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

MSc. Tania Bishola Tshitenge
Born in: Kinshasa, Democratic Republic of the Congo
Oral examination: 10th December 2021

Regulating a post-transcriptional regulator of gene expression, the RNA-binding protein 10, in *Trypanosoma brucei*

Referees: Prof. Dr. Christine Clayton (ZMBH)
Prof. Dr. Nina Papavasiliou (DKFZ)

“***Dans la vie, rien est à craindre, tout est à comprendre***”

Marie-Curie Skłodowska

Acknowledgements

My deepest and sincere gratitude first goes to my PhD supervisor, Prof. Christine Clayton, for giving me the opportunity to develop my PhD project in her lab. I am forever grateful for these four years of mentorship, moral support and for sharing brilliant ideas that helped me shape my work. I am thankful to my Thesis Advisory Committee (TAC) members, Prof. Nina Papavasiliou, Prof. Christian Janzen and Prof. Luise Krauth-Siegel for their guidance, the fruitful discussions and for developing my critical thinking. I am indebted to Dr. Susanne Kramer for her valuable supervision while spending the final stage of my work in her lab, and to Prof. Laurie Read for antibody donation and for friendly discussions on DRBD18.

My special thanks are directed to my former lab colleagues, Bin, Kevin and Esteban and all the current lab mates, Albina, Franziska, Fernanda and Larissa for creating a friendly environment in the lab. Each of you contributed to my personal development and I will forever treasure the moments we spent together. I deeply thank Franziska for the help in writing my Zusammenfassung. My gratitude is also extended to Francisco, Anastasia and Monica for creating a pleasant environment in the Papavasiliou's lab. I am grateful to the students that I supervised, Juyeop Kim and Lena Reichert. I deeply thank Claudia Helbig and Ute Leibfred for technical assistance and their kind support. I thank Bernardo and Paula for their help and the friendship shared during my research stay in Uni-Würzburg.

My gratitude is extended to Sabine Merker and Thomas Ruppert of the ZMBH mass spectrometry facility and David Ibberson of the Bioquant deep sequencing facility for processing our samples. The deepest expression of my gratitude is addressed to the Heidelberg Biosciences International Graduate School (HBIGS) from which I got my PhD fellowship. I really appreciate the support of Dr. Rolf Lutz, Martina Galvan and Sandra Martini of the HBIGS office during these past four years. I am also grateful to the Excellence Scholarship Program "BEBUC" for the support since my bachelor studies and until now.

I thank all the friends that I met during these years in Heidelberg. Thanks a lot for listening to my drama stories and my frustrations from the lab, thanks for sharing funny moments and laughter with me. Without you, this journey would be unbearable. In particular, my thanks are directed to Lucia, Tabeth and Mendi for spending weekly time prayers together.

I would like to deeply thank my family and my friends in Congo, in Germany and all over the world for supporting my efforts and for being strongly with me even from a distance. Your ever-unending moral support and your love help me to accomplish this work. Above all, a huge thank to my mother Agathe for everything she has done and for sacrificing a lot for me. And to my heavenly father, Jesus-Christ, who is the giver of perfect grace and excellent gift, I thank him for guiding me and sustaining my faith during all these years.

Table of contents

Acknowledgements ..... **iv**

List of abbreviations ..... **viii**

List of figures ..... **xi**

List of tables ..... **xiii**

Summary ..... **xiv**

Zusammenfassung ..... **xv**

1. Introduction **1**

1.1 Background **1**

1.2 Diseases caused by Trypanosoma brucei **1**

1.2.1 Human African Trypanosomiasis 1

1.2.2 Animal African Trypanosomiasis (AAT) 3

1.3 Life cycle of T. brucei **4**

1.4 T. brucei cell structure **6**

1.5 Gene regulation in T. brucei **7**

1.5.1 Transcription 7

1.5.2 mRNA processing 8

1.5.3 Nuclear export of mRNA 11

1.5.4 Translation 12

1.5.5 mRNA decay 13

1.5.6 Localization of mRNAs 14

1.6 Role of RNA-binding proteins in regulating gene expression **15**

1.6.1 ALBA and Pumilio domain proteins 15

1.6.2 Zinc-finger domain proteins 16

1.6.3 RNA-recognition motif (RRM) domain containing proteins 17

1.7 Role of the 3'-UTRs in post-transcriptional regulation **19**

1.8 Aims of the study **21**

2. Material and Methods **22**

2.1 Trypanosome cell culture methods **22**

2.1.1 Bloodstream form cell culture 22

2.1.2 Procyclic form cell culture 23

2.1.3 Transfection of bloodstream/procyclic trypanosomes 23

2.1.4 Antibiotics used for selection of transgenic trypanosomes 25
2.1.5 Storage of trypanosomes cell lines 25
2.1.6 Growth curves 25
2.1.7 Trypanosome differentiation 26
2.1.8 Cold shock conditions 26
2.2 DNA methods 26
2.2.1 Genomic DNA isolation 26
2.2.2 Plasmid construction and isolation 27
2.2.3 Endogenous tagging of DRBD18 and ZC3H28 27
2.3 RNA methods 28
2.3.1 RNA extraction, agarose gel electrophoresis and Northern blotting 28
2.3.1 RNA immunoprecipitation 29
2.3.2 High throughput RNA sequencing and bioinformatic analysis 31
2.3.3 Affymetrix single molecule mRNA fluorescence in situ hybridization (FISH) 32
2.4 Protein methods 33
2.4.1 SDS-PAGE and Western Blotting 33
2.4.2 Co-immunoprecipitation (Co-IP) 34
2.4.3 Yeast two-hybrid assay 36
2.4.4 Tandem Affinity Purification 37
2.4.5 Affinity purification of biotinylated proteins 38
2.4.6 Mass spectrometry analysis 39
2.5 Subcellular fractionation 40
2.6 CAT assay 40
2.7 Oligonucleotide list and plasmid list 41
2.7.1 Oligonucleotide list 42
2.7.2 Plasmid list 47
2.7.3 Plasmid maps 49
2.8 Web resources 58
3. Results 59
3.1 The RBP10 3'-UTR is sufficient for RBP10 developmental regulation 59
3.1.1 RBP10 mRNA is developmentally regulated 59
3.1.2 RBP10 mRNA is stable in bloodstream forms 60
3.1.3 RBP10 expression level is regulated during differentiation 61
3.1.4 The RBP10 3'-UTR regulates RBP10 developmental expression 62
3.1.5 The RBP10 3'-UTR contains several regulatory elements 64
3.1.6 Identification of motifs in the RBP10 3'-UTR 67

3.1.7 $(AU)_{10}$ element affects reporter translation in bloodstream forms 67
3.2 Validation of RNA-Protein interaction detection in T. brucei 68
3.3 Identification of proteins that regulate RBP10 expression 72
3.3.1 The role of HNRNPFH in regulating RBP10 expression 73
i. Depletion of HNRNPFH does not affect RBP10 expression levels 73
ii. HNRNPFH associates with mRNAs encoding various proteins 76
iii. HNRNPFH does not bind to RBP10 3'-UTR 79
3.3.2 The role of ZC3H28 in trypanosome bloodstream forms 81
i. ZC3H28 is essential for growth in bloodstream forms 81
ii. ZC3H28 increases the abundance of an attached reporter mRNA 82
iii. ZC3H28 interacts with the MKT1 complex and several ribosomal proteins 84
iv. ZC3H28 associates with long mRNAs that have low ribosome occupancy 89
v. ZC3H28 depletion affects trypanosome transcriptome 92
3.3.3 The role of DRBD18 in regulating processing and nuclear export of the RBP10 mRNA 96
i. DRBD18 is essential in bloodstream forms and affects RBP10 mRNA processing 96
ii. DRBD18 affects processing of other mRNAs 101
iii. DRBD18 mRNA binding does not correlate with RNAi effects 102
iv. The nuclear export factor MEX67 does not affect RBP10 mRNA processing 104
v. DRBD18 is localized in the nucleus and in the cytosol 105
vi. DRBD18 associates with proteins of the outer ring of the nuclear pore 106
vii. RBP10 long transcripts are trapped in the nucleus after DRBD18 depletion 108
4. Discussion 112
4.1 Developmental regulation of RBP10 is mediated by the 3'-UTR of its mRNA 112
4.2 HNRNPFH does not regulate RBP10 expression in bloodstream forms 115
4.3 ZC3H28 might stabilize long mRNAs through its interaction with MKT1-complex 116
4.4 DRBD18 affects processing and nuclear export of RBP10 mRNA 118
5. References 122
6. Supplementary information 139

List of abbreviations

Abbreviation	Description
3'-RACE	Rapid Amplification of cDNA ends (3'-ends of the mRNAs)
3'-UTR	3'-untranslated region
5'-UTR	5'-untranslated region
°C	Degree Celsius
A	Adenine
AA	Amino acid
AAT	Animal African Trypanosomiasis
ALBA	Acetylation lowers binding affinity domain
ALPH1	ApaH-like phosphatase 1
ARE	Adenylate/uridylate-rich element
ATP	Adenosine triphosphate
BARP	Brucei Alanine-Rich protein
BBB	Blood-brain barrier
bp	Base pairs
BSF	Bloodstream forms
BSA	Bovine Serum Albumin
BSD	Blasticidin S-Deaminase
C	Cytosine
CAT	Chloramphenicol Acetyltransferase
CBC	Cap binding complex
CCR4	Carbon Catabolite Repressor Protein
cDNA	Complementary DNA
CDS	Coding Sequence
CNS	Central Nervous System
cpm	Counts per minute
dCTP	Deoxy Cytosine Triphosphate
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
DRBD	Double-stranded RNA-binding domain
DTT	Dithiothreitol
DYRK	Dual-specificity tyrosine (Y) phosphorylation-regulated kinase
E. coli	*Escherichia coli*
ECL	Enhanced Chemiluminescence
EDTA	Ethylenediamine Tetraacetic Acid
EGTA	Ethylene glycol Tetraacetic Acid
EP	Glu/Pro repeat-containing protein

et. al.	And others
FCS	Fetal Calf Serum
FISH	Fluorescence *in situ* hybridization
G	Guanine
gDNA	Genomic DNA
gHAT	Gambiense Human African Trypanosomiasis
GPI	Glycosylphosphatidylinositol
GPEET	Gly/Pro/Glu/Glu/Thr repeat-containing protein
HMI-9	Hirumi's Modified Iscove's medium-9
HYG	Hygromycin
IgG	Immunoglobulin G
kB	Kilobase
mRNA	Messenger RiboNucleic Acid
mRNP	Messenger RiboNucleoprotein
mVSGs	Metacyclic Variant Surface Glycoproteins
NPT	Neomycin Phosphotransferase
nt	Nucleotide
PAC	Puromycin N-acetyltransferase
PAD	Proteins associated with differentiation
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
RBP	RNA-binding protein
rHAT	Rhodesiense Human African Trypanosomiasis
RNAi	RNA interference
RRM	RNA recognition Motif
rRNA	Ribosomal RNA
rpm	Revolution per minute
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SDS	Sodium dodecyl sulfate
SIF	Stumpy induction factor
SL	Spliced leader
SL RNA	Spliced leader RNA
snRNA	Small nuclear RNA
Spp.	species
T	Thymine
T. b.	*Trypanosoma brucei*
tRNA	Transfer RNA
TAP	Tandem affinity purification
U	Uracil
VSG	Variant Surface Glycoproteins

WHO	World Health Organization
WT	Wild Type

List of figures

Figure 1.1: Geographical distribution of Human African Trypanosomiasis. 2
Figure 1.2: Life cycle of *T. brucei* and its developmental forms. 6
Figure 1.3: Trypanosome cell structure and morphology 7
Figure 1.4: mRNA processing in trypanosomes. 10
Figure 3.1: RBP10 mRNA is developmentally regulated in T. brucei 60
Figure 3.2: Half-life of RBP10 mRNA in bloodstream and procyclic forms. 61
Figure 3.3: Regulation of RBP10 protein expression during differentiation 62
Figure 3.4: The RBP10 3'-UTR is sufficient for developmental regulation. 63
Figure 3.5: The RBP10 3'-UTR contains several regulatory sequences 66
Figure 3.6: Analysis of specific motifs 68
Figure 3.7: RNA-protein interaction detection in trypanosomes 69
Figure 3.8: Knockdown of HNRNPFH does not affect RBP10 expression levels and cell growth 74
Figure 3.9: Knockdown of HNRNPFH does not affect RBP10 mRNA levels 75
Figure 3.10: HNRNPFH binds mRNAs which encode a wide variety of proteins 77
Figure 3.11: Depletion of HNRNPFH does not affect the CAT reporter bearing RBP10 3'-UTR fragment 1.1. 80
Figure 3.12: Depletion of ZC3H28 affects cell growth in trypanosome bloodstream forms. 82
Figure 3.13: ZC3H28 increases expression levels of an attached CAT reporter mRNA 83
Figure 3.14: Tandem affinity purification of N-TAP ZC3H28, C-TAP GFP and C-TAP DRBD18 85
Figure 3.15: ZC3H28 associates with the MKT1 complex and several ribosomal proteins. 86
Figure 3.16: Interactions of ZC3H28 with the MKT1-complex 88
Figure 3.17: Interactions of ZC3H28 with mRNAs 91
Figure 3.18: Time course study for ZC3H28 RNAi. 93
Figure 3.19: Effect of ZC3H28 depletion on the transcriptome 94
Figure 3.20: DRBD18 depletion by RNAi affects RBP10 mRNA processing 97
Figure 3.21: Transcriptomics data showed that DRBD18 RNAi influences RBP10 alternative polyadenylation. 99
Figure 3.22: Effect of DRBD18 depletion by RNAi on the acetyl-coA synthetase mRNA 100
Figure 3.23: Effect of DRBD18 depletion on DRBD12 mRNA. 101
Figure 3.24: Interactions of DRBD18 with mRNAs 103
Figure 3.25: MEX67 depletion does not affect RBP10 mRNA processing. 104
Figure 3.26: Subcellular fractionation in bloodstream forms reveals localization of DRBD18. 105
Figure 3.27: DRBD18 associates with proteins of the outer ring of the nuclear pore 107

Figure 3.28: Long RBP10 mRNAs are trapped in the nucleus after DRBD18 depletion..... 109
Figure 4.1: Proposed model of regulation by the RBP10 3'-UTR..... 113
Figure 4.2: Proposed model for ZC3H28 function..... 117
Figure 4.3: Proposed model of regulation of RBP10 mRNA by DRBD18 121

List of tables

Table 2.1: Amaxa buffer 24
Table 2.2: Zimmerman's Post Fusion Medium (ZPFM) buffer 24
Table 2.3: Antibiotics concentrations 25
Table 2.4: EB Buffer 27
Table 2.5: 2X RNA loading buffer 29
Table 2.6: Hybridization solution for DNA probes 29
Table 2.7: Lysis Buffer 30
Table 2.8: IP Buffer 31
Table 2.9: 2X Laemmli Buffer 33
Table 2.10: 1x SDS-running buffer 33
Table 2.11: Blotting buffer 34
Table 2.12: TBS-T Buffer 34
Table 2.13: Antibodies used for western blot 34
Table 2.14: 200x DNAse Buffer 35
Table 2.15: Stop solution 36
Table 2.16: Drop solution (10X) 36
Table 2.17: SD medium, 5 L, sterile filtered 37
Table 2.18: Lysis buffer for biotinylation 38
Table 2.19: Wash buffer 2 for biotinylation 39
Table 2.20: Wash buffer 3 for biotinylation 39
Table 2.21: Wash buffer 4 for biotinylation 39
Table 2.22: List of all oligonucleotides used in this study 42
Table 2.23: List of all the plasmids made in this study 47
Table 3.1: Selected proteins associated with the PSSA-2 3'-UTR and the actin 3'-UTR 71
Table 3.2: List of potential regulators of RBP10 mRNA in bloodstream forms 72
Table 3.3: Selected mRNA targets of HNRNPFH identified by RNA Sequencing 78
Table 3.4: Details FISH results 111
Table 6.1: Affymetrix probes used for the FISH experiment 146
Table 6.2: Measurements of CAT reporter mRNAs containing the RBP10 3'-UTR fragment and the actin 3'-UTR 148

Summary

African trypanosomes are flagellated parasites that cause sleeping sickness disease in humans and nagana in cattle. In the bloodstream of the mammalian host, *Trypanosoma brucei* exist as proliferative long slender forms or as non-dividing stumpy forms; the latter differentiate to procyclic forms in the midgut of the tsetse fly. Since the majority of protein coding genes are arranged in polycistronic transcription units, trypanosomes rely almost exclusively on post-transcriptional mechanisms - mRNA processing, decay and translation - to regulate gene expression. The RNA-binding proteins assume the burden that is carried by transcription factors in others eukaryotes and therefore they are critical at all stages. The RNA-binding protein 10 (RBP10), which is exclusively expressed in bloodstream forms, binds the motif $UA(U)_6$ present in the 3'-untranslated regions of procyclic-specific mRNAs and targets them for destruction. This work focuses on understanding the mechanisms that regulate the differential expression of RBP10 in *Trypanosoma brucei*.

Results of this study show that the developmental regulation of RBP10 expression is mediated by the long 3'-untranslated regions of its mRNA. Different regulatory sequences that enhance stability and translation of the *RBP10* mRNA in bloodstream forms or destabilize and repress the mRNA in procyclics are scattered throughout the *RBP10* 3'-untranslated region. The only motif that was found to be involved in *RBP10* mRNA translation in bloodstream forms is the $(AU)_{10}$ element. RNAi-mediated depletion of the RNA-binding proteins, HNRNPFH, ZC3H28, ZC3H40 and ZC3H45 did not affect *RBP10* mRNA levels in bloodstream forms. Nevertheless, ZC3H28 associates with MKT1 complex, the translation initiation factor complex EIF4E4/EIF4G3 and several ribosomal proteins. The protein binds long mRNAs that have low ribosome occupancy and presumably might stabilize them. On the other hand, the depletion of the double RNA-binding protein 18 (DRBD18) in trypanosome bloodstream-forms affected cell growth and led to reduction of RBP10 protein levels. This was accompanied by accumulation of several shortened *RBP10* mRNA isoforms, and loss of the longer mRNA species. The effect was seen in over 100 other mRNAs. DRBD18 associated with the proteins of the outer ring of the nuclear pore complex as well as those of the exon junction complex among others. Depletion of the nuclear export factor MEX67 did not impair RBP10 processing, suggesting that the effect of DRBD18 on the *RBP10* mRNA was rather specific. After depletion of DRBD18, the longer RBP10 transcripts were trapped in the nucleus while the shorter ones were able to migrate to the cytosol. The proposed model implies that DRBD18 binds suboptimal processing signals in the 3'-UTRs of its target mRNAs, preventing their use and simultaneously allowing export of the bound mRNAs. In the absence of DRBD18, suboptimal splicing signals are bound by the splicing/polyadenylation machinery generating alternative polyadenylated mRNAs, which are exported in the cytosol.

Zusammenfassung

Afrikanische Trypanosomen sind begeißelte Parasiten, die beim Menschen die Schlafkrankheit und bei Nutztieren Nagana-Seuche verursachen. Im Blutkreislauf des Säugetierwirts kommt *Trypanosoma brucei* als proliferative, lange „Slender-Formen“ oder als sich nicht teilende „Stumpy-Formen“ vor; letztere differenzieren sich im Mitteldarm der Tsetsefliege zu prozyklischen Formen. Da die meisten Protein-kodierenden Gene in polycistronischen Transkriptionseinheiten angeordnet sind, verlassen sich Trypanosomen fast ausschließlich auf posttranskriptionelle Mechanismen – mRNA-Prozessierung, -Zerfall und -Translation – um die Genexpression zu regulieren. RNA-bindende Proteine übernehmen die Last, die von Transkriptionsfaktoren in anderen Eukaryoten getragen wird, und sind daher in allen Stadien wichtig. Das ausschließlich in Blutkreislauf-Formen exprimierte RNA-bindende Protein 10 (RBP10) bindet das Motiv $UA(U)_6$, das in den 3’-untranslatierten Regionen bestimmter prozyklischer mRNAs vorhanden ist, und zielt darauf ab, diese zu zerstören. Diese Arbeit konzentriert sich auf das Verständnis der Mechanismen, die die differentielle Expression von RBP10 in *Trypanosoma brucei* regulieren.

Die Ergebnisse dieser Arbeit zeigen, dass die Entwicklungsregulation der RBP10-Expression durch die langen 3’-untranslatierten Regionen seiner mRNA vermittelt wird. Verschiedene regulatorische Sequenzen, die die Stabilität und Translation der RBP10-mRNA in Blutkreislaufformen verbessern oder die mRNA in prozyklischen Formen destabilisieren und reprimieren, sind über die RBP10-3'-untranslatierte Region verstreut. Das einzige Motiv, über das man weiß, dass es an der RBP10-mRNA-Translation in Blutkreislaufformen beteiligt ist, ist das $(AU)_{10}$-Element. Die RNAi-vermittelte Depletion der RNA-bindenden Proteine HNRNPFH, ZC3H28, ZC3H40 und ZC3H45 beeinflusste die RBP10-mRNA-Level in Blutkreislaufformen nicht. Dennoch assoziiert ZC3H28 mit dem MKT1-Komplex, dem Translationsinitiationsfaktor-Komplex EIF4E4/EIF4G3 und mehreren ribosomalen Proteinen. Das Protein bindet lange mRNAs mit geringer Ribosomendichte und stabilisiert diese vermutlich. Andererseits beeinflusste die Depletion des Doppel-RNA-bindenden Proteins 18 (DRBD18) in Blutstromformen das Zellwachstum und führte zu einer Verringerung des RBP10-Proteinlevels. Dies ging mit einer Anhäufung mehrerer verkürzter RBP10-mRNA-Isoformen und einem Verlust der längeren mRNA-Varianten einher. Dieser Effekt wurde bei über 100 anderen mRNAs beobachtet. DRBD18 ist unter anderem mit den Proteinen des äußeren Rings des Kernporenkomplexes sowie denen des Exon-Junction-Komplexes assoziiert. Der Verlust des nuklearen Exportfaktors MEX67 beeinträchtigte die RBP10-Prozessierung nicht, was darauf hindeutet, dass die Wirkung von DRBD18 auf die RBP10-mRNA relativ spezifisch war. Nach der Depletion von DRBD18 waren die längeren RBP10-Transkripte im Zellkern gefangen, während die kürzeren in das Zytosol wandern konnten.

Das vorgeschlagene Modell impliziert, dass DRBD18 suboptimale Verarbeitungssignale in den 3'-UTRs seiner Ziel-mRNAs bindet, deren Verwendung verhindert und gleichzeitig den Export der gebundenen mRNAs ermöglicht. In Abwesenheit von DRBD18 werden suboptimale Spleiß-Signale von der Spleiß-/Polyadenylierungsmaschinerie gebunden, wodurch alternative polyadenylierte mRNAs erzeugt werden, die in das Zytosol exportiert werden.

1. Introduction

1.1 Background

Trypanosomes are extracellular eukaryotic parasites that belong to the protozoan *Trypanosomatidae* family within the *Kinetoplastida* class. Members of the *Kinetoplastida* order, commonly refer as Kinetoplastids, are characterized by the presence of the kinetoplast, a massive network of concatenated circular DNA (kDNA) that makes the mitochondrial genome (Stuart 1983; Lukes et al. 2002). They also have their genes arranged in giant polycistronic clusters and their glycolytic enzymes are compartmentalized into a microbody-like organelle called the glycosomes (Opperdoes and Borst 1977). The order contains a large number of free-living protists species, pathogens of humans, animals and plants (Simpson et al. 2006). *Bodonidae* and *Cryptobiidae* families have two flagella and mainly consist of free-living aquatic species; *Trypanosomatidae* family is identified as having a single flagellum (Moreira et al. 2004). The three subspecies of *T. brucei*, *T. b. gambiense*, *T. b. rhodesiense* and *T. b. brucei,* are transmitted by Tsetse fly (*Glossina spp*) and cause human and animal African Trypanosomiasis across sub-Saharan Africa, within the tsetse-fly infested areas. *Trypanosoma cruzi*, on the other hand, is transmitted by *Triatominae* bugs and cause Chagas disease (also known as American trypanosomiasis). Close relatives of trypanosomes include *Leishmania* species, transmitted by *Phlebotomus* and *Lutzomyia* spp (Sand fly), responsible for the visceral, cutaneous or muco-cutaneous leishmaniasis (Stuart et al. 2008).

1.2 Diseases caused by *Trypanosoma brucei*

1.2.1 Human African Trypanosomiasis

Human African Trypanosomiasis or sleeping sickness is classified as a neglected tropical disease by the World Health Organization (WHO), affecting 36 countries in sub-Saharan Africa where the tsetse fly is found (Figure 1.1). The disease is caused by two subspecies of *T. brucei*, *T. b. gambiense* and *T. b. rhodesiense*, which infect not only humans but also animals (Cecchi et al. 2009). *T. b. gambiense* is responsible for causing the chronic form of the disease, known as the gambiense form of HAT (g-HAT), endemic to the West and Central Africa. gHAT is more adapted to humans with a minor role for animal reservoirs, the disease progresses slowly to the fatality. In contrast, *T. b. rhodesiense* causes the acute fast-progressing form of the disease, endemic to the East and Southern Africa, termed rhodesiense HAT (r-HAT). This is a zoonotic disease, affecting mostly animals (livestock and cattle) with less incidence in humans (Franco et al. 2014). In the last decade, gHAT accounts for more than 98% of morbidities while rHAT causes only 2%. With continued efforts, in 2009, the reported HAT cases were below 10, 000 (~9878) for the first time in 50 years. The decline in the number of cases has continued with only 997 reported in 2018 (World Health Organization 2020).

As the number of HAT cases dropped by 95 % between 2000 and 2018, it is therefore possible to achieve the WHO's target of eliminating HAT as a public health problem by 2030. However, for that, the increased efforts of vector control approaches, the treatment of HAT cases and the coordination of stakeholders for active surveillance of HAT must be sustained (Franco et al. 2018; Sutherland et al. 2017).

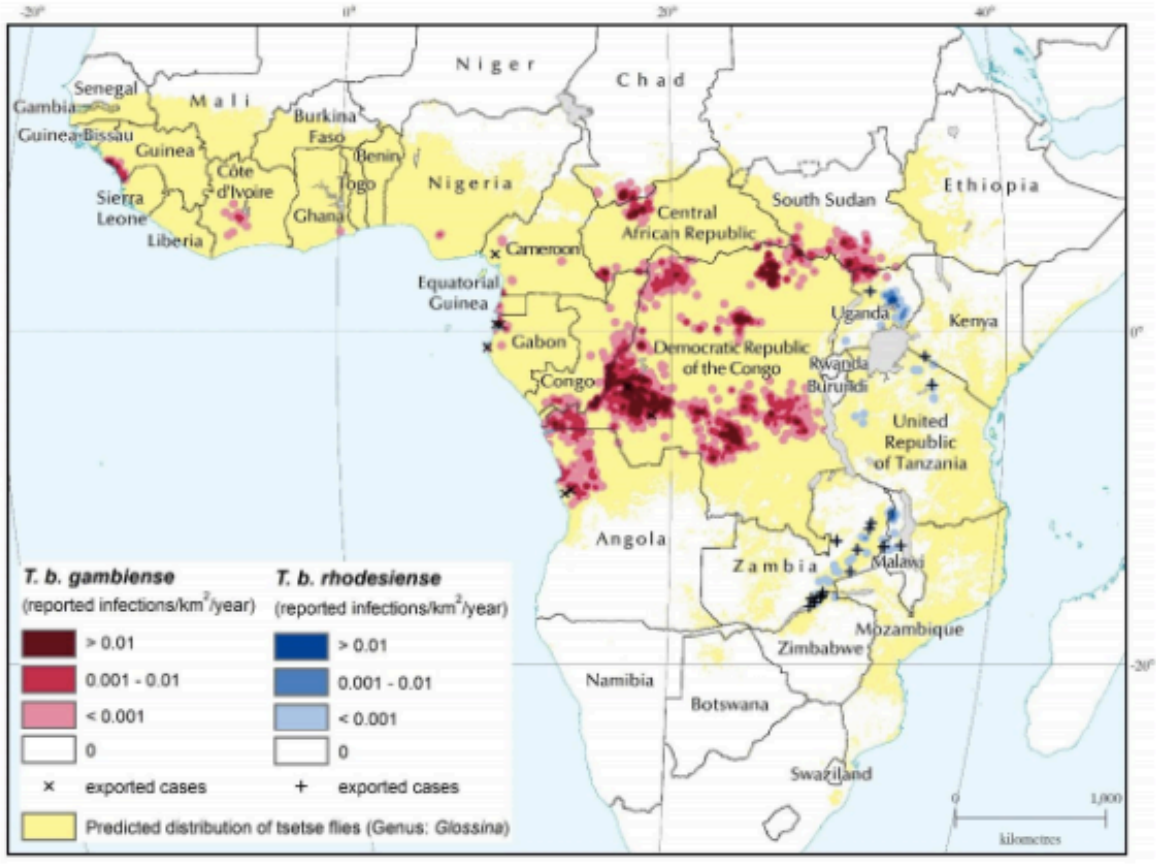


Figure 1.1: Geographical distribution of Human African Trypanosomiasis. The number of reported infections/km2/year by *T. b. gambiense* and *T. b. rhodesiense* are plotted as density in red and blue respectively, while the predicted distribution of the tsetse fly is shown in yellow. The data plotted were reported in the WHO atlas of HAT (Simarro et al. 2010). Image source (Büscher et al. 2017) license number 4410690838783.

In both forms of HAT, the disease evolves in two stages. In the early/first stage also known as haemolymphatic stage, the parasites are mainly found in the blood and lymphatic system. Patients in this stage develop no specific clinical signs, and symptoms include fever, headache, adenopathy, joint pains, pruritus, weakness and weight loss (Kuepfer et al. 2011; Stuart et al. 2008).

Although the rapid parasite growth is countered by host immune responses, the parasite antigenic variation enables immune evasion resulting in waves of parasitemia. Thereafter, the parasites cross the blood-brain barrier (BBB) into the cerebrospinal fluid and invade the

central nervous system (CNS) causing the late/second stage of the disease, known as the meningo-encephalitic stage. At this stage, the disease causes progressive neurological damages. Symptoms include motor weaknesses, walking difficulties, sensory disorders, tremors, visual impairments, sleep disturbances that subsequently culminate in coma, multiple organ failure and death in the absence of treatment (Fevre et al. 2008; Franco et al. 2014). The parasite antigenic variation has prevented the development of vaccines against HAT, leaving chemotherapy and vector control, the only options for disease management.

The early stage of rhodesiense HAT is treated with suramin, which is administered by slow intravenous infusion every 3–7 days for four weeks. Although effective, the drug is associated with urticarial rash in 90% of patients. Melarsoprol, an arsenical drug, is used for the late stage and is administered intravenously as a 3.6 % solution in propylene glycol in a 10-day regimen (Brun et al. 2011). It causes a reactive encephalopathy in approximately 10% of patients and fatalities in 5% of the reported cases (Nagle et al. 2014).

