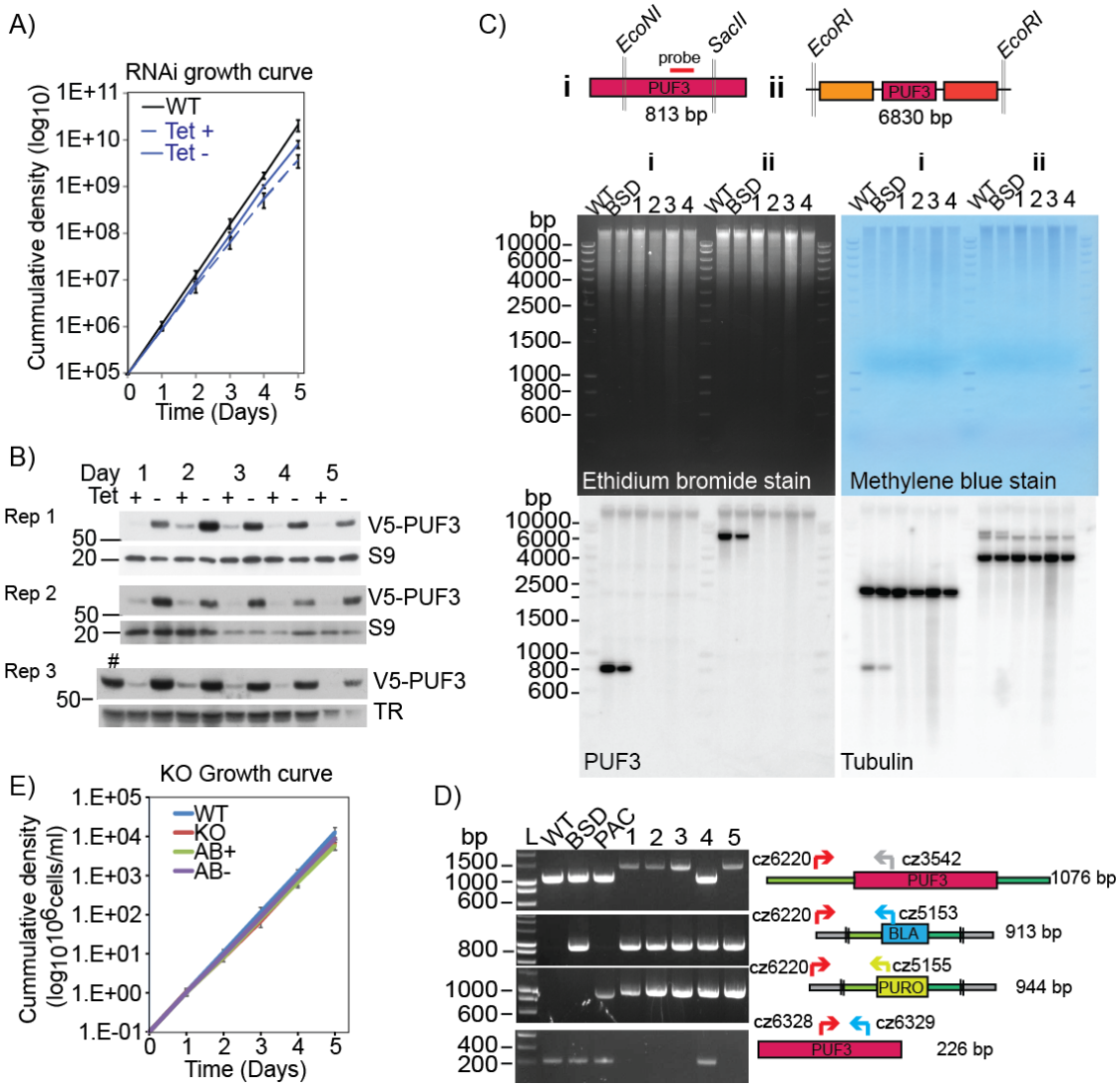


Figure 16: PUF3 is not essential for growth and survival of pleomorphic cells. (**A**) Growth curve of bloodstream cells induced for RNAi against *PUF3* in triplicate. Tetracycline induction and splitting of cell cultures were done daily. (**B**) Western blots showing depletion of PUF3 in the three replicates in (A). Trypanothione reductase (TR) or ribosomal protein S9 were used as loading controls. Lane marked with '#' indicate the parental cell line of the RNAi clones. (**C**) Southern blot to detect presence of *PUF3*. The upper panel represents the design of two southern blot experiments. Coding sequences of tandem genes are represented by rectangular blocks and vertical double-strokes show restriction sites. Genomic DNA of cells selected for double knockout (transfected with plasmid pHD2851 and pHD2852) were digested with *EcoRI* (experiment i) or *EcoNI* and *SacII* (experiment ii). (**D**) Polymerase chain reaction to test for *PUF3* knockout. Genomic DNA from wild-type, single knockouts (BSD and PAC) and knockout clones from (C) were used in a polymerase chain reaction using the primers shown (prefixed with "cz"). Double stroke lines represent the ends of the cloning fragment used for transfections. (**E**) Growth curve of knockout cells and add-back (AB+/-) expressing un-tagged PUF3 in the presence/absence of tetracycline.

In contrast to the monomorphic cells, some protein could still be detected even after 4 days of RNAi indicating that the RNAi was not adequate. Knockout of *PUF3* was therefore attempted. Just like in monomorphic cells, both PCR and southern blotting confirmed that *PUF3* is not essential for survival of the bloodstream cells. Furthermore a growth curve of the KO mutants did not show any growth defect in contrast to those depleted of *PUF3* by RNAi. This discrepancy can be as a result of adaptation by the *PUF3* null stable cell lines.

Because of the possible adaptation to life without PUF3, depletion by RNAi is short-term and may therefore be used to study the immediate effect of depleting *PUF3* before adaptation mechanisms ensue. With this assumption, a transcriptome profile was done following 24 hr depletion of *PUF3* by RNAi and fold changes compared between induced and un-induced samples (Fig 3).

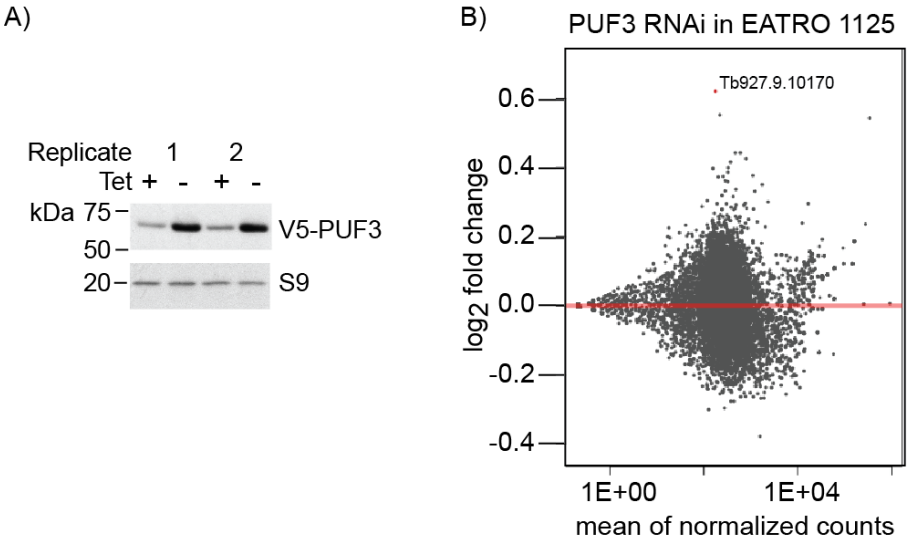


Figure 17: Transcriptome analysis of *PUF3* depleted pleomorphic EATRO 1125 bloodstream cells. (**A**) Western blot to detect PUF3. RNA interference against *PUF3* was induced for 24 hours in duplicate and depletion confirmed by Western blotting. Ribosomal protein S9 served as a loading control. (**B**) Total RNA from cells in (A) was depleted of rRNAs and given for library preparation and sequencing. RNA reads that aligned to coding sequences were analysed by DEseq2 and fold changes represented on a scatter plot.

Unexpectedly, only one gene (*Tb*927.9.10170) was differentially expressed at an adjusted P-value <0.1, and such a result is probably not significant. The gene is annotated as an anaphase-promoting complex subunit 11 (APC11). This gene was modestly upregulated by 1.5 fold. APC is an E3 ubiquitin ligase complex that ubiquitinates cell cycle regulators securin and cyclin B in mitosis for elimination, thus enabling cell cycle progression from metaphase to anaphase [174]. APC11 is one of the ten core components of APC in *T. brucei* [175,176]. Its

depletion by RNAi in either bloodstream or procyclic forms had no effect on the cell cycle [175] possibly due to the RNAi inefficiency. Ideally, this transcriptome profiling data would highlight a category of genes that are upregulated/downregulated and provide a hint to the function of PUF3. With the repressive function of PUF3, *APC11* could be a target of PUF3 during the cell cycle but given the lack of an obvious cellular phenotype when *APC11* is depleted [176], it was not pursued further.

Collectively, depletion of *PUF3* by RNAi in bloodstream forms causes a slight growth defect that does not significantly alter the transcriptome, at least in pleomorphic cells. Its depletion and knockout in bloodstream forms had no effect on cell growth and survival. PUF3 could therefore be required in other life stages hence its role in differentiation was explored.

3.14 PUF3 is required for normal differentiation of bloodstream forms

To investigate whether PUF3 is required for differentiation of bloodstream to procyclic forms, *PUF3* knockout (KO) cells were monitored for the formation of the glutamate-proline (EP) rich procyclin surface coat that is characteristic of procylic (insect midgut) forms. First, as a proof of concept and out of convenience, bloodstream monomorphic *PUF3* KO cell lines were used in the differentiation experiments because they were readily available. Pleomorphic cells would give a more realistic outlook on matters differentiation. The KO cells were grown to high density (1.8-2×10^6 cells/ml), treated with 6mM cis-aconitate and let to grow at 27°C. Samples were collected at various time-points and examined for EP/GPEET procyclin on a Western blot (Figure 18).

The KO cells showed a delay in synthesis of EP/GPEET procyclin. Traces of procyclins in the KO were visible after 24 hrs of differentiation. A *PUF3* KO cell line complemented with *PUF3* (Add-back, AB) had a modest rescue of this defect compared to wild-type (WT). This defect has also been reported in cells lacking zinc finger protein ZC3H18 [96] and *Tb*4EIP. Strangely, PUF3 with a C-terminal tag was unable to rescue the KO defect in multiple attempts (Figure 18C). In addition, detection of N-terminal tagged version of PUF3 was unreliable possibly due to some technical challenge. Therefore AB clones were made with untagged PUF3, and expression was confirmed by Northern blotting (Figure 18B).

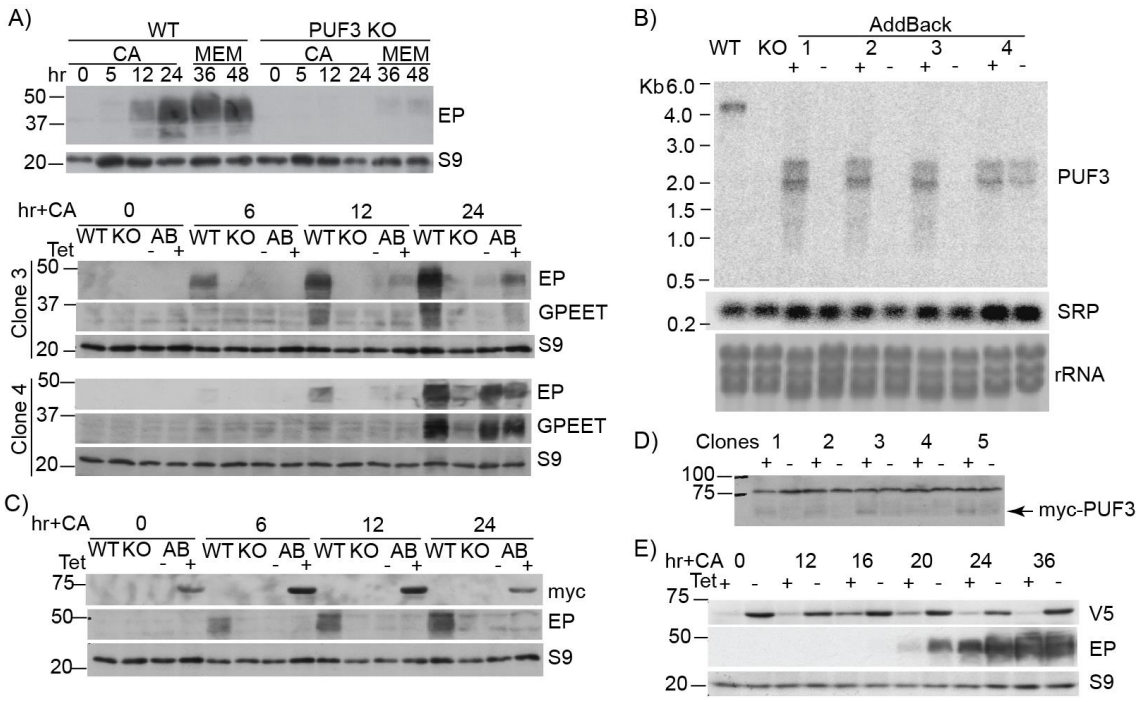


Figure 18: PUF3 is required for normal differentiation in monomorphic cells. (**A**) Western blots showing expression of EP/GPEET. Monomorphic bloodstream cells were grown to 1.8-2.0 $\times 10^6$ cells/ml and differentiated to procyclic by adding 6mM cis-aconitate (CA) to the culture and growing at 27°C. The time shown in hours is time after CA treatment. EP/GPEET were used to monitor differentiation and S9 represents the load. (**B**) Northern blot showing inducible expression of *PUF3* add-back (AB) in clones 1-4. Signal recognition particle SRP and rRNA represent load. (**C**) Western blot of an experiment conducted as in (A) but using AB clones expressing PUF3 that has a C-terminal myc tag. (**D**) Western blot analysis of putative AB clones 1-5 expressing PUF3 that contains N-terminal myc tag. (**E**) Expression of EP cells depleted of *PUF3* by RNAi. RNAi was induced for 24 hr, and cells at ~1.8×106 cells/ml were differentiated by adding 6mM CA to the culture and growing at 27°C.

The defect in differentiation was also observed when pleomorphic cells depleted of *PUF3* by RNAi were put to differentiate to procyclic forms. Cells were grown in semi-solid media containing methyl cellulose and meanwhile RNAi of *PUF3* was induced for 48 hrs before initiating cells differentiation to procyclic forms. *PUF3* depleted cells showed a delayed expression of EP (Figure 19A). In addition, PAD1 (marker for stumpy forms) was conspicuously depleted in RNAi conditions compared to similarly treated non-induced and wild-type cells (Figure 19B and C). PAD1 seems to appear later at six hours into differentiation and disappears as expected between 12-24 hrs. Notably, over-expression (Figure 19D, see 0hr) and RNAi (Figure 19B, see 6hr) are abolished in high density cells because both strategies are dependent on RNA polymerase I, which gets turned off. RNAi and over-expression were therefore induced for 48 hrs prior to growth arrest.