All the drugs used for the treatment of sleeping sickness are associated with severe side effects, difficulty in administration and emerging resistance. Thus, the need to develop new anti-trypanosomatid drugs with more efficacy and less side effects. On July 2021, fexinidazole, a 5'-nitroimidazole derivative, that was in clinical trials (Tarral et al. 2014), was approved as an all-oral treatment for both stages of the gambiense sleeping sickness (SANOFI 2021). Fexinidazole is indicated as a 10-day once-a-day treatment for adults and children aged ≥ 6 years and weighing ≥ 20 kg (Deeks 2019). Another promising candidate is acoziborole (tradename SCYX-7158 or AN5568), which is an orally available molecule that showed to cure both stages of HAT in pre-clinical studies (Jacobs et al. 2011)

1.2.2 Animal African Trypanosomiasis (AAT)

Subspecies of *T. brucei*, *T.b. rhodesiense* and *T. b. brucei* are minor causes of Animal African trypanosomiasis (AAT), called Nagana in cattle. The main causative agents of the disease in livestock are *T. vivax* and *T. congolense* (Auty et al. 2015). Other pathogens include *T. simiae, T. godfreyi,* and *T. suis,* which cause high mortality in domestic pigs (Rodrigues et al. 2020; Isaac et al. 2016). *T. b. evansi* infects horses, camels, water buffalo, cattle; and *T. b. equiperdum* is particularly infectious to horses and donkeys (Desquesnes et al. 2013; Ahmed et al. 2018). All these pathogens are transmitted cyclically by the bite of tsetse flies; but mechanical transmission by other biting flies has been reported for some species (Desquesnes and Dia 200; Desquesnes et al. 2009). For instance, *T. vivax* and *T. b. evansi* are mechanically transmitted by tabanids and stomoxyines (Desquesnes and Dia 2003; Desquesnes et al. 2013). These modes of transmission enable the disease to be spread beyond the tsetse infected regions of sub-Saharan Africa (Desquesnes et al. 2013; Jones and Dávila 2001).

Animal African Trypanosomiasis remains a major economic constraint especially in rural areas where it leads to annual losses of billions of dollars to livestock producers and consumers (Meyer et al. 2016). The disease is controlled via several methods such as selective bush clearing, sequential aerial spraying, use of insecticides in traps or as live baits, and in some cases breeding of naturally resistant cattle (Meyer et al. 2018). In areas outside of vector control operations, a number of trypanocides are employed in cattle, with diminazene aceturate and isometamidium chloride, the most largely used. However, resistance to these drugs has been increasing with a range of 35 and 70 million doses of trypanocides used across sub-Saharan Africa annually (Giordani et al. 2016). Despite this high demand, development of new drugs has lagged behind because of the lack of interest from pharmaceutical companies to be involved due to the low anticipated profits. However, recent support from the Bill and Melinda Gates Foundation and the UK department for International Development (through the Global Alliance for Livestock Veterinary Medicines, GALVmed) has emerged to help in the development of new anti-trypanosomes therapies, both therapeutic and prophylactic, in order to tackle Nagana.

1.3 Life cycle of *T. brucei*

The life cycle of *T. brucei* is digenetic, alternating between a mammalian host (humans, game animals and livestock) and the blood-feeding insect, tsetse fly (*Glossina* spp.). The parasite undergoes different metabolic, morphological and motility changes to survive in the disparate environments of the two hosts (Rotureau and Van Den Abbeele 2013). In the mammalian host, trypanosomes exist as proliferative long slender forms, which transform to stumpy forms at the peak of parasitaemia. The infection in the tseste fly starts when the trypanosome "stumpy forms" are taken from an infected mammalian host during the fly bloodmeal (Figure 1.2). In the tsetse midgut, these non-dividing forms differentiate into proliferative procyclic forms after 2-5 days (Gibson and Bailey 2003). Transition from mammals to tsetse fly entails a decrease in temperature from 37°C to around 20-32°C and a switch from glucose to amino acids as the main source of energy (Vickerman 1985). Another major change is the replacement of the variant surface glycoproteins (VSG) coat by a dense coat composed of procyclins; GPEET procyclin being expressed within 2 hours and repressed after 7-9 days post-infection, while EP procyclin is expressed in late procyclics (Vassella et al. 2000; Urwyler et al. 2005). After proliferation, procyclic forms cross the peritrophic matrix of the fly and translocate in the foregut as non-proliferative mesocyclic trypomastigotes (not shown in Figure 1.2) (Sharma et al. 2008). Assymetric division of these trypomastigotes produce both long and short epimastigotes, characterized by a kinetoplast anterior to the nucleus. Epimastigotes express a different type of surface proteins, known as BARPs (Brucei Alanine Rich Proteins) (Urwyler et al. 2005). Through elaborated flagellar membrane, the short epimastigotes parasites get attached to the salivary gland epithelium

where they generate growth-arrested metacyclic forms, which move freely in the salivary gland lumen (Van Den Abbeele et al. 1999).

As a prerequisite for survival in the mammalian host, metacyclic forms acquire a new coat made of metacyclic VSG (mVSGs) dimers, which provide initial protection from the mammalian immune system upon transmission (Turner et al. 1988; Tetley et al. 1987). With a tsetse fly meal, metacyclic trypomastigotes are released into the bloodstream of the mammalian host where they resume cell division and differentiate to long slender forms, the mVSG coat being replaced by a VSG coat. VSGs are linked to the surface membrane by a glycosylphosphatidylinositol (GPI) anchor (Graham et al. 1998) and the GPI anchors were described for the first time in trypanosomes (Ferguson and Williams 1988). The long slender trypomastigotes divide in the mammalian bloodstream and further go on to cross the blood-brain barrier and colonize the central nervous system (CNS) (Figure 1.2). Long slender trypanosomes possess a basic mitochondrion and make use of the substrate-level phosphorylation for energy generation in a specialized organelle called the glycosome (Opperdoes and Borst 1977).

In the mammalian blood and tissue fluids, the parasite lives extracellularly; hence it is exposed to the host adaptive immune responses. The dense homogenous coat of VSGs is highly immunogenic, but the parasite escapes the adaptive immunity by periodically switching to a different new variant of VSG, a process known as antigenic variation. There are approximately thousands of VSG genes, and only one is expressed at any time as guaranteed by an allelic exclusion mechanism (Horn 2014; Mugnier et al. 2016; McCulloch et al. 2017).

The trypanosome infection is therefore characterized by waves of parasitaemia; once the immune response catches a parasite population, they are eliminated but those expressing a different VSG survive and re-establish a new infection (Stockdale et al. 2008; Monica et al. 2015; Pinger et al. 2017). At the peak of parasitaemia, the long slender trypanosomes respond to a stumpy induction factor (SIF) via an oligopeptide quorum-sensing mechanism and they differentiate to shorter, fatter non-dividing “stumpy forms”, which are cell-cycle arrested at G1/G0 (Reuner et al. 1997; Vassella et al. 2000). As a pre-adaptation to the tsetse fly, stumpy cells express some mitochondrial enzymes necessary for proline metabolism and carboxylate transporters called PAD (proteins associated with differentiation), which convey the differentiation signal citrate/ *cis*-aconitate (Czichos et al. 1986; Boshart and Engstler 2004). Once a fly feeds on an infected mammalian host, stumpy forms are ingested and the cycle resumes.

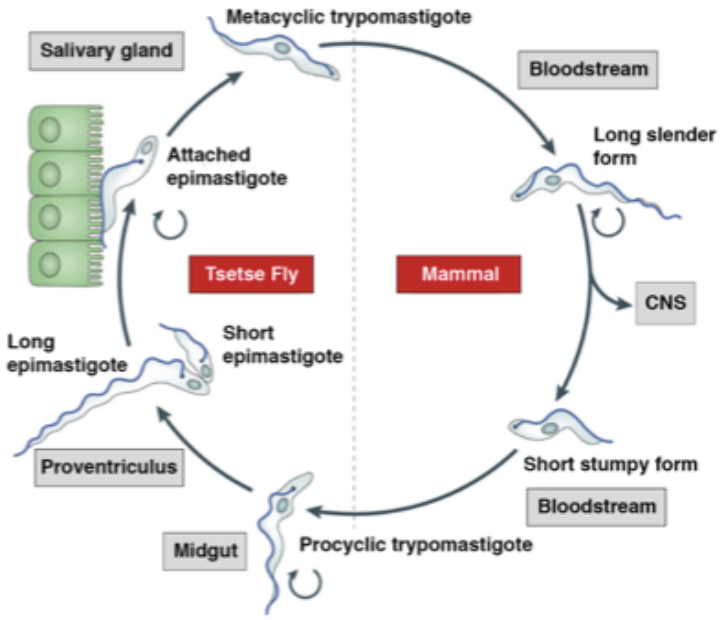


Figure 1.2: Life cycle of *T. brucei* and its developmental forms. Trypanosomes shuttle between a mammalian host and the tsetse fly vector and they have to adapt to different environmental conditions throughout their life cycle. Figure taken from (Langousis and Hill 2014).

1.4 *T. brucei* cell structure

A part from being disease agents, the structure of trypanosomes makes them an excellent model for the study of evolutionary cell biology. They possess all organelles of a typical eukaryotic cell including a nucleus, an endoplasmic reticulum (ER), the Golgi apparatus, the cytoskeleton, endosomes, lysosomes, a mitochondrion and an endo- and exocytosis system (Figure 1.3).

Trypanosome cell structure is highly polarized with microtubules running from the anterior (minus ends) to the posterior (plus ends) and defining the cell shape (Robinson et al. 1995). At the posterior end, the microtubules open to form the flagellar pocket, from which emerges the single flagellum that runs laterally alongside the cell body in a left-handed spiral towards the anterior pole (Field and Carrington 2009). The flagellum, a typical characteristic of trypanosomes, has a conventional axonemal structure but with an associated paraflagellar rod (Vaughan and Gull 2003). The flagellum is held attached to the cell surface by the flagellum attachment zone (FAZ) (Matthews 2005). It is entirely responsible for the parasite motility: they swim with the flagellum tip leading. Not only is the flagellum involved in motility but also in cytokinesis and has multiple sensory functions (Langousis and Hill 2014). The flagellar pocket is the only site of exocytosis and endocytosis, and is involved in membrane recycling, making trypanosome an excellent model for protein trafficking and

sorting studies (Field and Carrington 2009). The hydrodynamic flow in the bloodstream acts on trypanosome motility, causing the IgG-VSG immune complexes to move backward to the posterior pole where they are endocytosed in the flagellar pocket and degraded (Engstler et al. 2007). Another typical organelle is the single mitochondrion of the parasite, which is developmentally regulated: while in bloodstream forms, it is a simple tubular structure, reflecting the absence of mitochondrial respiration in this life stage, procyclics have a more elongated and complete mitochondrion (Matthews 2005). The mitochondrial genome is organized in a disk-like structure called the kinetoplast. It is comprised of two types of concatenated circular DNAs, the maxicircles encoding mitochondrial proteins and rRNAs, and the minicircles encoding short guide RNAs, which serve as templates for editing the maxicircles transcripts (Simpson et al. 1987).

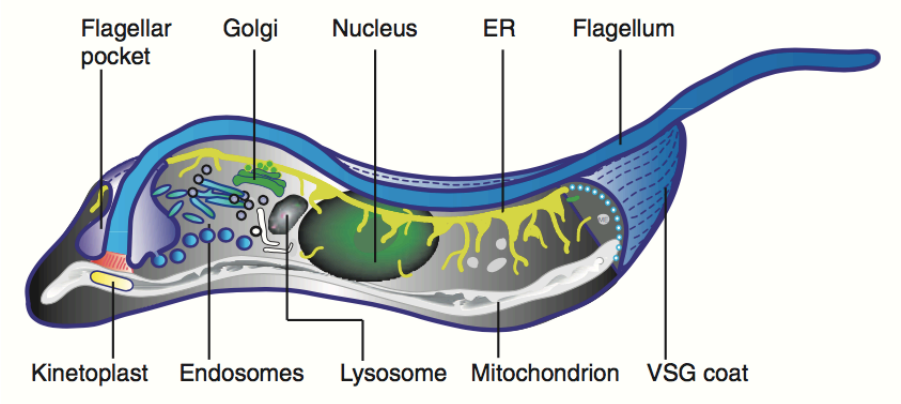


Figure 1.3: Trypanosome cell structure and morphology. A simplified representation of the location of the major structural features of the trypanosome bloodstream form cell. Figure taken from (Overath and Engstler 2004).

1.5 Gene regulation in *T. brucei*

1.5.1 Transcription

In contrast to other eukaryotes, trypanosomes have unique mechanisms for transcribing and processing mRNAs. Nearly all protein coding genes are transcribed as clusters in a polycistronic manner by the RNA polymerase II. These long transcription units contain up to 100 or more different open reading frames, encoding unrelated functional proteins (Clayton and Shapira 2007). Transcription initiation occurs in DNA sequences separating gene arrays orientated in opposite and divergent directions, so called strand-switch regions (Palenchar and Bellofatto 2006; Wedel et al. 2017). There are no introns in trypanosomes and initiation is regulated by histone variants and histone modifications rather than by transcription

factors (Siegel et al. 2009). The only exception includes transcription of the splice leader RNA where specific promoters and specific transcription factors are required (Lee et al. 2007). With no promoters from individual genes, there is no control of transcription initiation. Although transcription by RNA polymerase II was reported to happen at uniform rate (Florini et al. 2019), there are large differences in mRNA and protein levels between genes belonging to the same polycistronic transcription unit and between developmental stages (Jensen et al. 2014). This is mainly due to post-transcriptional regulation operating at the levels of pre-mRNA processing, mRNA export and transport, localization, mRNA decay as well as translation initiation and elongation (Clayton 2019).

During mRNA processing, individual mature mRNAs are generated from the primary transcript by *trans*-splicing of the capped spliced leader mini-exon at the 5'-end and the addition of a poly(A) tail at the 3'-end (Michaeli 2011). As a result of *trans*-splicing, each mature mRNA possesses the same 39 nucleotide splice leader RNA (SL RNA) at its 5'-end, which provides the mRNA with the hypermethylated cap4 structure, protecting the mRNA from degradation and improving its stability and translation (Zamudio et al. 2009). The hypermethylated cap of the processed mRNAs is then bound by the cap-binding complex (CBC), which consists of five components: CBP20 subunit (binds directly to the cap), the importin-α and three novel uncharacterized proteins (Li and Tschudi 2005). Termination of transcription by RNA polymerase II happens either at convergence regions where two transcription units converge or at regions transcribed by other polymerases (Clayton 2019).

In *T. brucei*, *cis*-splicing was found for only two genes encoding the poly(A) polymerase (Mair et al. 2000) and a putative RNA helicase (Jaé et al. 2010). The most abundant genes encoding the rRNA as well as the stage specific proteins (VSGs and procyclins) are transcribed by RNA polymerase I, which gives ten times more RNA than RNA pol II (Clayton 2002; Ridewood et al. 2017). RNA polymerase III transcribes tRNAs, the 7SL RNA and all U-rich snRNA genes (Vanhamme and Pays 1995).

1.5.2 mRNA processing

As previously mentioned, transcription in trypanosomes is polycistronic and mature mRNAs are created by *trans*-splicing of a 39-nucleotide splice leader sequence. This reaction is coupled with polyadenylation of the preceding mRNA (Michaeli 2011; Ullu et al. 1993). *Trans*-splicing mechanism is literally similar to the well-described eukaryotic *cis*-splicing, with the only exception that the two *trans*-esterification reactions happen between the precursor mRNA and the SL RNA. There is no consensus sequence for the branch point (Figure 1.4), but evidence showed the use of A residues closest to the polypyrimidine tract (Hummel et al. 2000; Patzelt et al. 1989).

Similar to *cis*-splicing in other organisms, a polypyrimidine tract of between 8-25 nucleotides and with varying composition, upstream of the AG splice acceptor site, is the main determinant for *trans*-splicing sites (Kolev et al. 2010). Sequences within the 5'-untranslated regions and the composition of the polypyrimidine tract can regulate the efficiency of splicing and therefore affect the abundance of mature mRNAs (Michaeli 2011). In *T. brucei*, there is a preference for U over C in the polypyrimidine tracts and longer polypyrimidine tracts give more efficient processing (Siegel et al. 2005). Previous work by Siegel et al. 2005 also revealed that inserting a cytidine in a purely uridine-containing polypyrimidine tract results in increased *trans*-splicing efficiency (strong splice sites) whereas interspacing with purines (A or G) leads to a large decrease (weak splice sites). There is also strong evidence that addition of the SL RNA occurs at the first AG dinucleotide following the polypyrimidine tract. The polyadenylation site specification is not conventional but it depends solely on *trans*-splicing of the downstream mRNA. Studies showed that polyadenylation happens with one or more A residues located between 80 and 140 nt upstream of the polypyrimidine tract. The polyadenylation site specification is not conventional but it depends solely on *trans*-splicing of the downstream mRNA. Studies showed that polyadenylation happens with one or more A residues located between 80 and 140 nt upstream of the polypyrimidine tract (Hug et al. 1994; Benz et al. 2005). More than one polypyrimidine tract can be found in trypanosome intergenic regions. The 3'-untranslated regions of the mRNAs are also riddled with low complexity sequences and can contain various polypyrimidine tracts. As a consequence, several alternative splicing signals can be used by the splicing machinery, which could generate different encoded proteins. In some cases, these alternative proteins have distinct targeting subcellular signals (Rettig et al. 2012).

As illustrated in Figure 1.4, the simplest mRNA processing scenario happens when only one polypyrimidine tract is present in a given intergenic region (a), resulting in two mRNAs (b): the polyadenylation of the first mRNA (ORF1) depends on *trans*-splicing of the second mRNA (ORF2). However, when two splicing signals, A and B, are present in a given intergenic region (c), both of them can be used independently. This will result in two mRNAs and an additional processed RNA from the intergenic region with no coding region. The first mRNA (ORF1) has a short 3'-UTR (i) and its polyadenylation does not depend on *trans*-splicing of the second mRNA (ORF2), which contains a short 5'-UTR (ii). If only the upstream splicing signal A is used (e), the first mRNA (ORF1) bears a short 3'-UTR (i) while the second (ORF2) has an extended 5'-UTR (iv). If the extended 5'-UTR contains an AUG and an open reading frame, the second mRNA (ORF2) will be weakly translated. When the downstream splicing signal B is used (f), the first mRNA (ORF1) has an extended 3'-UTR (v) with a retained splicing signal A, which could be bound by alternative splicing factors and exported with the mRNA out of the nucleus. The extended 3'-UTR could also contain additional regulatory sequences that were not present in the 3'-UTR obtained in the previous possibilities (i). There is no evidence

why some processing signals are preferred over the others but reporter assays showed that when several polypyrimidine tracts are present in an intergenic region, the closest to the downstream AUG is usually preferred (Hartmann et al. 1998; López-Estraño et al. 1998)

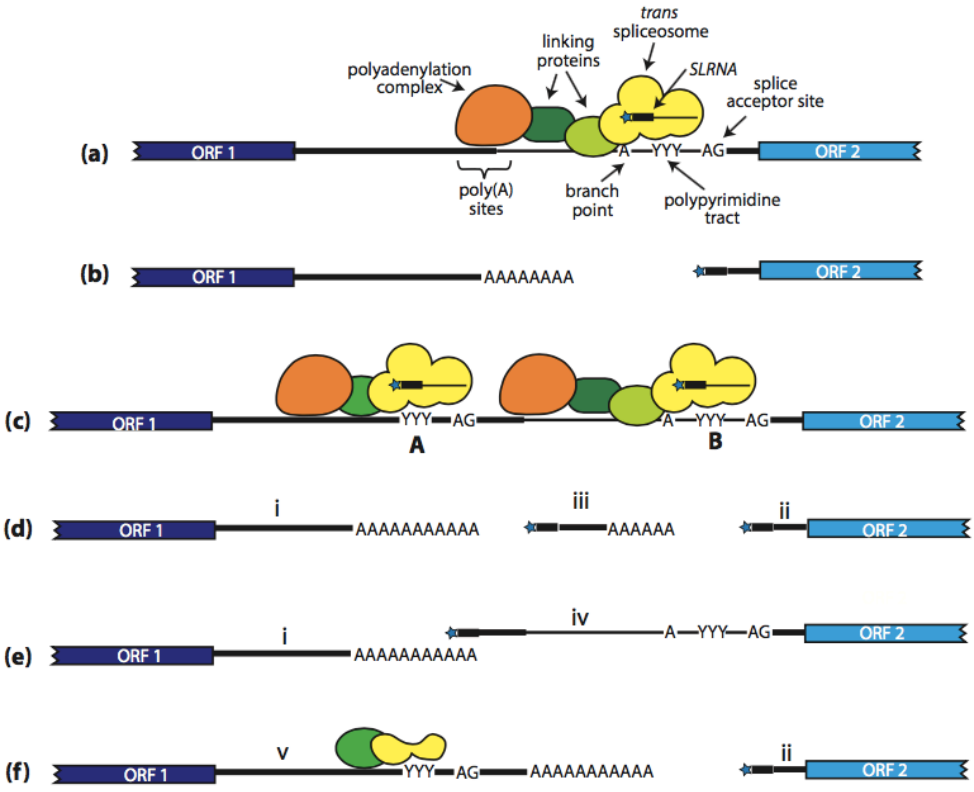


Figure 1.4: mRNA processing in trypanosomes. (a) A cartoon showing the region between two open reading frames, ORF1 and ORF2 with all the processing sites. Hypothetical proteins of the splicing and polyadenylation complex are illustrated; the nature of interaction between both complexes is unknown. (b) Illustration of the simplest way to process the primary transcript, when there is only one single splicing signal: two mRNAs are generated. (c) A precursor mRNA with two possible splicing signals, shown as A and B. (d) Cartoon showing the mRNAs resulted from a primary transcript where both splicing signals are used: two mRNAs with ORF1 and ORF2 are generated; an addition processed RNA with no open reading frame from the intergenic region. (e) Two mRNAs are generated when only splicing signal A is used: the second mRNA (ORF2) has a long 5'-UTR with a splicing signal. (f) If the splicing site B is used, two mRNAs are produced: the first (ORF1) has a long 3'-UTR which contains the unused splicing signal. Picture adapted from (Clayton 2019).

Many proteins are involved in splicing, with the Sm proteins being a crucial part of the *trans*-splicing machinery. They associate with the SL RNA and the uridine-rich small nuclear RNAs (U snRNAs)(Palfi et al. 2009). A part from the basal splicing factor U2AF65, which recognizes and binds the polypyrimidine tract and the branch point; two serine-arginine rich proteins, TSR1 and TSR-IP, have been identified as playing a role in *T. brucei* splicing (Gupta et al. 2014). Several other RNA-recognition motifs containing proteins were described as splicing

factors, including DRBD3 (PTB1), DRBD4 (PTB2), HNRNPFH (De Gaudenzi et al. 2005; Gupta et al. 2013) but they are also involved in mRNA stability (discussed later). Among proteins associated with the spliceosome, components of the polyadenylation were also identified such as the polyadenylation factor CPSF73 (Clayton and Michaeli 2011). The polyadenylation machinery is linked to the spliceosome, and their activity is coupled. When splicing is inhibited, polyadenylation does not take place (Ullu et al. 1993; Matthews et al. 1994).

1.5.3 Nuclear export of mRNA

As soon as mRNAs are transcribed in the nucleus, they are bound by a wide variety of RNA-binding proteins and modifying proteins to form messenger ribonucleoproteins (mRNPs) complexes, which are then exported from the nucleus to the cytoplasm. In trypanosomes, the nuclear export is done by a RanGTP-dependent mechanism. Similar to other eukaryotes, trypanosomes have a TRAMP-like complex and a nuclear RNA exosome, which degrade unspliced RNA (Etheridge et al. 2009; Kramer et al. 2016). They also have the major nuclear export factor MEX67 (Schwede et al. 2009; Dostalova et al. 2013); the helicase MTR2/p15, the Gle2/Rae1 and the Yra1/Sub2 adaptor (Serpeloni et al. 2011). The MEX67 complex includes a transportin-like protein, IMP1 which is required for mRNA export (Dostalova et al. 2013).

In contrast to yeast, the trypanosome MEX67 contains a N-terminal CCCH-domain, which is essential for function and cell survival. Depletion of both MEX67 and its interacting partners MTR2 and importin 1 is lethal in trypanosomes and leads to retention of poly (A)+ mRNAs in the nucleus (Dostalova et al. 2013). In addition, or in parallel to the MEX67/MTR2 pathway, trypanosomes also use the export factors NMD3 and XPO1, mainly for the export of ribosomal RNAs and procyclin-associated gene (PAG) transcripts, which are transcribed by RNA polI (Bühlmann et al. 2015). Homologues of the yeast proteins that act as export adaptors or control factors at the nuclear basket are absent in trypanosomes (Serpeloni et al. 2011; Obado et al. 2017).

In addition, the nuclear pore structure is also different from that of Opisthokonts. The outer components of the nuclear pore are more diverged, giving the pore a more symmetrical structure (Obado et al. 2016). Like in other eukaryotes, control of nuclear export is also poorly understood in trypanosomes. Recently, Goos et al. 2019 found that trypanosomes can initiate nuclear export before they have completed transcription; suggesting that it may depend on the recognition of the cap at the 5'-end (probably through the cap-binding proteins) or through some subunits of the exon junction complex (Y14, Magoh and an NTF2 like protein). DRBD18 is an RNA-binding protein that was recently found to be implicated in nuclear export in trypanosomes.

Mishra et al. 2021 reported that DRBD18 interacts directly with MTR2 and promotes the nuclear export of a subset of mRNAs in procyclic forms. RNAi-mediated depletion DRBD18 in procyclic forms caused partial nuclear retention of MTR2 and MEX67 and different effects on cytoplasmic and whole cell mRNA populations (Mishra et al. 2021).

1.5.4 Translation

Once in the cytosol, the fate of each mRNA is determined by dynamic and combinatorial effect of RNA-binding proteins that are recruited via the *cis*-acting elements present in the 3'-untranslated regions or via protein-protein interactions (Clayton 2019). The mRNA can then be efficiently translated, silenced or degraded in the cytosol. General eukaryotic translation starts when the initiation factor eIF4E binds to the 5'-cap and interacts with the scaffolding protein eIF4G, which in turn recruits the RNA helicase eIF4A. eIF4E, eIF4G and eIF4A form the eIF4F complex (Gingras et al. 1999). The scaffolding protein eIF4G interacts with the poly(A) binding protein (PABP) bound to the poly A tail at the 3'-end of the mRNA. This allows circularization of the mRNA, which facilitates translation efficiency and ribosome recycling (Sonenberg and Hinnebusch 2009). eIF4G also provides binding sites for several regulator factors such as the MAP kinase-interacting serine/threonine kinases 1/1 and eIF3. Through its interaction with eIF3, eIF4G recruits the 40S ribosomal subunit, which together with the Met-tRNA and the others initiation factors form the 43S pre-initiation complex (PIC). eIF4A unwinds secondary structures present in the 5'-UTR and together with eIF4B, facilitates the movement of the PIC along the target mRNA, a pre-requisite for the scanning of the 5'-UTR and the identification of the first AUG, which defines the beginning of the protein-coding region (Hinnebusch and Lorsch 2012). After the AUG initiation, the eIF4F is released, the large ribosomal subunit is recruited and the first bond of the nascent polypeptide is formed (Sonenberg and Hinnebusch 2009).

T. brucei has six isoforms of eIF4E (EIF4E1-6), five eIF4G homologs (EIF4G1-5) and two eIF4A homologs (EIF4A1-2), of which only one is involved in translation (Freire, Malvezzi, et al. 2014; Dhalia et al. 2006). Two paralogs of poly (A) binding proteins, PABP1 and PABP2 are also present in *T. brucei* genome (Zoltner et al. 2018). PABP1 interacts with a small subset of proteins and does not localize to granules after starvation. In contrast, it is preferentially localized in a fraction of small polysomes. On the other hand, PABP2 interacts with a wide range of proteins and is translocated to granules during starvation (Zoltner et al. 2018). Not all EIF4E homologs are involved in translation initiation and their sequences are very diverged from those in other eukaryotes, maybe to accommodate for the hypermethylated cap 4 structure.

They can be divided in three different groups. Group I is formed of EIF4E1 and EIF4E2, which cannot form the EIF4F-like complexes and therefore are not involved in general translation (Freire et al. 2017). *T. brucei* EIF4E1 does not interact with any EIF4G but instead associates with 4E-binding protein (4-EIP), which is involved in translation repression and is required for normal differentiation (Terrao et al. 2018). Even in the absence of 4-EIP, EIF4E1 does not associate with any other translation initiation factors; instead might repress translation through the interaction of 4-EIP with the cytosolic terminal uridylyl transferase, TUT3 (Falk et al. 2021). The role of EIF4E2 is still unclear (Freire et al. 2018). Group II contains EIF4E3 and EIF4E4, which can form EIF4F complexes and are involved in general translation initiation in trypanosomes (Freire et al. 2017). EIF4E3 interacts with EIF4G4 while EIF4E4 associates with EIF4G3; both complexes are associated with EIF4A1. In addition, the EIF4E4-EIF4G3 complex interacts with both PABP1 and PABP2 (Zoltner et al. 2018). Group III is made of EIF4E5 and EIF4E6, which interact with EIF4G and form EIF4F complexes. EIF4E5 associates with both EIF4G1 and EIF4G2 (Freire, Vashisht, et al. 2014) while EIF4E6 interacts with EIF4G5 (Freire, Malvezzi, et al. 2014).

Analysis of protein synthesis by ribosome profiling in *T. brucei* reveals that translational efficiency varies between different life cycles and for up to 100-fold between genes. This implies that translation to a similar extent as mRNA stability contributes to the regulation of gene expression (Vasquez et al. 2014). Since there was a poor correlation between mRNA abundance and translation efficiency, it is thought that the latter can be regulated independently of mRNA stability (Vasquez et al. 2014; Urbaniak et al. 2012).

1.5.5 mRNA decay

Translation and degradation of mRNAs play an important role in control of gene expression; and the two mechanisms compete with each other. In eukaryotes, degradation of mRNA usually starts with shortening of the 3'-poly A tail by the major deadenylation complex, the CCR4/CAF1/NOT complex; followed by removal of the 5'-cap by the decapping enzyme, which consists of two subunits, Dcp1 and Dcp2. Afterwards, the mRNA can be degraded from both ends: in the 3'-5' direction by the exosome and from the 5'-end by the major cytoplasmic exoribonuclease, XRN1 (Parker and Song 2004). Other eukaryotic deadenylating complexes, the PARN and the PAN2/PAN3 complexes have also been reported to play secondary or specialized roles in mRNA decay (Yamashita et al. 2005; Parker and Song 2004).

Trypanosomes have conserved all types of deadenylation complexes: the CAF1/NOT complex (Schwede et al. 2008; Färber et al. 2013), the PAN2/PAN3 (Schwede et al. 2009; Fadda et al. 2013) and they have three proteins related to PARN (Utter et al. 2011). The CAF1/NOT complex of trypanosomes is much simpler with no CCR4 homologs (Erben et al. 2013).

T. brucei also has an exosome responsible for the 3'-5' degradation (Estévez et al. 2001; Haile et al. 2003) and a cytosolic exoribonuclease (XRNA) which digests the mRNA from the 5'-ends. Three other exoribonucleases are present in trypanosomes (XRNB, XRNC and XRND), but their precise role is not known (Li et al. 2006). Trypanosomes lack homologs of decapping enzymes. Instead (Kramer 2017a) showed that the 5'-caps are removed by ALPH1, an ApaH-like phosphatase which leaves the mRNA with a 5'-diphosphate. RNAi-mediated depletion of ALPH1 causes an increase in mRNAs that are deadenylated but have not yet been degraded by the 5'-3' decay.

Analysis of mRNA decay by transcription inhibition coupled to RNA-Seq in trypanosomes revealed that in bloodstream forms, the median half-life of mRNAs is 12 minutes while in procyclics the mRNAs have on average half-lives of 20 minutes (Fadda et al. 2014). Developmentally regulated mRNAs showed more often regulated decay rates. Fadda et al. 2014 also found that steady state mRNA levels are determined by both splicing and polyadenylation; however, this seems to depend on the competition between the rates of RNA processing and co-transcriptional mRNA precursor destruction.

1.5.6 Localization of mRNAs

Another mechanism for controlling gene expression at the post-transcriptional level is the storage of mRNAs that are not engaged in translation in membranes free granules/aggregates in association with different proteins. Different types of such aggregates exist in trypanosomes (Cassola 2012; Kramer 2014). Examples include processing bodies (p-bodies) and stress granules. P-bodies were thought to be sites of mRNA degradation because they contain enzymes of 5'-3' degradation pathway and proteins of translation initiation repression (Decker and Parker 2012) including the decapping enzymes, the enhancer of decapping RNA helicase DHH1 and the 5'-3' exoribonuclease XRN1 (Krüger et al. 2013). P-bodies are constitutively present in unstressed cells and they are proportional to non-translating mRNAs. Their amount increases when global translation is repressed. However, when the dissociation of polysomes is blocked by cycloheximide, the amount of p-bodies decreases (Kramer 2014).