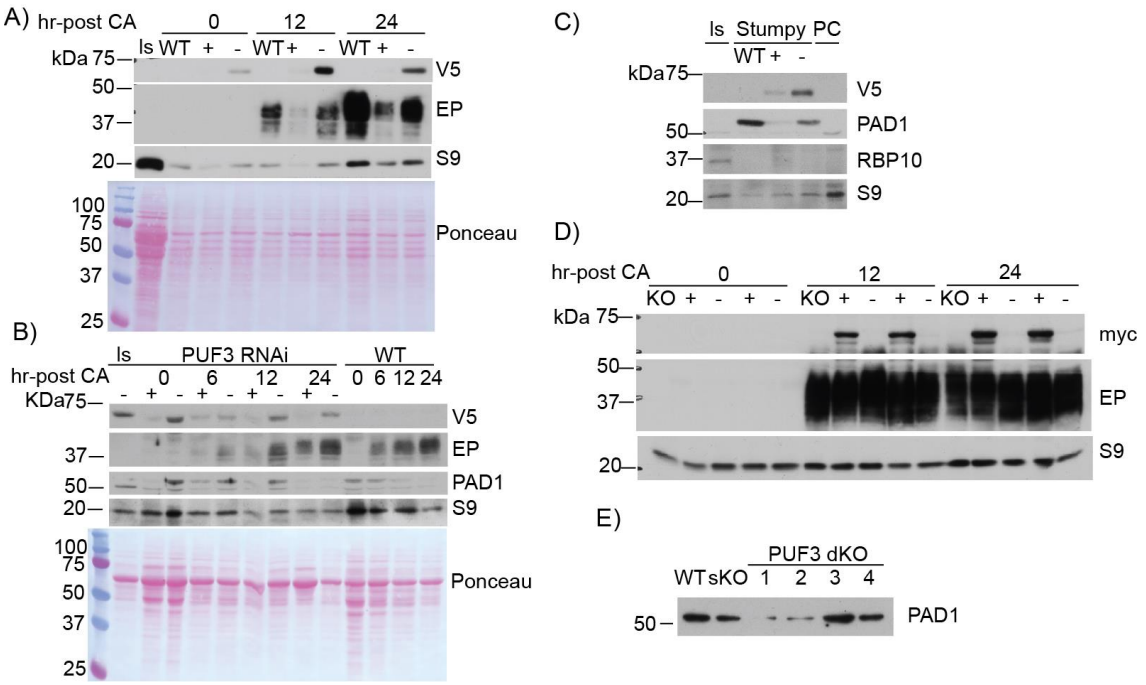


Figure 19: Differentiation of *PUF3* depleted and deleted pleomorphic bloodstream cells. (**A**) Expression of EP procyclin during differentiation of *PUF3* RNAi. EATRO1125 bloodstream cells were induced for RNAi for 48 hours while overgrowing in methyl cellulose containing media. The resulting stumpy cells at a density ~10^6 cells/ml were differentiated by adding cis-aconitate (CA) and let to grow at 27°C. ls: slender cells, WT; wild-type; +/-; *PUF3* RNAi with/without tetracycline. V5 represents expression of V5-PUF3, Ponceau and S9 represents load. (**B**) A replica of the experiment in (A) but showing in addition, expression of PAD1 – the marker for stumpy forms. (**C**) Expression of PAD1 in *PUF3* RNAi conditions. Stumpy cells were developed as in (A). Procyclic cells, PC, at a density of ~3×10^6 cells/ml and ls were included for comparison. V5 represents expression of V5-PUF3, RBP10 expression represents slender cells, S9 represents the load. (**D**) Expression of EP in PUF3 knockout (KO) cells and two add-back clones during differentiation from stumpy to procyclic forms. Stumpy cells were made by growing the cells to high densities in methyl cellulose containing media and collected for analysis at a density ~10^6 cells/ml. The add-back clones inducibly express PUF3 with a c-terminal myc tag. (**E**) Expression of PAD1 in stumpy cells obtained as in (D) for WT, *PUF3* single knockout (sKO) and double knockout (KO) clones 1-4.

Having obtained *PUF3* knockout pleomorphic cells, the expression of EP was tested. Surprisingly, these cells as well as add-backs express EP procyclin without any defect (Figure 19D). In addition, *PUF3* KO cells express the stumpy marker PAD1 similar to WT and single knockout (sKO) cells (Figure 19E). This is in complete contrast to the effect observed with depletion of PUF3 by RNAi and those from experiments using monomorphic cells. Depletion of *ZC3H18* also delays EP expression and in contrast to PUF3, the defect persisted even in the ZC3H18 knockout EATRO1125 cell lines [96]. These result could be caused by adaptation of the cells to life without PUF3 similar to adaptation already observed with the absence of defective growth in KO cells in contrast to cells depleted of PUF3 by RNAi (Figure 16). In

addition, the KO cells were capable of expressing PAD1 contrasting PUF3 depleted cells. A summary of the conditions that cause a defect in PAD1 and procyclin expression is shown in Table 2

Table 2: Comparison of the effect of *PUF3* depletion/deletion on expression of EP/PAD1 in monomorphic and pleomorphic cells

	Monomorphic cells	Pleomorphic cells
WT	Normal EP	Normal EP and PAD1
RNAi +	Delayed EP	Delayed EP and PAD1
RNAi -	Normal EP	Normal EP and PAD1
KO	EP delayed	Normal EP and PAD1
AB +	Normal EP	Normal EP and PAD1
AB -	EP delayed	Normal EP and PAD1

Collectively, the results suggest that normally, bloodstream cells require PUF3 for normal differentiation kinetics. The absence of *PUF3* in the short term (by RNAi) caused a delay in progression of differentiation. In contrast, complete knockout of *PUF3* in pleomorphic cells did not stall differentiation, possibly because of some adaptation mechanism absent in monomorphic cells. I therefore sought to determine the converse; the effect of over-expressing PUF3 in procyclic forms.

3.15 Over-expression of PUF3 is lethal to procyclic cells

The effect of over-expressing PUF3 was tested during differentiation of bloodstream to early procyclic forms and further as late procyclic forms. *PUF3* add-back (AB) cells were generated from the pleomorphic *PUF3* knockout (KO) bloodstream cells and these expressed an inducible *PUF3* under the control of a RNA polymerase I promotor, and was therefore expected to be over-expressed in the presence of tetracycline. Since the AB cells expressed untagged PUF3, and an antibody against PUF3 was unavailable, the inducible expression could not be confirmed by Western blotting. Nonetheless, KO and AB cells were differentiated from slender/stumpy to procyclic cells. Bloodstream AB cells were grown for 24 hrs (slender) or 48 hrs (stumpy) in tetracycline prior to differentiation. After differentiation, both KO and AB cell cultures without tetracycline grew to high densities but surprisingly, add-back cells over-expressing PUF3 had stunted growth and eventually died (Figure 20A and B). Cells

differentiated from slender cells started to divide from the third day while those differentiated from stumpy forms started dividing immediately on the first day in procyclic media (MEM) confirming that differentiation from stumpy forms is more efficient than from slender forms. The time point for cell division of KO and non-induced AB cells appeared to coincide with the density-decline time point of cells over-expressing PUF3. These results indicated a growth disadvantage for procyclic cells that supposedly over-express PUF3.

To determine whether this defect was persistent in late procyclic cells, the early procyclic cells were grown in procyclic conditions for over 4 weeks and their growth kinetics monitored daily. The cells grew at a similar rate for 2-3 days and the AB over-expressing PUF3 stopped dividing. Wild-type (WT) cells grow to much higher densities indicating that procyclic cells lacking *PUF3* also have a growth defect (Figure 20C).

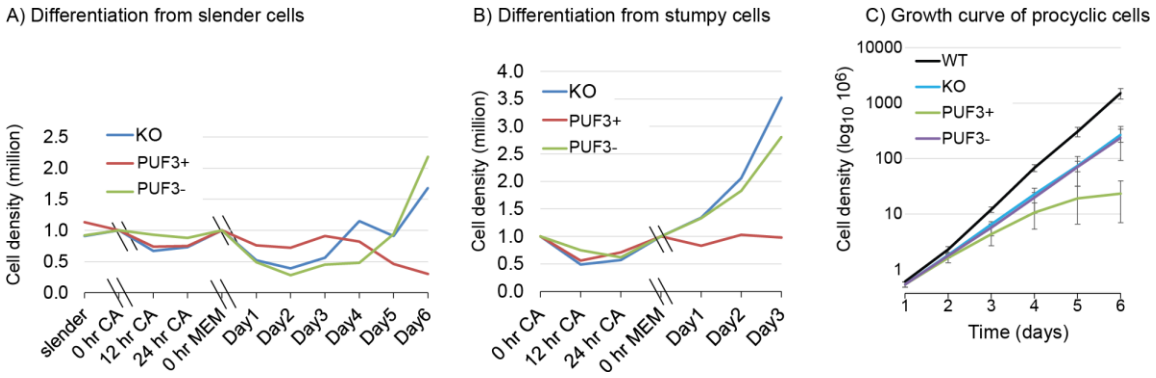


Figure 20: Growth curves of pleomorphic cells over-expressing PUF3. (**A**) Growth curve of slender bloodstream cells differentiating to procyclic forms. Slender cells were treated with/without tetracycline for 24 hrs and their densities set to 10^6 cells/ml, treated with 6mM cis-aconitate (CA) and put to grow at 27°C – 0hr CA. After 24 hr in CA, the cells were transferred to MEM procyclic media at equal densities of 10^6 cells/ml. The double stroke lines indicate change of time scale. (**B**) Growth curve of stumpy cells differentiating to procyclic forms. Slender cells were grown with/without tetracycline to high densities in methyl cellulose containing media to form stumpy forms. These were transferred to methyl cellulose-free media at a density of 10^6 cells/ml, treated with 6mM CA and grown at 27°C. After 24 hr in CA, the cells were transferred to MEM procyclic media at equal densities of 10^6 cells/ml. (**C**) Growth curve of late procyclic cells. Differentiated bloodstream cells were grown for at least four weeks as procyclic forms before the experiment. Their density was monitored daily in the presence/absence of tetracycline. WT; wild-type, KO; PUF3 double knockout, PUF3+/-; KO with inducible PUF3 add-back in the presence/absence of tetracycline.

Together, PUF3 seems to be essential for growth in procyclic forms but an excess of it is also lethal. Over-expression of PUF5 has also been reported to cause a growth defect in procyclic but not bloodstream trypanosomes [101]. The authors proposed that over-expressed protein may outcompete other RBPs in binding some non-targets of low affinity. The similarities in PUF domains could allow a PUF protein to bind targets of other PUF proteins even just with low

affinity. Over-expression therefore could make up for the low affinity binding of a non-target and disrupt regulation. Elsewhere, it was impossible to obtain procyclic cells with over-expressed-tagged version of ZC3H12, while procyclic cells over-expressing ZC3H13 experienced impaired growth [94]. In the same vein, over-expression of PUF3 could therefore affect other PUF targets and cause the growth defect observed. Though over-expression of PUF3 could not be demonstrated, the actin 3'UTR used in the over-expression plasmid (pHD2895) may disrupt regulation of *PUF3* and favour mRNA stability or translation, hence over-expression. The northern blot in Figure 18B however shows roughly equal amounts of *PUF3* mRNA for the WT and AB clones.

3.16 Putative targets of PUF3 identified by RNA immunoprecipitation

To determine the mRNAs that are bound by PUF3, RNA immunoprecipitation in tandem with RNAseq (RIP-seq) were carried out. Bloodstream pleomorphic cells with one *PUF3* allele deleted and the remaining tagged on its N-terminus with a tandem affinity purification (TAP) tag were used. Cells were UV-irradiated, lysed and incubated with IgG beads. RNA was extracted from the unbound and the bound fraction of two replicates, and given for sequencing. Reads were mapped to the genome of *T. brucei* and the normalized read counts (reads per million, RPM) determined. The ratios between bound and unbound fractions were determined and those with ratios ≥2 are shown in table 1.

Table 3: List of mRNA co-purified with PUF3

GeneID	Annotation	Category	Bound (B) average RPM	Unbound (U) average RPM	log2 (B/U)
*Tb*927.10.11280	Hypothetical protein, *T brucei* only	Unknown	1.9	0.4	2.1
*Tb*927.11.4440	Hypothetical protein	Unknown	2.5	0.8	1.6
*Tb*927.1.4210	Hypothetical protein	Unlikely	1.7	0.7	1.3
*Tb*927.11.16190 †	Hypothetical protein	Unknown	1.7	0.7	1.2
*Tb*927.10.1790	Hypothetical protein	Unknown	1.2	0.5	1.2
*Tb*927.11.17690	ESAG3 pseudogene	ESAG pseudo	5.4	2.4	1.1
*Tb*11.v5.0375 †	Hypothetical protein, conserved	Unknown	8.1	3.9	1.1
*Tb*927.9.8910	Hypothetical protein	Unlikely	1.2	0.5	1.1
*Tb*927.11.12710 †	Variant surface glycoprotein (VSG)-related, putative	VSG	5.3	2.5	1.1

*Tb*927.9.2610	Hypothetical protein	Unlikely	4.5	2.1	1.1
*Tb*927.1.2780 †	Hypothetical protein	Unknown	2.8	1.3	1.1
*Tb*927.7.2660 †	ZC3H20	RNA binding	76.8	35.4	1.1
*Tb*927.1.3510	Hypothetical protein	Unlikely	1.5	0.7	1
*Tb*927.3.5780	Hypothetical protein	Unknown	2.6	1.3	1
*Tb*927.11.13140 †	Cytochrome oxidase subunit X	Mito electron transport	20.2	10.1	1
*Tb*927.9.14660	SLACS reverse transcriptase, putative	Unknown	7.7	3.9	1
*Tb*927.1.810	Hypothetical protein	Unlikely	3.1	1.5	1
*Tb*927.11.18660	Expression site-associated gene 9	ESAG	1.7	0.9	1
*Tb*927.1.940	Hypothetical protein	Unlikely	1	0.5	1
*Tb*927.1.2170	Hypothetical protein	Unlikely	1.4	0.7	1

Genes marked with "†" have annotated 3'UTRs and were used for motif searches.

Generally, apart from ZC3H20, the genes had very low RPMs and according to our lab experience, such candidates are likely to be non-specific. Reverse transcription followed by PCR did not detect ZC3H20 or *Tb*927.10.11280 in the bound RNA. These two were chosen because ZC3H20 had most read representation while *Tb*927.10.11280 was the best bound candidate. Motif search on the 3'UTR of these genes did not yield any significant result. Variants of the consensus pumilio motif UGUANAUG [107] appeared in both bound and unbound genes. The high number of PUF proteins which supposedly share a consensus motif would make it difficult and biased to get a PUF3 motif judging only from the e-values.

3.17 Protein partners of PUF3

To determine the proteins that associate with PUF3 and therefore acquire a hint of how PUF3 represses gene expression, a cell line expressing a fusion of TAP tag and PUF3 (pHD 2893, TAP-PUF3) from the endogenous locus was used for tandem affinity purification. The experiment was done in triplicate and the bound proteins analyzed by mass spectrometry. A control pull-down with GFP-TAP together with a list of common contaminants from previous analyses from our lab were used to obtain the list of possible targets (Table 4). For inclusion, a protein had to be absent in the GFP control and have less than three peptides in 59 previous runs from our lab.

Table 4: List of proteins and number of peptides co-purified with PUF3 in triplicate

GeneID	Description	rep1	rep2	rep3
*Tb*927.10.310	PUF3	28	18	16
*Tb*927.11.7510	Luminal binding protein 1 (BiP), putative	16	4	2
*Tb*927.11.4100	Variant surface glycoprotein (VSG), putative	6	3	1
*Tb*927.11.4820	60S ribosomal protein L29	3	2	2
*Tb*927.10.8980	Hypothetical protein, conserved	2	1	2

PUF3 was successfully purified in all the replicates and was the most abundant protein. Other precipitated proteins were present in rather low amount and were likely non-specific based on experiences with previous pull-downs in our lab, and were therefore not pursued. This does not however mean that PUF3 does not bind any proteins. Instead, interactions with PUF3 could be unstable thus unsustainable throughout the two-step long procedure of TAP purification. Transient interactions are also conceivable and would not be detected by this procedure. However, technical challenges could also be a reason for the fewer number of peptides detected in replicate 2 and 3 as compared to replicate 1.

3.18 Targets of PUF3 identified by RNA editing

As an alternative method to identify RNA targets of PUF3, I used a method called TRIBE (Targets of RNA Identified By Editing) [159]. It exploits the *Drosophila* ADAR catalytic domain (adenosine deaminase acting on RNA), which deaminates adenosine (A) to inosine (I). This mutation is detected as guanosine (G) by ribosomes and RNAseq [159]. Native ADAR consists of an N-terminus double-stranded RNA-binding domain and a C-terminus catalytic domain. I expressed a fusion between the ADAR catalytic domain (henceforth simply referred to as ADAR) and PUF3 from the endogenous locus of PUF3 (Figure 21A). PUF3 would therefore direct editing by ADAR, which marks PUF3 targets with irreversible A>I mutations approximately within 500 bp [159] from PUF3 binding site, and targets are identified by RNAseq and bioinformatics. ADAR however, has a preference for adenosines flanked by 5' uridine and 3' guanosine in a double-stranded RNA sequence and lacks processivity, and as a result may lead to failure to identify some RBP targets [177].

With the premise that PUF3 regulates gene expression during differentiation to stumpy (Figure 19BC), cells were treated for 36hrs with 10µM of a hydrolysable cAMP analog (8-pCPT-2'-O-Me-cAMP) which induces stumpy formation [33–35], and their RNA sequenced. An alternative

would have been to grow the cells to high densities in methyl cellulose containing media, but this method never gave intact RNA in a previous experiment (see Figure 8C). Edited sites were detected as in the original description of TRIBE [159] and as described in the methods section. A>G mutations were defined as loci with A>80% and G=0% in genomic DNA and G>0% in RNA. For genes in the reverse strand, the reverse complement T>C was evaluated. The following threshold - used in the original description of TRIBE - was also used here i.e. loci with alignment depth < 20 and with < 10% edited residues (edited bases/total bases) were discarded as endogenous editing sites. Edited loci were compared against identical loci from a control cell line expressing ADAR-GFP fusion. PUF3 with ADAR on its C-terminus was also included because it was hypothesized from KO complementation studies (Figure 15) that C-terminal tags inactivate PUF3, and would therefore act as an additional control (Figure 21A). All the fusion proteins used were knockins into the *PUF3* locus with the other *PUF3* allele deleted.

3.18.1 Editing by ADAR is directed by an RNA binding domain

I first looked at the raw data before removing sites that had not meet the editing threshold i.e. sites with depth <20 and editing events <10%. I included TRIBE datasets from *Drosophila* (details in the methods) for comparison, and examined all possible mutations in addition to A>G (Figure 21B). In both *Trypanosoma* and *Drosophila*, the number of A>G edited sites were more in samples expressing ADAR compared to those lacking ADAR, indicating that ADAR is indeed responsible for the additional A>G mutations. ADAR-PUF3 had more A>G editing than in the GFP control (see aG vs aP, Figure 21B). C>G mutations were the least frequent in the two species but the most frequent varied.

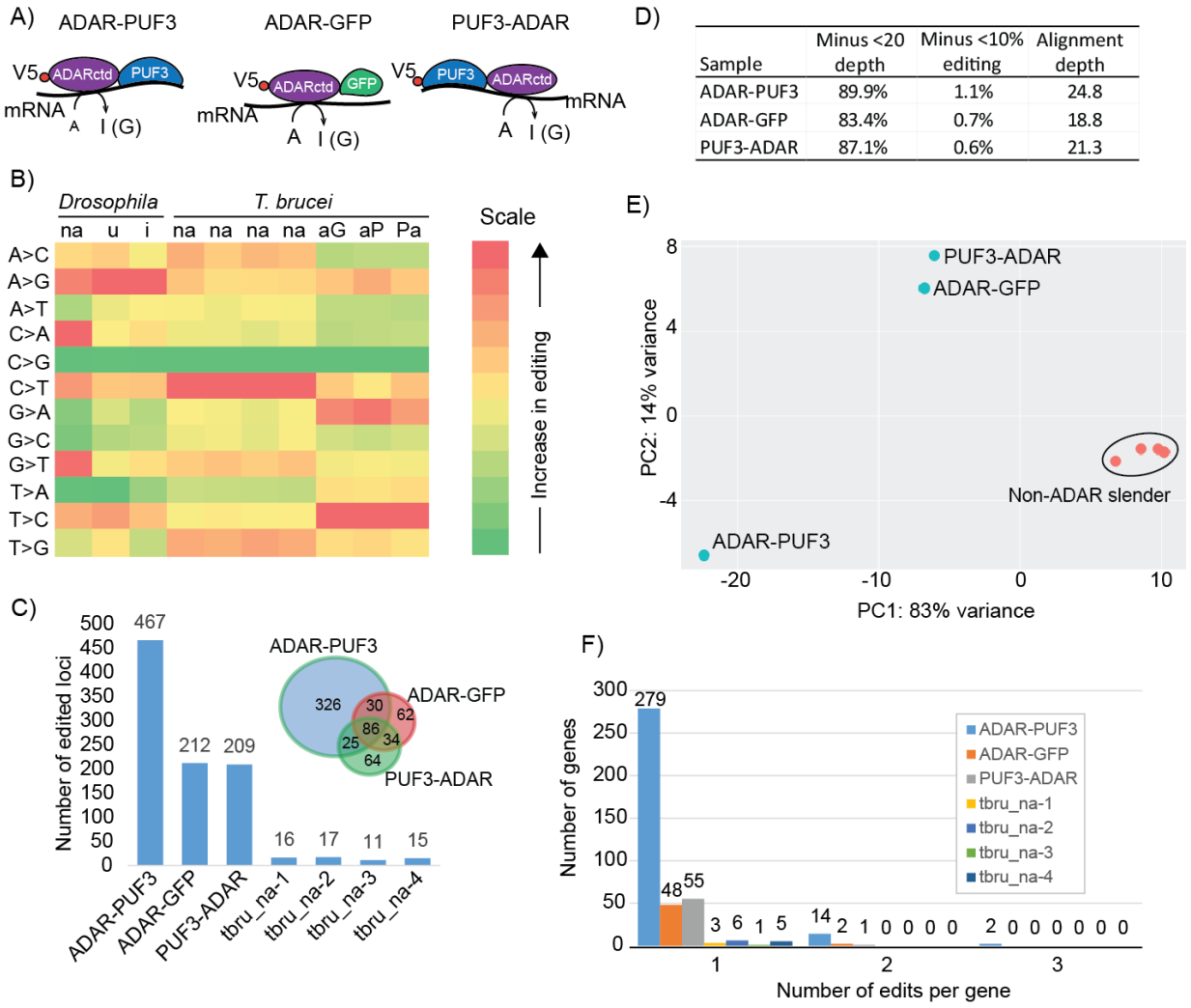


Figure 21: Editing by ADAR catalytic domain is directed by an RNA binding PUF3. (**A**) Models of the fusions between ADAR catalytic domain (ADARctd) and PUF3/GFP. (**B**) Comparison of number of all detected edited sites for *Drosophila* and *T. brucei*. na; Non-ADAR, u; uninduced, i; induced, aG; ADAR-GFP, aP; ADAR-PUF3, Pa; PUF3-ADAR. (**C**) Number of edited sites in ADAR containing samples of *T. brucei* following filters in (B). The insert Venn diagram represents the overlap between ADAR containing samples. (**D**) Table shows the percentage editing events retained after filtering for depth and percentage editing. Alignment depths are average of all loci used in the alignment and were determined using samtools. (**E**) Principle component analysis of the samples in (C). The non-ADAR samples were slender forms while the ADAR samples were differentiated to stumpy forms. (**F**) Frequency of edited sites per gene. *Tb*ru_na 1-4 represent the four *T. brucei* samples lacking ADAR in the same order shown in (A).

On removing endogenous editing events from the A>G raw data of trypanosomes, more than 80% of the editing events were discarded for lack of at least 10% editing and another 10% for lack of depth (Figure 21D). These samples had roughly the same alignment depth (Figure 21D) thus excluding bias due to depth differences. ADAR-PUF3 had more than 5-times more

edited loci than the GFP control (Figure 21C) indicating that editing by ADAR is preferably directed by an RNA binding protein.

As predicted earlier, PUF3 with a C-terminal fusion appears inactive; it might be expected to either direct ADAR editing similar to PUF3 with ADAR on the N terminus or be inactive. In fact, it had similar number of sites as the control cell-line. Furthermore, most of the A>G loci in the PUF3-ADAR cell line (120/209) were shared with the GFP control, 86 of which were common in all TRIBE samples and were probably ADAR biased sites (These are double stranded RNA sites that are preferentially edited by even ADAR lacking its dsRNA binding domain). In addition, PUF3-ADAR and ADAR-GFP also cluster together in a principle component analysis plot (Figure 21E). It may be argued that in the PUF3-ADAR cell-line, it was ADAR which was inactive and not PUF3 because of tagging its N-terminus. However, the native ADAR in *Drosophila* is preceded by an RNA-binding domain at its N-terminus therefore an N-terminal tag would not inactivate ADAR [159,178]. ADAR on PUF3 C-terminal end could compromise RNA binding because PUF3 pumilio domains are in the C-terminus (Figure 13) and the addition of ~50kDa ADAR-fusion on the distal end may sterically impair access to RNA, resulting in non-specific editing. This cell-line was henceforth used as a control together with GFP-ADAR. PUF3-ADAR and ADAR-GFP cell lines could essentially be considered as PUF3 knockout cells since they have no functional PUF3. The lack of association between ADAR-PUF3 cells with the controls in the PCA plot (Figure 21E) is suggestive of a difference in the transcriptomes. However, differential expression analysis could not be done because that would need replicates.

Edited sites were mapped to genes models using ANNOVAR [162] and the number of editing per gene determined (Figure 21E). I identified 295 genes in the ADAR-PUF3 sample which are six-times more than the TRIBE controls (GFP-ADAR and PUF3-ADAR). This increase in the number of genes identified indicates that editing by ADAR is faithfully directed by the RNA-binding protein. This would assume that PUF3-ADAR is defective in binding RNA resulting to its similarities with the GFP control. Most of the genes were edited on a single loci consistent with ADAR's lack of processivity and similar to TRIBE in *Drosophila* [159].

Overall, the results indicate high RNA editing/mutations across cells even in the absence of ADAR. However, more edited sites on RNA were observed in the sample that contained a functional RNA binding domain (ADAR-PUF3) and less in samples expressing ADAR that lacked an RNA binding capability (GFP-ADAR or PUF3-ADAR). Figure 22 shows examples of

screenshots of select sites for PUF3-TRIBE in *Trypanosomes* and Hrp48-TRIBE in *Drosophila*.

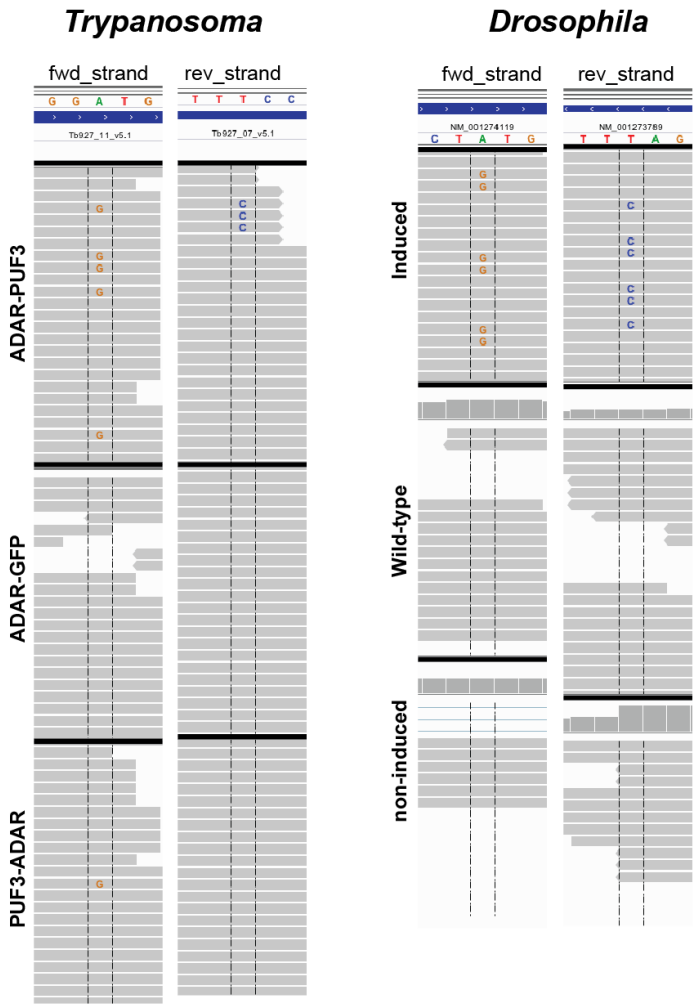


Figure 22: Example screenshots of edited sites in PUF3-TRIBE in *Trypanosoma* and Hrp48-TRIBE in *Drosophila*. Horizontal grey bars represent RNA reads aligned to genes with mismatches shown by nucleotide letters. A>G mutations were detected in the forward strand and T>C mutations in the reverse strand. The name of the sample is indicated on the left of the reads. Inducible Hrp48-ADAR was used in Hrp48-TRIBE in *Drosophila* and the Wild-type sample lacked the TRIBE construct. The datasets are described in the methods section.