Another type of messenger ribonucleoprotein (mRNP) aggregates are the stress granules that are much larger than the p-bodies. The stress granules are normally induced by different type of stress, such as heat shock or starvation and they typically contain components of translation initiation machinery (Decker and Parker 2012). In trypanosomes, Kramer 2014 showed that starvation stress granules contain proteins such as the DEAD box RNA helicase DHH1, the 5'-3' exoribonuclease homologue XRNA, the two poly(A) binding proteins, PABP1 and PABP2, the scaffold protein SCD6, the U-rich RNA binding protein, the translation initiation factors like EIF4E1, 2 and 3 together with other RNA-binding proteins.

However, ribosomal subunits and the EIF4G scaffold proteins were not detected in trypanosomes stress granules (Cassola 2012). Based on their composition, stress granules are thought to be aggregates of mRNPs stalled in the process of translation initiation. Since they contain polyadenylated mRNAs, it is assumed that they just served as temporal storage of mRNAs during the stress (Cassola et al. 2007).

1.6 Role of RNA-binding proteins in regulating gene expression

RNA-binding proteins (RBPs) play an important role in post-transcriptional regulation of gene expression. They are known to bind specific sequences within the mRNAs and to regulate the localization, stability and translation of the target mRNAs (Clayton 2019).

The majority of regulatory *cis*-acting elements are found in the 3'-untranslated regions (3'-UTRs) of the mRNAs. RBPs are categorized based on the types of structural domains involved in the RNA-binding. *T. brucei* genome encodes for over 75 RNA-recognition motif (RRM) proteins (Wurst et al. 2009); about 48 CCCH-type zinc finger domain proteins (Kramer et al. 2010); 4 proteins with acetylation lower binding affinity (ALBA) domain (Mani et al. 2011) and approximately 11 pumilio/fem-3 domain (PUF) proteins (Caro et al. 2006).

1.6.1 ALBA and Pumilio domain proteins

All the four *T. brucei* ALBA domain proteins can form homo and heterodimeric complexes. All are cytoplasmic and regulate translation of a reporter mRNA in a sequence-specific fashion (Mani et al. 2011). ALBA1/2 and ALBA3/4 dimers can localize in starvation granules with poly(A) RNA. In addition, dimers of ALBA2/3 interact with the translation initiation factor EIF4E4 and partially associate with polysomes (Mani et al. 2011). Localization of ALBA3/4 in stress granules was further confirmed by Subota et al. 2011 who showed that the dimers co-localize with DHH1 and poly(A) RNA. Subota et al. 2011 also found that ALBA3/4 is expressed throughout the developmental stages of trypanosomes in the tsetse fly, except during the mesocyclic to epimastigote transitions in proventriculus. A recent study by Bevkal et al. 2021 showed that ALBA3 and ALBA4 are redundant in bloodstream forms but only ALBA3 is essential for the differentiation of the stumpy form to the procyclic forms in the tsetse host.

Among the 11 pumilio-domain proteins encoded by trypanosome genome, PUF7, PUF8 and PUF10 are localized in the nucleolus and the rest are cytosolic. PUF7 and PUF10 are essential for rRNA maturation and together with their interacting partners BOP1 and NRG1, they regulate the mRNA and expression levels of GPEET (Droll et al. 2010; Schumann Burkard et

al. 2013). PUF9 stabilizes a small subset of mRNAs that show a peak in the mid-to-late S-phase and contain the motif UUGUAC in their 3'-UTRs (Archer et al. 2009). PUF2 is required for cell growth (Jha et al. 2014) while PUF5 appears not to be (Jha et al. 2013). In a study by Hoek et al. 2002, PUF1 was found to interact with the expression-site associated gene 8 (ESAG8) mRNA and protein. PUF1 also preferentially binds retroproson RNAs (Luu et al. 2006); and PUF3 specifically binds mRNAs containing the classical pumilio domain motif UGUA[U/C]AUU (Marucha and Clayton 2020).

1.6.2 Zinc-finger domain proteins

Zinc-finger domain proteins are the second largest group of RNA-binding proteins in trypanosomes. As for now, a few of them have been characterized and research work is still needed to characterize the role of the others. The proteins ZFP1, ZFP2 and ZFP3 mutually interact. ZFP1 is upregulated during differentiation from bloodstream forms to procyclics and highly expressed in procyclics while ZFP2 and ZFP3 are constituvely expressed (Paterou et al. 2006; Hendriks et al. 2001). Bloodstream form cells lacking ZPF1 and cells depleted of ZFP2 are unable to differentiate to procyclics (Hendriks et al. 2001; Hendriks and Matthews 2005). Overexpression of ZFP2 in procyclics as well as the ectopic expression of ZFP3 in bloodstream forms result in the "nozzle" phenotype, which is caused by polar extension of cytoskeleton in early stage cells (Hendriks and Matthews 2005).

On the same note, the proteins ZC3H20, ZC3H21 and ZC3H22 all share a region containing two CCCH-zing finger domains. Their mRNAs contain respectively seven, five and two copies of the RBP10 consensus recognition motif, $UA(U)_6$ in their 3'-UTRs (Mugo and Clayton 2017). ZC3H20 is present in low levels in bloodstream forms and it becomes more abundant in procyclics while ZC3H21 and ZC3H22 are only expressed in procyclics (Dejung et al. 2016). ZC3H20 and ZC3H21 are activators of gene expression and essential for trypanosome procyclic cell growth (Erben et al. 2014; Ling et al. 2011). They both also have a variant of the MKT1-binding motif (Y/W/T)(R/T/Q)H(N/D)PY near their C-termini (Liu et al. 2019). Liu et al. 2019 showed that ZC3H20 is not only essential for growth of the procyclics but it is also required for the expression of the stumpy marker, PAD1 and for differentiation of stumpy to procyclic forms. Cayla et al. 2020 showed that the regulation of trypanosome quorum sensing by ZC3H20 requires the integrity of its TbDYRK phosphorylation site. Liu et al. 2019 also revealed that both ZC3H20 and ZC3H21 stabilize a number of mRNAs encoding membrane and mitochondrial proteins. Stabilization of the bound mRNAs is probably through recruitment of the complex containing MKT1, PBP1 and PABP2 (which is discussed below) (Melo do Nascimento et al. 2020). ZC3H22, on the other hand is a repressor of gene expression and its specific role is still unknown (Erben et al. 2021). Finally, ZC3H11, which contains a single CCCH zinc finger domain, is essential for growth in bloodstream forms and its depletion in procyclic forms renders cells susceptible to heat shock (Droll et al. 2013).

ZC3H11 binds and stabilizes mRNAs encoding trypanosomes chaperones, required for protein refolding after heat shock. The N-terminal domain of ZC3H11 specifically binds the UAA repeats in the 3'-UTRs of the bound mRNAs while the C-terminal domain has the stabilizing activity (Droll et al. 2013). ZCH311 was the first RNA-binding protein from which a mechanism of action was shown: It recruits the MKT1 protein via the interaction motif, HNPY. MKT1, in turn interacts with PBP1, which recruits LSM12 and the PABPs (Singh et al. 2014). Recent work from our lab showed that the expression activator 1 (XAC1) protein is associated not only with the MKT1 complex (MKT1-PBP1-LSM12-PABP) but also with a complex containing an MKT1-like protein (MKT1L-PBP1-LSM12-PABP) (Melo do Nascimento et al. 2020). This study reveals that MKT1 complex specifically recruits the PABP2 and the cap binding translation initiation complex EIF4E6-EIF4G5. MKT1 interacts with various numerous RNA-binding proteins often (not always) via the (Y/W/T)(R/T/Q)H(N/D)PY motif and it is therefore believed that such proteins would stabilize their target mRNAs and or promote their translation (Melo do Nascimento et al. 2020; Melo do Nascimento et al. 2021).

1.6.3 RNA-recognition motif (RRM) domain containing proteins

The RRM is one of the most commonly abundant protein domains in eukaryotes (Cléry et al. 2008). The motif consists of four-stranded β-sheets and two α-helices, which can bind 2-8 nucleotide sequences on single stranded RNA, but it can also interact with DNA and other proteins (Nagai et al. 1990; Lunde et al. 2007). In trypanosomes, approximately half of RRM-containing proteins are required for cell growth in at least one life cycle. However, only few have orthologues in eukaryotes. These include PABP1, PABP2, the translation initiation factor eIF3B and the splicing factors, U2AF35 and HNRNPFH (Kramer and Carrington 2011). Many others are unique to Kinetoplastids but only few of them have been characterized.

The uridine binding proteins, UBP1 and UBP2, have a single RRM and have been shown to affect the mRNA levels of CFB1 and CFB2 (proteins containing the cyclin F-box domain), which are involved in cell cycle in bloodstream forms (Hartmann et al. 2007). RBP42 is partially associated with polysomes and it binds preferentially the coding sequences of the target mRNAs, including some encoding proteins involved in energy metabolism (Das et al. 2012). DRBD3 (also called PTB1) and DRBD4 (known as PTB1) contain each two RRMs and they are essential for trypanosome growth. Both of them play a role in *trans*-splicing of genes containing C-rich polypyrimidine tract (Stern et al. 2009). In addition, DRBD3 regulates the stability of a subset of mRNAs (Estévez 2008). HNRNPFH contains a single RRM and is localized both in the nucleus and the cytosol. It acts as a repressor of *trans*-splicing and it affects mRNA stability in bloodstream and procyclic forms (Gupta et al. 2013). A purine-rich motif AAGAA was found to be enriched in the mRNAs affected by HNRNPFH (Gupta et al.

2013). DRBD18, another double RRM-containing protein is expressed both in bloodstream and procyclic forms (Dejung et al. 2016). It is an abundant protein in procyclics where it is required for normal growth. DRBD18 is a substrate of the arginine methyltransferase, PRMT1. Depending on the degree of methylation, DRBD18 was found to be associated with different ribonucleoprotein complexes and methylation was shown to trigger mRNA stabilization by DRBD18 (Lott et al. 2015). As mentioned earlier, DRBD18 is also involved in the nuclear export of a subset of mRNAs in procyclic forms (Mishra et al. 2021).

Finally, two RRM-containing proteins, RBP6 and RBP10 are involved in the regulation of trypanosome differentiation (Kolev et al. 2012; Wurst et al. 2012; Mugo and Clayton 2017). Overexpression of RBP6 in procyclic forms for ten days leads to differentiation to epimastigotes and further to infective metacyclic cells. After infecting the mice with metacyclic forms, the cells differentiate to bloodstream form cells. However, *in vitro*, the metacyclic forms were not able to progress to bloodstream form stage (Kolev et al. 2012). Another study by Shi et al. 2019 showed that an overexpression of the RBP6 with a single point mutation (Q109K) in procyclic forms was able to promote differentiation not only to metacyclic but also to bloodstream forms *in vitro*. This differentiation skipped the epimastigote forms.

RBP10, which is the focus of my doctoral thesis, is a cytosolic RBP with a single RRM domain. It is exclusively expressed in long slender bloodstream forms where it is essential (Mugo and Clayton 2017). In stumpy forms, *RBP10* mRNA and protein levels are very low and they become undetectable in procyclic forms (Dejung et al. 2016; Silvester et al. 2018). RNAi-mediated depletion of RBP10 in bloodstream forms leads to a downregulation of bloodstream form specific mRNAs with an upregulation of mRNAs that are abundant in procyclics (Wurst et al. 2012). Further studies by Mugo and Clayton 2017 showed that tethering the RBP10 C-terminus to a reporter mRNA mediates degradation of the mRNA and its translation repression. There was no evidence that RBP10 associates with any known mRNA degradation complex. However, RBP10 was shown to mediate the negative regulation in bloodstream forms. When RBP10 was depleted for 17 hours in bloodstream forms and further transfer to conditions adequate for procyclic forms growth, the cells were able to grow as procyclics forms. In addition, overexpression of RBP10 in procyclics for 24 hours promotes differentiation to bloodstream forms. RNA immunoprecipitation coupled to deep sequencing (RIP-Seq) analysis shows that RBP10 binds procyclic specific mRNAs containing the recognition motif UAUUUUUU in their 3'-UTRs. These include the mRNAs encoding EP procyclin, the zinc finger domain proteins ZC3H20 and ZC3H22, as well as several regulatory proteins. The authors suggest that RBP10 binds procyclic specific mRNAs, by recognizing the above motif, promoting mRNA degradation and translation repression.

1.7 Role of the 3'-UTRs in post-transcriptional regulation

The 3'-UTRs regulate the stability, degradation, translation and localization of mRNA through the binding of specific RNA-binding proteins. The RBPs bind to *cis*-elements present in the 3'-UTRs of the mRNAs and they interact with other proteins, which in turn recruit either the degradation machinery or the translation complex (Mayr 2017). Therefore, the functions of a given 3'-UTR depend on the RBPs that directly bind to it and the other effector proteins that are in its vicinity at a given state. For example, a protein that binds to a given 3'-UTR and then recruits the deadenylation complex, might destabilize the mRNA. On the other hand, a protein that interacts with the PABPs might promote stability and translation of the bound mRNA. The binding of RBPs to a particular regulatory *cis*-element is dynamic and depends on the local environment regulating for example posttranslational modifications, the cell type and the cellular state, the expression of other RBPs and the interactions with membranes and cytoskeleton (Kocabas et al. 2015). In addition to that, the secondary and tertiary RNA structure formation influence the accessibility of the 3'-UTRs and regulate the binding of RBPs to the *cis*-elements (Mukherjee et al. 2019; Khong and Parker 2020).

The regulatory proteins also compete for binding on the same motifs present in the 3'-UTRs of mRNAs. As an example, the AU-rich elements (ARE) in the 3'-UTRs of mammalian cells are known to mediate mRNA degradation (Beisang and Bohjanen 2009). Binding of the proteins tristetrapolin (TTP), Brf-1, the KH-type splicing regulatory protein (KSRP) and some isoforms of the poly(U) binding degradation factor 1 (AUF1) to the AU-sequences results in mRNA decay. The TTP protein recruits the CCR4-CAF1-NOT and the decapping complexes, necessary for degradation. However, the HuR family proteins prevent binding of the destabilizing proteins and they instead stabilize the mRNA. Some 3'-UTRs are also involved in mRNA localization: one example includes the *ASH1* mRNA which contains four *cis*-elements, bound by two RBPs, She2p and She3p, which binds a myosin motor protein, Myo4p, responsible for transport of the mRNP along the actin microfilaments to the bud tip (Niednery et al. 2014).

As in other eukaryotic cells, identification of *cis*-elements in trypanosomes has relied on reporter assays, in which deletions within a particular 3'-UTR are tested for their ability to mediate regulation. This approach is prone to mis-interpretations since fragmentations of the UTR can compromise the RNA conformation and therefore the protein binding (Clayton 2014). Nevertheless, some reporter assays have given great insights into the role of 3'-UTR in trypanosomes. For instance, the *EP procyclin* mRNA contains both a 16-mer and a 26-mer elements in its 3'-UTR that contribute to the stability of the mRNA and translation efficiency in procyclic forms (Hehl et al. 1994; Hotz et al. 1997). In addition, the 26-mer element, which consists of a double repeated U-rich motif, UAUUUUUUU, mediates the procyclin mRNA

degradation and translation suppression in bloodstream forms (Schürch et al. 1997; Quijada et al. 2002). This motif $UA(U)_6$ was also implicated in the negative regulation of mRNAs encoding the B isoform of phosphoglycerate kinase (Quijada et al. 2002), the pyruvate phosphate dikinase (Quijada et al. 2002), and the cytochrome oxidase subunit COX V (Mayho et al. 2006) in bloodstream forms. The *GPEET* mRNA contains a *cis*-element in the 3'-UTR, which confers regulation in response to glycerol and hypoxia (Vassella et al. 2000). For bloodstream-form specific mRNAs, the only motif that had been implicated in regulation is a 16-mer element in the 3'-UTR of the variant surface glycoprotein (VSG), responsible for mRNA abundance and stability in bloodstream forms (Ridewood et al. 2017; Berberof et al. 1995). The 16-mer is also required for the m^6A modification of the poly(A) tail, which plays a significant role in the VSG mRNA stabilization (Viegas et al. 2020).

Another way of identifying regulatory sequences in the 3'-UTRs is by finding mRNA targets of a specific RNA-binding protein through RNA immunoprecipitation. The protein is cross-linked *in vivo* to RNAs using the UV light, then it is affinity purified and the bound RNAs are sequenced. The binding motifs on the 3'-UTRs of the target mRNAs can also be identified by addition of RNAse to digest any RNA that is not protein-protected. Our lab has revealed for example, after affinity purification of RBP10 that it binds procyclic-specific mRNAs containing the $UA(U)_6$ motif. Through reporter assays, it was confirmed that RBP10 mediate translation repression of the *EP procyclin* mRNA by binding to the same motif (Mugo and Clayton 2017).

After finding regulatory elements in the 3'-UTRs of mRNAs, the most difficult task is to identify the proteins that bind to them. When the regulatory element is less than 30 nt, the sequence can be used for affinity purification. As an example, a 25-nt element in the GPEET 3'-UTR is involved in the regulation of the *GPEET* mRNA during differentiation of the parasite from early to late procyclic form (Vassella et al. 2004). This RNA sequence element was incubated with procyclic extracts and proteins interacting specifically with the 25-nt element were identified by mass-spectrometry. The ALBA proteins were specifically enriched in the pulldown although the authors did not find evidence of regulating the *GPEET* mRNA (Mani et al. 2011). An ideal approach for identification of regulatory proteins is to cross-link the proteins to RNAs *in vivo*, then purify the mRNA using either an affinity column bearing oligonucleotides which hybridize to the mRNA or specific tags such as biotinylated oligomers that hybridize to the mRNAs and allow purification using the streptavidin beads (Theil et al. 2019). Using this method, Rico-Jiménez et al. 2021 found proteins that bind to the purine responsive element within the 3'-UTR of the NT8 nucleobase transporter mRNA in trypanosomes. One could also fuse the sequence of interest to an RNA aptamer that directly binds a ligand (Leppek and Stoecklin 2014). One can also make use of the tetracycline-induction of the biotin ligase tethered to the specific regulatory sequence for capturing proteins in the nearby environment (Ramanathan et al. 2018).

All these approaches work well for abundant, stable RNA-protein complexes. When the mRNA is unstable or in fewer copies, the method would not be appropriate. Using the RNA affinity purification with biotinylated oligonucleotides complementary to the *VSG* mRNA, Melo do Nascimento et al. 2021 identified proteins that are associated with the *VSG* mRNA relative to the *tubulin* mRNA. An unconventional RNA-binding protein with an F-box domain, CFB2 was reported to bind to the 16-mer of the *VSG* mRNA, mediating stability of the mRNA through recruitment of the MKT1 complex (MKT1-PBP1-LSM12), the PABP2 and the translation initiation factor EIF4E6/EIF4G5.

As of now, no other specific sequences have been identified for bloodstream-form specific mRNAs. The RBP10 binds the $UA(U)_6$ of the procyclic-specific mRNAs, targeting them for destruction and translation repression. By regulating other mRNAs, it turns out that the RBP10 must also be tightly regulated. The protein is not detectable in stumpy or procyclic forms. The levels of the mRNA encoding RBP10 are also relatively low in procyclics (Wurst et al. 2012) and probably also in the salivary gland trypanosomes (Telleria et al. 2014). Therefore, regulation seems to be not only at the mRNA level but also at the translational level but the mechanisms involved in this regulation are not known. The *RBP10* mRNA has an unusual long 3'-UTR of approximately 7.3 kb. This feature is typical to mRNAs encoding RNA-binding proteins (Clayton 2019). This 3'-UTR might harbor multiple regulatory sequences, with specific conformations. Recruitment of different mRNP complexes via the regulatory elements might influence stability and/or translation of *RBP10* mRNA in the different stages of trypanosome life cycle. Finding the responsible *trans*-acting factors through an RNA purification will provide more information of the developmental regulation of the protein.

1.8 Aims of the study

The main goal of my PhD project was to understand how the expression of the master regulator of differentiation, the RNA-binding protein 10, is regulated in the life cycle of *Trypanosoma brucei*.

The specific objectives assigned to this work included:

- Identification of regulatory sequences responsible for the developmental regulation of RBP10 using reporter assays;
- Investigation of the role of RNA-binding proteins in regulating the expression of RBP10 through an RNA pulldown assay and RNA interference;
- Investigating the mechanisms of action of proteins regulating the expression of RBP10 using deep RNA sequencing (transcriptomics) and mass spectrometry.

2. Material and Methods

2.1 Trypanosome cell culture methods

Some experiments in this study were carried out using monomorphic *T. brucei* Lister 427 bloodstream or procyclic forms. For differentiation experiments, the pleomorphic cell line EATRO 1125 was used. All cell lines constitutively express the tetracycline repressor but are referred to as wild type (Alibu et al. 2005).

2.1.1 Bloodstream form cell culture

Monomorphic Lister 427 bloodstream-form trypanosomes were cultured at densities between 0.1-1.5x10^6/ml in supplemented HMI-9 medium at 37°C with 5% CO_2, in vented flasks. The pleomorphic cell lines "EATRO 1125" were cultured in the same conditions and diluted every day to keep the cell density below 5x10^5/ml. For long culture periods, these cells were maintained in HMI-9 medium containing 1.1% of methyl cellulose (Sigma, M0512) to preserve their pleomorphism. All work was carried out under sterile conditions in a laminar flow hood. The cell density was determined by loading 10 µl of the culture into an improved Neubauer counting chamber and counting cells in 16 squares. The number was then multiplied by 10, 000 to give the number of parasites present in 1 ml of the culture.

Composition of HMI-9 medium (Hirumi and Hirumi 1989)

HMI-9 medium contained: 17.66 g/l Iscove's modified Dulbecco's medium, 3.024 g/l NaHCO3, 136 mg/l hypoxanthine, 110 mg/l sodium pyruvate, 39 mg/l Thymidine, and 28 mg/l Bathocuprono disulfonic acid disodium salt (pH 6.3). The media was filter sterilized and stored at 4°C in 450 ml aliquots. Prior to use, the media was supplemented with the following components: 10% (v/v) fetal bovine serum (heat inactivated at 55°C for 30 min), 50 U/ml penicillin-streptomycin (Gibco), 1.5 mM L-Cysteine-HCL.H_2O and 7 µl of 0.2mM ß-mercaptoethanol (14.1 M stock, Sigma).

Before diluting the cells, this complete media was pre-warmed in a 37°C water bath for half an hour.

HMI-9 medium with 1.1% methylcellulose, 1 litre:

To make 1 liter of HMI-9 medium containing 1.1% of methylcellulose, the following ingredients were mixed:

465 ml of 2x HMI-9 medium
435 ml of 2.5% Methylcellulose (Sigma)
10% (v/v) fetal bovine serum (heat inactivated at 55°C for 30 min)
50 U/ml penicillin-streptomycin (Gibco)
1.5 mM L-Cysteine-HCL.H_2O
0.2mM ß-mercaptoethanol (14.1 M stock, Sigma)

2.5 % methylcellulose was prepared in 500 ml of distilled water, then dissolved by stirring at 4°C for 2 days. The media was then sterilized by autoclaving, thereafter stirred at 4°C to re-dissolve. The volume was topped up to 600 ml using sterile water under sterile conditions, then further stirred at 4°C (1-2 days) to make sure that it is completely dissolved.

2.1.2 Procyclic form cell culture

Monomorphic Lister 427 and pleomorphic procyclic-form trypanosomes were cultured in supplemented MEM-Pros medium at 27°C. The caps of the cell culture flasks were tightly closed and the density of the cells was maintained between 0.5-2 x 10^6 cells/ml. As mentioned above, the cell concentration was also determined using a Neubauer counting chamber and the culture work was done in a laminar flow hood.

Composition of MEM-Pros Medium

MEM-Pros medium contained: 16.55 g/l MEM-pros mixture (Biochrom), 1% (v/v) MEM non-essential amino acids (sigma), 1% (v/v) MEM vitamins (sigma) and 10mg/l phenol red. The pH was adjusted to 7.4 using NaOH, and thereafter the media was filter sterilized and stored in 450 ml aliquots either frozen at -20°C or at 4°C. Prior to use, the media was supplemented with 10% (v/v) fetal bovine serum (heat inactivated at 55°C for 30 min), 50 U/ml penicillin-streptomycin (Gibco) and 7.5 mg/L Hemin. The complete media was briefly warmed in a 37°C water bath and then placed in the 27°C room for 20 minutes before diluting the cells.

2.1.3 Transfection of bloodstream/procyclic trypanosomes

Prior to transfection, 20-30 µg of the construct was linearized in a 50-100 µl restriction enzyme digestion. Linearization was confirmed on a 1% w/v agarose gel prepared with 1x TAE (Tris-acetate EDTA) buffer. The DNA construct was then ethanol precipitated with 2.5 volumes of ice-cold 100% ethanol and 1/10 volume of 3M Sodium acetate at -20°C for 2 hours. After centrifugation at 13,000g for 15 minutes, the DNA construct was washed twice with 70% ethanol and 98% ethanol, then air-dried and resuspended in 20 µl of distilled water.

For each transfection, approximately 1-2 x 10^7 cells growing at logarithmic phase were collected after centrifugation at 2300 rpm for 8 minutes. For bloodstream forms, the cell pellet was resuspended in 130 µl of AMAXA buffer (components listed in Table 2.1) and mixed with 5-10 µg of the linearized plasmid. The solution was transferred to an electroporation cuvette (2 mm electrodes gap, Peqlab). Transfection was carried out using an electroporation machine with the program X-001 (Amaxa Biosystems, Nucleofactor II) as described in Burkard et al. 2007. For the procyclic forms, the cells were washed twice with 0.4 ml of the ZPFM buffer (components listed in Table 2.2), resuspended with 0.4 ml of the same transfection buffer, then mixed with 5-10 µg of the linearized plasmid. The mixture was transferred to an electroporation cuvette (2 mm electrodes gap, Peqlab). The transfection was performed using 1.5 kV and resistance R2 of the BTX electroporation machine (Electro Cell manipulator, Harvard apparatus). After transfection, the cells were transferred to 25 ml of pre-warmed complete medium without drugs and they were left for 6-7 hours to recover. Thereafter, the selective antibiotic was added to the cells and they were plated in serial dilution on a 24-well plate. The different antibiotics and the respective concentrations are listed in section 2.1.4. The positive proliferating clones were identified 5 days later for bloodstream forms or 7-10 days later for procyclic forms by microscopy and left to grow in the presence of the respective selective antibiotic. After scaling up the cell numbers, the transgene expression was confirmed using western blotting for at least three independent clones.

Transfection buffers

Table 2.1: Amaxa buffer

Quantity	Ingredients
90 mM	Sodium phosphate buffer (NaH_2PO_4), pH 7.3
5 mM	Potassium chloride (KCl)
0.15 mM	Calcium chloride ($CaCl_2$)
50 mM	HEPES, pH 7.3

Table 2.2: Zimmerman's Post Fusion Medium (ZPFM) buffer

Quantity	Ingredients
132 mM	Sodium Chloride (NaCl)
8 mM	Potassium chloride (KCl)
8 mM	Sodium phosphate (Na_2HPO_4)
1.5 mM	KH_2PO_4
1.5 mM	MgOAcx4H_2O
90 µM	Ca $(OAc)_2$
pH is adjusted to 7.0 with NaOH	

2.1.4 Antibiotics used for selection of transgenic trypanosomes

For selection of transgenic monomorphic bloodstream and procyclic forms, the appropriate antibiotics were added in the following concentrations listed below (Table 2.3). For pleomorphic EATRO 1125 cells, 8 µg/ml hygromycin and 2 µg/ml blasticidin were used for selection. All growth experiments were performed in the absence of antibiotics. For inducible expression of genes, tetracycline was added to a final concentration of 500 ng/ml.

Table 2.3: Antibiotics concentrations

Antibiotics	Bloodstream forms	Procyclic forms
Phleomycin	0.2 µg/ml	0.2 µg/ml
G418	5 µg/ml	15 µg/ml
Hygromycin	15 µg/ml	50 µg/ml
Puromycin	0.2 µg/ml	1 µg/ml
Blasticidin	5µg/ml	10 µg/ml

2.1.5 Storage of trypanosomes cell lines

For storage of trypanosome cells, 500 µl of the cell culture growing in exponential phase were mixed with 500 µl of the respective growth media (HMI-9 or MEM-Pros) containing 20 % of glycerol. The mixture was done in a cryovial tube and stored at -80°C for one or two days. The tubes were then placed into the liquid nitrogen tanks for longer storage. To thaw the cells, the cryovial was picked from the liquid nitrogen and allowed to thaw at room temperature. The culture was transferred into 5 ml of medium appropriate for the life-cycle stage, centrifuged at 2300 rpm for 8 minutes. Afterwards, the cells were again transferred to 5 ml of medium and placed in conditions appropriate for the life-cycle stage. The following day, the appropriate selective drugs for the cell line were added and the experiments were done as the cells grow exponentially.

2.1.6 Growth curves

To make RNAi growth curves in bloodstream forms, the cell cultures growing in logarithmic phase were diluted to $1x10^5$ cells/ml and tetracycline was added to a final concentration of 500 ng/ml to induce the RNAi. Every 24 hours, the uninduced and induced cells were diluted to the starting cell density in order to prevent the parasites from reaching the stationary phase. Tetracycline was also added daily to the induced cultures. In procyclic forms, the cell cultures growing in log phase were diluted to $5x10^5$ cells/ml before overexpressing the

biotin ligase (Myc-BirA*). Tetracycline was also added daily in the same concentration as in bloodstream forms.

2.1.7 Trypanosome differentiation

For differentiation of bloodstream forms to procyclic forms, pleomorphic bloodstream forms were cultured in HMI-9 with 1.1% methylcellulose and left to grow to high density ($3x10^6$ cells/ml). The cell culture was then diluted 1:5 with pre-warmed 1X PBS (phosphate buffered saline), filtered through MN616 1/4 filter papers (Macherey Nagel) followed by centrifugation at 1400 g for 10 min. The cells were resuspended in HMI-9 medium without methylcellulose; the cell density was checked and diluted to $1.0x10^6$ cells/ml. The differentiation was induced using 6mM *cis*-aconitate (Sigma) followed by incubation of the cells at 27°C. After 17h, the cells were harvested (5-7 x 10^5 cells/ml) in MEM-pros medium and maintained at 27 °C for at least two weeks.

2.1.8 Cold shock conditions

Trypanosome bloodstream forms were cooled by incubation in the 27°C room or at 20°C (temperature was monitored with a thermometer) for 24 hours. Cells were taken after 6, and 24 hours.

2.2 DNA methods

2.2.1 Genomic DNA isolation

Genomic DNA from *T. brucei* was isolated using $1-2x10^8$ cells. The cell pellet was collected by centrifugation (2300 rpm, 8 minutes), washed once in cold 1x PBS, and lysed in 0.5 ml of EB buffer (Table 2.4) with addition of 12 µl RNAse A (1mg/ml stock solution). After 1 hour of incubation at 37°C, 200 µl ice-cold 5M ammonium acetate was added to the mixture and vortexed briefly. The precipitated proteins and cell debris were pelleted by centrifugation (maximum speed, 5 minutes) and the supernatant was transferred to a new tube. The DNA was then precipitated with 0.7x isopropanol followed by centrifugation (maximum speed, 15 minutes). The pellet was then washed once with 70 % ethanol to remove salts and then again with 100 % ethanol. The DNA pellet was then dried for approximately 5 minutes and dissolved in TE buffer at 37°C. The concentration was measured using Nanodrop.