3.18.2 Putative targets of PUF3 identified by TRIBE

To obtain target genes of PUF3, all the A>G edited loci from the control cells and from those lacking ADAR were subtracted from the ADAR-PUF3 list to obtain 295 possible targets. These genes were clustered in functional categories (Figure 23). The majority of the genes were of unknown function, followed by cytoskeleton, protein kinases, RNA binding, leucine rich repeat

proteins (LRRP), expression-site associated genes (ESAG), chaperones and translation. Retrotransposons have high mutation rates and always detected in SNP calling, so were considered as false targets [159].

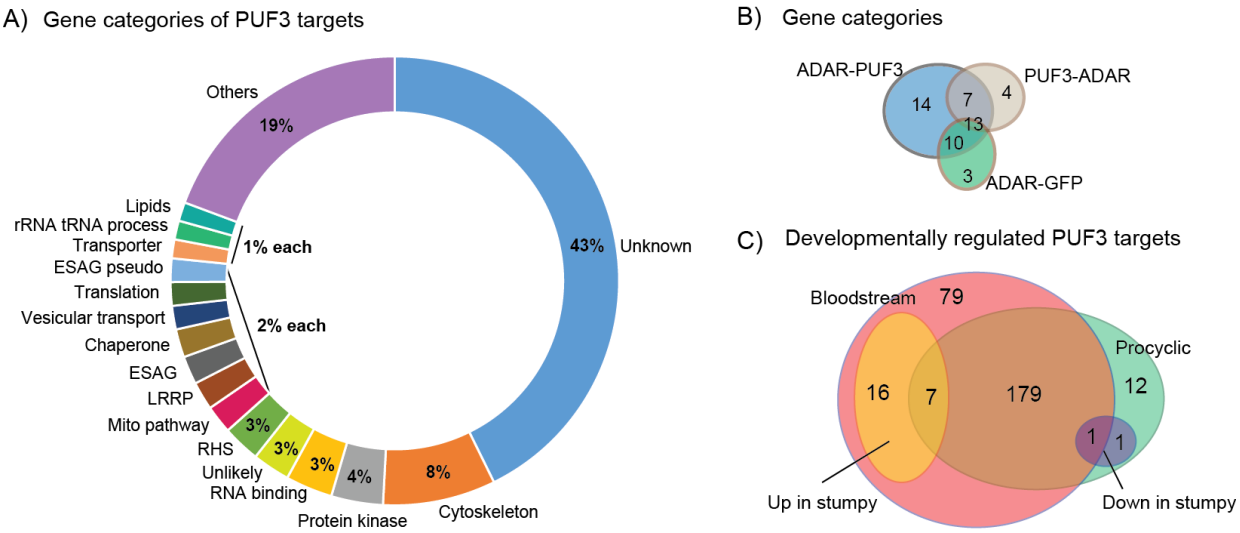


Figure 23: Putative targets of PUF3. (**A**) Functional categories of putative PUF3 targets. Categories with less than 1% representation were clustered together under 'Others' for demonstration purposes. (**B**) Number of categories of genes unique to three ADAR samples. (**C**) Numbers of developmentally regulated targets. These are genes that are more than 2-fold regulated in slender bloodstream, procyclic or stumpy forms.

Out of the 295 putative PUF3 target genes, 43% encode proteins of unknown function (Figure 23A). Interesting categories include cytoskeleton proteins (8%), protein kinases (4%) and RNA binding proteins (3%), mitochondria pathways (2%), since these are likely to be involved in differentiation [45]. In addition, ADAR-PUF3 had more unique gene categories than both controls (Figure 23B).

PUF3 is predicted to participate in developmental regulation and its targets would therefore be regulated between different forms and during differentiation. Genes that are regulated at least 2-fold between bloodstream and procyclic forms [45] and those that are regulated during stumpy formation [179] were matched with the list of ADAR-PUF3 targets. Interestingly, 79 of these were enriched in bloodstream forms consistent with PUF3 role as a repressor during differentiation. However, dominance of bloodstream expressed genes can arguably be circumstantial because the morphology of cAMP treated cells is not entirely stumpy and remnant slender bloodstream cells exist. Further cAMP treated cells are "unhealthy" [35] and this could explain the high number of cytoskeleton related genes. Nevertheless, having bloodstream enriched genes as targets is consistent with a role of PUF3 as a repressor of bloodstream-type mRNAs

during differentiation to stumpy forms. These targets include the repressors RBP10, DRBD5 and a conserved hypothetical protein. High confidence genes edited more than once are listed in Table 5.

Table 5: List of mRNA edited more than once by ADAR-PUF3

Gene ID	Annotation	#edits	Region	Category
*Tb*927.3.720	ZFP3	2	3'UTR;3'UTR	RNA binding
*Tb*927.10.1540	ZC3H30	2	3'UTR;3'UTR	RNA binding
*Tb*927.10.13490	serine/threonine kinase, putative	2	3'UTR;3'UTR	Protein kinase
*Tb*927.8.6260	hypothetical protein, conserved	2	Exonic; Exonic	Unknown
*Tb*927.8.4500	eIF4G5	3	3'UTR;3'UTR;3'UTR	Translation
*Tb*927.8.3690	Isocitrate dehydrogenase [NADP], mitochondrial precursor (IDH)	2	3'UTR;3'UTR	Citric acid cycle
*Tb*927.8.3250	IAD-1beta inner arm dynein heavy chain	2	Exonic; Exonic	Cytoskeleton
*Tb*927.4.360	1,2-Dihydroxy-3-keto-5-methylthiopentene dioxygenase, putative	2	3'UTR;5'UTR	Amino acids
*Tb*927.9.11050	4EIP, 4E-interacting protein	2	3'UTR;5'UTR	Translation
*Tb*927.11.5850	RBP38	2	3'UTR;exonic	RNA binding
*Tb*927.11.7570	ATP-grasp domain containing protein, putative	2	3'UTR;exonic	Unknown
*Tb*927.11.15240	Ras-related protein RAB2B, putative	2	3'UTR;3'UTR	Vesicular transport
*Tb*927.5.3790	Tetraspanin family, putative	2	3'UTR;exonic	Unknown
*Tb*927.7.350	hypothetical protein	2	Exonic; Exonic	Unknown
*Tb*927.7.3580	NEK11 protein kinase	3	3'UTR;3'UTR	Protein kinase
*Tb*927.6.4800	Hypothetical protein, conserved, tetratricopeptide repeat domain	2	Exonic; Exonic	Unknown

There was no overlap between PUF3 RNA-immunoprecipitation (PUF3-RIP) and PUF3-TRIBE. An overlap was not however expected because slender cells were used in PUF3-RIP and stumpy cells for PUF3-TRIBE. The use of TRIBE enabled the study of PUF3 in stumpy cells where RNA immunoprecipitation would otherwise be extremely challenging.

A high number of protein kinases and RNA binding proteins (RBPs) were identified among the target mRNA products. These might be involved in perception of the differentiation signal and the subsequent signaling pathway. Protein kinases and RBPs however, have long 3'UTRs and have more chances of getting edited. To check whether there was bias of length in the genes edited, I compared the lengths of 3'UTR and coding sequence (CDS) of the ADAR targets verses all trypanosome genes. Because of the differences in numbers of targets in each sample (Figure 21D), I assumed non-parametric test - Kruskal Wallis statistics. In all cases, the lengths of CDS and 3'UTRs of ADAR-PUF3 edited genes were longer compared to all trypanosome genes (Figure 24A and B) (adjusted P<0.0001). In addition, PUF3-ADAR edited genes had longer 3'UTRs. Since lengths of GFP-ADAR edited genes were not different from the total genes, the length differences observed here were characteristic of PUF3 and not ADAR.

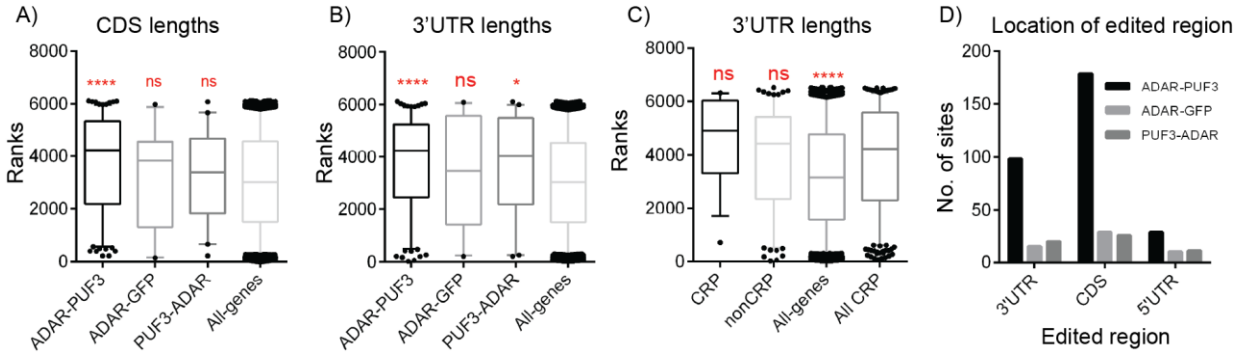


Figure 24: Characteristics of ADAR edited genes. (**A**) Box and whiskers plot comparing lengths of coding sequences (CDS) of ADAR-PUF3, ADAR-GFP and PUF3-ADAR with lengths of all genes using one-way ANOVA and Kruskal Wallis test. Boxes represent the upper quartile, median and lower quartile, while whiskers represent the 5 and 95 percentiles. (****) indicate significance at adjusted p<0.0001, ns; not significant, (*) indicate significance at adjusted p=0.04. (**B**) Similar parameters as in (A) to compare lengths of 3'UTRs. (**C**) Similar comparison as in (B) between lengths of CRP (cytoskeleton, RBP and protein kinases genes), non-CRP and all genes with lengths of all CRP. (**D**) Number of sites in each 3'UTR, coding sequence (CDS) and 5'UTR of edited genes. In all cases, only genes with annotated 3'UTR were considered.

In the categories of ADAR-PUF3 targets, cytoskeleton, RBP and protein kinases genes (here collectively abbreviated to CRP) are the most represented and in addition, are known to have long 3'UTRs. To determine whether CRP dominate the long 3'UTRs in the targets, I compared 3'UTR lengths of CRP, non-CRP and all genes against all CRP in the genome using the same statistics as before. CRPs had longer 3'UTR lengths as expected but their lengths were not different from other non-CRP targets, further confirming that longer 3'UTR targets are a dominant character of ADAR-PUF3 (Figure 24C).

RBPs affect their targets by mostly binding elements in the 3'UTR, though there are also examples of CDS targeting [77]. Using 2-way ANOVA and Turkey's multiple comparison test, I examined whether there was a preference for editing by recombinant ADAR in the 5'UTR, CDS or 3'UTR. There was no significant difference ($P<0.05$) between editing frequency in any region. ADAR-PUF3 had more sites in the CDS followed by the 3'UTR but these differences were not significant when regions were compared to each other within samples. There was equally no significant difference in the number of edited region between samples.

Motif searches were performed on the 295 ADAR-PUF3 targets using the GFP-ADAR targets as controls in DREME [155]. 1000 nucleotides upstream and downstream of the edited sites were used as input sequences. Edited sites in TRIBE are found within 500 nucleotides of the RBP-binding site [159]. No motifs were detected with an e-values <0.05 and none was detected using 3'UTRs of all targets, bloodstream enriched targets or the most abundant gene categories.

These targets could however not be validated by RNA immunoprecipitation because cells treated with cAMP analogs are unhealthy and it would not be viable to obtain intact RNA following lengthy immunoprecipitation procedures. This had already been predicted from efforts to obtain intact RNA from stumpy forms (Figure 8C). In addition, large amounts of cAMP analog would be required for such an experiments and was therefore not an economically viable choice.

4.0 Discussion

4.1 4E-interacting protein represses expression during differentiation

The initiation steps of translation offer multiple opportunities for control. Mechanisms range from the use of eIF4E-like cap binding proteins that do not recruit eIF4G [119,141,180,181], to use of 4E-binding proteins (4EBPs) that compete with eIF4G for binding eIF4E [125,140]. The genomic potential to encode multiple eIF4Gs, eIF4E, eIF4E-like and 4EBPs enables segmentation of general translation into fractions that enables dedicated translation of a subset of mRNAs in response to stress conditions and during development. Trypanosomes and related kinetoplastids have two hosts – a mammal (homoeothermic) and an invertebrate insect (poikilothermic). While inside their hosts, trypanosomes survive various stresses including immune and nutritional, and as they move from one body compartment to another, they exhibit pleomorphism, a manifest of their robust gene regulation mechanisms. Consistently, they encode a relatively high number of translation initiation factors; six eIF4Es, five eIF4Gs and two eIF4A which have potential for multiple eIF4F-like complexes. In addition, they encode a novel kinetoplastid-conserved 4E-binding protein, 4E-interacting protein (*Tb*4EIP) which was the focus of this work.

The results show that *Tb*4EIP is a repressor of gene expression independent of *Tb*EIF4E1. Tethering of *Tb*EIF4E1 in the absence of *Tb*4EIP had no effect on a reporter mRNA, suggesting repression by *Tb*EIF4E1 is dependent on *Tb*4EIP. These results are similar to those of mammalian GIGYF2 and 4E-T, which functions independent of 4EHP [141,142], and *Drosophila* Cup which functions independent of eIF4E [135]. GIGYF2 binds to mammalian 4EHP and is recruited to a subset of mRNA by various RNA binding proteins, including zinc finger protein ZNF598 [116] and tristetraprolin (TTP) [117,118]. Immunoprecipitation of *Tb*4EIP did not identify possible ligand that would suggest recruitment to particular mRNAs, but identified *Tb*EIF4E1 as the only reproducible binding partner [152]. However, yeast-2 hybrid assay identified the zinc finger protein ZC3H14 as a possible ligand of *Tb*4EIP [152]. However, ZC3H14 was not identified as an activator/repressor of gene expression in a tethering screen [84] and does not seem to bind RNA according to a poly(A) mRNA-bound proteome [78]. In fact, there is no evidence that ZC3H14 is expressed in the forms available in culture (TriTrypDB), and it may therefore be active in other parasite forms. *Tb*4EIP protein partners other than *Tb*EIF4E1 therefore remain to be identified.

According to the tethering results, *Tb*EIF4E1 is not a repressor and is unlikely to promote translation. These result should however be read with caution because of the artificial nature of the mRNA-protein interaction in a tethering assay. *Tb*EIF4E1 was tethered by its N-terminus to the 3'UTR of a reporter mRNA and it may not necessarily bind the mRNA cap in that state. In addition its cap binding affinity which is already 3 times worse than *Tb*EIF4E4/3 [148] may be further lowered on tethering. The alternative interpretation therefore is that *Tb*EIF4E1 is a repressor in the absence of *Tb*4EIP by merely occupying the cap, an effect that tethering may not report. However, the flexibility of mRNA could facilitate proximity and binding between the 3'UTR-tethered *Tb*EIF4E1 and the mRNA cap. Tethering of proteins to the 5'UTR is not possible since this would prevent movement of the translation pre-initiation complex. It is therefore still not known whether *Tb*EIF4E1 is a functional translation factor/repressor, or whether it is only relevant in the presence of *Tb*4EIP.

*Tb*EIF4E1 could be a translation factor if it were to bind *Tb*EIF3 like in *Leishmania*. *Leishmania Lm*EIF4E1 was found to interact with other translation factors including subunits of eIF2 and eIF3 in promastigotes [150]. In addition, the protein levels of *Lm*EIF4E1 were unchanged during heat shock (this induces differentiation to amastigotes) and in amastigotes, while those of *Lm*EIF4E4 and *Lm*EIF4G3 decreased in amastigotes. Together, it was predicted that *Lm*EIF4E1 could be a translation factor in amastigotes where *Lm*4EIP interaction is lacking [150]. A similar mechanism could therefore be speculated in trypanosomes where binding of *Tb*4EIP would prevent binding of eIF3 and dislodge *Tb*EIF4E1 from the mRNA cap to suppress translation. Functions of *Leishmania* and *Trypanosoma* 4EIP may however be different because they are regulated differently in different life stages [150].

Another possible mechanism could assume that *Tb*EIF4E1 is a repressor and binding of *Tb*4EIP to *Tb*EIF4E1 would free the cap and then enable a translationally active eIF4E to initiate translation. This would be consistent with a proposed model in *Leishmania* where binding of *Lm*4EIP decreased cap binding affinity of *Lm*EIF4E1 [151]. This however depicts *Tb*4EIP as a passive translational repressor and ignores its repressive role independent of *Tb*EIF4E1.

*Tb*4EIP could also act independent of *Tb*EIF4E1 to repress expression. It is not known where *Tb*4EIP binds its target mRNA and what its RNA binding domain is. The glutamine-proline rich C-terminus could facilitate RNA binding by ionic interactions. If it were to bind the 5'UTR, then *Tb*4EIP-*Tb*EIF4E1 complex could spatially block binding of a translationally active eIF4E or block scanning by the pre-initiation complex hence repress translation. However, the tethering

assay also showed that *Tb*4EIP is a repressor of mRNA, suggesting recruitment of mRNA degradation machineries in addition to repressing translation.

*Tb*4EIP interaction partners identified by tandem affinity purification (TAP) and yeast two hybrid screens could not be validated by co-immunoprecipitation with the exception of *Tb*EIF4E1. In addition to TAP purification, a single-step myc pull-down was done (by colleagues) and identified the terminal uridylyl transferase TUT3. I was able to validate the *Tb*4EIP and TUT3 interaction by co-immunoprecipitation. The amount co-purified even in the reciprocal pull-down was small suggesting that the interaction was very weak (Figure 6). This could be the reason why TUT3 was undetectable in the two-step TAP purifications. This result highlights the existence of weak interactions that are undetectable by standard two-step immunoprecipitation experiments that are often missed resulting in false negative results.

TUT3 function is unknown. It has been shown to catalyze uridylation *in vitro* [170]. Uridylated GPEET mRNA has been detected in early procyclic forms and is absent in late procyclic forms [168], consistent with the exclusive expression of GPEET in early procyclic forms. This uridylation could facilitate GPEET mRNA degradation similar to mammalian TUT4/7, which uridylate mRNA with short poly(A) tails, marking them for deadenylation/ decapping and degradation [167]. Uridylation could therefore be a signal for mRNA degradation in trypanosomes. Since TUT3 is neither a repressor nor an activator of expression (according to a genome-wide tethering screen [84]), and it does not bind poly(A) mRNA [78], it is conceivable that *Tb*4EIP recruits TUT3 to mRNA for uridylation that is followed by degradation. Uridylation by TUT3 could therefore form one of the mechanisms of *Tb*4EIP. TUT3 tethering experiment and *in vivo* studies on the effects of uridylation are needed to strengthen or disqualify this hypothesis. To test this hypothesis, the fate of GPEET mRNA could be analyzed in *Tb*4EIP KO cells, assuming that TUT3 is responsible for GPEET uridylation in early procyclic forms. The inability of *Tb*4EIP to repress expression in TUT3 KO would certainly support such a hypothesis.

*Tb*4EIP is not essential in bloodstream or procyclic forms but is required during transformation from the former to the latter. Activity of *Tb*4EIP however may not be exclusive during differentiation as growing cells experienced a mild growth defect (Figure 7). As recently shown together with my colleagues, *Tb*4EIP in growing bloodstream cells binds to unstable mRNAs that have low ribosome occupancy [152]; this points to a function in fine-tuning the transcriptome besides its role in differentiation. This function may however be dispensable because upon prolonged culture, cells lacking *Tb*4EIP seem to adapt and recover normal growth (Figure 7).

Nevertheless, *Tb*4EIP mutants were unable to repress translation during stumpy formation and this generated defective cells that could not differentiate to viable procyclic cells (Figure 8). In addition, over-expression of *Tb*4EIP increased longevity at high cell densities and speeded the expression of the stumpy marker PAD1, suggesting a protective role of *Tb*4EIP in stumpy forms. *Leishmania Lm*4EIP is also thought to be involved in stage-specific functions as it is expressed mainly in the insect stage promastigote forms [150]. Its interaction with *Lm*EIF4E1 is also restricted to promastigotes contrary to trypanosome *Tb*4EIP whose interaction with *Tb*EIF4E1 is present in growing forms of both mammals and insects (Figure 4).

4.2 PUF3 is required for normal differentiation

PUF3 was identified as a strong repressor in a genome-wide tethering screen and in a subsequent tethering screen of a mini-ORFeome [78,84]. It is conserved among trypanosomatids, has similar domain organization as its orthologs and is likely to have similar function to its orthologs. PUF proteins are generally involved in cell differentiation and organelle biogenesis [107], and they act by repressing mRNA stability or translation.

The results here suggest that PUF3 is required for normal differentiation. Knockout mutants exhibited a short delay in expression of differentiation markers PAD1, EP and GPEET. This defect was demonstrated in pleomorphic cells under PUF3 RNAi conditions but was conspicuously absent in PUF3 knockout pleomorphic cells. It could be argued that PUF3 RNAi was indiscriminate and would target other PUF mRNAs resulting in the observed defect. However, the defect was reproduced in monomorphic cells with *PUF3* knocked-out. In addition, the defect was rescued by adding back PUF3 to the knockouts therefore demonstrating that the defect was as a result of PUF3 deficiency. A similar phenotype has been reported in cells depleted of ZC3H18 [96]. In contrast to PUF3, the phenotype in ZC3H18 depleted pleomorphic cells was short-lived – detectable only up to 12 hours after induction of differentiation, and PAD1 expression was unaffected. The role of ZC3H18 in differentiation is unknown. ZC3H18, PUF3 and *Tb*4EIP represent a class of genes which are dispensable in growing cells but are specifically required in differentiation. These RNA binding proteins likely fine-tune the transcriptome/proteome in readiness for differentiation.

Putative protein partners of PUF3 could not be confidently identified using tandem affinity purifications. The identified ones were mostly ubiquitous proteins or proteins unrelated to

PUF3 role as a repressor. This could possibly be due to a general challenge to recover unstable degradation-associated complexes as compared to recovering stable ones: assembly of degradation machinery lead to rapid destruction of mRNA-protein complexes while stabilizing complexes persist to protect mRNA and these are more likely to be detected. Examples of other RBPs for which no RNA decay proteins could be identified using TAP purification included repressors such as RBP10 [45] and 4EIP [152]. On the other hand, stabilizing complexes involving MKT1 [91] were readily found. Some abundant targets e.g. translation factors could be falsely discarded as contaminants. This challenge could extend to identifying mRNA targets of repressors by RNA immunoprecipitation due to the rapid loss of targets.

Initial experiments on growth kinetics in pleomorphic bloodstream cells showed that PUF3 depletion by RNAi caused a slight growth defect, while PUF3 knockout did not. This could be attributed to an adaptation mechanism to life without PUF3 similar to that observed in pleomorphic bloodstream cells lacking *Tb*4EIP. Add-back of PUF3 similarly had no effect on the growth of pleomorphic bloodstream cells lacking PUF3. Interestingly, a growth defect was observed on forced expression of PUF3 add-back during differentiation to procyclic forms and further in late procyclic forms. The add-back gene was intended to be over-expressed, though this could not be verified for lack of an anti-PUF3 antibody and useful tagging options. Over-expression of PUF5 was also shown to impair growth of procyclic but not bloodstream cells, and the authors attributed this defect to disrupted gene regulation by non-specific binding of the excess PUF5 [101]. In addition, over-expression of ZC3H12 or ZC3H13 was also toxic in procyclic cells [94]. Over-expression in *T. brucei* has however been used in library screens and other applications with acceptable successes [35,45,84,173,182]. In a recent study, expression of RBP6 to form epimastigotes subsequently increased PUF5 transcripts. This would be consistent with the over-expression results of PUF5 that disrupts growth of procyclic forms – PUF5 could be involved in epimastigote biology. Since over-expression of PUF3 in bloodstream forms caused no growth defect (Figure 16), the defect observed in procyclic forms could not be entirely dismissed as non-specific. Procyclic forms lacking PUF3 showed a slight growth defect while no effect on cell growth was observed by PUF3 over-expression in bloodstream forms. These results suggest that PUF3 levels could be tightly regulated below some quota in procyclic forms.

4.3 Identifying targets of RBP using TRIBE

The standard method to identify targets of RNA binding proteins is by cross-linking followed by immunoprecipitation (CLIP) [183,184]. There are numerous variants of CLIP which improve various steps of the original method including cross-linking, cell lysis, RNA fragmentation, purification of protein-RNA complexes, RNA extraction and library preparations [185]. For example, in PAR-CLIP (Photoactivatable ribonucleoside-enhanced CLIP), cells are pre-incubated with 4-thiouridine (4SU) or 6-thioguanosine (6SG), which crosslink to bound protein when irradiated at a specific longer wavelength of 365 nm instead of 254 nm. This method has the highest resolution to date. However, longer incubation with the nucleosides can be toxic to cells, and RNA with many U and G residues would be preferentially crosslinked. Generally, CLIP requires a large amount of starting material which can be limited in the case of trypanosomes, which are approximately 100 times smaller than mammalian cells, and in most experiments cannot be grown to very high densities. The whole CLIP procedure is extremely lengthy and also has a bias of identifying long and highly expressed genes [186].

TRIBE (Targets of RNA-binding proteins Identified by Editing) is a method that could overcome many CLIP constraints. Targets of RNA binding proteins are permanently marked with novel RNA editing events that can be identified by Ranse and bioinformatics. The method does not require an antibody: it involves cloning of a fusion between ADAR (adenosine deaminase acting on RNA) and an RNA binding protein, standard library preparation and RNAseq, thus only a small amount of starting material is needed. This is especially useful in differentiation studies since many cells do not complete the differentiation process *in vitro* thus limiting the RNA yield as already seen in dense cells (Figure 8C). In terms of media required for cell growth, an RNA-immunoprecipitation experiment would need 3-6 litres whereas 20-50 ml are enough for TRIBE.

Using TRIBE, mRNA targets of PUF3 in cells treated with 10µM cAMP analog (8-pCPT-2'-O-Me-cAMP) were identified, the majority of which were enriched in slender bloodstream forms, consistent with PUF3 as a repressor of bloodstream genes during differentiation. However, because the cAMP analog generates cells that are not fully stumpy (stumpy-like) [35], this could result in a slender-bloodstream biased transcriptome. Nevertheless, a principle component analysis did not cluster the cAMP treated cells with slender bloodstream cells (Figure 21E).

There was no overlap between RNA-immunoprecipitation (RIP) and PUF3-TRIBE targets, perhaps because different parasite forms were used in each of the experiments. In addition, the pull-down in the RIP was not reliable, evidenced by the few reads that represented

precipitated RNAs, and the failure to detect the best candidates by RT-PCR. Further, no motifs were detected in the TRIBE targets, but this was also seen with *Tb*4EIP suggesting that other factors apart from the sequence could be involved in binding e.g. secondary structures or cooperation with other RBPs. However, the increased number of gene targets relative to the controls indicates that the TRIBE targets are not random but result from PUF3 directed editing.

The ADAR catalytic domain has also been exploited in other editing contexts where guide RNAs are used to direct precise editing of particular bases in RNA by ADAR [187]. Another setup involves RNA tagging, where an RBP is fused to *C. elegans* poly(U) polymerase, which adds U-tails on RNA. However, this method requires special library preparation procedures. In contrast, TRIBE follows standard library preparation procedures [159]. However, TRIBE too has some limitations including a false negative problem because ADAR has a preference for editing sites in double stranded regions even in the absence of its double-stranded RNA binding domain. In particular, it has a preference for an adenosine flanked by a 5' uridine and a 3' guanosine (U<u>A</u>G) in a double stranded RNA, and lacks processivity. This was evidenced by the many genes identified with only one edited site (Figure 21F). However, an improvement of TRIBE called HyperTRIBE has recently been used with promising outcomes [177]. HyperTRIBE uses an E488Q mutant of ADAR which has reduced specificity i.e. a hyperactive ADAR that catalyzes more edit events per binding site [177]. HyperTRIBE identified more targets and had a higher overlap with CLIP data indicating a reduced false-negative. I have embarked on using HyperTRIBE in trypanosomes. To enable subsequent validation, I use RBP10 and ZC3H11 whose targets are published, and in addition, PUF2 as a test protein. If successful, this method could be applied to other RBPs and enable venturing into novel stage-specific targets with unprecedented sensitivity.

5.0 Conclusion

The complex life cycle of trypanosomes requires robust mechanisms that modulate gene expression as the parasite differentiates and is transmitted from one host to another. In addition to ZFP1, ZFP2, ZFP3, ZC3H18, RBP10 and RBP6, PUF3 and 4EIP include the list of RNA binding proteins that modulate differentiation. PUF3 and 4EIP are both repressors of gene expression and possibly involved in fine-tuning the transcriptome/proteome of

differentiating cells. The role of PUF3 however appears dispensable as pleomorphic cells seem to adapt to life without it.

The results from PUF3-TRIBE depict a specific editing pattern by ADAR directed by PUF3. If validated, this method would open limitless horizons to easily identify targets of RBPs in trypanosomes.