Table 2.4: EB Buffer

Quantity	Ingredients
10 mM	Tris-HCl pH 8.0
10 mM	NaCl
10 mM	EDTA
0.5 %	SDS

2.2.2 Plasmid construction and isolation

All the primers used in this study (listed in section 2.7) were ordered from Biomers.net. Full-length or fragments of ORFs, inserts with site-directed mutagenesis were obtained by PCR using either high fidelity Phusion DNA polymerase (NEB) or Q5 DNA polymerase (NEB) following the manufacturer's instructions. Amplification of the *RBP10* 3'-UTR fragments were done with Taq DNA polymerase (NEB). Sequences of the *RBP10* 3'-UTR full length and the *RBP10* 3'-UTR fragments are listed in Supplementary text 1. Products of PCR and restriction digests were separated on 1% agarose gels prepared with 1x TAE buffer and supplemented with ethidium bromide (final concentration 0.2-0.5 ug/ml).

The PCR amplicons and vectors were digested with the appropriate restriction enzymes (NEB or Fermentas) and ligated to a plasmid with the same sticky ends using T4 DNA ligase (NEB). The ligation product was then transformed into competent *Escherichia coli* DH5-alpha cells by heat shock at 42°C for 45 seconds and then put back on ice for 2 minutes. 200 µL of LB broth medium were added to the cells and the mixture was incubated for 1 hour at 37°C with constant shaking. The entire volume was then spread on a LB agar plate containing the appropriate selective antibiotic and grown overnight. Single colonies were picked and grown in LB broth medium with the appropriate antibiotics for 16 hours. Positive clones were identified by colony PCR using GoTaq DNA polymerase (promega) and further confirmed by restriction digests after plasmid purification (Nucleospin Plasmid). One or two plasmids were further selected for the final verification by Sanger DNA sequencing (Microsynth Seqlab). The details of the plasmids and oligonucleotides used in this study are provided in section 2.7. The maps of the plasmids are also given in section 2.8.

2.2.3 Endogenous tagging of DRBD18 and ZC3H28

For the endogenous tagging of DRBD18, a cell line with *in situ* TAP-DRBD18 was generated by replacing one endogenous copy of DRBD18 with a gene encoding a C-terminally TAP tagged DRBD18. For that, a construct with neomycin resistance gene plus a TAP tag cassette was flanked on the 3'-end with a fragment of DRBD18 3'-UTR. Upstream on the 5'-end, the C-terminal region of the DRBD18 ORF without the stop codon was cloned in frame with the

TAP tag. Prior to transfection, the plasmid (pHD 3200) was cut with ApaI and XbaI enzymes to allow homologous recombination. Using the cell lines with *in situ* C-TAP DRBD18, it was not possible to knock-out the other copy of DRBD18.

A cell line with *in situ* TAP-ZC3H28 gene was generated by replacing one endogenous copy of ZC3H28 with a gene encoding N-terminally TAP tagged ZC3H28. For this purpose, a construct with puromycin resistance gene plus TAP tag cassette was flanked on the 5'-end with a fragment of ZC3H28 5'-UTR. Also, downstream on the 3'-end, the N terminal region of ZC3H28 ORF was cloned in frame with the TAP tag. Prior to transfection, the plasmid (pHD3236) was cut with SacI and ApaI enzymes to allow homologous recombination. Using the cell lines expressing the *in situ* N-TAP ZC3H28, the other copy of ZC3H28 could not be deleted.

2.3 RNA methods

2.3.1 RNA extraction, agarose gel electrophoresis and Northern blotting

Total RNA was extracted from ~ $1x10^8$ bloodstream-form or $5x10^7$ procyclic-form cells using PeqGold Trifast (peqLab) according to the manufacturers' instructions. The purified RNA (5-10 µg) was mixed with 2x RNA loading dye (components below, Table 2.5), denatured for 10 minutes at 65°C and then resolved on either formaldehyde agarose gel or urea gel. The RNA was afterwards blotted onto Nylon membranes (Amersham Hybond-N+, GE Healthcare, RPN203B) with 10x SSC buffer (saline-sodium citrate buffer) by capillary transfer overnight. In the case of urea gel, the transfer was done by electroblotting in 0.5x TBE buffer at 600 mA for 1hour. The RNA was then cross-linked to positively charged membranes using a UV-crosslinker (Stratagene UV Stratalinker 2400, 2x240 mJoules) and stained with methylene blue (SERVA) for 10 minutes. The northern blots were pre-hybridized in hybridization solution (Table 2.6) for 1h at 65°C and then hybridized in the same solution with the appropriate (α-^{32}P) dCTP radioactively labelled DNA probes for overnight at 65°C. Labelling of DNA probes was done with Prime-IT RmT Random Primer Labelling Kit, Stratagene. Following incorporation of (α-^{32}P) dCTP, the probes were purified using the QIA quick Nucleotide Removal kit (Qiagen, CAT N. 28304). The following day, the blot was washed twice for 10 minutes at room temperature with wash solution 1 (2x SSC, 0.1% SDS) and once for 30 minutes at 65°C with wash solution 2 (0.1x SSC, 0.1% SDS). Afterwards, the blots were exposed to autoradiography films and the signals were detected with the phosphoimager (Fuji, FLA7000). The images were processed using ImageJ.

Table 2.5: 2X RNA loading buffer

Quantity	Ingredients
1.6x	MOPS (Morpholine Propane Sulfonic Acid)
7%	Formaldehyde
65%	Formamide
50 µg/ml	Ethidium bromide
0.025%	Bromophenol blue

Table 2.6: Hybridization solution for DNA probes

Quantity	Ingredients
2.5 ml	20x SSC buffer
5.9 ml	Water
0.5 ml	10% SDS buffer
1 ml	50x Denhardt's Solution
0.1 ml	10g/ml Salmon Sperm (denatured at 95°C for 5 min)

2.3.1 RNA immunoprecipitation

Approximately $1x10^9$ cells expressing *in situ* C-TAP DRBD18*, in situ* N-TAP HNRNPFH and *in situ* N-TAP ZC3H28 with a concentration of $1x10^6$ cells/ml (in case of Lister 427) or a concentration of ~$7 x10^5$ cells/ml (in case of pleomorphic "EATRO 1125" cells) were pelleted by centrifugation at 3000 rpm for 13 minutes at 4°C. The pellet was resuspended in 50 ml 1x PBS, then transferred to a 50 ml falcon tube. The cells were again pelleted by centrifugation at 2300 rpm for 8 minutes at 4°C. The pellet was washed twice with ice cold 1x PBS and snap frozen in liquid nitrogen. The RNA immunoprecipitation was done as described in (Archer et al. 2009), with the exception that the magnetic beads were used in place of sepharose beads. The cell pellet was lysed in 1 ml of the lysis buffer (components listed below in table 2.7) by passing 20 times through a 21G x ½ needle using a 1 ml syringe and 20 times through a 27G x ¾ needle using a 1 ml syringe. The lysate was cleared by centrifugation at 15,000g for 15 minutes at 4°C. Afterwards, the supernatant was transferred to a new Eppendorf tube and the salt concentration was adjusted to 150 mM KCl. 2% of the lysate was collected for western blot as the input fraction (IN) and 10% was used for RNA extraction. 40 µl of magnetic beads (Dynabeads™ M-280 Tosylactivated, Invitrogen, Thermo Fischer Scientific 14203) coupled with Rabbit Gamma globulin antibodies (Jackson Immuno Research Labratories, Inc) were washed three times with 500 µl of IP buffer (Table 2.8), mixed slowly on a rotator in the cold room for 5 minutes and then let to settle down on a magnetic rack (Dyna-mag magnet) for 1 minute.

The cell lysate was added to the washed beads and incubated at 4°C for 2-3 hours with gentle rocking. 2% of the unbound fraction was collected for western blot (UN) and the rest was used for RNA extraction. Subsequently, the beads were washed thrice with IP buffer, then re-suspended in 500 µl of TEV cleavage buffer (IP buffer with 150 units of TEV protease) and incubated at 15-20°C for 2 hours with gentle rocking (or at 4°C for overnight). The eluate was collected by magnetic separation. Approximately 2% of the TEV eluate sample was used for western blotting and the rest for RNA extraction using TriFast-FL reagent (Peqlab, GMBH) or Tri Reagent LS (Sigma-Aldrich). RNA was purified from the input, the unbound and the eluate fractions using the peqGOLD Trifast FL (Peqlab, GMBH) according to the manufacturers' instructions. The quality of the purified RNAs was assessed on formaldehyde agarose gels which indicate the integrity of the ribosomal RNAs (rRNAs). Total RNA from the unbound and the eluate fraction were depleted of ribosomal RNAs (rRNA) using a cocktail of 131 DNA oligos (50 bases) complementary to the trypanosome rRNAs. The rRNAs hybridized to the oligonucleotides were digested with RNAse H (NEB, M0297S) as previously described in Minia and Clayton 2016 and Mugo and Clayton 2017.

Following rRNA depletion, the samples were subjected to DNAse I treatment in order to remove any trace of oligonucleotides using the Turbo™ DNAse kit (Invitrogen, ThermoScientific). The RNA samples were afterward purified using the RNA Clean & Concentrator™ -5 kit (ZYMO RESEARCH) following the manufacturers' instructions. The recovered purified RNAs from both bound and unbound samples were then analyzed by RNA-Sequencing.

Table 2.7: Lysis Buffer

Quantity	Ingredients
20mM	Tris pH 7.5
5 mM	MgCl2
0.1% (v/v)	IGEPAL
1mM	DTT
100U	RNasin (Promega)
10 µg/ml	Leupeptin
10 µg/ml	Aprotinin

Table 2.8: IP Buffer

Quantity	Ingredients
20mM	Tris pH 7.5
5 mM	MgCl2
0.1% (v/v)	IGEPAL
1mM	DTT
200U	RNasin (Promega)
150 mM	KCl
10 µg/ml	Leupeptin
10 µg/ml	Aprotinin

2.3.2 High throughput RNA sequencing and bioinformatic analysis

RNA Sequencing was performed at the Cell Networks Deep Sequencing Core Facility of the University of Heidelberg. The library preparation was done using the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina® (NEB, E7760S). The libraries were multiplexed (six or eight samples per lane) and sequenced with a NextSeq 550 system, generating single-end sequencing reads of about 75 bp. RNA sequencing analysis was done according to Leiss pipeline Analysis (https://github.com/klprint/DESeqUI/tree/v1.0).

Before analysis, the quality of the raw sequencing data was checked using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Cutadapt (Martin 2011) was used to remove sequencing primers. After primer removal, the sequencing data were aligned to *T. brucei* 927 reference genome using Bowtie (Langmead et al. 2009), allowing 1 alignment per read, then sorted and indexed using SAMtools (Li et al. 2009). The reads that mapped to the open reading frames of the TREU 927 genome were counted. For RNAi analysis, the reads that mapped to open reading frames, 3'-UTRs and to non-coding RNA in the TREU 927 genome were counted. For comparative analysis and enrichment of functional characteristics, a list of unique genes modified from (Siegel et al. 2010) was used in order to account for repeated genes and multigene families. To generate reads per millions for the unique gene set, the list of genes was extracted together with the reads, then the reads were multiplied by the gene copy number. For the RIP-Seq, the reads per millions were counted and the ratios eluate versus flow-through were calculated. An mRNA was considered as "bound mRNA" if the ratios from all the three pulldowns were higher than 2 or 3. The 3'-UTR motif enrichment search was done using MEME in the relative enrichment mode (Bailey et al. 2015). Annotated 3'-UTRs were downloaded from TritrypDB. For other 3'-UTRs, the manual annotation was performed using the RNA-Seq reads and polyA site data in TritrypDB (Siegel et al. 2010); Kolev et al. 2010). Analysis of differentially expressed genes after DRBD18 and ZC3H28 RNAi was done in R using the DESeqUI

(https://github.com/klprint/DESeqUI/tree/v1.0), a customized version of DESeq2 package (Love et al. 2014) adapted for trypanosome transcriptomes. Statistical analyses were done using R and Microsoft excel. Raw data are available at Array express with accession numbers E-MTAB-10735 (RNAs associated with DRBD18), E-MTAB-9783 (effect of DRBD18 RNAi), E-MTAB-10674 (ZC3H28-associated RNAs), E-MTAB-10751 (effect ZC3H28 RNAi) and E-MTAB-10708 (HNRNPFH bound mRNAs).

2.3.3 Affymetrix single molecule mRNA fluorescence *in situ* hybridization (FISH)

The Affymetrix single molecule mRNA FISH was carried out as described in Kramer 2017b. A total of 200 ml bloodstream-form trypanosomes at ~5-8 x10^5 cells/ml were harvested by centrifugation (8 minutes, 1400g), resuspended in 1ml 1x PBS and pelleted again by centrifugation (5 minutes, 1400g). The cell pellet was resuspended again in 1ml 1x PBS, followed by addition of 1 ml formaldehyde (8% in PBS) for cell fixation. The mixture was incubated at room temperature for 10 minutes with an orbital mixer. A total of 13 ml 1x PBS were added and the cells were harvested by centrifugation (5 minutes, 1400 g). The fixed cell pellet was resuspended in 1 ml 1x PBS and spread on glass microscopy slides (previously incubated at 180°C for 2h for RNAse removal) within circles of hydrophobic barriers (PAP pen, Sigma). The cells were allowed to settle at room temperature for 20 minutes. The slides were then washed twice in 1x PBS. Permeabilization of the fixed cells was done with addition of 50 µl of detergent solution QC in each circle on the slides. This was followed by a 2-step washing in 1x PBS. The protease solution was diluted 1:1000 in 1x PBS and briefly vortexed to allow complete dissolution. 100 µl of the protease solution was added to each circle and incubated exactly for 10 minutes at 25°C. The slides were then washed twice in 1x PBS and used for Affymetrix FISH experiments as described in the manual of the QuantiGene® ViewRNA ISH Cell Assay (Affymetrix), protocol for glass slide format. The only modification from the kit protocol is that the protease digestion was done at 25°C rather than the normal room temperature and we used a self-made washing buffer (0.1x SSC buffer, 0,1% SDS) instead of the washing buffer from the kit. All Affymetrix probe sets used in this work are described in the Supplementary information/section 2. For visualization, the labelled cells were mounted with 4′,6-diamino-2-phenylindole dihydrochloride (DAPI) solution, diluted 1:1000 in 1x PBS. Images were taken with a fluorescent inverted wide-field microscope Leica DMI6000B (Leica Microsystems GmbH, Wetzlar, Germany) equipped with 100x oil immersion (NA 1.4) and a Leica DFC365 camera (6.45 m/pixel). Deconvolution was done using Huygens Essential software (SVI, Hilversum, The Netherlands) and images are presented as Z-stack projection (sum slices). The image analysis was carried out using the available tools in Image J software and 50 cells for each slide were selected for quantifying the number of mRNAs present in the cytosol and in the nucleus.

2.4 Protein methods

2.4.1 SDS-PAGE and Western Blotting

$5x10^6$ cells were collected per sample, harvested by centrifugation (3000 rpm, 5 minutes), washed twice in cold 1x PBS, resuspended in 2x Laemmli buffer (components listed in Table 2.9) and heated at 95°C for 10 minutes. The samples were subjected to SDS-PAGE gel electrophoresis using 10-12% polyacrylamide gels (standard recipe). The gel was run at 100 volts with the Biorad system in 1x SDS-running buffer (Table 2.10). The SDS-gels were then blotted on a 0.45 µm nitrocellulose blotting membrane (Neolabs) for 1h at 100 volts in blotting buffer (Table 2.11) at 4°C. The protein transfer was verified by staining the membrane with Ponceau S (Serva). The membrane with the visualized transferred proteins was scanned for further quantitative analysis. The membrane was then blocked in 3% (w/v) milk in TBS-T buffer (Table 2.12) for 1 hour. Afterwards, the first antibody diluted in 3% milk (in TBS-T buffer) was incubated with the membrane for either 1 hour at room temperature or overnight at 4°C. After washing the membrane thrice with TBS-T buffer, the appropriate secondary antibody coupled to horse radish peroxidase (HRP) (1:2000, GE Healthcare) diluted in 3% milk (in TBS-T buffer) was incubated with the membrane at room temperature for 1 hour. Western Lightning Ultra® (Perkin Elmer) was used as a chemiluminescence system and the signal was detected using X-ray films or the Amersham Imager 600 (GE Healthcare). The antibodies used for this study are listed in Table 2.13.

Table 2.9: 2X Laemmli Buffer

Quantity	Ingredients
125 mM	Tris-HCl pH 6.8
4%	SDS
15 mM	EDTA
10%	ß-Mercaptoethanol
20%	Glycerol
0.1%	Bromophenol Blue
1 tablette	Protease inhibitor

Table 2.10: 1x SDS-running buffer

Quantity	Ingredients
25 mM	Tris-HCl pH 6.8
192 mM	Glycine
0.1 %	SDS

Table 2.11: Blotting buffer

Quantity	Ingredients
25 mM	Tris-HCl pH 6.8
192 mM	Glycine
20 %	Methanol

Table 2.12: TBS-T Buffer

Quantity	Ingredients
50 mM	Tris-HCl, pH 6.8
150 mM	NaCl
0.1 %	Tween 20

Table 2.13: Antibodies used for western blot

Antibody	Company or publication	Product number	Host	Dilution
Anti-myc (9E10)	Santa Cruz Biotechnology	B0614	Mouse	1:1000
Anti-DRBD18	Lott et al. 2015		Rabbit	1:2500
Anti-XRND	Li et al. 2006		Rabbit	1:2500
Anti-RBP10	Wurst et al. 2012		Rat	1:2000
Anti-V5	Biorad	MCA1360	Mouse	1:2000
PAP (Peroxidase Anti-Peroxidase)	Sigma	P-2026	Rabbit	1:5000
Anti-S9	Clayton lab		Rat	1:1000
ECL Anti mouse IgG	GE Healthcare	NA931V		1:2000
ECL Anti rabbit IgG	GE Healthcare	NA934V		1:2000
ECL Anti rat IgG	GE Healthcare	NA935V	Goat	1:2000
Streptavidin-horseradish peroxidase conjugate	GE Healthcare	RPN1231V		1:5000

2.4.2 Co-immunoprecipitation (Co-IP)

For confirming the interaction of ZC3H28 with MKT1, the cells used for this experiment expressed *in situ* N-TAP ZC3H28 and *in situ* V5-MKT1. The cells expressing only *in situ* V5-MKT1 were used as controls. For the interaction between ZC3H28 and PBP1, the cells used were expressing both *in situ* N-TAP ZC3H28 and *in situ* V5-PBP1. The cells expressing only *in situ* V5-PBP1 were used as controls.

For the experiment, approximately $2x10^8$ bloodstream-form cells with a concentration of $1x10^6$ were collected by centrifugation at 3000 rpm for 13 minutes at 4°C. The pellet was resuspended in 20 ml 1x PBS and concentrated by centrifugation at 2300 rpm for 8 minutes at 4°C. The cell pellet was washed again once in ice cold 1x PBS and snap frozen in liquid nitrogen. The rest of the protocol was done in the cold room. The cells were lysed in 1 ml lysis buffer (components listed in Table 2.7) by passing them 20 times through a 21G x 1 ½ needle using a 1 ml syringe and 20 times through a 27G x ¾ needle using a 1ml syringe as well. The salt concentration was then adjusted to 300 mM KCl. In order to remove any trace of DNA, 1x DNase buffer (components listed below, Table 2.14) and 10 µl DNaseI (NEB, M0303S) was added to the samples. The cell debris were then incubated at 37°C for 10 minutes. The reaction was stopped by addition of 5 mM stop solution (components listed below, Table 2.15), followed by incubation on ice for 15 minutes. The lysates were precleared by centrifugation at 15,000 g for 15 minutes at 4°C. Afterwards, the supernatant was transferred to a new Eppendorf tube. The Bradford assay was used to determine the protein concentration. 500 µg of the protein in wash buffer was used for the Co-IP. 20 µg of the lysate sample was collected as the input fraction (IN). 30 µl of the magnetic beads (Dynabeads™ M-280 Tosylactivated, Invitrogen, Thermo Fischer Scientific 14203) coupled with Rabbit Gamma globulin antibodies (Jackson Immuno Research Laboratories, Inc) were washed 3 times with 500µl of IP buffer (same as Table 2.8), mixed gently on a rotator for 5 minutes each time. The beads were collected by settling the mixture in a magnetic rack (DynaMag-2 magnet, Invitrogen, 12321D) for 1 minute. The protein samples (500 µg) were then incubated with the magnetic beads for 1h at 4°C with gentle rocking on a rotator. The unbound fraction was removed and 20 µg of this fraction was collected for western blot analysis. The magnetic beads were washed 3 times with the IP buffer by rotating for 5 minutes each. They were then resuspended in 500 µl of TEV cleavage buffer (IP buffer with 150 units of TEV protease) and incubated at 4°C with gentle rocking for overnight. The supernatant was removed and the proteins were eluted by addition of 50 µl 2x Laemmli and boiling at 95°C for 10 minutes. Tubes were attached to the magnetic rack to remove the eluate. 20 µg of the total protein, 20 µg unbound sample and half of the eluate were analyzed by SDS-PAGE and Western Blotting.

Table 2.14: 200x DNAse Buffer

Quantity	Ingredients
500 mM	$MgCl_2$
100 mM	$CaCl_2$

Table 2.15: Stop solution

Quantity	Ingredients
250 mM	EDTA
250 mM	EGTA

2.4.3 Yeast two-hybrid assay

For testing direct interactions of ZC3H28 with the MKT1 complex, the Matchmaker Yeast Two-Hybrid system (Clontech) was used according to the manufacturer's instructions. The coding sequence of ZC3H28 as well the N-terminal region of ZC3H28 were PCR-amplified from genomic DNA and cloned into pGBKT7 and pGADT7 plasmids. The prey and the baits plasmids of ZC3H28 as well as those of the MKT1complex proteins were co-transformed pairwise into AH109 yeast strains. Selection was done initially on double dropout (DDO) plates (i.e., SD medium lacking Tryptophan and Leucine) to check expression of both bait and prey. The growth was then checked on quadruple drop-out (QDO) plates (i.e., lacking Tryptophan, Leucin, Histidine and Adenine) that indicates positive interactions. As negative controls, the pGBKT7 plasmid containing Lamin and the pGADT7 containing the SV40 large T antigen were used. The interaction between p53 and SV40 large T-antigen was used as positive control. The composition of the SD-medium are described below (Table 2.16 and 2.17).

Table 2.16: Drop solution (10X)

Quantity	Ingredients
300 mg/L	L-isoleucine
1500 mg/L	L-valine
200 mg/L	L-arginine HCl
300 mg/L	L-lysine HCl
200 mg/L	L-methionine
500 mg/L	L-phenylalanine
2000 mg/L	L-threonine
300 mg/L	L-tyrosine
200 mg/L	L-uracil

Table 2.17: SD medium, 5 L, sterile filtered

Quantity	Ingredients
33.5 g	Yeast nitrogen base without amino acids
500 ml	Dropout solution (10X)
100 g	Glucose

2.4.4 Tandem Affinity Purification

Approximately $1x10^9$ cells expressing either *in situ* C-TAP DRBD18 (pHD3200), *in situ* N-TAP ZC3H28 (pHD3236) or tet-inducible GFP-TAP (pHD1743) with a concentration of $1x10^6$ cells/ml were harvested by centrifugation at 3000 rpm for 13 minutes at 4°C. For each cell line, three technical replicates were done without RNAse A treatment. The cell pellet was resuspended in 10 ml of cold 1x PBS and centrifuged at 2800 rpm for 8 minutes at 4°C.

The cell pellet was lysed in 1 ml of the lysis buffer (components listed below in table 2.7) by passing 20 times through a 21G x ½ needle using a 1 ml syringe and 20 times through a 27G x ¾ needle using a 1 ml syringe. The lysate was cleared by centrifugation at 15,000 g for 15 minutes at 4°C. Afterwards, the supernatant was transferred to a new Eppendorf tube and the salt concentration was adjusted to 150 mM KCl. 2% of the lysate was collected for western blot as the input fraction (IN). 40 µl of magnetic beads (Dynabeads™ M-280 Tosylactivated, Invitrogen, Thermo Fischer Scientific 14203) coupled with Rabbit Gamma globulin antibodies (Jackson Immuno Research Laboratories, Inc) were washed three times with 500µl of IP buffer (Table 2.8), mixed slowly on a rotator in the cold room for 5 minutes and then let to settle down on a magnetic rack (Dyna-mag magnet for 1 minute. The cell lysate was added to the washed beads and incubated at 4°C for 2-3 hours with gentle rocking. 2% of the unbound fraction was collected for western blot (UN) and the rest was discarded. Subsequently, the beads were washed four times with IP buffer, and 2% of each washing step was collected by magnetic separation. After washing, the beads were re-suspended in 25 µl of TEV cleavage buffer (IP buffer containing 5 units of TEV protease). The TEV cleavage was done at 4°C for overnight with gentle rotation. The eluate was collected by magnetic separation, 5 µl was used for western blot analysis and the rest was subjected to removal of the TEV protease using the Nickel-Nitriloacetic acid (Ni-NTA) beads. 6x laemmli buffer was added to the final eluate which was separated on a 12% SDS-polyacrylamide gel. The gel was run up to 2cm and the proteins were visualized by colloidal Coomassie staining. Two gel areas per lane were excised and analyzed by mass spectrometry.

2.4.5 Affinity purification of biotinylated proteins

$5x10^8$ (or $1x10^9$) procyclic cells constitutively expressing either the PSSA-2 3'-UTR (pHD 3093) or the actin 3'-UTR (pHD 2277) flanked by BoxB elements and inducibly expressing the tet-inducible lambda N-myc-BirA* (pHD 3108) were grown in presence of tetracycline (100 ng/ml) and freshly made D-biotin solution (50 µM) for 24 hours as described in (Morriswood et al. 2013). The cells were then harvested by centrifugation (1800g, 10 minutes, 4°C), washed twice with 25 ml of cold 1x PBS and transferred to Eppendorf tubes. The cells were then lysed in 1 ml of lysis buffer (Table 2.18) for 30 minutes with gentle mixing at RT. The supernatants were transferred to new Eppendorf tubes and 5% was taken as the input sample (IN). 500 µl of Streptavidin-coated Dynabeads (MyOne™ Streptavidin C1, Invitrogen) were then added to the supernatants and incubated with gentle agitation for 4 hours at 4°C. The beads were separated from the unbound fraction by magnetic separation using a Dyna-mag magnet. A 5% sample was taken as the unbound fractions (UN).

Afterwards, the dynabeads were collected by magnetic separation and washed extensively to break the biotin-avidin interaction as described in Roux et al. 2012. Briefly, the dynabeads were washed twice for 8 minutes in 1 ml of wash buffer 1 (2% SDS in distilled water) at 25°C. This step was repeated once with wash buffer 2 (Table 2.19), then once more with wash buffer 3 (Table 2.20) and twice with wash buffer 4 (Table 2.21). 5% of the sample was taken from each washing step. The bound proteins were eluted from the beads by incubation with 50 µl of 2x Laemmli buffer containing 4 mM biotin and 20 mM DTT at 98°C for 20 minutes. For the samples to be analyzed by mass spectrometry, the beads were further washed twice in 50 mM NH_4HCO_3 and once with 8 mM urea solution before elution. The final eluate was mixed with 6x laemmli buffer and separated on a 12% SDS-polyacrylamide gel. The gel was run up to 3 cm and the proteins were visualized by colloidal Coomassie staining. Five gel areas per lane were excised and analyzed by mass spectrometry.

Table 2.18: Lysis buffer for biotinylation

Quantity	Ingredients
50 mM	Tris-HCl pH 7.4
0.4%	SDS
500 mM	NaCl
5 mM	EDTA
1 mM	DTT
10 µg/ml	Leupeptin
10 µg/ml	Aprotinin

Table 2.19: Wash buffer 2 for biotinylation

Quantity	Ingredients
1 mM	EDTA
0.1%	Deoxycholate
500 mM	NaCl
1%	Triton X-100
50 mM	Hepes, pH .5

Table 2.20: Wash buffer 3 for biotinylation

Quantity	Ingredients
1 mM	EDTA
0.5%	Deoxycholate
250 mM	LiCl
0.5%	IGEPAL
10 mM	Tris, pH 8.1

Table 2.21: Wash buffer 4 for biotinylation

Quantity	Ingredients
50 mM	NaCL
50 mM	Tris, pH 7.4

2.4.6 Mass spectrometry analysis

Proteins that co-purified with DRBD18 and ZC3H28 were analyzed in three independent experiments in the ZMBH Mass Spectrometry facility via the Ultimate 3000 liquid chromatography system directly coupled to an Orbitrap Elite mass spectrometer (Thermo Fisher). Cell lines expressing TAP-GFP served as control. MS spectra (m/z 400–1600) were acquired in the Orbitrap at 60,000 (m/z 400) resolution. Fragmentation in Collision-induced dissociation (CID) was performed for up to 10 precursors. MS2 spectra were acquired at rapid scan rate. Raw data were analyzed using MaxQuant 1.5.8.30, with label-free quantification (LFQ), match between runs (between triplicates), and the iBAQ algorithm enabled for peptide identification and quantification. The identified proteins were filtered for known contaminants and reverse hits, as well as hits without unique peptides. Statistical analysis was performed using the Perseus software (Tyanova et al. 2016). Data were filtered for at least two valid values in at least one condition and remaining missing values were imputed with a normal distribution based on the whole data set (width = 0.3; shift = 1.8). Significant outliers were determined with t-test statistics (permutation-based false

discovery rate of 1% and S0 of 1). MS2 spectra data were searched against the TriTrypDB-8.1 TREU 927 – Annotated Proteins-1 database, which contains 11567 sequences. The mass spectrometry proteomics data are deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Perez-Riverol et al. 2019) with the dataset identifier PXD027792. The proteins associated with the PSSA-2 3'-UTR and the actin 3'-UTR were also analyzed in the ZMBH Mass spectrometry facility with the same settings as previously described.

2.5 Subcellular fractionation

Nuclear-cytosolic fractionation was performed as described in Biton et al. 2006. Briefly, $3x10^8$ Lister 427 bloodstream form wild-type cells as well as those expressing the RNAi stem loop vector targeting DRBD18 were harvested by centrifugation at 3000 rpm for 13 minutes at 4°C. The cell pellet was washed twice with cold 1x PBS and collected by centrifugation at 2300 rpm for 8 minutes at 4°C. Afterwards, the pellet was resuspended in hypotonic buffer (10 mM HEPES pH 7.9, 1.5 mM MgCl2, 10 mM KCl, 0.5 mM dithiothreitol, 5 µg/ml leupeptin and 100 U RNasin) and lysed in presence of 0,1% IGEPAL (Nonidet P-40) by passing 20 times through a 21G x ½ needle using a 1 ml syringe and 20 times through a 27G x ¾ needle using a 1 ml syringe. The cells were allowed to rest on ice for 20 minutes. Then, the nuclei fraction was pelleted by centrifugation at 15,000 g for 15 minutes at 4°C. The supernatant containing the cytoplasmic fraction was transferred to a new tube. 20 % of the cytoplasmic fraction was used for western blotting and the rest was used for RNA extraction (RNA data is not shown). The pellet containing the nuclei fraction was resuspended in TBS containing 0.1 % SDS; 20 % of this fraction was used for western blotting while the rest was used for RNA extraction (RNA data is not shown).

2.6 CAT assay

Chloramphenicol acetyltransferase (CAT) activity was measured in a kinetic assay that involves the transfer of an acetyl group from radiolabeled ^{14}C butyryl-CoA to chloramphenicol as described in Smale 2010. The ^{14}C butyryl-chloramphenicol becomes water insoluble and moves from the aqueous to the organic phase of the scintillation fluid (Ultima Gold™ F, Perkin Elmer, 6013171).

Sample preparation

For each sample, 1-2 x 10^7 cells expressing the CAT reporter gene were harvested (2300 rpm, 8 minutes, 20°C), washed twice with 1ml of 1x PBS. Thereafter, the cell pellets were resuspended in 200 µl of CAT buffer (100mM Tris-HCL, pH 7.8) and lysed by three cycles of freeze-thawing using liquid nitrogen and a 37°C heating block. After centrifugation

(15,000g, 3 minutes, 4°C), the supernatants were then transferred to a new tube and total protein concentration was measured. For longer storage, the samples were kept and stored at -80°C until use.

Determination of protein concentration (Bradford assay)

The total protein concentration of a sample was measured using the Bradford method (Bradford 1976). This assay is based on the proportional binding of the Coomassie dye to the proteins. When bound to proteins, the dye formerly in the doubly protonated red cationic form is converted to a more unprotonated blue form that has a maximum absorption at 595 nm, detectable using a spectrophotometer or a microplate reader.

To generate the Bradford standard curve, 0, 1, 2, 4, 8, 12, 16, 20 µg (100 µg/ml stock) of bovine serum albumin (BSA) in 150 µl of water were used. 146 µl of water was mixed with 4 µl of each protein sample. Then, 50 µl of Bradford reagent dye (Protein Assay Dye Reagent Concentrate Biorad, Cat#500-0006) was added to all protein samples as well as to the BSA standards. The reaction mixture was incubated for 5 minutes at room temperature and the optical density (OD) was measured using a microplate reader (SPECTROstar^{Nano}, BMG Labtech) at 595 nm. A BSA standard curve was calculated in Excel and the linear regression line was determined.

CAT Assay

To measure the activity of the CAT enzyme in all samples, equal amounts of the protein (0.5-1 µg) were taken for each sample and scaled up to 50 µl using the CAT buffer (100 mM Tris-HCl, pH 7.8) in a Wheaton scintillation tube HDPE (neoLab #9-0149). After addition of 200 µl of CAT buffer, 2 µl of 40 mg/ml chloramphenicol (Serva, 16785.03), 10 µl of ^{14}C butyryl-CoA, 4ml of the scintillation cocktail (Ultima Gold™ F, Perkin Elmer, 6013171) was added. The reactions were then measured every 12-16 minutes for 1.5 hour in a scintillation counter (Beckman LS6000IC), using program 7 that detect ^{14}C. Once plotted, the data slope at the linear range (before saturation) was compared between samples to estimate the relative CAT protein levels.

2.7 Oligonucleotide list and plasmid list

2.7.1 Oligonucleotide list

Table 2.22: List of all oligonucleotides used in this study

Oligo No.	Description	Restriction site	Sequence	Purpose
Cloning primers				
CZ 6270	Fwd full length RBP10 3'-UTR (pHD 2860)	HindIII	ATC**AAGCTT**TGGCACAGAGGGTAACGAAG	CAT Reporter for RBP10 3'-UTR
CZ 6271	Rev full length RBP10 3'-UTR (pHD 2860)	ApaI	ATC**GGGCCC**TCCACCACCTCTCCTCCTTA	CAT Reporter for RBP10 3'-UTR
CZ 6887	Fwd Actin 3'-UTR (pHD 3021)	XhoI	ATA**CTCGAG**TAACACCGGGTTGTGTG	CAT Reporter for Actin 3'-UTR
CZ 6888	Rev Actin 3'-UTR (pHD 3021)	HpaI	TAT**GTTAAC**AATACTGCATAGATAAC	CAT Reporter for Actin 3'-UTR
CZ 6680	Fwd RBP10 3'-UTR-F2 and F2.1 (pHD 2961, pHD 3048)	XhoI	ATA**CTCGAG**CAGCGGAAGCAGACAGTAAA	CAT Reporter for RBP10 3'-UTR
CZ 6679	Rev RBP10 3'-UTR-F2, F2.2, F2.2.2 and F2.2.4 (pHD 2961, pHD 3044, pHD 3045, pHD 3085)	SalI	TAT**GTCGAC**CTCTCAAATCGATGGGATGC	CAT Reporter for RBP10 3'-UTR
CZ 6682	Fwd RBP10 3'-UTR-F1 and F1.1 (pHD 3042, pHD 3087)	XhoI	ATA**CTCGAG**TGGCACAGAGGGTAACGA	CAT Reporter for RBP10 3'-UTR
CZ 6681	Rev RBP10 3'-UTR-F1 and F1.4 (pHD 3042, pHD 3084)	SalI	TAT**GTCGAC**AATGAAAAGCGACAAAATCATATTA	CAT Reporter for RBP10 3'-UTR
CZ 6678	Fwd RBP10 3'-UTR-F3 and F3.1 (pHD 2962, pHD 3052)	XhoI	ATA**CTCGAG**TTATAAAAGACGAGGAAAAGCGGA	CAT Reporter for RBP10 3'-UTR
CZ 6677	Rev RBP10 3'-UTR-F3, F3.2, F3.2.2 (pHD 2962, pHD 3046, pHD 3047)	SalI	TAT**GTCGAC**TGTCAGTAATAATAAGAAATGATAGCC	CAT Reporter for RBP10 3'-UTR
CZ 6816	Fwd RBP10 3'-UTR-F4 and F3 (pHD 3043, pHD 3088)	XhoI	ATA**CTCGAG**TTATAGCTATTGTTATTGTTTACTG	CAT Reporter for RBP10 3'-UTR
CZ 6817	Rev RBP10 3'-UTR-F4 (pHD 3043)	SalI	TAT**GTCGAC**AAAAACGGCAGGTGAGATGTAAAG	CAT Reporter for RBP10 3'-UTR
CZ 6818	Fwd RBP10 3'-UTR-F2.2 and F2.2.1 (pHD 3044, pHD 3049)	XhoI	ATA**CTCGAG**GATCCGCTTTGCCTTTGCAT	CAT Reporter for RBP10 3'-UTR
CZ 6819	Fwd RBP10 3'-UTR F2.2.2 AND F2.2.3 (pHD 3045, pHD 3086)	XhoI	ATA**CTCGAG**CTTCATCCATGCCGACTCTGC	CAT Reporter for RBP10 3'-UTR
CZ 6820	Fwd RBP10 3'-UTR-F3.2 and F3.2.1 (pHD 3046, pHD 3053)	XhoI	ATA**CTCGAG**CGCTTCTATCCATCTGTCTTCC	CAT Reporter for RBP10 3'-UTR
CZ 6821	Fwd RBP10 3'-UTR-F3.2.2 (pHD 3047)	XhoI	ATA**CTCGAG**TACCGCATTTGTGAGCCACC	CAT Reporter for RBP10 3'-UTR

Oligo No.	Description	Restriction site	Sequence	Purpose
CZ 6986	Rev RBP10 3'-UTR-F2.1 (pHD 3048)	SalI	TAT**GTCGAC**ATGCAAAGGCAAAGCGGATC	CAT Reporter for RBP10 3'-UTR
CZ 7002	Rev RBP10 3'-UTR-F2.2.1 (pHD 3049)	SalI	TAT**GTCGAC**TGAAAAATAATAAGAGTC	CAT Reporter for RBP10 3'-UTR
CZ 7022	Rev RBP10 3'-UTR F3.1, F3.1.1 (pHD 3052, pHD 3202 and pHD 3203)	SalI	TAT**GTCGAC**TGGATAGAAGCGTCATCA	CAT Reporter for RBP10 3'-UTR
CZ 7025	Fwd RBP10 3'-UTR-F (pHD 3084)	XhoI	ATA**CTCGAG**GTTTTTACTTCTCCTCGTCCA	CAT Reporter for RBP10 3'-UTR
CZ 7070	Rev RBP10 3'-UTR F3.2.1 (pHD 3053)	SalI	TAT**GTCGAC**TGCGGTAAGGAAAGCACCTT	CAT Reporter for RBP10 3'-UTR
CZ 7125	Fwd RBP10 3'-UTR F2.2.4 (pHD 3085)	XhoI	ATA**CTCGAG**ATGTTTCCACACCATTGGG	CAT Reporter for RBP10 3'-UTR
CZ 7124	Rev RBP10 3'-UTR F2.2.3 (pHD 3086)	SalI	TAT**GTCGAC**CATTTTCTGCCTCACTTTA	CAT Reporter for RBP10 3'-UTR
CZ 7205	Rev RBP10 3'-UTR F3.4 (pHD 3088)	SalI	TAT**GTCGAC**CAAAGTGAACAGAAGTCCGC	CAT Reporter for RBP10 3'-UTR
CZ 7208	Rev RBP10 3'-UTR F1.1 (pHD 3087)	SalI	TAT**GTCGAC**TTTCTTAAACTGAAGAATCA	CAT Reporter for RBP10 3'-UTR
CZ 7209	Rev RBP10 3'-UTR F1.3 (pHD 3118)	SalI	TAT**GTCGAC**TAAAAACTTTTCAATTTAAT	CAT Reporter for RBP10 3'-UTR
CZ 7246	Fwd RBP10 3'-UTR F1.2 (pHD 3117)	XhoI	ATA**CTCGAG**TGATTCTTCAGTTTAAG	CAT Reporter for RBP10 3'-UTR
CZ 7247	Rev RBP10 3'-UTR F1.2 (pHD 3117)	SalI	TAT**GTCGAC**CCTTCTTCTCCTTTCCTCTTTCC	CAT Reporter for RBP10 3'-UTR
CZ 7248	Fwd RBP10 3'-UTR F1.3 (pHD 3118)	XhoI	ATA**CTCGAG**GGGAAAGAGGAAAGGAGAAGAAGG	CAT Reporter for RBP10 3'-UTR
CZ 7422	Fwd RBP10 3'-UTR F1.2 with (AU) deleted (pHD 3162)	Site directed mutagenesis	TTTGTATTCATATTTATGTCTTATATAC	CAT Reporter for RBP10 3'-UTR
CZ 7423	Rev RBP10 3'-UTR F1.2 with (AU) deleted F1.2 (pHD 3162)	Site directed mutagenesis	TTAGATTTATATAAATCGCTCTAC	CAT Reporter for RBP10 3'-UTR
CZ 7563	Fwd RBP10 3'-UTR F3.1.2 with AAGG(A)10 deleted (pHD 3203)	Site directed mutagenesis	ATA**GGATCC**CGAAAAAGAAAAAAGGAAAGCACCCGAGTG	CAT Reporter for RBP10 3'-UTR
CZ 7238	Fwd PSSA-2 3'-UTR (pHD 3093)	XhoI	ATA**CTCGAG**TTAACGCGGTTAGG	CAT Reporter for PSSA-2 3'-UTR
CZ 7239	Rev PSSA-2 3'-UTR (pHD 3093)	SalI	TAT**GTCGAC**TTTTTTTAGTTCACAATC	CAT Reporter for PSSA-2 3'-UTR
CZ 7361	Fwd ZC3H45-Partial ORF (pHD 3161)	HindIII, BglII	ATA**AAGCTTAGATCT**TGATGACTCCCCAAATCGCC	RNAi Stem loop for ZC3H45
CZ 7362	Rev ZC3H45-Partial ORF (pHD 3161)	EcoRI, SalI	TAT**GTCGACGAATTC**AGCCGCCGTTATTTCTGGAA	RNAi Stem loop for ZC3H45
CZ 7383	Fwd ZC3H40 Partial ORF (pHD 3179)	HindIII, BglII	ATA**AAGCTTAGATCT**TGCCTTCGAGTGTATGGTCG	RNAi Stem loop for ZC3H40
CZ 7384	Rev ZC3H40 Partial ORF (pHD 3179)	EcoRI, SalI	TAT**GTCGACGAATTC**TTGAGGTGATGTGTCGGGTG	RNAi Stem loop for ZC3H40

Oligo No.	Description	Restriction site	Sequence	Purpose
CZ 7385	Fwd DRBD18 Partial ORF (pHD 3175)	HindIII, BglII	ATA**AAGCTTAGATCT**CAGAGTCCGAACCACTGCAT	RNAi Stem loop for DRBD18
CZ 7386	Rev DRBD18 Partial ORF (pHD 3175)	EcoRI, SalI	TAT**GTCGACGAATTC**TGGCGGAGGTTGTTGTTCC	RNAi Stem loop for DRBD18
CZ 7387	Fwd HNRNPFH Partial ORF (pHD 3166)	HindIII, BglII	ATA**AAGCTTAGATCT**GGCCATGAGTCGATGATGGT	RNAi Stem loop for HNRNPFH
CZ 7388	Rev HNRNPFH Partial ORF (pHD 3166)	EcoRI, SalI	TAT**GTCGACGAATTC**TGACTGCCCAGGGAAAACTG	RNAi Stem loop for HNRNPFH
CZ 7607	Fwd ZC3H28 Partial ORF (pHD 3234)	HindIII, BglII	GATC**AAGCTTAGATCT**CTAACTTGTCGCCACCACCT	RNAi Stem loop for ZC3H28
CZ 7608	Rev ZC3H28 Partial ORF (pHD 3234)	EcoRI, SalI	GATC**GTCGACGAATTC**TACGCCATGCTTCACAACCT	RNAi Stem loop for ZC3H28
CZ 7770	Fwd HNRNPFH big fragment of the ORF (pHD 3244)	EamI	ATGGAACAGTCTCCTGAGCA	RNAi using 2 opposing T7 promoters
CZ 7771	Rev HNRNPFH big fragment of the ORF (pHD 3244)	EamI	GGACTGTAGTGATGAAACTGTTGG	RNAi using 2 opposing T7 promoters
CZ 7302	Fwd HNRNPFH ORF (pHD 3182)	ApaI	TAGT**GGGCCC**ATGGAACAGTCTCC	HNRNPFH tethering
CZ 7303	Rev HNRNPFH ORF (pHD 3182)	HpaI	CACC**GTTAAC**ACAGAACATCTGTGGTACCA	HNRNPFH tethering
CZ 5720	Fwd ZC3H28 ORF (pHD 2699)	AscI	GGCCGCATGGCGCGCCAT	ZC3H28 tethering
CZ 5721	Rev ZC3H28 ORF (pHD 2699)	AscI	GGCCATGGCGCGCCATGC	ZC3H28 tethering
CZ 7523	Fwd DRBD18 5'-UTR (pHD 3195)	XhoI	ATA**CTCGAG**GAAGAATTGTGAAGTAGGCGG	Single knock-out DRBD18-5'
CZ 7524	Rev DRBD18 5'-UTR (pHD 3195)	HindIII	ATA**AAGCTT**TATCTTTCTCTTTTATTTTCTTTTC	Single knock-out DRBD18-5'
CZ 7525	Fwd DRBD18 3'-UTR (pHD 3195)	BamHI	ATA**GGATCC**AATACCTGAGCATTGGGTATATG	Single knock-out DRBD18-3'
CZ 7526	Rev DRBD18 3'-UTR (pHD 3195)	XbaI	TAT**TCTAGA**CCACTTAAACGCAACATTGAC	Single knock-out DRBD18-3'
CZ 7652	Fwd ZC3H28 3'-UTR (pHD 3294)	BamHI	AGTC**GGATCC**TTAACAATGGAGTAAGCATTGACATC	Single knock-out ZC3H28-3'
CZ 7653	Rev ZC3H28 3'-UTR (pHD 3294)	XbaI	ATGC**TCTAGA**CATGGACGAGAAAGAAACAAACCA	Single knock-out ZC3H28-3'
CZ 7707	Fwd ZC3H28 5'-UTR (pHD 3294)	KpnI	TAT**GGTACC**TAGAGGGGAAATATTAAGAG	Single knock-out ZC3H28-5'
CZ 7708	Rev ZC3H28 5'-UTR (pHD 3294)	XhoI	AGTC**CTCGAG**TGGGCTAGCCTGCCGTAAAAAAGCAAC	Single knock-out ZC3H28-5'
CZ 7519	Rev DRBD18 5'-UTR (pHD 3196)	NdeI	ATA**CATATG**TATCTTTCTCTTTTATTTTCTTTTC	*In situ* N-TAP-DRBD18
CZ 7520	Fwd DRBD18 5'-UTR (pHD 3196)	SacI	ATA**GAGCTC**GAAGAATTGTGAAGTAGGCGGA	*In situ*-N-TAP-DRBD18
CZ 7554	Fwd DRBD18 mini-ORF (pHD 3196)	HindIII	AGA**AAGCTTCC**ATGGAACAGTCTCCTGAGCA	*In situ* N-TAP-DRBD18
CZ 7522	Rev DRBD18 mini-ORF (pHD 3196)	ApaI	ATA**GGGCCC**TTTGTGGCATCGGTTGCTGTC	*In situ* N-TAP-DRBD18

Oligo No.	Description	Restriction site	Sequence	Purpose
CZ 7552	Fwd DRBD18 mini-ORF (pHD3200)	ApaI	CAGA**GGGCCC**ATCTTTGTTCTTATTGAAGTGT	*In situ* C-TAP DRBD18
CZ 7553	Rev DRBD18 mini-ORF (pHD 3200)	XhoI	ATC**CTCGAG**GTGCTGAACCATTTTCCCCAGCACCTTCATTTA	*In situ* C-TAP DRBD18
CZ 7655	Rev ZC3H28 mini-ORF (pHD3236)	ApaI	TATG**GGGCCC**TGGTGATGGTGGTGATGGTGGTGAGGCACAC	*In situ* N-TAP ZC3H28
CZ 7693	Fwd ZC3H28 mini-ORF (pHD3236)	HindIII	ATGC**AAGCTTCC**ATGTATTCTAGCGAAAAGGAGAAGG	*In situ* N-TAP ZC3H28
CZ 7649	Rev ZC3H28 5'-UTR (pHD3236)	NdeI	ATGC**CATATG**TGGGCTAGCCTGCCGTAAAAAAGCAAC	*In situ* N-TAP ZC3H28
CZ 7656	Fwd ZC3H28 5'-UTR (pHD3236)	SacI	ATGC**GAGCTC**CCCCATATAGAGGGGAAATATTAAG	*In situ* N-TAP ZC3H28
CZ 7827	Fwd ZC3H28 ORF (pHD 3295, pHD 3296, pHD 3297, pHD 3298)	EcoRI	AGA**GAATTC**ATGTATTCTAGCGAAAAGGAGAAG	Y2H bait/prey (ZC3H28)
CZ 7868	Rev ZC3H28 ORF (pHD 3295, 3297)	BamHI	AGTC**GGATCC**TCACGAACGACCCGGATGCTTTGAGG	Y2H bait/prey (ZC3H28)
CZ 7869	Rev middle in ZC3H28 ORF (pHD 3296, pHD 3298)	BamHI	AGTC**GGATCC**TATGTGGTGGTGATGGTGGTGGTGG	Y2H bait/prey (N-terminus ZC3H28)
Primers for Northern Blots				
CZ 2724	Fwd Tubulin mini ORF		TGACTCGCCGCAACCTCGAT	Tubulin probe
CZ 2581	Rev Tubulin mini ORF		CCTTTGGCACAACGTCACCACGG	Tubulin probe
CZ 4615	Fwd CAT mini ORF		ATGGAGAAAAAAATCACTGGATAT	CAT Probe
CZ 2698	Rev CAT mini ORF		GAAAGACGGTGAGCTGGT	CAT Probe
CZ 7066	Fwd NPT ORF		ACGCAGGTTCTCCGGCCGCTT	NPT Probe
CZ 7067	Rev NPT ORF		AATCGGGAGCGGCGATACCGT	NPT Probe
CZ 7593	Fwd RBP10 mini ORF		GAATTACTTCCAGGGCACGCT	RBP10 Probe
CZ 7594	Rev RBP10 mini ORF		AGAACGATGGAGCTGTCGAG	RBP10 Probe
CZ 7670	Fwd Acetyl-CoA ORF		ATGACAGAGGGCCGCGATGC	Acetyl-CoA probe
CZ 7671	Rev Acetyl-CoA ORF		CTAGTGTCTCCGCCGTGGA	Acetyl-CoA probe
CZ 7709	Fwd DRBD12 ORF		ATGAGCTTATTTTCTTTTACGCCAA	DRBD12 probe
CZ 7710	Rev DRBD12 ORF		TAGGTAACACTGGAGGCAAA	DRBD12 probe
Semi-quantitative RT-PCR primers				
CZ 7181	Fwd CAT ORF		GCCAATCCCTGGGTGAGTTT	RT-PCR (polyadenylation site)
CZ 7301	Fwd RBP10 3'-UTR F3		TGTCTGCTTCGTGTTGACGT	RT-PCR (polyadenylation site)

Oligo No.	Description	Restriction site	Sequence	Purpose
CZ 7308	Rev Oligo-dT Mix		GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTT	RT-PCR (polyadenylation site)
CZ 7309	Rev Universal Primer AUAP		ACGCGTATCGATGTCGACTT	RT-PCR (polyadenylation site)
CZ 5725	Fwd Splice leader		5'-ACGCTATTATTAGAACAGTTTCTGTAC-3'	RT-PCR RBP10
CZ 7594	Rev RBP10 mini ORF		5'-AGAACGATGGAGCTGTCGAG-3	RT-PCR RBP10
CZ 6702	Fwd alpha-tubulin		5'-TCACTTCTGGAGCACACCGAT-3'	RT-PCR alpha-tubulin
CZ 6703	Rev alpha-tubulin		5'-GCTCAAACACAGCGTTCGAGA-3'	RT-PCR alpha-tubulin

2.7.2 Plasmid list

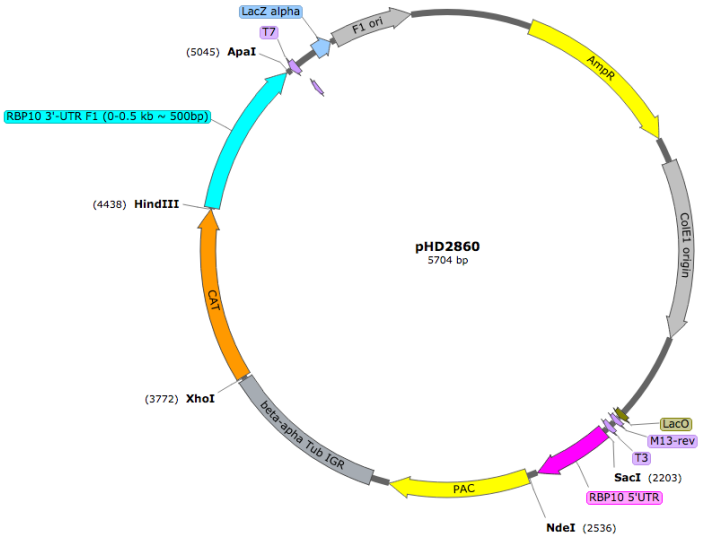
Table 2.23: List of all the plasmids made in this study

pHD number	Description	Primers	Selection marker	Plasmid backbone
CAT Reporter integrated in RBP10 locus				
2860	CAT Reporter, ß-Tub 5'-UTR, RBP10 3'-UTR	CZ 6270, CZ 6271	Puromycin	pHD 2854
CAT Reporters integrated in tubulin locus				
3021	CAT Reporter, EP 5'-UTR, Actin 3'-UTR	CZ 6887, CZ 6888	G418	pHD 2164
2961	CAT Reporter, EP 5'-UTR, F2 RBP10 3'-UTR	CZ 6680, CZ 6679	G418	pHD 2164
2962	CAT Reporter, EP 5'-UTR, F3 RBP10 3'-UTR	CZ 6678, CZ 6677	G418	pHD 2164
3042	CAT Reporter, EP 5'-UTR, F1 RBP10 3'-UTR	CZ 6682, CZ 6681	G418	pHD 2164
3043	CAT Reporter, EP 5'-UTR, F4 RBP10 3'-UTR	CZ 6816, CZ 6817	G418	pHD 2164
3044	CAT Reporter, EP 5'-UTR, F2.2 RBP10 3'-UTR	CZ 6818, CZ 6679	G418	pHD 2164
3045	CAT Reporter, EP 5'-UTR, F2.2.2 RBP10 3'-UTR	CZ 6819, CZ 6679	G418	pHD 2164
3046	CAT Reporter, EP 5'-UTR, F3.2 RBP10 3'-UTR	CZ 6820, CZ 6677	G418	pHD 2164
3047	CAT Reporter, EP 5'-UTR, F3.2.2 RBP10 3'-UTR	CZ 6821, CZ 6677	G418	pHD 2164
3048	CAT Reporter, EP 5'-UTR, F2.1 RBP10 3'-UTR	CZ 6680, CZ 6986	G418	pHD 2164
3049	CAT Reporter, EP 5'-UTR, F2.2.1 RBP10 3'-UTR	CZ 6818, CZ 7002	G418	pHD 2164
3052	CAT Reporter, EP 5'-UTR, F3.1 RBP10 3'-UTR	CZ 6678, CZ 7022	G418	pHD 2164
3053	CAT Reporter, EP 5'-UTR, F3.2.1 RBP10 3'-UTR	CZ 6820, CZ 7070	G418	pHD 2164
3084	CAT Reporter, EP 5'-UTR, F1.4 RBP10 3'-UTR	CZ 7025, CZ 6681	G418	pHD 2164
3085	CAT Reporter, EP 5'-UTR, F2.2.4 RBP10 3'-UTR	CZ 7125, CZ 6679	G418	pHD 2164
3086	CAT Reporter, EP 5'-UTR, F2.2.3 RBP10 3'-UTR	CZ 6819, CZ 7124	G418	pHD 2164
3087	CAT Reporter, EP 5'-UTR, F1.1 RBP10 3'-UTR	CZ 6682, CZ 7208	G418	pHD 2164
3117	CAT Reporter, EP 5'-UTR, F1.2 RBP10 3'-UTR	CZ 7246, CZ 7247	G418	pHD 2164
3118	CAT Reporter, EP 5'-UTR, F1.3 RBP10 3'-UTR	CZ 7248, CZ 7209	G418	pHD 2164
3130	CAT Reporter, EP 5'-UTR, F1.4.1 RBP10 3'-UTR	Mlu/SalI restriction	G418	pHD 3084
3131	CAT Reporter, EP 5'-UTR, F1.4.2 RBP10 3'-UTR	Mlu/XhoI restriction	G418	pHD 3084
3132	CAT Reporter, EP 5'-UTR, F2.2.5 RBP10 3'-UTR	SalI/Eco72I restriction	G418	pHD 3085
3162	CAT Reporter, EP 5'-UTR, F1.2 with $(AU)_{10}$ deleted RBP10 3'-UTR	CZ 7422, CZ 7423	G418	pHD 3117
3202	CAT Reporter, EP 5'-UTR, F3.1.2 RBP10 3'-UTR	CZ 7564, CZ 7022	G418	pHD2164
3203	CAT Reporter, EP 5'-UTR, F3.1.2 with $AAGG(A)_{10}$ deleted RBP10 3'-UTR	CZ 7563, CZ 7022	G418	pHD 3202

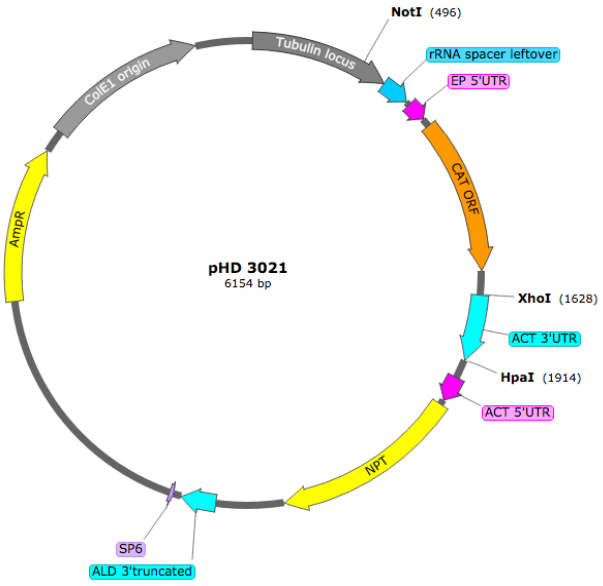
pHD number	Description	Primers	Selection marker	Plasmid backbone
CAT Reporters integrated in rRNA locus				
2277	CAT Reporter, EP 5'-UTR, 5 BoxB elements, ACT 3'-UTR	By Dr. Esteban Erben	G418	pHD 1991
3093	CAT Reporter, EP 5'-UTR, 5 BoxB elements, PSSA-2 3'-UTR	CZ 7238, CZ 7239	G418	pHD 2277
Ectopic inducible expression from rRNA locus (RNAi constructs)				
3161	Stem loop RNAi targeting ZC3H45	CZ 7361, CZ 7362	Hygromycin	pHD 1146
3166	Stem loop RNAi targeting HNRNPFH	CZ 7387, CZ 7388	Hygromycin	pHD 1146
3175	Stem loop RNAi targeting DRBD18	CZ 7386, CZ 7385	Hygromycin	pHD 1146
3179	Stem loop RNAi targeting ZC3H40	CZ 7383, CZ 7384	Hygromycin	pHD 1146
3234	Stem loop RNAi targeting ZC3H28	CZ 7607, CZ 7608	Hygromycin	pHD 1146
3244	RNAi using 2 opposing T7 promoters for HNRNPFH	CZ 7770, CZ 7771	Hygromycin	P2T7
Ectopic inducible expression from rRNA locus (overexpression and tethering constructs)				
3108	LambdaN-myc-BirA*	Dr. Esteban Erben	Hygromycin	pHD 617
3182	LambdaN-HNRNPFH-myc	CZ 7302, CZ 7303	Hygromycin	pHD 2202
2699	Lambda N-myc-ZC3H28	CZ 5720, CZ 5721	Hygromycin	pHD 2412
Single knock-outs				
3065	3'- and 5'-UTR HNRNPFH + Blasticidin resistance cassette	By Dr. Larissa Melo	Blasticidin	pHD 1748
3195	3'- and 5'-UTR DRBD18 + Blasticidin resistance cassette	CZ 7523, CZ7524 CZ7525, CZ7526	Blasticidin	pHD 1748
3294	3'- and 5'-UTR ZC3H28 + Blasticidin resistance cassette	CZ 7652, CZ 7653, CZ 7707, CZ 7708	Blasticidin	pHD 1748
***In situ* tagging constructs**				
2910	*In situ* N-TAP ZC3H20	(Liu et al. 2019)	Puromycin	pHD 1959
3094	*In situ* N-TAP HNRNPFH	By Dr. Larissa Melo	Puromycin	pHD 1959
3196	*In situ* N-TAP DRBD18	CZ 7554, CZ 7522, CZ 7519, CZ 7520	Puromycin	pHD 1959
3200	*In situ* C-TAP DRBD18	CZ 7522, CZ 7523	G418	pHD 3195
3236	*In situ* N-TAP ZC3H28	CZ 7693, CZ 7655, CZ 7649, CZ 7656	Puromycin	pHD 1959
Yeast-two hybrids constructs				
3295	pGADT7+ Tb927.9.9450 (AD-ZC3H28)	CZ 7827, CZ 7868	Ampicillin	pGADT7
3296	pGADT7+ Tb927.9.9450 N-terminal (AD-N terminal ZC3H28)	CZ 7827, CZ 7869	Ampicillin	pGADT7
3297	pGBKT7+ Tb927.9.9450 (BD-ZC3H28)	CZ 7827, CZ 7868	Kanamycin	pGBKT7
3298	pGBKT7+ Tb927.9.9450 N-terminal (BD-N terminal ZC3H28)	CZ 7827, CZ 7869	Kanamycin	pGBKT7