6.0 References

1. Opperdoes FR, Borst P. Localization of nine glycolytic enzymes in a microbody-like organelle in Trypanosoma brucei: The glycosome. *FEBS Lett*. 1977;80(2):360-364. doi:10.1016/0014-5793(77)80476-6
2. Campbell DA, Thornton DA, Boothroyd JC. Apparent discontinuous transcription of trypanosoma brucei variant surface antigen genes. *Nature*. 1984;311(5984):350-355. doi:10.1038/311350a0
3. Michaeli S. Trans-splicing in trypanosomes: Machinery and its impact on the parasite transcriptome. *Future Microbiol*. 2011;6(4):459-474. doi:10.2217/fmb.11.20
4. Stevens JR, Noyes HA, Schofield CJ, Gibson W. The Molecular Evolution of Trypanosomatidae. *Adv Parasitol*. 2001;48:1-56.
5. Dollet M. Plant diseases caused by flagellate protozoa (Phytomonas). *Annu Rev Phytopathol*. 1984;22(1):115-132. doi:10.1016/B978-0-08-047378-9.50022-1
6. Franco JR, Cecchi G, Priotto G, et al. Monitoring the elimination of human African trypanosomiasis: Update to 2014. *PLoS Negl Trop Dis*. 2017;11(5):1-26. doi:10.1371/journal.pntd.0005585
7. Fèvre EM, Wissmann B V, Welburn SC, Lutumba P. The burden of human African trypanosomiasis. *PLoS Negl Trop Dis*. 2008;2(12):e333. doi:10.1371/journal.pntd.0000333
8. Welburn SC, Maudlin I, Simarro PP. Controlling sleeping sickness - a review. *Parasitology*. 2009;136(14):1943-1949. doi:10.1017/S0031182009006416
9. Welburn SC, Coleman PG, Maudlin I, Fèvre EM, Odiit M, Eisler MC. Crisis, what crisis? Control of Rhodesian sleeping sickness. *Trends Parasitol*. 2006;22(3):123-128. doi:10.1016/j.pt.2006.01.011
10. Yun O, Priotto G, Tong J, Flevaud L, Chappuis F. NECT is next: implementing the new drug combination therapy for Trypanosoma brucei gambiense sleeping sickness. *PLoS Negl Trop Dis*. 2010;4(5):e720. doi:10.1371/journal.pntd.0000720
11. Balasegaram M, Young H, Chappuis F, Priotto G, Raguenaud M-E, Checchi F. Effectiveness of melarsoprol and eflornithine as first-line regimens for gambiense sleeping sickness in nine Médecins Sans Frontières programmes. *Trans R Soc Trop Med Hyg*. 2009;103(3):280-290. doi:10.1016/j.trstmh.2008.09.005
12. Politi C, Carrín G, Evans D, Kuzoe FA, Cattand PD. Cost-effectiveness analysis of alternative treatments of African gambiense trypanosomiasis in Uganda. *Health Econ*.

1995;4(4):273-287.

13. Kennedy PGE. Diagnostic and neuropathogenesis issues in human African trypanosomiasis. *Int J Parasitol*. 2006;36(5):505-512. doi:10.1016/j.ijpara.2006.01.012
14. Connor RJ. The impact of nagana. *Onderstepoort J Vet Res*. 1994;61(4):379-383.
15. Holmes P. Tsetse-transmitted trypanosomes - Their biology, disease impact and control. *J Invertebr Pathol*. 2013;112 Suppl:S11-4. doi:10.1016/j.jip.2012.07.014
16. Borst P, Fase-Fowler F, Gibson WC. Kinetoplast DNA of Trypanosoma evansi. *Mol Biochem Parasitol*. 1987;23(1):31-38. doi:10.1016/0166-6851(87)90184-8
17. Mehlhorn H. Trypanosoma equiperdum. In: *Encyclopedia of Parasitology*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2016:2937-2939. doi:10.1007/978-3-662-43978-4_4456
18. Blattner J, Helfert S, Michels P, Clayton C. Compartmentation of phosphoglycerate kinase in Trypanosoma brucei plays a critical role in parasite energy metabolism. *Proc Natl Acad Sci U S A*. 1998;95(20):11596-11600. doi:10.1073/PNAS.95.20.11596
19. Vickerman K. Antigenic variation in trypanosomes. *Nature*. 1978;273(5664):613-617. doi:10.1038/273613a0
20. Barbour AG, Restrepo BI. Antigenic variation in vector-borne pathogens. *Emerg Infect Dis*. 2000;6(5):449-457. doi:10.3201/eid0605.000502
21. Horn D. Antigenic variation in African trypanosomes. *Mol Biochem Parasitol*. 2014;195(2):123-129. doi:10.1016/j.molbiopara.2014.05.001
22. Vassella E, Reuner B, Yutzy B, Boshart M. Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. *J Cell Sci*. 1997;110:2661-2671.
23. Rico E, Rojas F, Mony BM, Szoor B, Macgregor P, Matthews KR. Bloodstream form pre-adaptation to the tsetse fly in Trypanosoma brucei. *Front Cell Infect Microbiol*. 2013;3:78. doi:10.3389/fcimb.2013.00078
24. Fenn K, Matthews KR. The cell biology of Trypanosoma brucei differentiation. *Curr Opin Microbiol*. 2007;10(6):539-546. doi:10.1016/j.mib.2007.09.014
25. Silvester E, McWilliam K, Matthews K. The Cytological Events and Molecular Control of Life Cycle Development of Trypanosoma brucei in the Mammalian Bloodstream. *Pathogens*. 2017;6(3):29. doi:10.3390/pathogens6030029
26. Nolan DP, Rolin S, Rodriguez JR, Van Den Abbeele J, Pays E. Slender and stumpy bloodstream forms of Trypanosoma brucei display a differential response to extracellular acidic and proteolytic stress. *Eur J Biochem*. 2000;267(1):18-27.

doi:10.1046/j.1432-1327.2000.00935.x

27. Dean S, Marchetti R, Kirk K, Matthews KR. A surface transporter family conveys the trypanosome differentiation signal. *Nature*. 2009;459(7244):213-217. doi:10.1038/nature07997
28. Shi H, Ramey-butler K, Tschudi C. A single-point mutation in the RNA-binding protein 6 generates Trypanosoma brucei metacyclics that are able to progress to bloodstream forms in vitro. *Mol Biochem Parasitol*. 2018. doi:10.1016/j.molbiopara.2018.07.011
29. Szöor B, Ruberto I, Burchmore R, Matthews KR. A novel phosphatase cascade regulates differentiation in Trypanosoma brucei via a glycosomal signaling pathway. *Genes Dev*. 2010;24(12):1306-1316. doi:10.1101/gad.570310
30. Vanhollebeke B, Uzureau P, Monteyne D, Pérez-Morga D, Pays E. Cellular and molecular remodeling of the endocytic pathway during differentiation of Trypanosoma brucei bloodstream forms. *Eukaryot Cell*. 2010;9(8):1272-1282. doi:10.1128/EC.00076-10
31. Vassella E, Krämer R, Turner CMR, et al. Deletion of a novel protein kinase with PX and FYVE-related domains increases the rate of differentiation of Trypanosoma brucei. *Mol Microbiol*. 2001;41(1):33-46. doi:10.1046/j.1365-2958.2001.02471.x
32. Dejung M, Subota I, Bucerius F, et al. Quantitative Proteomics Uncovers Novel Factors Involved in Developmental Differentiation of Trypanosoma brucei. *PLoS Pathog*. 2016;12(2):1-20. doi:10.1371/journal.ppat.1005439
33. Laxman S, Riechers A, Sadilek M, Schwede F, Beavo JA. Hydrolysis products of cAMP analogs cause transformation of Trypanosoma brucei from slender to stumpy-like forms. *Proc Natl Acad Sci U S A*. 2006;103(50):19194-19199. doi:10.1073/pnas.0608971103
34. Mony BM, MacGregor P, Ivens A, et al. Genome-wide dissection of the quorum sensing signalling pathway in Trypanosoma brucei. *Nature*. 2014;505(7485):681-685. doi:10.1038/nature12864
35. MacGregor P, Matthews KR. Identification of the regulatory elements controlling the transmission stage-specific gene expression of PAD1 in Trypanosoma brucei. *Nucleic Acids Res*. 2012;40(16):7705-7717. doi:10.1093/nar/gks533
36. Batram C, Jones NG, Janzen CJ, Markert SM, Engstler M. Expression site attenuation mechanistically links antigenic variation and development in Trypanosoma brucei. *Elife*. 2014;2014(3):2324. doi:10.7554/eLife.02324
37. Besteiro S, Barrett MP, Rivière L, Bringaud F. Energy generation in insect stages of

Trypanosoma brucei: metabolism in flux. *Trends Parasitol*. 2005;21(4):185-191. doi:10.1016/J.PT.2005.02.008

38. Hao Z, Kasumba I, Lehane MJ, Gibson WC, Kwon J, Aksoy S. Tsetse immune responses and trypanosome transmission: implications for the development of tsetse-based strategies to reduce trypanosomiasis. *Proc Natl Acad Sci U S A*. 2001;98(22):12648-12653. doi:10.1073/pnas.221363798
39. Vassella E, Van Den Abbeele J, Bütikofer P, et al. A major surface glycoprotein of Trypanosoma brucei is expressed transiently during development and can be regulated post-transcriptionally by glycerol or hypoxia. *Genes Dev*. 2000;14(5):615-626. doi:10.1101/gad.14.5.615
40. Gibson W, Bailey M. The development of Trypanosoma brucei within the tsetse fly midgut observed using green fluorescent trypanosomes. *Kinetoplastid Biol Dis*. 2003;2(1):1.
41. Czichos J, Nonnengaesser C, Overath P. Trypanosoma brucei: cis-Aconitate and temperature reduction as triggers of synchronous transformation of bloodstream to procyclic trypomastigotes in vitro. *Exp Parasitol*. 1986;62(2):283-291. doi:10.1016/0014-4894(86)90033-0
42. Bass KE, Wang CC. The in vitro differentiation of pleomorphic Trypanosoma brucei from bloodstream into procyclic form requires neither intermediary nor short-stumpy stage. *Mol Biochem Parasitol*. 1991;44(2):261-270. doi:10.1016/0166-6851(91)90012-U
43. Vassella E, Probst M, Schneider A, Studer E, Kunz Renggli C, Roditi I. Expression of a Major Surface Protein of Trypanosoma brucei Insect Forms Is Controlled by the Activity of Mitochondrial Enzymes. *Mol Biol Cell*. 2004;15:3986-3993. doi:10.1091/mbc.E04
44. Schumann Burkard G, Käser S, de Araújo PR, et al. Nucleolar proteins regulate stage-specific gene expression and ribosomal RNA maturation in Trypanosoma brucei. *Mol Microbiol*. 2013;88(4):827-840. doi:10.1111/mmi.12227
45. Mugo E, Clayton C. Expression of the RNA-binding protein RBP10 promotes the bloodstream-form differentiation state in Trypanosoma brucei. *PLoS Pathog*. 2017;13(8):e1006560. doi:10.1371/journal.ppat.1006560
46. Vassella E, Boshart M. High molecular mass agarose matrix supports growth of bloodstream forms of pleomorphic Trypanosoma brucei strains in axenic culture. *Mol Biochem Parasitol*. 1996;82(1):91-105. doi:10.1016/0166-6851(96)02727-2
47. Urwyler S, Studer E, Renggli CK, Roditi I. A family of stage-specific alanine-rich proteins on the surface of epimastigote forms of Trypanosoma brucei. *Mol Microbiol*.

2007;63(1):218-228. doi:10.1111/j.1365-2958.2006.05492.x

48. Nolan DP, Jackson DG, Biggs MJ, et al. Characterization of a novel alanine-rich protein located in surface microdomains in Trypanosoma brucei. *J Biol Chem*. 2000;275(6):4072-4080. doi:10.1074/JBC.275.6.4072
49. VanDenAbbeele J, Claes Y, VanBockstaele D, LeRay D, Coosemans M. Trypanosoma brucei spp. developmet in the tsetse fly: characterization of teh post-mesocyclic stages int eh foregut and proboscis. *Parasitol Today*. 1999;118:416-478.
50. Kolev NG, Ramey-Butler K, Cross GAM, Ullu E, Tschudi C. Developmental progression to infectivity in Trypanosoma brucei triggered by an RNA-binding protein. *Science (80)*. 2012;338(6112):1352-1353. doi:10.1126/science.1229641
51. Figueiredo RC, Rosa DS, Soares MJ. Differentiation of Trypanosoma cruzi epimastigotes: metacyclogenesis and adhesion to substrate are triggered by nutritional stress. *J Parasitol*. 2000;86(6):1213-1218. doi:10.1645/0022-3395(2000)086[1213:DOTCEM]2.0.CO;2
52. Romaniuk MA, Frasch AC, Cassola A. Translational repression by an RNA-binding protein promotes differentiation to infective forms in Trypanosoma cruzi. Goldenberg S, ed. *PLoS Pathog*. 2018;14(6):e1007059. doi:10.1371/journal.ppat.1007059
53. Trindade S, Rijo-Ferreira F, Carvalho T, et al. Trypanosoma brucei Parasites Occupy and Functionally Adapt to the Adipose Tissue in Mice. *Cell Host Microbe*. 2016;19(6):837-848. doi:10.1016/j.chom.2016.05.002
54. Mair G, Shi H, Li H, et al. A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. *RNA*. 2000;6(2):163-169.
55. Krause M, Hirsh D. A trans-spliced leader sequence on actin mRNA in C. elegans. *Cell*. 1987;49(6):753-761. doi:10.1016/0092-8674(87)90613-1
56. Rajkovic A, Davis RE, Simonsen JN, Rottman FM. A spliced leader is present on a subset of mRNAs from the human parasite Schistosoma mansoni. *Proc Natl Acad Sci U S A*. 1990;87(22):8879-8883. doi:2247461
57. Tessier LH, Keller M, Chan RL, Fournier R, Weil JH, Imbault P. Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by trans-splicing in Euglena. *EMBO J*. 1991;10(9):2621-2625. doi:10.1002/J.1460-2075.1991.TB07804.X
58. Vandenberghe AE, Meedel TH, Hastings KE. mRNA 5'-leader trans-splicing in the chordates. *Genes Dev*. 2001;15(3):294-303. doi:10.1101/gad.865401
59. Zhang H, Hou Y, Miranda L, et al. Spliced leader RNA trans-splicing in dinoflagellates. *Proc Natl Acad Sci*. 2007;104(11):4618-4623. doi:10.1073/pnas.0700258104

60. Gilinger G, Bellofatto V. Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. *Nucleic Acids Res*. 2001;29(7):1556-1564. doi:10.1093/nar/29.7.1556
61. Siegel TN, Hekstra DR, Kemp LE, et al. Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei. *Genes Dev*. 2009;23(9):1063-1076. doi:10.1101/gad.1790409
62. Wedel C, Förstner KU, Derr R, Siegel TN. GT-rich promoters can drive RNA pol II transcription and deposition of H2A.Z in African trypanosomes. 2017:1-14. doi:10.15252/embj
63. Manful T, Fadda A, Clayton C. The role of the 5′-3′ exoribonuclease XRNA in transcriptome-wide mRNA degradation. *RNA*. 2011;17(11):2039-2047. doi:10.1261/rna.2837311
64. Ouellette M, Papadopoulou B. Coordinated gene expression by post-transcriptional regulons in African trypanosomes. *J Biol*. 2009;8(11):100. doi:10.1186/jbiol203
65. Siegel TN, Hekstra DR, Wang X, Dewell S, Cross GAM. Genome-wide analysis of mRNA abundance in two life-cycle stages of Trypanosoma brucei and identification of splicing and polyadenylation sites. *Nucleic Acids Res*. 2010;38(15):4946-4957. doi:10.1093/nar/gkq237
66. Nilsson D, Gunasekera K, Mani J, et al. Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of Trypanosoma brucei. Parsons M, ed. *PLoS Pathog*. 2010;6(8):21-22. doi:10.1371/journal.ppat.1001037
67. Kolev NG, Franklin JB, Carmi S, Shi H, Michaeli S, Tschudi C. The Transcriptome of the Human Pathogen Trypanosoma brucei at Single-Nucleotide Resolution. Beverley SM, ed. *PLoS Pathog*. 2010;6(9):e1001090. doi:10.1371/journal.ppat.1001090
68. Preußer C, Jaé N, Bindereif A. mRNA splicing in trypanosomes. *Int J Med Microbiol*. 2012;302(4-5):221-224. doi:10.1016/j.ijmm.2012.07.004
69. Fadda A, Ryten M, Droll D, et al. Transcriptome-wide analysis of trypanosome mRNA decay reveals complex degradation kinetics and suggests a role for co-transcriptional degradation in determining mRNA levels. *Mol Microbiol*. 2014;94(2):307-326. doi:10.1111/mmi.12764
70. Schwede A, Ellis L, Luther J, Carrington M, Stoecklin G, Clayton C. A role for Caf1 in mRNA deadenylation and decay in trypanosomes and human cells. *Nucleic Acids Res*. 2008;36(10):3374-3388. doi:10.1093/nar/gkn108