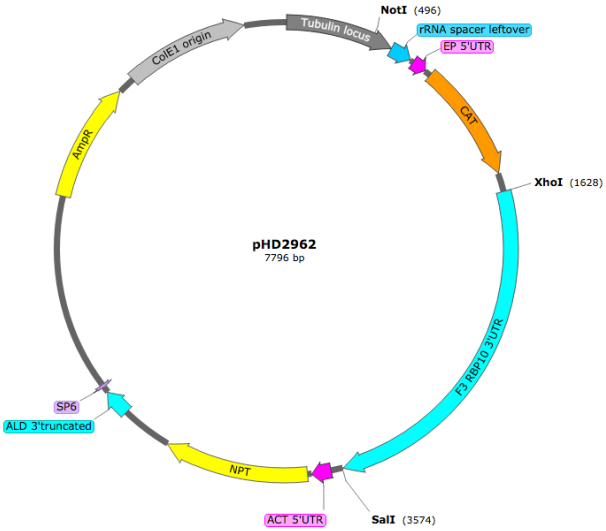
2.7.3 Plasmid maps

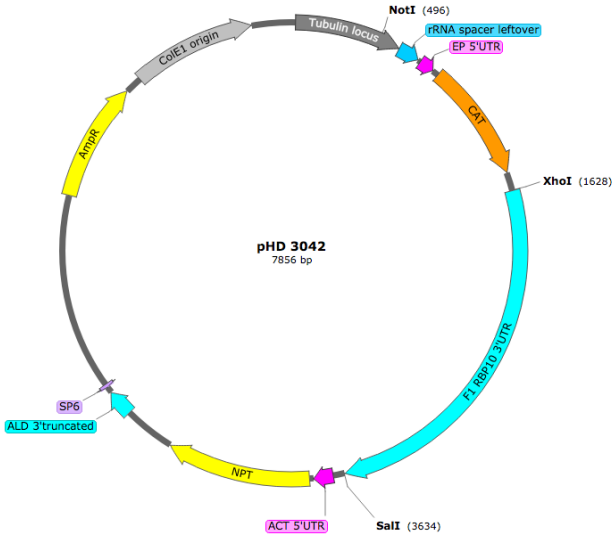
CAT Reporter integrated in RBP10 locus

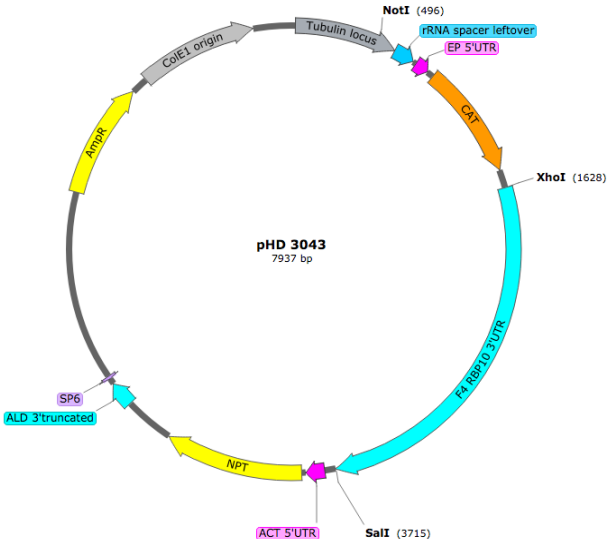


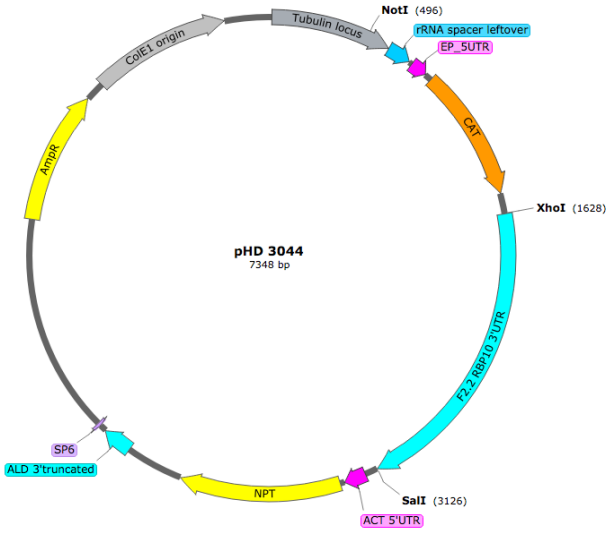
CAT Reporter integrated in the tubulin locus

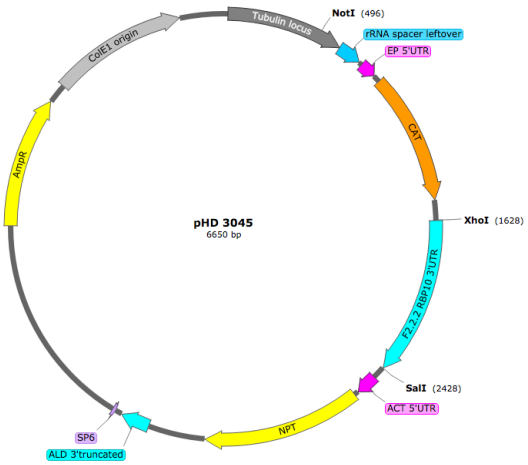


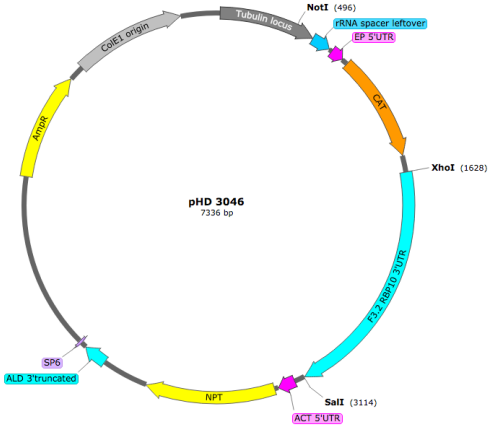

pHD2962
7796 bp
NotI (496)
Tubulin locus
rRNA spacer leftover
EP 5'UTR
CAT
XhoI (1628)
F3 RBP10 3'UTR
SalI (3574)
ACT 5'UTR
NPT
ALD 3'truncated
SP6
AmpR
ColE1 origin

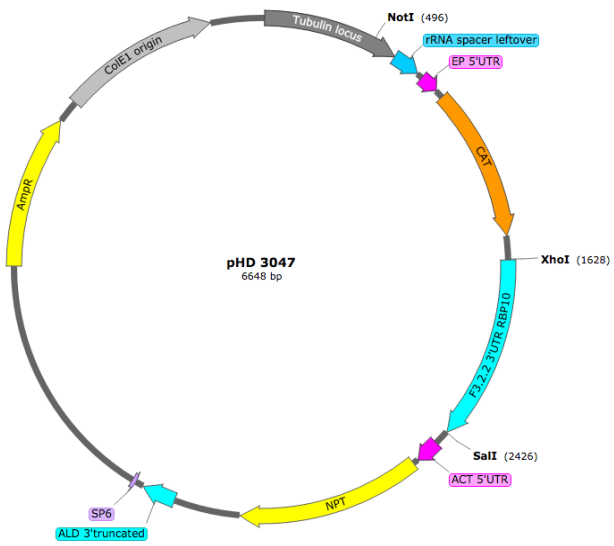

pHD 3042
7856 bp
NotI (496)
Tubulin locus
rRNA spacer leftover
EP 5'UTR
CAT
XhoI (1628)
F1 RBP10 3'UTR
SalI (3634)
ACT 5'UTR
NPT
ALD 3'truncated
SP6
AmpR
ColE1 origin

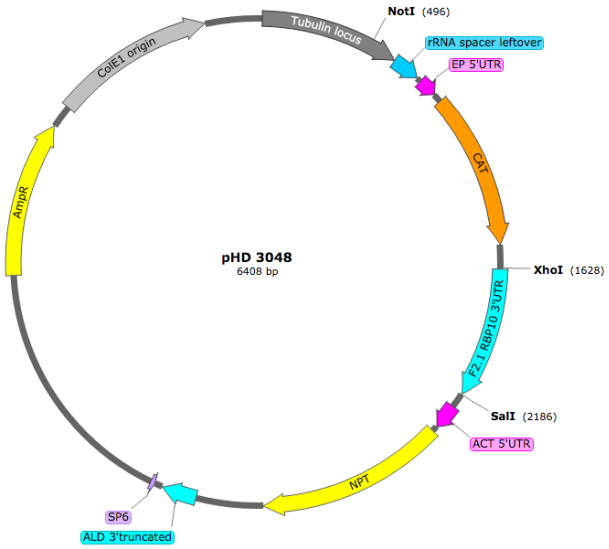

pHD 3043
7937 bp
NotI (496)
Tubulin locus
rRNA spacer leftover
EP 5'UTR
CAT
XhoI (1628)
F4 RBP10 3'UTR
SalI (3715)
ACT 5'UTR
NPT
ALD 3'truncated
SP6
AmpR
ColE1 origin

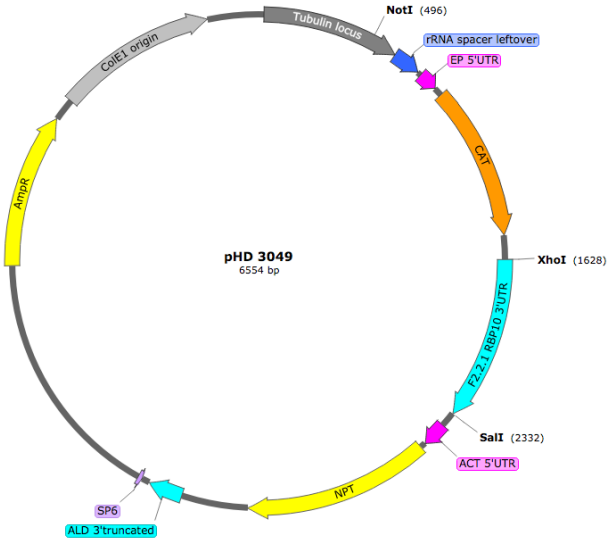

pHD 3044
7348 bp
NotI (496)
Tubulin locus
rRNA spacer leftover
EP_5UTR
CAT
XhoI (1628)
F2.2 RBP10 3'UTR
SalI (3126)
ACT 5'UTR
NPT
ALD 3'truncated
SP6
AmpR
ColE1 origin

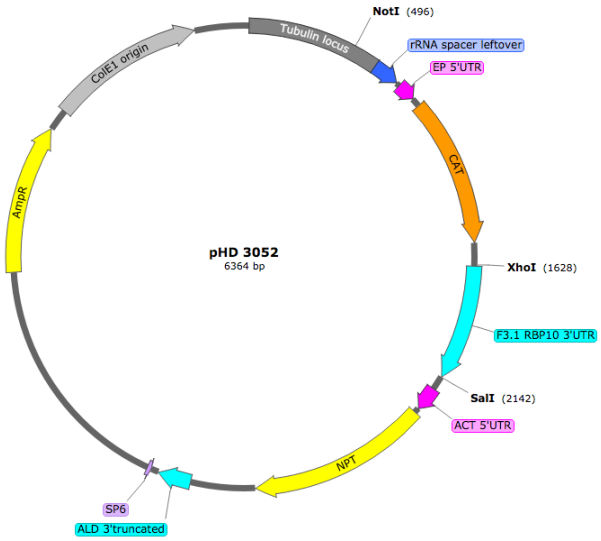

pHD 3045
6650 bp
NotI (496)
Tubulin locus
rRNA spacer leftover
EP 5'UTR
CAT
XhoI (1628)
F2.2 RBP10 3'UTR
SalI (2428)
ACT 5'UTR
NPT
ALD 3'truncated
SP6
AmpR
ColE1 origin


pHD 3046
7336 bp
NotI (496)
Tubulin locus
rRNA spacer leftover
EP 5'UTR
CAT
XhoI (1628)
F2.2 RBP10 3'UTR
SalI (3114)
ACT 5'UTR
NPT
ALD 3'truncated
SP6
AmpR
ColE1 origin


pHD 3047
6648 bp
NotI (496)
Tubulin locus
rRNA spacer leftover
EP 5'UTR
CAT
XhoI (1628)
F3.2.2 3'UTR RBP10
SalI (2426)
ACT 5'UTR
NPT
SP6
ALD 3'truncated
AmpR
ColE1 origin

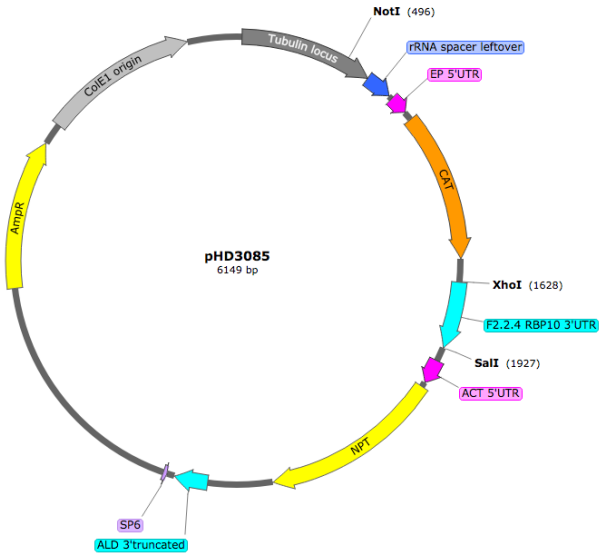

pHD 3048
6408 bp
NotI (496)
Tubulin locus
rRNA spacer leftover
EP 5'UTR
CAT
XhoI (1628)
F2.1 RBP10 3'UTR
SalI (2186)
ACT 5'UTR
NPT
SP6
ALD 3'truncated
AmpR
ColE1 origin

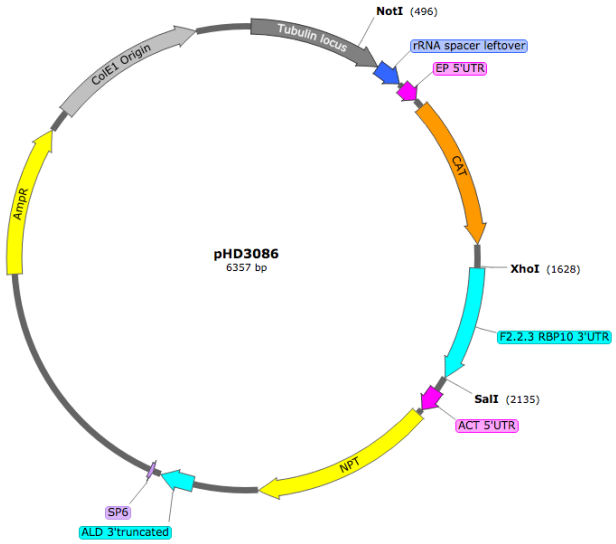

pHD 3049
6554 bp
NotI (496)
Tubulin locus
rRNA spacer leftover
EP 5'UTR
CAT
XhoI (1628)
F2.2.1 RBP10 3'UTR
SalI (2332)
ACT 5'UTR
NPT
SP6
ALD 3'truncated
AmpR
ColE1 origin

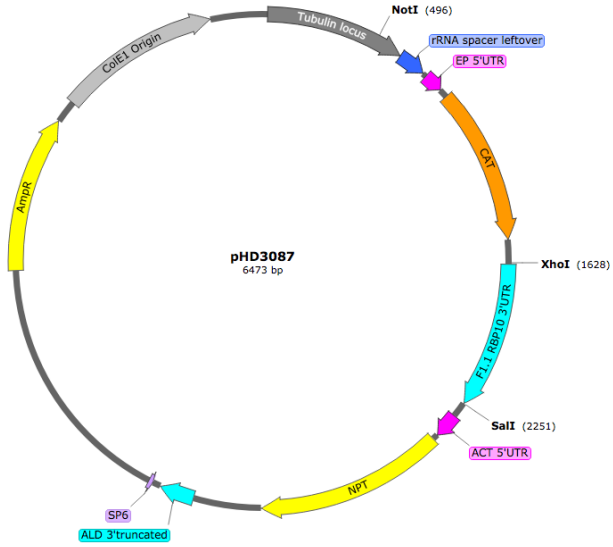

pHD 3052
6364 bp
NotI (496)
Tubulin locus
rRNA spacer leftover
EP 5'UTR
CAT
XhoI (1628)
F3.1 RBP10 3'UTR
SalI (2142)
ACT 5'UTR
NPT
SP6
ALD 3'truncated
AmpR
ColE1 origin

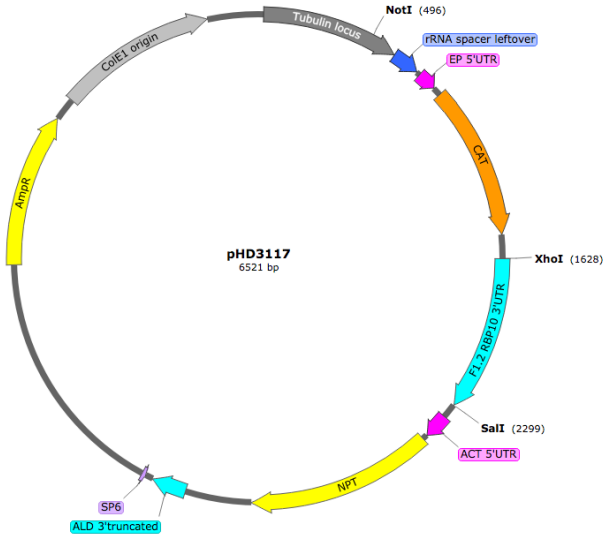

pHD 3053
6551 bp
NotI (496)
Tubulin locus
rRNA spacer leftover
EP 5'UTR
CAT
XhoI (1628)
F3.2.1 RBP10 3'UTR
SalI (2329)
ACT 5'UTR
NPT
SP6
ALD 3'truncated
AmpR
ColE1 origin

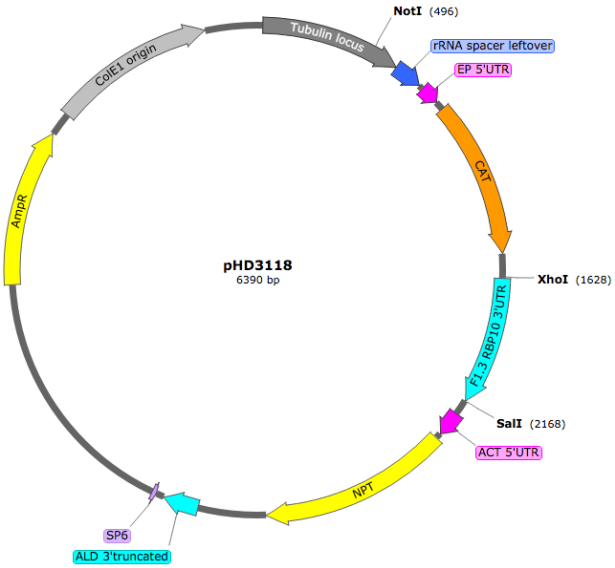

pHD3084
6091 bp
NotI (496)
Tubulin locus
rRNA spacer leftover
EP 5'UTR
CAT
XhoI (1628)
F1.4 RBP10 3'UTR
SalI (1869)
ACT 5'UTR
NPT
SP6
ALD 3'truncated
AmpR
ColE1 origin


NotI (496)
Tubulin locus
rRNA spacer leftover
EP 5'UTR
CAT
XhoI (1628)
F2.2.4 RBP10 3'UTR
SalI (1927)
ACT 5'UTR
NPT
SP6
ALD 3'truncated
AmpR
ColE1 origin
pHD3085
6149 bp

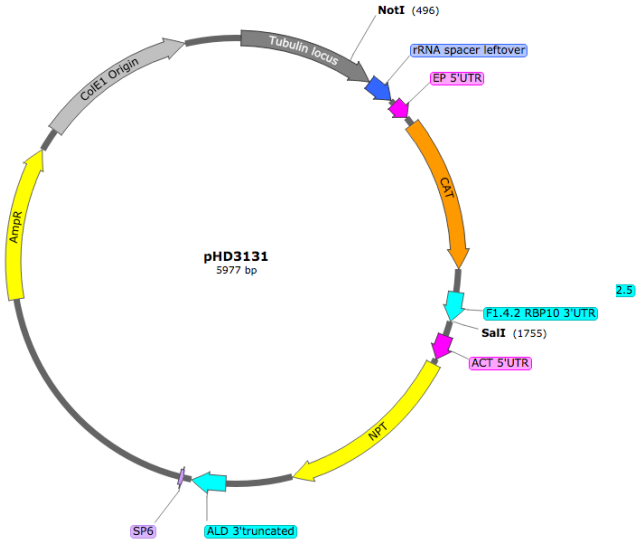

NotI (496)
Tubulin locus
rRNA spacer leftover
EP 5'UTR
CAT
XhoI (1628)
F2.2.3 RBP10 3'UTR
SalI (2135)
ACT 5'UTR
NPT
SP6
ALD 3'truncated
AmpR
ColE1 Origin
pHD3086
6357 bp

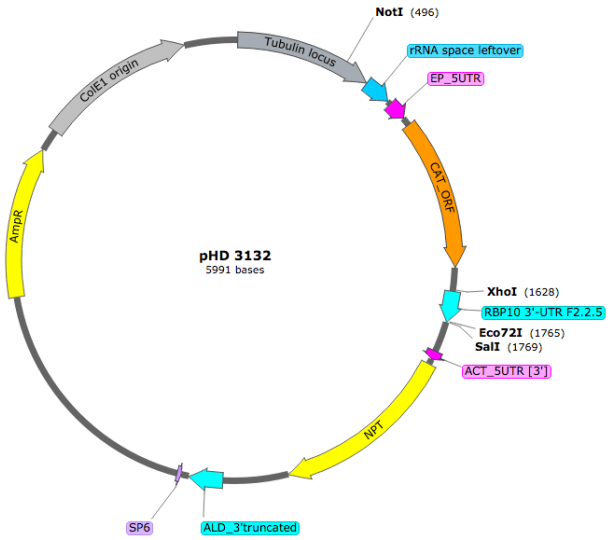

NotI (496)
Tubulin locus
rRNA spacer leftover
EP 5'UTR
CAT
XhoI (1628)
F1.1 RBP10 3'UTR
SalI (2251)
ACT 5'UTR
NPT
SP6
ALD 3'truncated
AmpR
ColE1 Origin
pHD3087
6473 bp

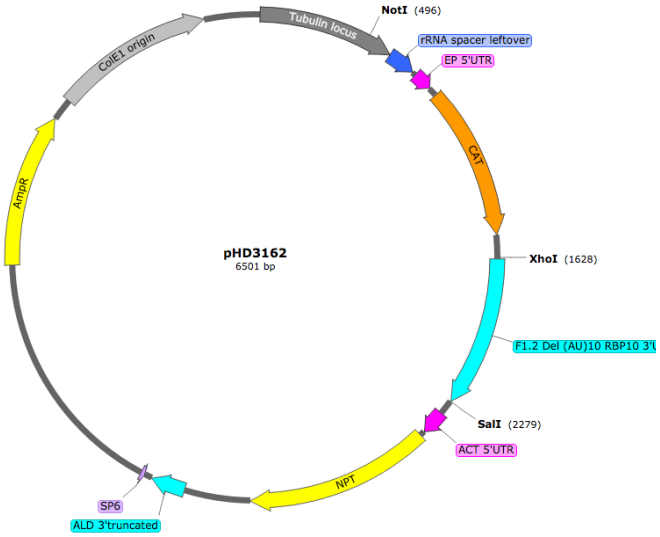

NotI (496)
Tubulin locus
rRNA spacer leftover
EP 5'UTR
CAT
XhoI (1628)
F1.2 RBP10 3'UTR
SalI (2299)
ACT 5'UTR
NPT
SP6
ALD 3'truncated
AmpR
ColE1 origin
pHD3117
6521 bp

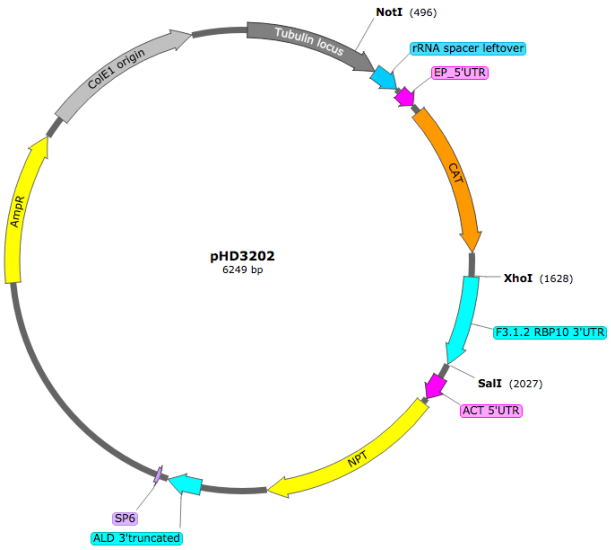

NotI (496)
Tubulin locus
rRNA spacer leftover
EP 5'UTR
CAT
XhoI (1628)
F1.3 RBP10 3'UTR
SalI (2168)
ACT 5'UTR
NPT
SP6
ALD 3'truncated
AmpR
ColE1 origin
pHD3118
6390 bp

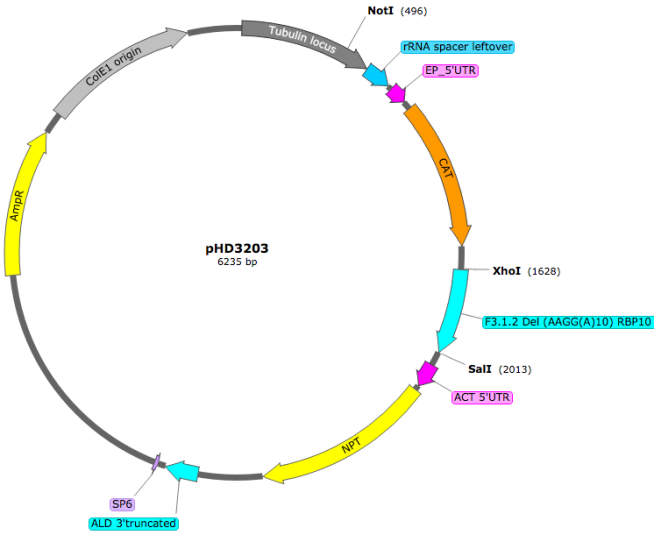

NotI (496)
Tubulin locus
rRNA spacer leftover
EP 5'UTR
CAT
XhoI (1628)
F1.4.1 RBP10 3'UTR
MluI (1742)
SalI (1748)
ACT 5'UTR
NPT
SP6
ALD 3'truncated
AmpR
ColE1 Origin
pHD3130
5970 bp


NotI (496)
Tubulin locus
rRNA spacer leftover
EP 5'UTR
CAT
F1.4.2 RBP10 3'UTR
SalI (1755)
ACT 5'UTR
NPT
SP6
ALD 3'truncated
AmpR
ColE1 Origin
pHD3131
5977 bp

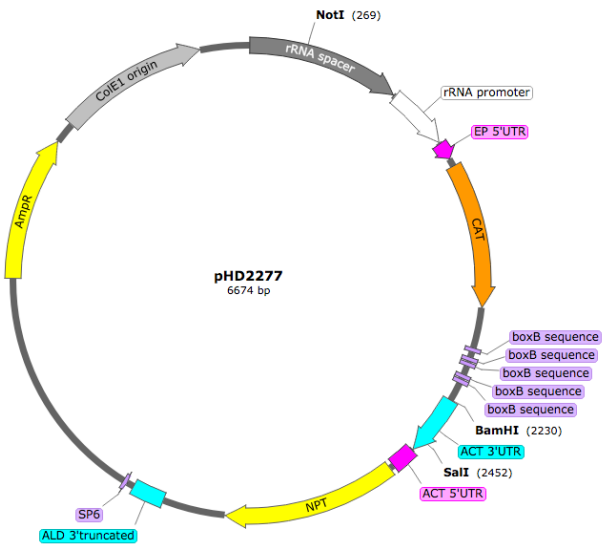

NotI (496)
Tubulin locus
rRNA space leftover
EP_5UTR
CAT ORF
XhoI (1628)
RBP10 3'-UTR F2.2.5
Eco72I (1765)
SalI (1769)
ACT_5UTR [3']
NPT
SP6
ALD_3'truncated
AmpR
ColE1 origin
pHD 3132
5991 bases

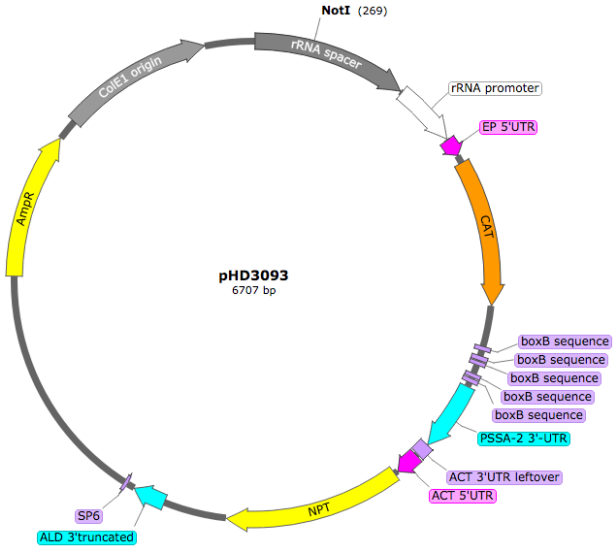

NotI (496)
Tubulin locus
rRNA spacer leftover
EP 5'UTR
CAT
XhoI (1628)
SalI (2279)
ACT 5'UTR
NPT
SP6
ALD 3'truncated
AmpR
ColE1 origin
pHD3162
6501 bp


NotI (496)
Tubulin locus
rRNA spacer leftover
EP_5'UTR
CAT
XhoI (1628)
F3.1.2 RBP10 3'UTR
SalI (2027)
ACT 5'UTR
NPT
SP6
ALD 3'truncated
AmpR
ColE1 origin
pHD3202
6249 bp

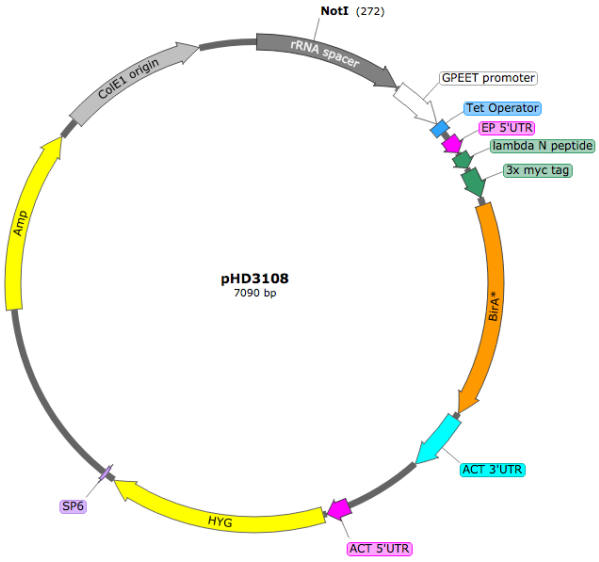

NotI (496)
Tubulin locus
rRNA spacer leftover
EP_5'UTR
CAT
XhoI (1628)
SalI (2013)
ACT 5'UTR
NPT
SP6
ALD 3'truncated
AmpR
ColE1 origin
pHD3203
6235 bp

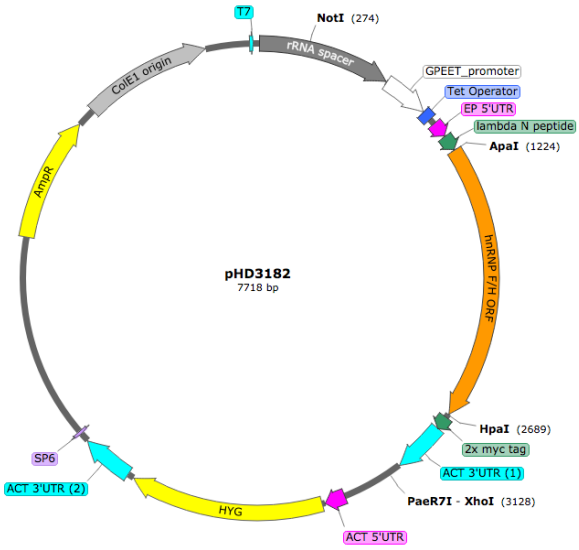
CAT Constructs integrated in the rRNA locus