71. Färber V, Erben E, Sharma S, Stoecklin G, Clayton C. Trypanosome CNOT10 is essential for the integrity of the NOT deadenylase complex and for degradation of many mRNAs. *Nucleic Acids Res*. 2013;41(2):1211-1222. doi:10.1093/nar/gks1133
72. Kramer S. The ApaH-like phosphatase TbALPH1 is the major mRNA decapping enzyme of trypanosomes. Hill KL, ed. *PLoS Pathog*. 2017;13(6):e1006456. doi:10.1371/journal.ppat.1006456
73. Haile S, Estévez AM, Clayton C. A role for the exosome in the in vivo degradation of unstable mRNAs A role for the exosome in the in vivo degradation of unstable mRNAs. 2003:1491-1501. doi:10.1261/rna.5940703.Parker
74. De Gaudenzi JG, Carmona SJ, Agüero F, Frasch AC. Genome-wide analysis of 3'-untranslated regions supports the existence of post-transcriptional regulons controlling gene expression in trypanosomes. *PeerJ*. 2013;1:e118. doi:10.7717/peerj.118
75. Archer SK, Luu V, Queiroz RA De, Brems S, Clayton C. Trypanosoma brucei PUF9 Regulates mRNAs for Proteins Involved in Replicative Processes over the Cell Cycle. *PLoS Pathog*. 2009;5(8). doi:10.1371/journal.ppat.1000565
76. Das A, Bellofatto V, Rosenfeld J, et al. High throughput sequencing analysis of Trypanosoma brucei DRBD3/PTB1-bound mRNAs. *Mol Biochem Parasitol*. 2015;199(1-2):1-4. doi:10.1016/j.molbiopara.2015.02.003
77. Lee EK, Gorospe M. Coding region: the neglected post-transcriptional code. *RNA Biol*. 2011;8(1):44-48. doi:10.4161/RNA.8.1.13863
78. Lueong S, Merce C, Fischer B, Hoheisel JD, Erben ED. Gene expression regulatory networks in Trypanosoma brucei: Insights into the role of the mRNA-binding proteome. *Mol Microbiol*. 2016;100(3):457-471. doi:10.1111/mmi.13328
79. D'Orso I, De Gaudenzi JG, Frasch ACC. RNA-binding proteins and mRNA turnover in trypanosomes. *Trends Parasitol*. 2003;19(4):151-155.
80. Fernández-Moya SM, Estévez AM. Posttranscriptional control and the role of RNA-binding proteins in gene regulation in trypanosomatid protozoan parasites. *Wiley Interdiscip Rev RNA*. 2010;1(1):34-46. doi:10.1002/wrna.6
81. Clayton C. The Regulation of Trypanosome Gene Expression by RNA-Binding Proteins. *PLoS Pathog*. 2013;9(11):9-12. doi:10.1371/journal.ppat.1003680
82. Baron-Benhamou J, Gehring NH, Kulozik AE, Hentze MW. Using the λN Peptide to Tether Proteins to RNAs. In: *MRNA Processing and Metabolism*. New Jersey: Humana Press; 2004:135-154. doi:10.1385/1-59259-750-5:135
83. Keryer-Bibens C, Barreau C, Osborne HB. Tethering of proteins to RNAs by

bacteriophage proteins. *Biol Cell*. 2008;100(2):125-138. doi:10.1042/BC20070067

84. Erben ED, Fadda A, Lueong S, Hoheisel JD, Clayton C. A genome-wide tethering screen reveals novel potential post-transcriptional regulators in Trypanosoma brucei. *PLoS Pathog*. 2014;10(6):e1004178. doi:10.1371/journal.ppat.1004178
85. Kramer S, Bannerman-Chukualim B, Ellis L, et al. Differential Localization of the Two T. brucei Poly(A) Binding Proteins to the Nucleus and RNP Granules Suggests Binding to Distinct mRNA Pools. *PLoS One*. 2013;8(1):e54004. doi:10.1371/journal.pone.0054004
86. Hartmann C, Benz C, Brems S, et al. Small trypanosome RNA-binding proteins TbUBP1 and TbUBP2 influence expression of F-box protein mRNAs in bloodstream trypanosomes. *Eukaryot Cell*. 2007;6(11):1964-1978. doi:10.1128/EC.00279-07
87. Das A, Morales R, Banday M, et al. The essential polysome-associated RNA-binding protein RBP42 targets mRNAs involved in Trypanosoma brucei energy metabolism. *RNA*. 2012;18(11):1968-1983. doi:10.1261/rna.033829.112
88. Stern MZ, Gupta SK, Salmon-Divon M, et al. Multiple roles for polypyrimidine tract binding (PTB) proteins in trypanosome RNA metabolism. *RNA*. 2009;15(4):648-665. doi:10.1261/rna.1230209
89. Mani J, Güttinger A, Schimanski B, et al. Alba-Domain Proteins of Trypanosoma brucei Are Cytoplasmic RNA-Binding Proteins That Interact with the Translation Machinery. El-Sayed NM, ed. *PLoS One*. 2011;6(7):e22463. doi:10.1371/journal.pone.0022463
90. Gupta SK, Kosti I, Plaut G, et al. The hnRNP F/H homologue of Trypanosoma brucei is differentially expressed in the two life cycle stages of the parasite and regulates splicing and mRNA stability. *Nucleic Acids Res*. 2013;41(13):6577-6594. doi:10.1093/nar/gkt369
91. Singh A, Minia I, Droll D, Fadda A, Clayton C, Erben E. Trypanosome MKT1 and the RNA-binding protein ZC3H11: interactions and potential roles in post-transcriptional regulatory networks. *Nucleic Acids Res*. 2014;42(7):4652-4668. doi:10.1093/nar/gkt1416
92. Droll D, Minia I, Fadda A, et al. Post-transcriptional regulation of the trypanosome heat shock response by a zinc finger protein. *PLoS Pathog*. 2013;9(4):e1003286. doi:10.1371/journal.ppat.1003286
93. Minia I, Clayton C. Regulating a Post-Transcriptional Regulator: Protein Phosphorylation, Degradation and Translational Blockage in Control of the Trypanosome Stress-Response RNA-Binding Protein ZC3H11. Horn D, ed. *PLoS Pathog*. 2016;12(3):e1005514. doi:10.1371/journal.ppat.1005514

94. Ouna BA, Stewart M, Helbig C, Clayton C. The Trypanosoma brucei CCCH zinc finger proteins ZC3H12 and ZC3H13. *Mol Biochem Parasitol*. 2012;183(2):184-188. doi:10.1016/J.MOLBIOPARA.2012.02.006
95. Ling AS, Trotter JR, Hendriks EF. A zinc finger protein, TbZC3H20, stabilizes two developmentally regulated mRNAs in trypanosomes. *J Biol Chem*. 2011;286(23):20152-20162. doi:10.1074/jbc.M110.139261
96. Benz C, Mulindwa J, Ouna B, Clayton C. The Trypanosoma brucei zinc finger protein ZC3H18 is involved in differentiation. *Mol Biochem Parasitol*. 2011;177(2):148-151. doi:10.1016/j.molbiopara.2011.02.007
97. Hendriks EF, Robinson DR, Hinkins M, Matthews KR. A novel CCCH protein which modulates differentiation of Trypanosoma brucei to its procyclic form. *EMBO J*. 2001;20(23):6700-6711. doi:10.1093/emboj/20.23.6700
98. Walrad PB, Capewell P, Fenn K, Matthews KR. The post-transcriptional trans-acting regulator, TbZFP3, co-ordinates transmission-stage enriched mRNAs in Trypanosoma brucei. *Nucleic Acids Res*. 2012;40(7):2869-2883. doi:10.1093/nar/gkr1106
99. Luu V, Brems S, Hoheisel D, Burchmore R, Guilbride DL, Clayton C. Functional analysis of Trypanosoma brucei PUF1. *Mol Biochem Parasitol*. 2006;150:340-349. doi:10.1016/j.molbiopara.2006.09.007
100. Jha BA, Fadda A, Merce C, Mugo E, Droll D, Clayton C. Depletion of the Trypanosome Pumilio domain protein PUF2 or of some other essential proteins causes transcriptome changes related to coding region length. *Eukaryot Cell*. 2014;13(5):664-674. doi:10.1128/EC.00018-14
101. Jha BA, Archer SK, Clayton CE. The Trypanosome Pumilio Domain Protein PUF5. Li Z, ed. *PLoS One*. 2013;8(10):e77371. doi:10.1371/journal.pone.0077371
102. Droll D, Archer S, Fenn K, Delhi P, Matthews K, Clayton C. The trypanosome Pumilio-domain protein PUF7 associates with a nuclear cyclophilin and is involved in ribosomal RNA maturation. *FEBS Lett*. 2010;584(6):1156-1162. doi:10.1016/j.febslet.2010.02.018
103. Goyal M, Banerjee C, Nag S, Bandyopadhyay U. The Alba protein family: Structure and function. *Biochim Biophys Acta - Proteins Proteomics*. 2016;1864(5):570-583. doi:10.1016/j.bbapap.2016.02.015
104. Lu G, Hall TMT. Alternate modes of cognate RNA recognition by human PUMILIO proteins. *Structure*. 2011;19(3):361-367. doi:10.1016/j.str.2010.12.019
105. Zhao Y-Y, Mao M-W, Zhang W-J, et al. Expanding RNA binding specificity and affinity of engineered PUF domains. *Nucleic Acids Res*. 2018;46(9):4771-4782.

doi:10.1093/nar/gky134

106. Wang M, Ogé L, Perez-Garcia MD, Hamama L, Sakr S. The PUF protein family: Overview on PUF RNA targets, biological functions, and post transcriptional regulation. *Int J Mol Sci*. 2018;19(2). doi:10.3390/ijms19020410
107. Miller MA, Olivas WM. Roles of Puf proteins in mRNA degradation and translation. *Wiley Interdiscip Rev RNA*. 2011;2(4):471-492. doi:10.1002/wrna.69
108. Kaye JA, Rose NC, Goldsworthy B, Goga A, L'Etoile ND. A 3′UTR Pumilio-Binding Element Directs Translational Activation in Olfactory Sensory Neurons. *Neuron*. 2009;61(1):57-70. doi:10.1016/j.neuron.2008.11.012
109. Forbes a, Lehmann R. Nanos and Pumilio have critical roles in the development and function of Drosophila germline stem cells. *Development*. 1998;125(4):679-690.
110. Zhang B, Gallegos M, Puoti A, et al. A conserved RNA-binding protein that regulates sexual fates in the C. elegans hermaphrodite germ line. *Nature*. 1997;390(6659):477-484. doi:10.1038/37297
111. Moore FL, Jaruzelska J, Fox MS, et al. Human Pumilio-2 is expressed in embryonic stem cells and germ cells and interacts with DAZ (Deleted in AZoospermia) and DAZ-Like proteins. *Proc Natl Acad Sci*. 2003;100(2):538-543. doi:10.1073/pnas.0234478100
112. Furuichi Y. Discovery of m(7)G-cap in eukaryotic mRNAs. *Proc Jpn Acad Ser B Phys Biol Sci*. 2015;91(8):394-409. doi:10.2183/pjab.91.394
113. Jackson RJ, Hellen CUT, Pestova T V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol*. 2010;11(2):113-127. doi:10.1038/nrm2838
114. Kahvejian A, Roy G, Sonenberg N. The mRNA closed-loop model: The function of PABP and PABP-interacting proteins in mRNA translation. In: *Cold Spring Harbor Symposia on Quantitative Biology*. Vol 66. Cold Spring Harbor Laboratory Press; 2001:293-300. doi:10.1101/sqb.2001.66.293
115. Hernández G, Proud CG, Preiss T, Parsyan A. On the diversification of the translation apparatus across eukaryotes. *Comp Funct Genomics*. 2012;2012. doi:10.1155/2012/256848
116. Morita M, Ler LW, Fabian MR, et al. A Novel 4EHP-GIGYF2 Translational Repressor Complex Is Essential for Mammalian Development. *Mol Cell Biol*. 2012;32(17):3585-3593. doi:10.1128/MCB.00455-12
117. Tao X, Gao G. Tristetraprolin Recruits Eukaryotic Initiation Factor 4E2 To Repress Translation of AU-Rich Element-Containing mRNAs. *Mol Cell Biol*. 2015;35(22):3921-

3932. doi:10.1128/MCB.00845-15

118. Fu R, Olsen MT, Webb K, Bennett EJ, Lykke-Andersen J. Recruitment of the 4EHP-GYF2 cap-binding complex to tetraproline motifs of tristetraprolin promotes repression and degradation of mRNAs with AU-rich elements. *RNA*. 2016;22(3):373-382. doi:10.1261/rna.054833.115
119. Cho PF, Poulin F, Cho-Park YA, et al. A new paradigm for translational control: Inhibition via 5′-3′ mRNA tethering by Bicoid and the eIF4E cognate 4EHP. *Cell*. 2005;121(3):411-423. doi:10.1016/j.cell.2005.02.024
120. Cho PF, Gamberi C, Cho-Park YA, Cho-Park IB, Lasko P, Sonenberg N. Cap-Dependent Translational Inhibition Establishes Two Opposing Morphogen Gradients in Drosophila Embryos. *Curr Biol*. 2006;16(20):2035-2041. doi:10.1016/j.cub.2006.08.093
121. Bush MS, Hutchins AP, Jones AME, et al. Selective recruitment of proteins to 5′ cap complexes during the growth cycle in Arabidopsis. *Plant J*. 2009;59(3):400-412. doi:10.1111/j.1365-313X.2009.03882.x
122. Ruud KA, Kuhlow C, Goss DJ, Browning KS. Identification and characterization of a novel cap-binding protein from Arabidopsis thaliana. *J Biol Chem*. 1998;273(17):10325-10330. doi:10.1074/jbc.273.17.10325
123. Dinkova TD, Keiper BD, Korneeva NL, Aamodt EJ, Rhoads RE. Translation of a small subset of Caenorhabditis elegans mRNAs is dependent on a specific eukaryotic translation initiation factor 4E isoform. *Mol Cell Biol*. 2005;25(1):100-113. doi:10.1128/MCB.25.1.100-113.2005
124. Costello J, Castelli LM, Rowe W, et al. Global mRNA selection mechanisms for translation initiation. *Genome Biol*. 2015;16(1):10. doi:10.1186/s13059-014-0559-z
125. Kamenska A, Simpson C, Standart N. eIF4E-binding proteins: new factors, new locations, new roles. *Biochem Soc Trans*. 2014;42(4):1238-1245. doi:10.1042/BST20140063
126. Mader S, Lee H, Pause A, Sonenberg N. The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4 gamma and the translational repressors 4E-binding proteins. *Mol Cell Biol*. 1995;15(9):4990-4997. doi:10.1128/MCB.15.9.4990
127. Kinkelin K, Veith K, Grünwald M, Bono F. Crystal structure of a minimal eIF4E-Cup complex reveals a general mechanism of eIF4E regulation in translational repression. *RNA*. 2012;18(9):1624-1634. doi:10.1261/rna.033639.112
128. Umenaga Y, Paku KS, In Y, Ishida T, Tomoo K. Identification and function of the

second eIF4E-binding region in N-terminal domain of eIF4G: Comparison with eIF4E-binding protein. *Biochem Biophys Res Commun*. 2011;414(3):462-467. doi:10.1016/j.bbrc.2011.09.084