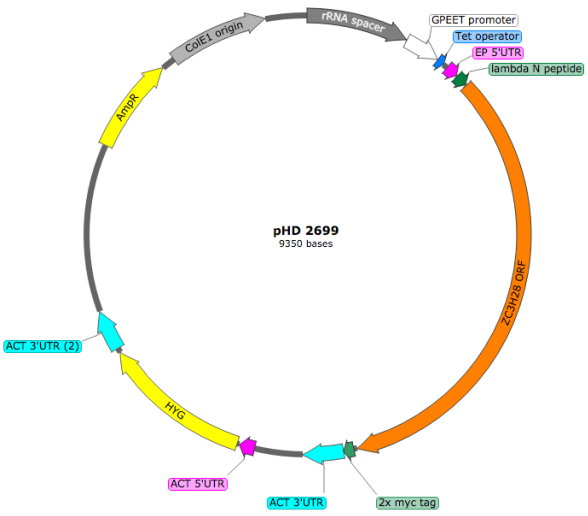




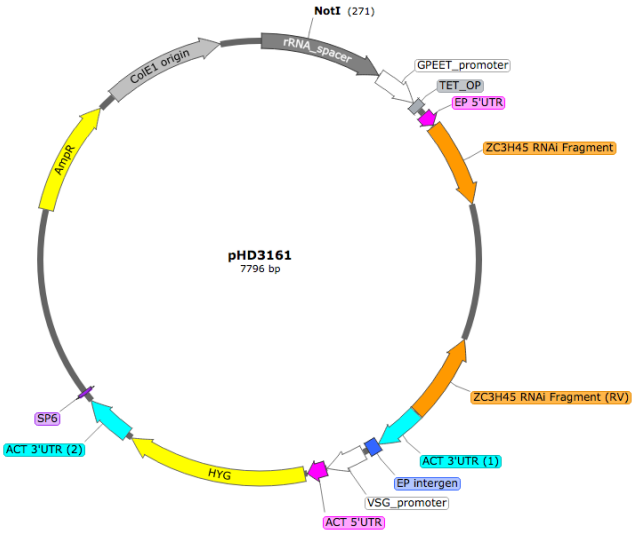
Tethering constructs

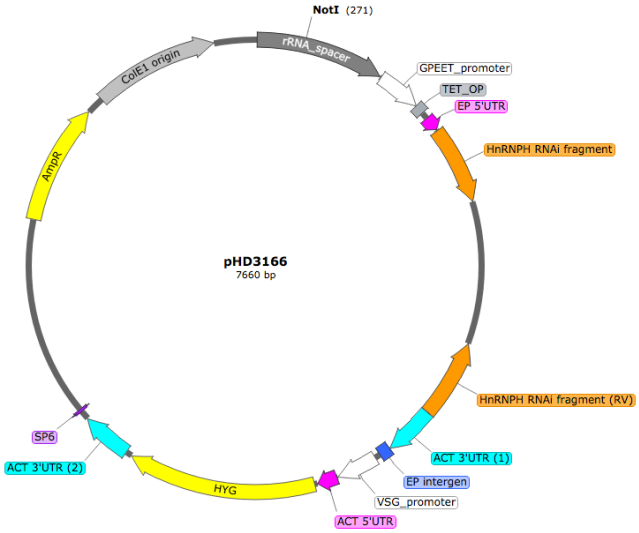


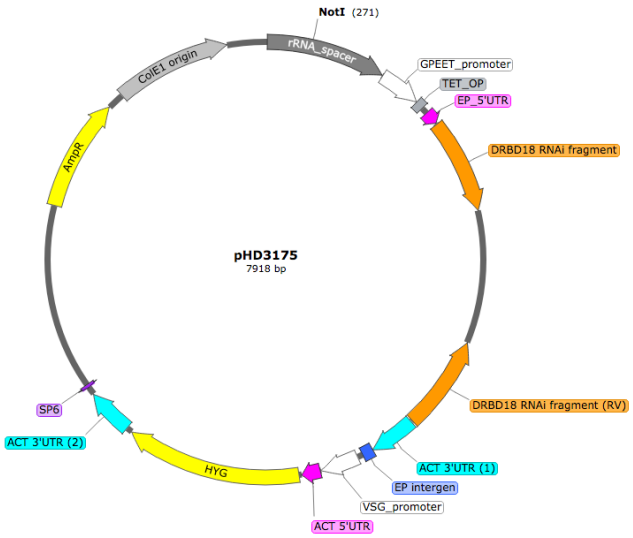


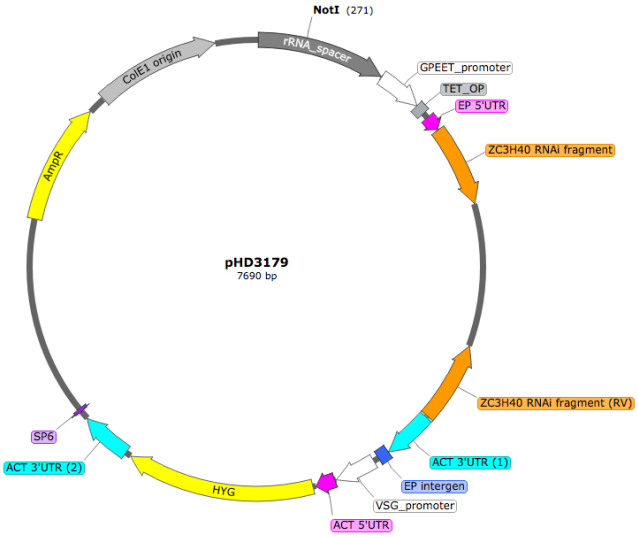


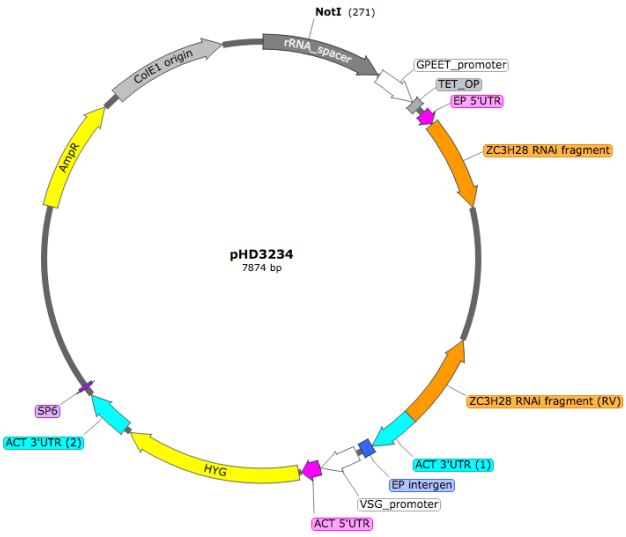
RNAi Constructs

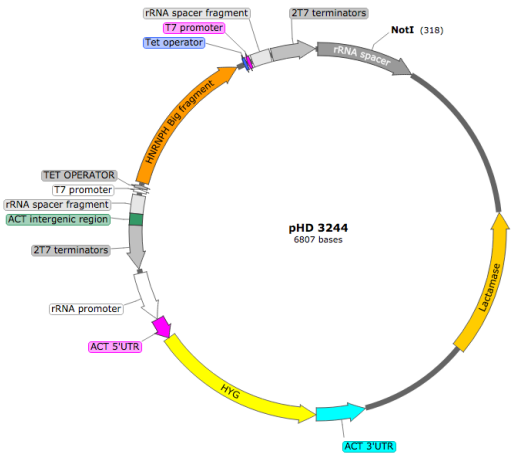




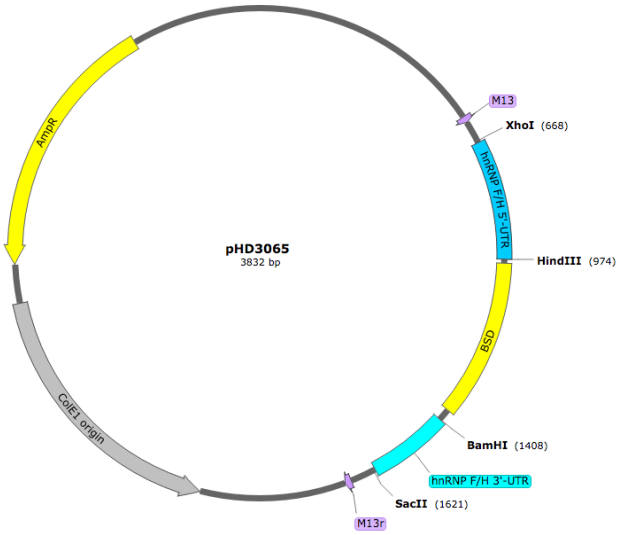


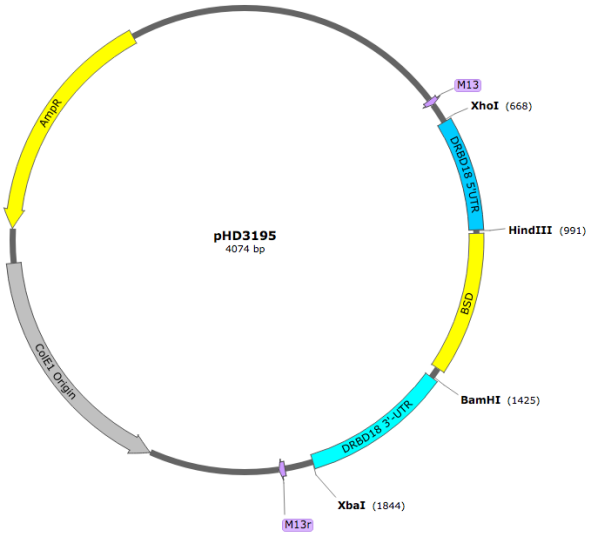


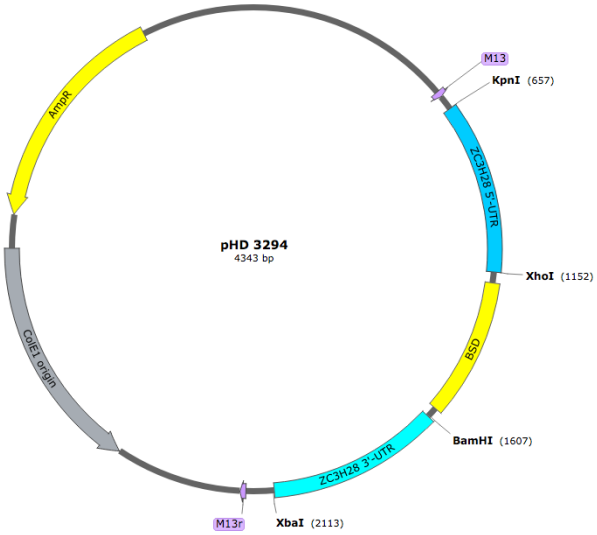




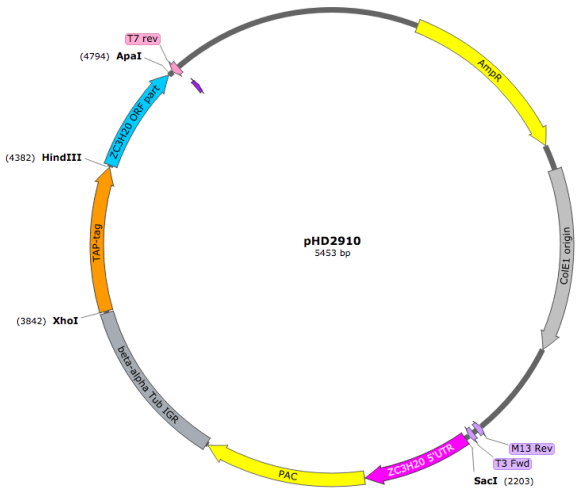
Single knockouts

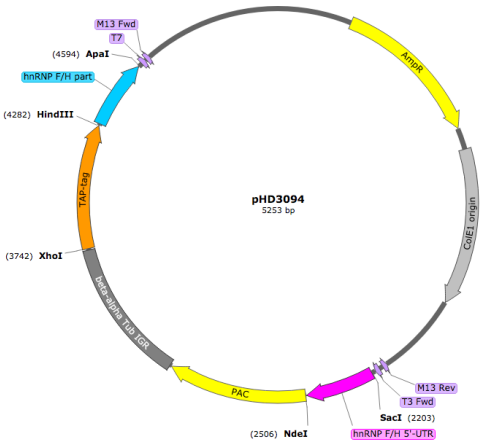


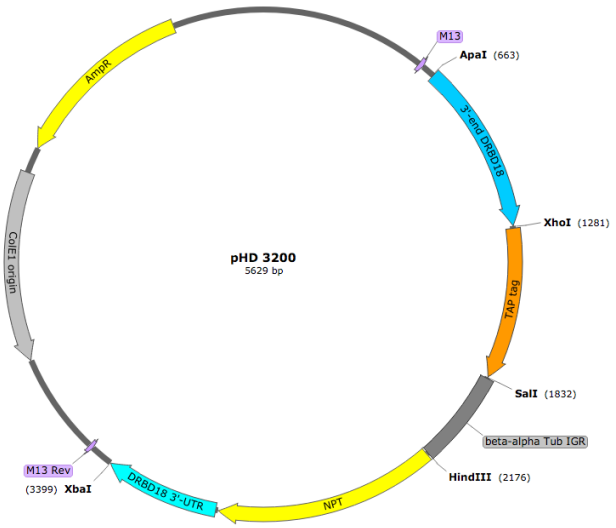


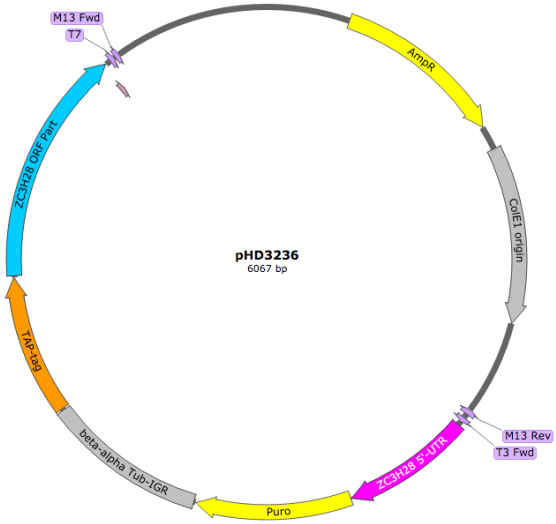


In situ tagging constructs

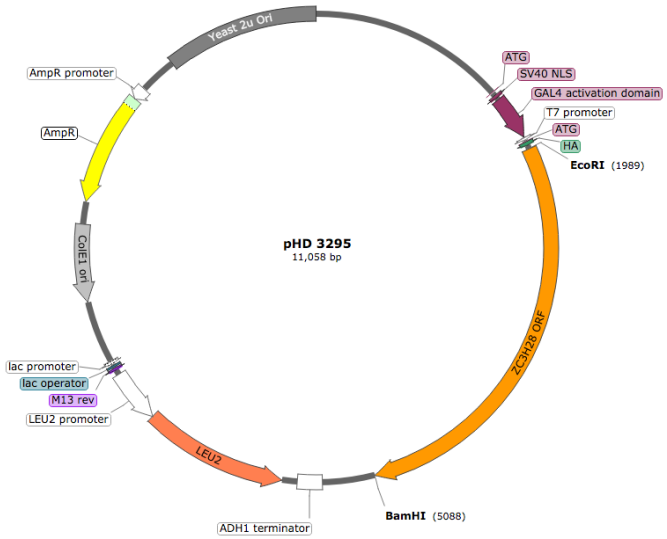


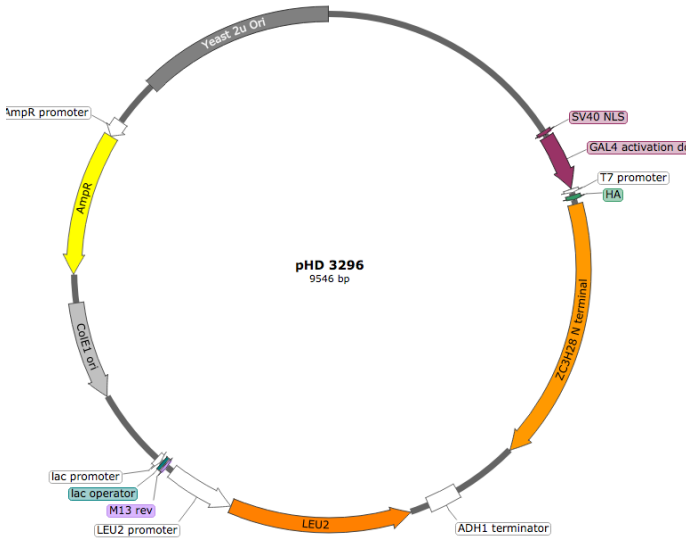


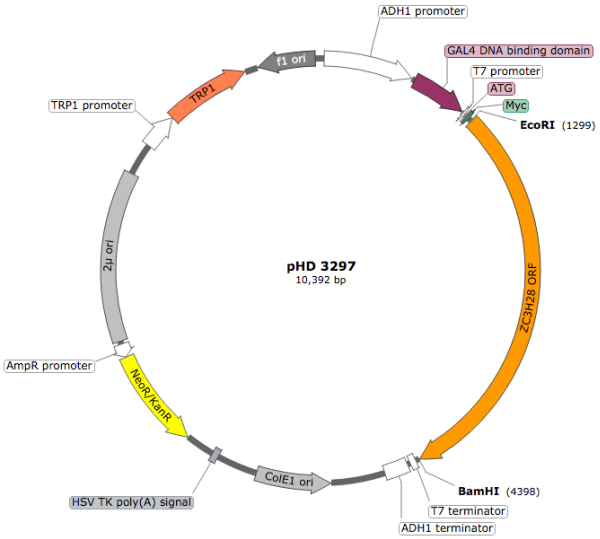


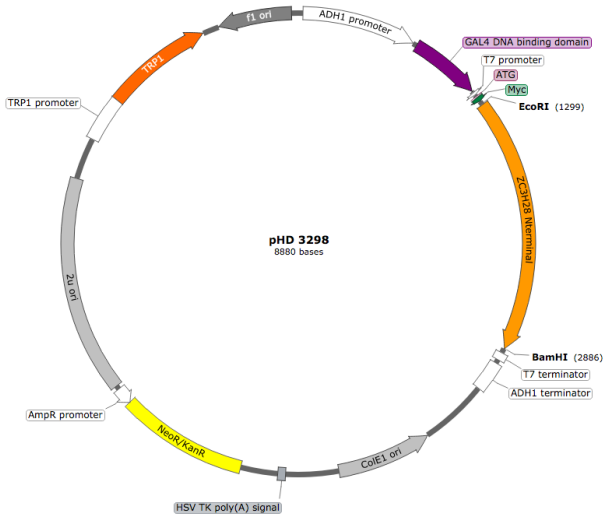


Yeast two-hybrid constructs









2.8 Web resources

Web resources	Links
BLAST NCBI	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Clustal Omega	https://www.ebi.ac.uk/Tools/msa/clustalo/
InterPro	http://www.ebi.ac.uk/interpro/enry/InterPro/IPR035979/
mfold Web server	http://unafold.rna.albany.edu/?q=mfold/download-mfold
MEME Motif Discovery	http://meme-suite.org/
RNAi tool	https://dag.compbio.dundee.ac.uk/RNAit/
TritryDB	https://tritrypdb.org

3. Results

3.1 The *RBP10* 3'-UTR is sufficient for RBP10 developmental regulation

The first objective of my thesis was to identify the regulatory sequences that are required for RBP10 developmental expression. Since most *cis*-acting elements are found in the 3'-UTRs of the mRNAs (Clayton 2014), I herein investigated the role of *RBP10* 3'-UTR in mediating stability and translation of the mRNA in bloodstream forms, and its degradation and repression in procyclic forms. In this part of my thesis, I also addressed the question of finding regulatory motifs within the *RBP10* 3'-UTR and to see whether they are shared with other co-regulated mRNAs.

3.1.1 *RBP10* mRNA is developmentally regulated

At first glance, high-throughput data were checked in order to obtain a detailed overview of RBP10 regulation. Previous studies from our lab (Wurst et al. 2012) together with quantitative proteomics (Dejung et al. 2016) reveal that RBP10 (Tb927.8.2780) is only expressed in long slender forms and barely detectable in stumpy or procyclic forms. The transcriptomics data indicate that there are about 4 copies of the *RBP10* mRNA per cell in bloodstream forms and approximately 1 per cell in procyclics while the ribosome occupancy on the coding region is 9 times higher in bloodstream forms as compared to procyclics (Fadda et al. 2014; Antwi et al. 2016). Results from RNA-Seq and ribosome profiling (Jensen et al. 2014) also suggest that *RBP10* 3'-UTR is approximately 7.3 kb long, which gives the *RBP10* mRNA a total length of about 8.5 kb (Figure 3.1 A). In order to confirm these profiles, Northern blots from the EATRO 1125 strain bloodstream and procyclic forms were carried out with probes from *RBP10* coding region (Figure 3.1 B). The *RBP10* mRNA migrated slower than the 6 kb marker (Figure 3.1 B) and extrapolation using a semi-logarithmic equation suggested a length of approximately 8.5 kb. The quantitative analysis of the northern blot also indicated that there is about 8-fold more *RBP10* mRNA in bloodstream forms than in procyclic forms (Figure 3.1 B), confirming previous transcriptomics data and clearly revealing that the *RBP10* mRNA is developmentally regulated. Finding the polyadenylation site (PAS) using the 3'-RACE with oligo $(dT)_{20}$ was prone to errors as the results gave position 6315-6318 after the stop codon as the potential RBP10 PAS. It is essential to know that the middle region of the *RBP10* 3'-UTR, giving grey colored reads in the alignment (Figure 3.1 A) is also found in another contiguous sequence in the TREU927 genome assembly but only once in the Lister 427 strain genome (Müller et al. 2018). The gene immediately downstream of RBP10 (Tb927.8.2790) is annotated as an acetyl-coA synthetase pseudogene (Figure 3.1 A) and reads alignment over that region suggests that it is present both in bloodstream and

procyclic forms. A complete coding region of the acetyl-coA synthetase is present elsewhere in the genome (Tb927.8.2520).

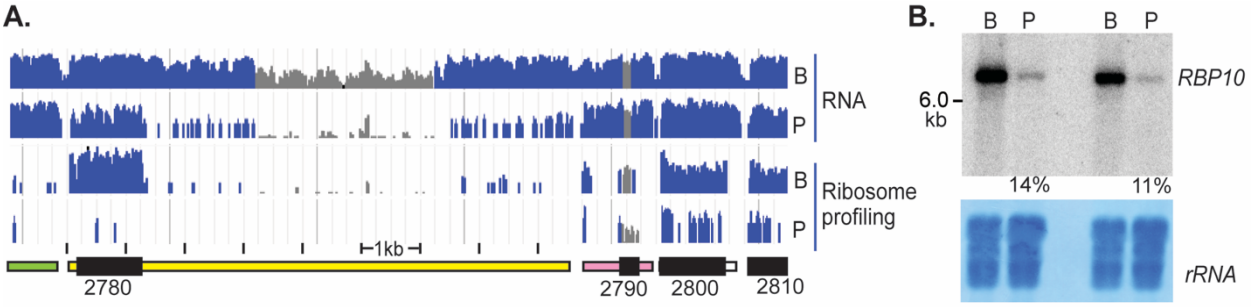


Figure 3.1: *RBP10* mRNA is developmentally regulated in *T. brucei*. (A) The panels are from TritryDB database and Jensen et al. 2014. They show RNA-sequencing data (RNA) and ribosome profiling reads from bloodstream forms (B) and procyclic forms (P) aligned to the relevant segment of the TREU927 reference genome. Unique reads are in blue and non-unique reads are in grey (this is on a log scale). The protein IDs are designated without Tb927.8 for simplicity: RBP10 is Tb927.8.2780; for simplicity, it is written 2780. Open reading frames are drawn as black segments and untranslated regions have different colors. The sequences indicated as "non-unique" in the middle of *RBP10* 3'-UTR (~2.5 – 4.5 kb) are also present in an TREU927 DNA contiguous sequence that has not been assigned to a specific chromosome; this segment is not present in the Lister 427 genome (2018). Transcription here is from left to right. This figure is from Prof. C. Clayton. **(B)** Northern blot for two independent cultures of EATRO 1125 showing the relative mRNA abundance of RBP10 in bloodstream (B) and in procyclic forms (P). A section of methylene blue staining is depicted to show the loading; and the measured amounts in procyclic forms relative to bloodstream forms are also shown.

3.1.2 *RBP10* mRNA is stable in bloodstream forms

Due to the importance of posttranscriptional regulation in trypanosome gene expression, the most probable mechanism for controlling mRNA abundance is via mRNA stability and degradation (Clayton 2014). Therefore, to gain insight into the stability of *RBP10* mRNA in bloodstream and in procyclic forms, I inhibited transcription and *trans*-splicing respectively with Actinomycin D (10 µg/ml) and Sinefungin (2 µg/ml) and followed the mRNA decay at different time points in each life cycle via northern blotting. For each case, the mRNA levels were quantified with respect to the amount of rRNA loaded (Figure 3.2 A). Measurements from three independent replicates in bloodstream forms suggest an *RBP10* mRNA half-life of about 2 hours (Figure 3.2 B) while in procyclics, the mRNA was barely detectable after 30 minutes (Figure 3.2 A). With rapid degradation of the *RBP10* mRNA in procyclics, extrapolation of its half-life in this life stage was rather difficult. Previous measurements of mRNA decay in trypanosomes by RNA-Seq suggested that *RBP10* mRNA has a half-life of over one hour in Lister 427 bloodstream forms while in procyclics, the data were poor, probably also due the low levels of the mRNA in this life stage (Fadda et al. 2014).

With regard to this result, *RBP10* mRNA reveals to be relatively stable in bloodstream forms and unstable, prone to rapid destruction in procyclic forms.

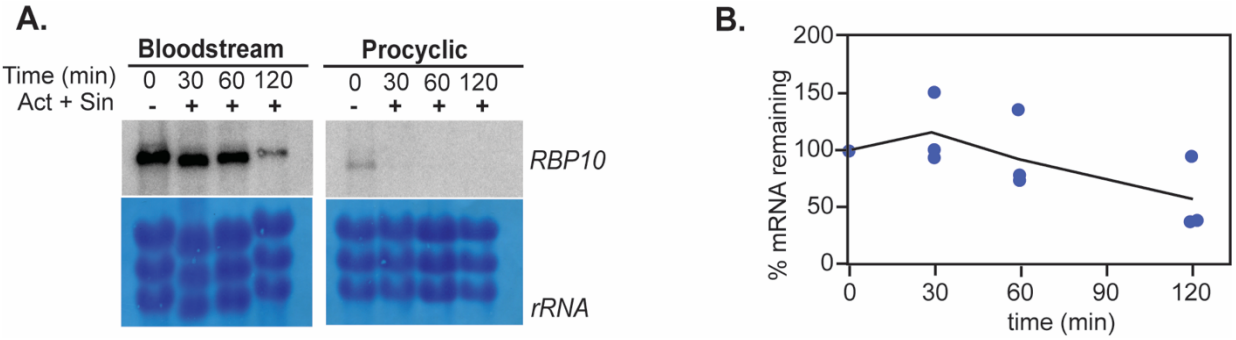


Figure 3.2: Half-life of *RBP10* mRNA in bloodstream and procyclic forms. *(*A) Representative Northern blots of *RBP10* mRNA from bloodstream and procyclic form cells. The cells were treated at indicated times with 10 µg/ml actinomycin D and 2 µg/ml Sinefungin, which stop both mRNA processing and transcription. RNA was isolated at the times shown and the untreated cells (-) were set as control. *RBP10* mRNA is stable in bloodstream forms while in procyclics, the mRNA was not detectable already after 30 minutes. **(B)** The amount of *RBP10* mRNA in bloodstream forms was measured in three independent replicates. The black line joins the arithmetic mean from the three replicates. There is an initial increase in mRNA abundance over time then the amount decreases after 1 hour. This is usually seen for stable trypanosomes mRNAs; the reason is unknown.

3.1.3 RBP10 expression level is regulated during differentiation

Knowing that *RBP10* mRNA and protein levels are developmentally regulated in trypanosomes, I next searched for inducers of trypanosome differentiation that correlate with the decline of RBP10 expression. Boshart and Engstler 2004 showed that a cold shock of 20°C was enough to induce the expression of EP procyclic surface protein; and at the same time this sensitizes trypanosomes to chemical inducers of cell differentiation. A cold shock of 20°C was therefore comparable to the standard *in vitro* differentiation protocol (Boshart and Engstler 2004). To search for differentiation inducers that correspond with the decline of RBP10 expression, pleomorphic bloodstream form cells EATRO 1125 were cultured for 6 and 24 hours at 27°C and at 20 °C with or without 6mM *cis*-aconitate. RBP10 protein was detected by Western blotting using anti-RBP10 polyclonal antibody (Figure 3.3 A). The experiment was done thrice using the same conditions. Quantification of RBP10 expression levels from the three independent experiments is plotted in Figure 3.1.3 B. Growing the cells at either 27°C or 20°C without the differentiation inducer, *cis*-aconitate, did not affect RBP10 expression levels although at 20°C, in one replicate there was a decrease of approximately 1.5-fold after 24 hours (Figure 3.3 B). However, in presence of *cis*-aconitate, RBP10 protein levels decrease about 1.2-fold already after 6 hours at either 27°C or 20°C. And after 24 hours, the decrease was more pronounced and reaching about 3-fold for both temperatures used. This result shows that the cold shock alone does not affect drastically RBP10 expression. Sensing the temperature drop and responding to the inducers of differentiation

trigger the reduction in RBP10 expression levels. Therefore, this differentiation condition was used for downstream experiments while searching for sequences involved in RBP10 developmental regulation.

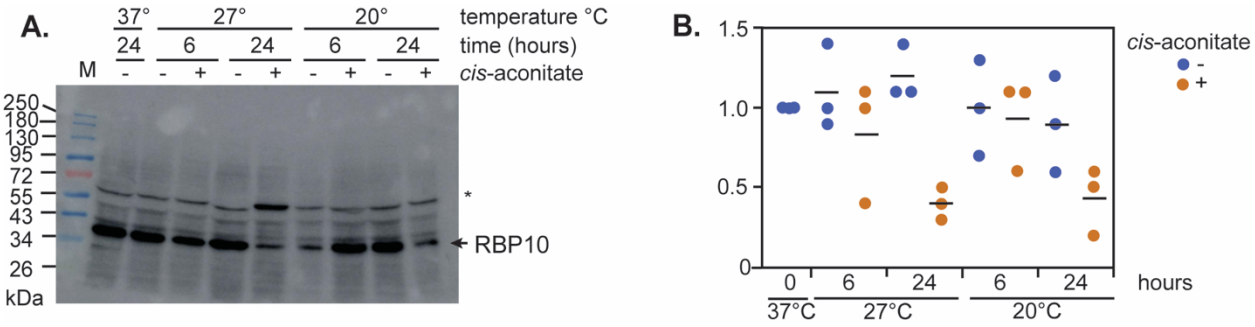


Figure 3.3: Regulation of RBP10 protein expression during differentiation. (A) Representative Western blot of RBP10 protein expression levels during differentiation. Bloodstream form cells EATRO 1125 growing at logarithmic phase ($5x10^5$) were cultured in HMI-9 medium with or without *cis*-aconitate at 27°C and 20°C. Cells were harvested after 6 and 24 hours. Cells incubated at 37°C (first lane) were used as control. Anti-RBP10 polyclonal antibody was used for detection of RBP10 and the asterisk (*), a non-specific band was used as a loading control. M is the marker. **(B)** The amount of RBP10 protein levels was measured from three independent replicates (similar to figure A) at time and temperatures shown, in presence or absence of cis-aconitate. The mean is indicated for each time point. Decline in RBP10 expression levels was more pronounced in presence of cis-aconitate regardless of the incubation temperature.