129. Bah A, Vernon RM, Siddiqui Z, et al. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature*. 2015;519(7541):106-109. doi:10.1038/nature13999
130. Sekiyama N, Arthanari H, Papadopoulos E, Rodriguez-Mias RA, Wagner G, Léger-Abraham M. Molecular mechanism of the dual activity of 4EGI-1: Dissociating eIF4G from eIF4E but stabilizing the binding of unphosphorylated 4E-BP1. *Proc Natl Acad Sci*. 2015;112(30):E4036-E4045. doi:10.1073/pnas.1512118112
131. Thoreen CC, Chantranupong L, Keys HR, Wang T, Gray NS, Sabatini DM. A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature*. 2012;485(7396):109-113. doi:10.1038/nature11083
132. Thoreen CC, Kang SA, Chang JW, et al. An ATP-competitive mammalian target of rapamycin inhibitor reveals rapamycin-resistant functions of mTORC1. *J Biol Chem*. 2009;284(12):8023-8032. doi:10.1074/jbc.M900301200
133. Nakamura A, Sato K, Hanyu-Nakamura K. Drosophila Cup Is an eIF4E Binding Protein that Associates with Bruno and Regulates oskar mRNA Translation in Oogenesis. *Dev Cell*. 2004;6(1):69-78. doi:10.1016/S1534-5807(03)00400-3
134. Smibert CA, Wilson JE, Kerr K, Macdonald PM. smaug protein represses translation of unlocalized nanos mRNA in the Drosophila embryo. *Genes Dev*. 1996;10(20):2600-2609. doi:10.1101/GAD.10.20.2600
135. Chekulaeva M, Hentze MW, Ephrussi A. Bruno Acts as a Dual Repressor of oskar Translation, Promoting mRNA Oligomerization and Formation of Silencing Particles. *Cell*. 2006;124(3):521-533. doi:10.1016/J.CELL.2006.01.031
136. Gosselin P, Martineau Y, Morales J, et al. Tracking a refined eIF4E-binding motif reveals Angel1 as a new partner of eIF4E. *Nucleic Acids Res*. 2013;41(16):7783-7792. doi:10.1093/nar/gkt569
137. Hernández G, Miron M, Han H, et al. Mextli is a novel eukaryotic translation initiation factor 4E-binding protein that promotes translation in Drosophila melanogaster. *Mol Cell Biol*. 2013;33(15):2854-2864. doi:10.1128/MCB.01354-12
138. Zanchin NI, McCarthy JE. Characterization of the in vivo phosphorylation sites of the mRNA.cap-binding complex proteins eukaryotic initiation factor-4E and p20 in Saccharomyces cerevisiae. *J Biol Chem*. 1995;270(44):26505-26510.

doi:10.1074/JBC.270.44.26505

139. Ferraiuolo MA, Basak S, Dostie J, Murray EL, Schoenberg DR, Sonenberg N. A role for the eIF4E-binding protein 4E-T in P-body formation and mRNA decay. *J Cell Biol*. 2005;170(6):913-924. doi:10.1083/jcb.200504039
140. Kamenska A, Lu WT, Kubacka D, et al. Human 4E-T represses translation of bound mRNAs and enhances microRNA-mediated silencing. *Nucleic Acids Res*. 2014;42(5):3298-3313. doi:10.1093/nar/gkt1265
141. Peter D, Weber R, Sandmeir F, et al. GIGYF1/2 proteins use auxiliary sequences to selectively bind to 4EHP and repress target mRNA expression. *Genes Dev*. 2017;31(11):1147-1161. doi:10.1101/gad.299420.117
142. Amaya Ramirez CC, Hubbe P, Mandel N, Béthune J. 4EHP-independent repression of endogenous mRNAs by the RNA-binding protein GIGYF2. *Nucleic Acids Res*. 2018;46(11):5792-5808. doi:10.1093/nar/gky198
143. Schopp IM, Amaya Ramirez CC, Debeljak J, et al. Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes. *Nat Commun*. 2017;8:15690. doi:10.1038/ncomms15690
144. Freire ER, Dhalia R, Moura DMN, et al. The four trypanosomatid eIF4E homologues fall into two separate groups, with distinct features in primary sequence and biological properties. *Mol Biochem Parasitol*. 2011;176(1):25-36. doi:10.1016/j.molbiopara.2010.11.011
145. Dhalia R, Reis CRS, Freire ER, et al. Translation initiation in Leishmania major: Characterisation of multiple eIF4F subunit homologues. *Mol Biochem Parasitol*. 2005;140(1):23-41. doi:10.1016/j.molbiopara.2004.12.001
146. Dhalia R, Marinsek N, Reis CRS, et al. The two eIF4A helicases in Trypanosoma brucei are functionally distinct. *Nucleic Acids Res*. 2006;34(9):2495-2507. doi:10.1093/nar/gkl290
147. da Costa Lima TD, Moura DMN, Reis CRS, et al. Functional characterization of three Leishmania poly(A) binding protein homologues with distinct binding properties to RNA and protein partners. *Eukaryot Cell*. 2010;9(10):1484-1494. doi:10.1128/EC.00148-10
148. Freire E, Sturm N, Campbell D, de Melo Neto O. The Role of Cytoplasmic mRNA Cap-Binding Protein Complexes in Trypanosoma brucei and Other Trypanosomatids. *Pathogens*. 2017;6(4):55. doi:10.3390/pathogens6040055
149. Freire ER, Moura DMN, Bezerra MJR, et al. Trypanosoma brucei EIF4E2 cap-binding protein binds a homolog of the histone-mRNA stem-loop-binding protein. *Current*

Genetics.

150. Zinoviev A, Léger M, Wagner G, Shapira M. A novel 4E-interacting protein in Leishmania is involved in stage-specific translation pathways. *Nucleic Acids Res*. 2011;39(19):8404-8415. doi:10.1093/nar/gkr555
151. Meleppattu S, Arthanari H, Zinoviev A, et al. Structural basis for LeishIF4E-1 modulation by an interacting protein in the human parasite Leishmania major. *Nucleic Acids Res*. 2018;46(7):3791-3801. doi:10.1093/nar/gky194
152. Terrao M, Marucha KK, Mugo E, et al. The suppressive cap-binding complex factor 4EIP is required for normal differentiation. *Nucleic Acids Res*. August 2018. doi:10.1093/nar/gky733
153. MacGregor P, Rojas F, Dean S, Matthews KR. Stable transformation of pleomorphic bloodstream form Trypanosoma brucei. *Mol Biochem Parasitol*. 2013;190(2):60-62. doi:10.1016/j.molbiopara.2013.06.007
154. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. doi:10.1186/s13059-014-0550-8
155. Bailey TL. DREME: Motif discovery in transcription factor ChIP-seq data. *Bioinformatics*. 2011;27(12):1653-1659. doi:10.1093/bioinformatics/btr261
156. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res*. 2018;46(D1):D493-D496. doi:10.1093/nar/gkx922
157. Chojnacki S, Cowley A, Lee J, Foix A, Lopez R. Programmatic access to bioinformatics tools from EMBL-EBI update: 2017. *Nucleic Acids Res*. 2017;45(W1):W550-W553. doi:10.1093/nar/gkx273
158. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25(9):1189-1191. doi:10.1093/bioinformatics/btp033
159. McMahon AC, Rahman R, Jin H, et al. TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. *Cell*. 2016;165(3):742-753. doi:10.1016/j.cell.2016.03.007
160. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120. doi:10.1093/bioinformatics/btu170
161. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-359. doi:10.1038/nmeth.1923
162. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants

from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164-e164. doi:10.1093/nar/gkq603

163. Dean S, Sunter JD, Wheeler RJ. TrypTag.org: A Trypanosome Genome-wide Protein Localisation Resource. *Trends Parasitol*. 2017;33(2):80-82. doi:10.1016/j.pt.2016.10.009
164. Fritz M, Vanselow J, Sauer N, et al. Novel insights into RNP granules by employing the trypanosome's microtubule skeleton as a molecular sieve. *Nucleic Acids Res*. 2015;43(16):8013-8032. doi:10.1093/nar/gkv731
165. Reuner B, Vassella E, Yutzy B, Boshart M. Cell density triggers slender to stumpy differentiation of Trypanosoma brucei bloodstream forms in culture. *Mol Biochem Parasitol*. 1997;90(1):269-280. doi:10.1016/S0166-6851(97)00160-6
166. Matthews KR. The developmental cell biology of Trypanosoma brucei. *J cell*. 2005;118:283-290. doi:10.1016/j.mib.2007.09.014
167. Lim J, Ha M, Chang H, et al. Uridylation by TUT4 and TUT7 Marks mRNA for Degradation. *Cell*. 2014;159(6):1365-1376. doi:10.1016/j.cell.2014.10.055
168. Knüsel S, Roditi I. Insights into the regulation of GPEET procyclin during differentiation from early to late procyclic forms of Trypanosoma brucei. *Mol Biochem Parasitol*. 2013;191(2):66-74. doi:10.1016/j.molbiopara.2013.09.004
169. Aphasizhev R, Aphasizheva I. Terminal RNA uridylyltransferases of trypanosomes. *Biochim Biophys Acta*. 2008;1779(4):270-280. doi:10.1016/j.bbagrm.2007.12.007
170. Aphasizhev R, Aphasizheva I, Simpson L. Multiple terminal uridylyltransferases of trypanosomes. *FEBS Lett*. 2004;572(1-3):15-18. doi:10.1016/j.febslet.2004.07.004\rS0014579304008555 [pii]
171. Toh H, Weiss BL, Perkin SAH, et al. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of Sodalis glossinidius in the tsetse host. *Genome Res*. 2006;16(2):149-156. doi:10.1101/gr.4106106.1
172. Vasquez J-J, Hon C-C, Vanselow JT, Schlosser A, Siegel TN. Comparative ribosome profiling reveals extensive translational complexity in different Trypanosoma brucei life cycle stages. *Nucleic Acids Res*. 2014;42(6):3623-3637. doi:10.1093/nar/gkt1386
173. Alsford S, Turner DJ, Obado SO, et al. High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. *Genome Res*. 2011;21(6):915-924. doi:10.1101/gr.115089.110
174. Zachariae W, Nasmyth K. Whose end is destruction: Cell division and the anaphase-promoting complex. *Genes Dev*. 1999;13(16):2039-2058. doi:10.1101/gad.13.16.2039

175. Kumar P, Wang CC. Depletion of anaphase-promoting complex or cyclosome (APC/C) subunit homolog APC1 or CDC27 of Trypanosoma brucei arrests the procyclic form in metaphase but the bloodstream form in anaphase. *J Biol Chem*. 2005;280(36):31783-31791. doi:10.1074/jbc.M504326200
176. Bessat M, Knudsen G, Burlingame AL, Wang CC. A Minimal Anaphase Promoting Complex/Cyclosome (APC/C) in Trypanosoma brucei. Li Z, ed. *PLoS One*. 2013;8(3):e59258. doi:10.1371/journal.pone.0059258
177. Xu W, Rahman R, Rosbash M. Mechanistic implications of enhanced editing by a HyperTRIBE RNA-binding protein. *RNA*. 2018;24(2):173-182. doi:10.1261/rna.064691.117
178. Rodriguez J, Menet JS, Rosbash M. Nascent-Seq Indicates Widespread Cotranscriptional RNA Editing in Drosophila. *Mol Cell*. 2012;47(1):27-37. doi:10.1016/j.molcel.2012.05.002
179. Mulindwa J, Leiss K, Ibberson D, et al. Transcriptomes of Trypanosoma brucei rhodesiense from sleeping sickness patients, rodents and culture: Effects of strain, growth conditions and RNA preparation methods. *PLoS Negl Trop Dis*. 2018;12(2). doi:10.1371/journal.pntd.0006280
180. Kropiwnicka A, Kuchta K, Lukaszewicz M, et al. Five eIF4E isoforms from Arabidopsis thaliana are characterized by distinct features of cap analogs binding. *Biochem Biophys Res Commun*. 2015;456(1):47-52. doi:10.1016/j.bbrc.2014.11.032
181. Yarunin A, Harris RE, Ashe MP, Ashe HL. Patterning of the Drosophila oocyte by a sequential translation repression program involving the d4EHP and Belle translational repressors. *RNA Biol*. 2011;8(5):904-912. doi:10.4161/rna.8.5.16325
182. Jones NG, Thomas EB, Brown E, et al. Regulators of Trypanosoma brucei Cell Cycle Progression and Differentiation Identified Using a Kinome-Wide RNAi Screen. Horn D, ed. *PLoS Pathog*. 2014;10(1):e1003886. doi:10.1371/journal.ppat.1003886
183. Ule J, Jensen K, Mele A, Darnell RB. CLIP: A method for identifying protein-RNA interaction sites in living cells. *Methods*. 2005;37:376-386. doi:10.1016/j.ymeth.2005.07.018
184. Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. CLIP identifies Nova-regulated RNA networks in the brain. *Science*. 2003;302(5648):1212-1215. doi:10.1126/science.1090095
185. Lee FCY, Ule J. Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Mol Cell*. 2018;69(3):354-369. doi:10.1016/j.molcel.2018.01.005

186. Ouwenga RL, Dougherty J. Fmrp targets or not: long, highly brain-expressed genes tend to be implicated in autism and brain disorders. *Mol Autism*. 2015;6(1):16. doi:10.1186/s13229-015-0008-1
187. Montiel-Gonzalez MF, Vallecillo-Viejo I, Yudowski GA, Rosenthal JJC. Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. *Proc Natl Acad Sci*. 2013;110(45):18285-18290. doi:10.1073/pnas.1306243110