3.1.4 The *RBP10* 3'-UTR regulates RBP10 developmental expression

To find out whether the developmental regulation of RBP10 is mediated by the 3'-UTR of its mRNA, the ability of *RBP10* 3'-UTR to appropriately regulate a heterologous reporter gene encoding for chloramphenicol acetyl transferase (CAT) was investigated. To achieve this purpose, one allele of the RBP10 open reading frame (ORF) in pleomorphic bloodstream forms was replaced with the CAT reporter gene (Figure 3.4 A). Transfectants were selected with puromycin using the *PAC* (Puromycin N-acetyltransferase) resistance marker. Integration of the *CAT* gene upstream of the *RBP10* 3'-UTR was checked using PCR with primers annealing the CAT open reading and the *RBP10* 3'-UTR (Figure 3.4 A, red arrows), giving an amplicon of about 840 bp. All transfectants show amplification above the 800 bp marker and no amplification was detected in the wild-type cell line or in the water control (Figure 3.4 B). The size of the *CAT* mRNA in all the clones was also confirmed using Northern blot. The *CAT* mRNA migrated slower than the 6 kb marker (Figure 3.4 C); semi-logarithmic equation suggests a length of about 8.4 kb. This indicates polyadenylation of the *CAT* mRNA at the proper site. To confirm whether the single knockout had an effect on growth, the cell growth of three different transfectants $RBP10^{+/-}$ were checked for 5 days together with a wild-type cell line. As shown in Figure 3.4 D, there was no drastic effect on growth. The bloodstream-form transfectants $RBP10^{+/-}$ were then differentiated to procyclic forms by

addition of 6 mM *cis*-aconitate and a temperature shift from 37°C to 27°C as described in Queiroz et al. 2009. The CAT activity and the *CAT* mRNA levels were measured by CAT assay and Northern blot, respectively. A down-regulation of about 3-fold was observed in the *CAT* mRNA levels in procyclics relative to bloodstream forms. On the other hand, the CAT activity was not detectable at all in procyclics (Figure 3.4 E). This result mimics the expression profile of RBP10, revealing that the *RBP10* 3'-UTR is sufficient for development regulation of *RBP10* mRNA and protein levels.

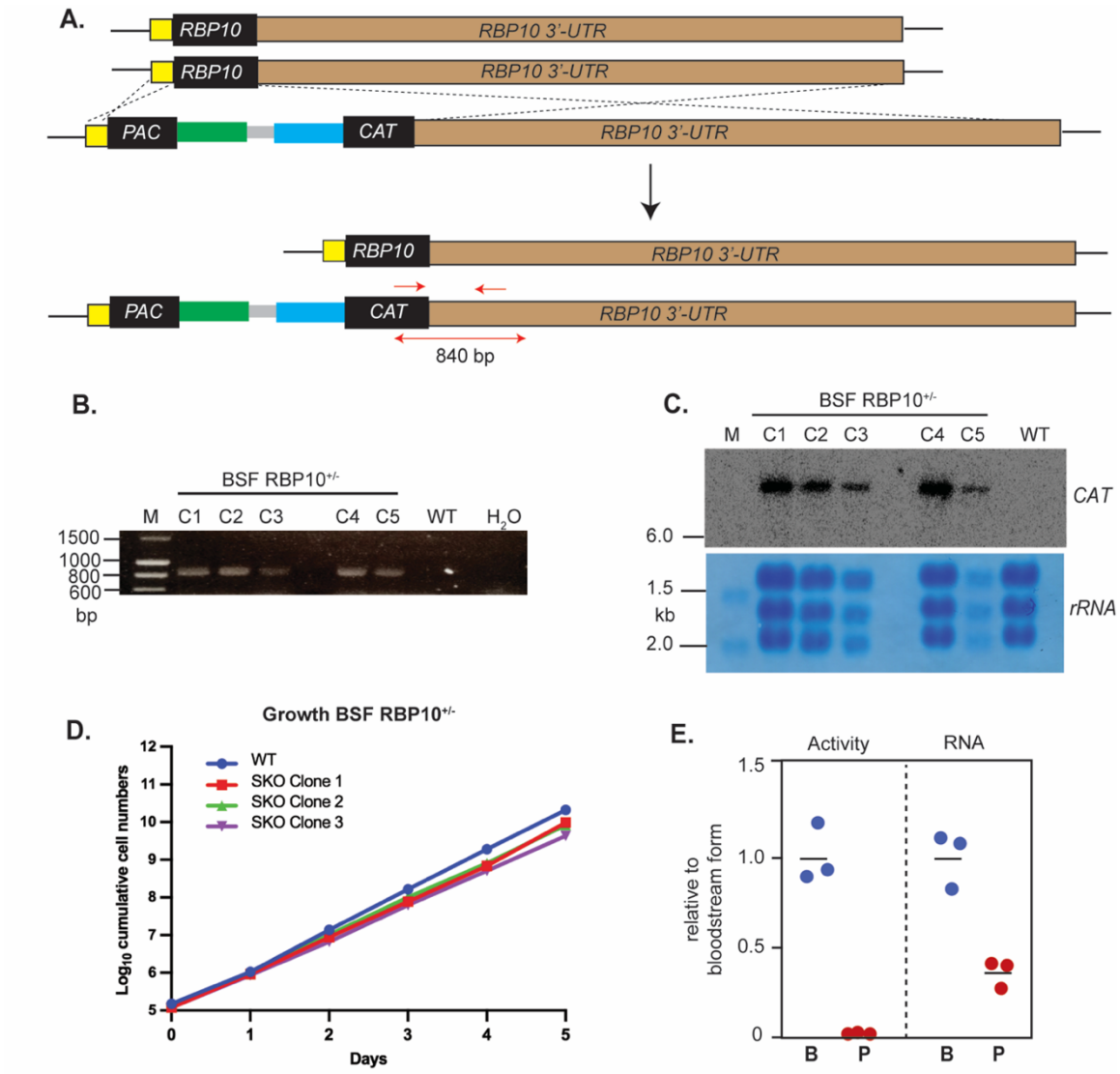


Figure 3.4: The *RBP10* 3'-UTR is sufficient for developmental regulation. (Legend continues on the next page) (A) Schematic representation of the single knock-out cell lines used. The crossover recombination results in replacement of one of the two RBP10 open reading frames by a dicistronic construct encoding the puromycin resistance gene (PAC) and the chloramphenicol acetyltransferase (CAT). For simplicity, the full linearized construct is not depicted. The red arrows indicate approximate binding sites of the used primers to confirm

proper integration of the construct. The diagram is on linear scale. (B) PCR amplification of the CAT coding sequence containing the *RBP10* 3'-UTR from genomic DNA of five different bloodstream forms (BSF) transfectants of RBP10$^{+/-}$. A wild-type cell line is used as control and water (H_2O) serves as a negative control for the PCR. No band was detected in the wild-type or in the water samples. M is the marker. (C) Northern blots of the five different transfectants RBP10$^{+/-}$ SKO and the wild type (EATRO 1125) showing the CAT mRNA migrating slower than the 6kb marker. As expected, there was no *CAT* mRNA detected in the wild-type control. A section of methylene blue staining is depicted to show the loading. M is the marker. (D) Cumulative growth curves of three different transfectants RBP10$^{+/-}$ together with a wild-type cell line. (E) Three independent transfectants of RBP10$^{+/-}$ bloodstream forms (mentioned as "B") were differentiated to procyclics (indicated as "P"). The panel shows measurements of CAT activity and mRNA for the three independent cell lines bloodstream (B) and procyclic (P) forms, normalized to the average of bloodstream forms (B).

3.1.5 The *RBP10* 3'-UTR contains several regulatory elements

In order to analyze regulatory elements inside the *RBP10* 3'-UTR that contribute to the stability and translation of *RBP10* mRNA in bloodstream forms and to its instability and translational repression in procyclic forms, I made use of a reporter plasmid (pHD 2164) that encodes for the *CAT* under the control of the *actin* 3'-UTR (Figure 3.5 A). This plasmid integrates into the tubulin locus, which results in read-through transcription by the RNA polymerase II. The stable transfectants are selected with G418 using the *NPT* (neomycin phosphotransferase) marker. The *CAT* reporter mRNA contains a 5'-UTR and a splice signal from an EP procyclin gene. The intergenic region between *CAT* and *NPT* is from between two actin (*ACT*) genes with a restriction site at the mapped polyadenylation site (Figure 3.5 A). The polyadenylation of the *CAT* mRNA bearing the *ACT* 3'-UTR is driven by the downstream splice site for *NPT* mRNA. The *RBP10* 3'-UTR fragments were cloned in place of the *ACT* 3'-UTR using *SalI* and *XhoI* restriction sites (shown in the map of pHD 2164, section 2.9). The sequences of the different *RBP10* 3'-UTR fragments are shown in Supplementary information/ section 3. Prior to transfection, the constructs were digested with *NotI* and the linearized plasmids were transfected into EATRO 1125 bloodstream forms. Two or three independent clones were then differentiated into procyclic forms. CAT activities were measured enzymatically in a CAT assay experiment and mRNA levels were measured by Northern blotting, which simultaneously enabled the measurements of the mRNA sizes. All values were normalized to the arithmetic mean results from the *ACT* 3'-UTR control. The sizes of the mRNAs are tabulated in Supplementary information/ section 3.

Due to the huge length of *RBP10* 3'-UTR, it was impossible to clone it in pHD 2164. Instead, the 3'-UTR was first segmented into 4 different fragments (1-4, in Figure 3.5 B), each of 2 kB long. The last fragment (F4) was extended beyond the RBP10 polyadenylation site, including the intergenic region before the downstream gene (Tb927.8.2790). The *CAT* mRNAs resulting from fragment 1-3 migrated slightly faster, giving sizes slightly shorter than expected (Supplementary information/ section 3) but the sizes of the mRNAs resulting from fragment F4 suggest that the genomic processing signals were used instead of the ones

provided in pHD 2164 (Supplementary information/ section 3). Surprisingly, all the four fragments of *RBP10* 3'-UTR decreased the *CAT* mRNA levels and abolished the CAT activity in procyclic forms, revealing a role in degradation and translation repression. In addition, fragments 2 and 3 increased the CAT activity by 2.5 and 1.5-fold, respectively in bloodstream forms, which suggest a role not only in mRNA stability but also in translation (Figure 3.5 B). This result implied that the regulatory sequences responsible for developmental expression of RBP10 are scattered throughout the *RBP10* 3'-UTR.

To further search for the sequences responsible for regulation, I dissected fragment 1 in six sub-fragments (F1.1, F1.2, F1.3. F1.4, F1.4.1 and F1.4.2). The sizes of these sub-fragments are respectively 618-nt, 664-nt, 533-nt, 234-nt, 113-nt and 126-nt long. Fragment 2 was divided in six sub-fragments: F2.1, F2.2, F2.2.1, F2.2.2, F2.2.3, F2.2.4 and F2.2.5. The sizes of the fragments are as follows: 551-nt, 1490-nt, 696-nt, 793-nt, 500-nt, 292-nt and 135-nt long. At last, I segmented fragment 3 in four smaller fragments; F3.1, F3.2, F3.2.1 and F3.2.2 of 507, 1479-nt, 694-nt and 791-nt long respectively (Figure 3.5 B). The *CAT* mRNAs derived from fragments F1.1, F1.3, F1.4.1, F1.4.2, F2.2.2, F2.2.5, F3.1 and F3.2.1 migrated slightly faster than expected (Supplementary information/ section 3), suggesting a polyadenylation upstream of the expected sizes. The other mRNAs migrated as expected, which suggest polyadenylation at the expected sizes. mRNAs derived from fragments 2.2.1 and 2.2.3 gave two variants, longer ones which migrated slightly as expected and shorter ones which migrated faster.

On one hand, fragments 1.1, 1.2, 1.4 and 2.2.2 (Figure 3.5 B) were each able to suppress the CAT reporter levels in procyclic forms and at the same time enhance the reporter levels in bloodstream forms. The mRNA levels were also regulated by all these fragments, except for fragment 1.2 which did not increase the mRNA levels in bloodstream forms. Hence, fragment 1.2 seems to be involved in translational regulation of the *RBP10* mRNA. Fragment 1.4 (Figure 3.5 B) is the shortest sequence that decrease the CAT expression in procyclic forms. On the other hand, fragments 2.1 and 2.2.4 each suppressed the CAT activity levels in procyclic forms with no increase in bloodstream forms. The *CAT* mRNA levels were also close to the control *ACT* 3'-UTR levels. The fragments 3.2 and 3.2.2 gave low CAT expression and mRNA levels in both stages. Fragment 1.3 gave low CAT expression levels but higher mRNA levels in both stages, revealing a strong regulation at the translation level.

It is important to notice that sometimes dissection of a sequence created an activity that was absent previously. For example, fragment 2.2 gave poor regulation but the sub-fragment 2.2.2 regulated the reporter levels in a similar way as fragment 2, fragment 2.2.3 gave the highest CAT activity in bloodstream forms close to fragment 3.1 (Figure 3.5 B) and fragment 2.2.4 gave suppression of the CAT activity in procyclics like fragment 2. Probably, the different fragments are bound by proteins that compete for binding on the *RBP10* 3'-UTR.

Therefore, a dissection might allow some proteins to bind efficiently while others are losing specificity in their binding. It is also noticeable that a regulation that was seen before a segmentation was lost after subsequent fragmentation. For example, fragment 1.4 regulated both the CAT activity and the mRNA levels but the developmental regulation was lost in the sub-fragments 1.4.1 and 1.4.2. This suggests that subsequent dissections might have interfered with secondary structures of the mRNA, disrupting the motifs that were previously bound by regulatory proteins.

Overall, these results show that the developmental regulation of RBP10 expression is achieved by numerous sequence elements. Some of these regulatory sequences have competing activities and several proteins might be involved in this regulation.

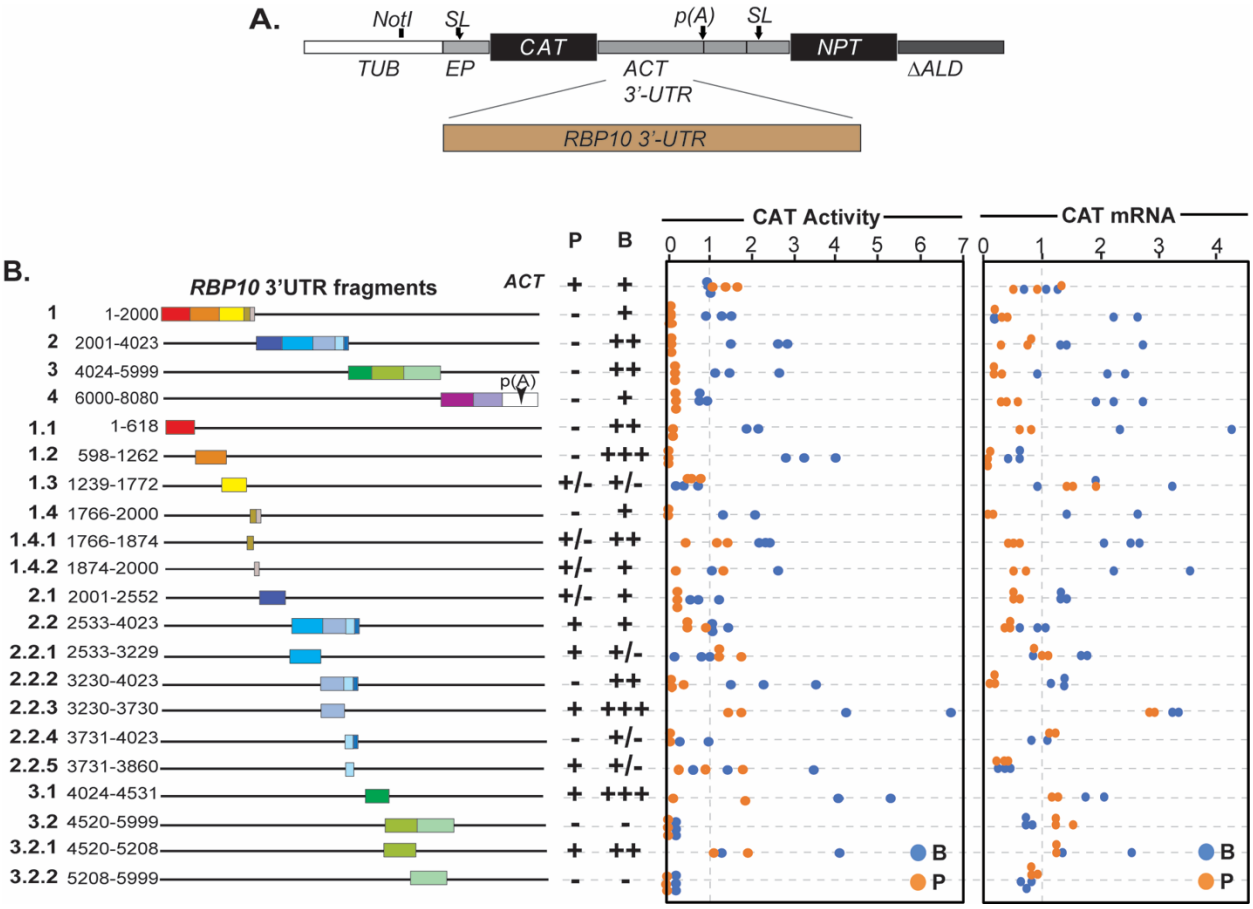


Figure 3.5: The *RBP10* 3'-UTR contains several regulatory sequences. (Legend continues on the next page) **(A)** A cartoon showing the CAT reporter construct (pHD 2164) used in this study. The different fragments of RBP10 3'-UTR were cloned downstream of the CAT reporter coding sequence. The polyadenylation site "p(A)" is specified by the NPT signal and is roughly on the boundary of the inserted fragment and the intergenic region. SL indicates the splice leader addition sites. ACT denotes actin. TUB indicates tubulin locus, EP is the procyclin surface protein and ΔALD is a truncated version of aldolase. **(B)** On the left, schematic diagrams of the *RBP10* 3'-UTR segments used to map the regulatory sequences. The fragments numbers are from 1 to 3.2.2. The length

of the different fragments is shown as a range of the nucleotide positions in the 3'-UTR (for example, F1 is from 1 to 2000 nt). Measurements of the CAT activity and the mRNA levels from two or three permanent cell lines are shown on the right. Each dot is a result for an independent clone, with blue and orange representing bloodstream (B) and procyclic (P) forms, respectively. The values were normalized to those from the ACT 3'-UTR control (pHD 2164). The average result for the control bloodstream (B) is set to 1. For easy indication of expression, "+" = 0.5-2X, "++" = 2-3x and "+++" = >3x. Formatting of some plasmid maps was done by Prof. C. Clayton.

3.1.6 Identification of motifs in the *RBP10* 3'-UTR

Knowing that *RBP10* 3'-UTR contains several regulatory sequences, I then investigated whether it contains repeated motifs. For that, *RBP10* regulatory sequences (fragments 1, 2, 3, 4, 1.1, 1.2, 1.4, 2.2, 2.2.2) were compared with those that lacked regulation, and motifs were searched using the motif search tool, MEME (Bailey et al. 2015) with default settings. No enriched motifs were found in these sequences. I then check whether some motifs were enriched in both the regulatory fragments and other co-regulated bloodstream-form specific mRNAs (Supplementary section 4). For this analysis, a set of non-regulated fragments and procyclic-specific mRNAs were used as controls. Again, no enriched motifs were found. This suggests that *RBP10* 3'-UTR contains several different sequences that are implicated in the developmental regulation, involving perhaps several RNA-binding proteins. Each protein might have a totally different binding site on the *RBP10* 3'-UTR.

3.1.7 $(AU)_{10}$ element affects reporter translation in bloodstream forms

Since there were no shared motifs between *RBP10* mRNA and the co-regulated genes, I examined two repetitive motifs that are known to be involved in regulation in mammalian cells, the AU repeats (Beisang and Bohjanen 2009) and the A-rich sequences (Vazquez-Pianzola et al. 2011). Deletion of the $(AU)_{10}$ element from the *RBP10* 3'-UTR fragment 1.2 (F1.2) resulted in a drastic decrease of the reporter expression levels (Figure 3.6 A). Several fragments of the *RBP10* 3'-UTR with good translation in bloodstream forms contain A-rich sequences, and it was thought that they might act by recruiting a poly(A) binding protein. However, deletion of the A-rich sequences from one of these fragments (Fragment 3.1.1) unexpectedly resulted in an increase in reporter expression levels, rather than a decrease (Figure 3.6 B). Therefore, the only motif found to be involved in enhancing the reporter expression levels is the $(AU)_{10}$ repeat. I do not know whether this effect is stage-specific or not. The A-rich sequences seem to be involved in translation repression in bloodstream forms. There is no evidence from the literature of any mechanisms that could explain this result.

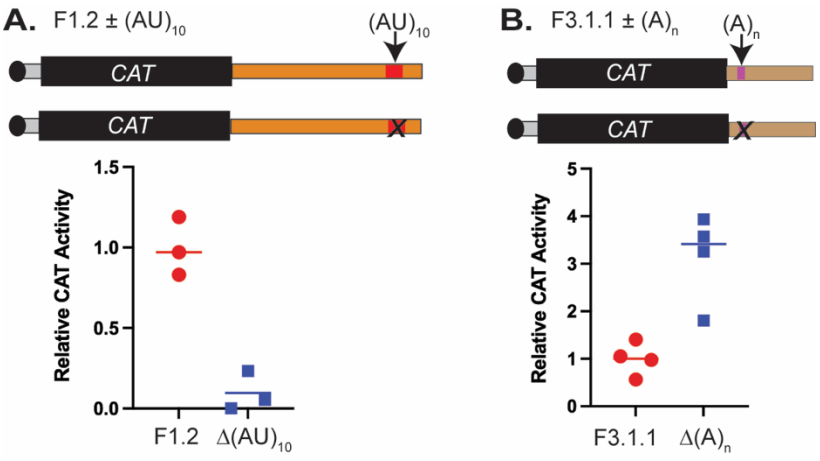


Figure 3.6: Analysis of specific motifs. **(A)** Schematic representation of the reporter plasmid containing *RBP10* 3'-UTR fragment 1.2 bearing $(AU)_{10}$ repeats and a mutation version without them. CAT activity was measured in three independent clones and the average result for the fragment with the repeats is set to 1. **(B)** Same as (A) but here showing deletion of A-rich sequences from fragment 3.1.1. CAT activity was measured in four independent clones.

3.2 Validation of RNA-Protein interaction detection in *T. brucei*

The second objective of my thesis was to identify the proteins that bind to *RBP10* mRNA and regulate its expression in bloodstream and procyclic forms. For that, I wanted to make use of the RNA-protein interaction detection (RaPID) method, which was described by Ramanathan et al. 2018. It relies on proximity-dependent labeling, based on the biotin ligase to capture all the proteins that bind to a specific RNA sequence in living cells. As a proof of concept, I attempted to validate this method in trypanosomes with a known RNA-protein interaction, the ZC3H20 which stabilizes a procyclic-specific membrane protein, PSSA-2 (Liu et al. 2019). To achieve that, the *Escherichia coli* biotin ligase (BirA*) fused with the lambda N peptide and a myc tag at the N-terminus was inducibly expressed in the procyclic cell lines constitutively expressing the CAT reporter containing the *PSSA-2* 3'-UTR upstream of five copies of the Box B sequences (Figure 3.7 A). Cell lines with the CAT reporter containing the *actin* 3'-UTR were used as controls (Figure 3.7 B). The lambda N (λN) peptide is bound with high affinity to the Box B sequences. Tethering of the biotin ligase to the *CAT* mRNA bearing the *PSSA-2* 3'-UTR is thought to rapidly biotinylates the proximal proteins binding to the *PSSA-2* 3'-UTR. Both cell lines express an *in situ* N-TAP tagged ZC3H20. The biotin ligase was successfully expressed in these cell lines as checked by Western blotting (Figure 3.7 C and D) and the overexpression did not affect cell growth (Figure 3.7 E and F). Biotinylation of proximal proteins binding to either the *PSSA-2* 3'-UTR or the *actin* 3'-UTR was induced by incubating the procyclic cells in media containing tetracycline (100 ng/ml) and freshly made D-biotin solution (50 µM) for 24 hours. Similar cultures were grown without tetracycline and D-biotin. The biotinylated proteins were purified using the streptavidin-coated dynabeads and they were eluted from the beads by incubation with 50 µL of laemmli SDS-sample buffer containing biotin at 98°C for 10 minutes.

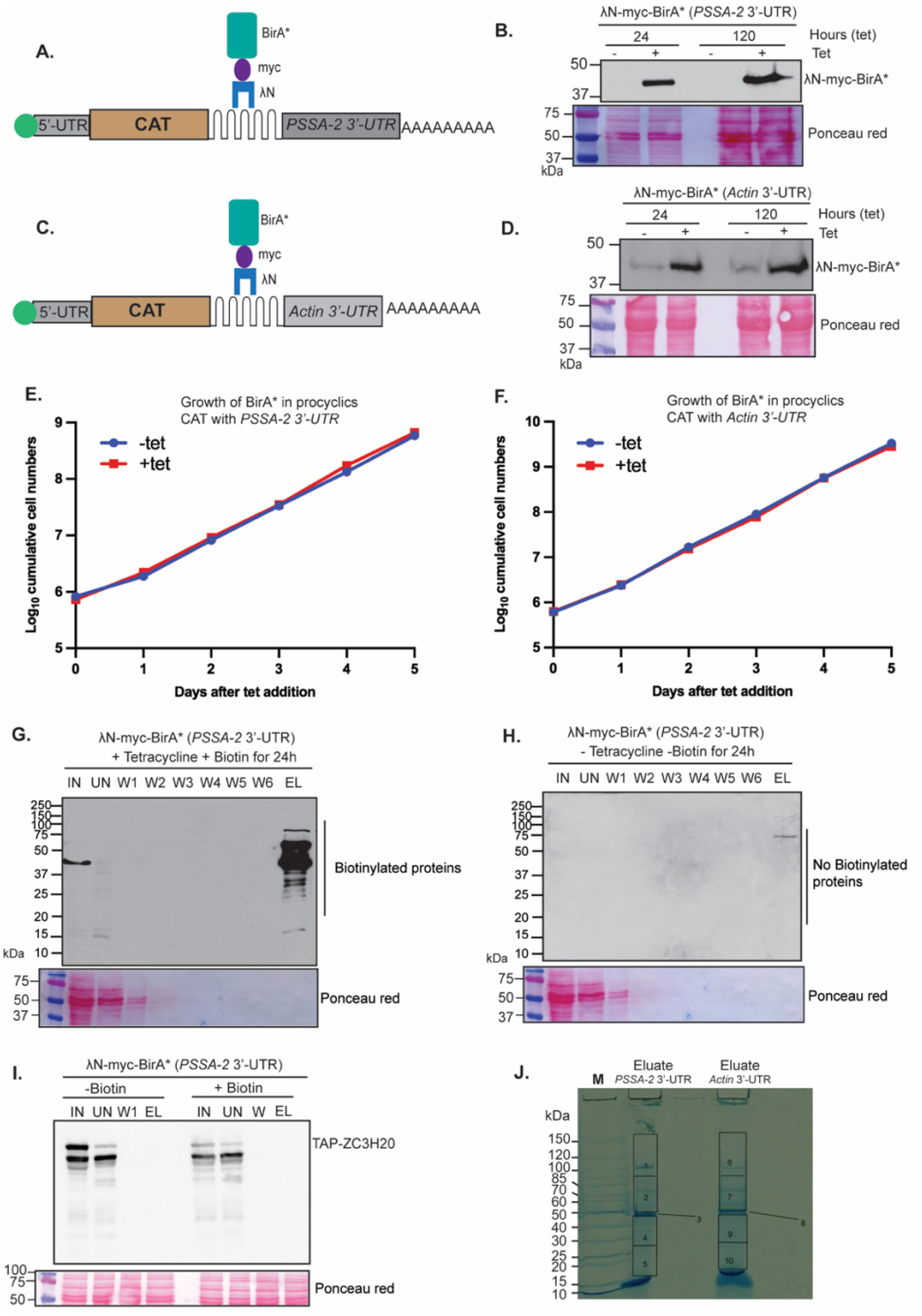


Figure 3.7: RNA-protein interaction detection in trypanosomes. (Legend continues on the next page) **(A)** Schematic representation of the cells used for this experiment. The biotin ligase (BirA*) is fused on the N-terminus with the lambda N (λN) peptide and the myc tag. The lambda N (λN) is bound with high affinity to 5 copies of the Box B sequences downstream of the CAT reporter coding sequence and upstream of the procyclic specific membrane *PSSA-2* 3'-UTR. **(B)** Expression of λN-myc-BirA* was induced for 24 and 120 hours with

tetracycline (Tet) in the cell lines constitutively expressing the CAT reporter bearing the *PSSA-2* 3'-UTR. The protein was detected with anti-myc antibody and the red ponceau staining is used as control. **(C)** The control cell lines contain the biotin ligase (BirA*) fused on the N-terminus with the lambda N (λN) peptide and the myc tag, tethered to the CAT reporter containing BoxB sequences and the *actin* 3'-UTR. **(D)** Expression of λN-myc-BirA* was induced for 24 and 120 hours with tetracycline (tet) in cell lines constitutively expressing the CAT reporter bearing the *actin* 3'-UTR. The protein was detected with anti-myc antibody and the red ponceau staining is used as loading control. **(E)** Cumulative growth curves of trypanosomes expressing the CAT reporter bearing the *PSSA-2* 3'-UTR with and without expression of the biotin ligase for 5 days. **(F)** Cumulative growth curves of trypanosomes expressing the CAT reporter bearing the *actin* 3'-UTR with and without expression of the biotin ligase for 5 days. **(G)** Cells were incubated with tetracycline and biotin for 24 hours to induce biotinylation of proximal proteins. The biotinylated proteins were purified and for each step of the purification, 20 µl sample was taken and mixed with 2x Laemmli buffer for analysis by Western blotting. IN= input, UN= unbound, W1-W6= wash fractions 1 to 6, EL= eluate. Streptavidin-horseradish peroxidase antibody was used for detection of biotinylated proteins. **(H)** Same as G but with no tetracycline and no biotin added to the cell cultures. **(I)** 2% of the input (IN), unbound (UN) and wash fraction 1 (W1) as well as all of the eluate (EL) were analyzed by western blotting for detection of the N-TAP tagged ZC3H20. The protein was detected with peroxidase-labelled anti-peroxidase (PAP) antibody. The red ponceau staining is used as loading control. **(J)** Coomassie stain gel of the eluates obtained from purifications of cells expressing the biotin ligase with CAT reporter containing either the *PSSA-2* 3'-UTR or the *actin* 3'-UTR. Each sample was sliced in five parts and later analyzed by quantitative mass spectrometry. M is the marker.

Western blot of aliquots from each step of the purification showed that several proteins were biotinylated after incubation of cells in presence of biotin and tetracycline for 24 hours (Figure 3.7 G). No biotinylation was detectable in the absence of biotin and tetracycline (Figure 3.7 H). Samples from the input, the unbound, the first wash fraction and the eluate were again analyzed by western blotting to detect the N-TAP tagged ZC3H20. Oddly, ZC3H20 was only detected in the input and the unbound fractions but not in the eluate (Figure 3.7 I). This suggests that ZC3H20 was not biotinylated as a candidate protein in the proximity of the *PSSA-2* 3'-UTR. Nevertheless, the eluate fractions containing the biotinylated proteins in proximity of the *PSSA-2* 3'-UTR and the *actin* 3'-UTR were loaded on SDS-PAGE gel and stained with Coomassie (Figure 3.7 J). Samples in each eluate replicate were divided into five slices (numbered boxes in Figure 3.7 J) and analyzed by quantitative liquid chromatography-tandem mass spectrometry (LC/MS). The experiment was only done once. In total, 168 proteins were identified in both purifications. No specific enrichment was found with the *PSSA-2* 3'-UTR. The proteins that were associated with the *PSSA-2* 3'-UTR were also related to the *actin* 3'-UTR. In addition, these are abundant proteins that are generally found in any purifications from trypanosomes extracts. Selected proteins are listed in Table 3.1. A procyclic cell contains approximately $4x10^8$ protein molecules and if the cells used in my experiment contain for example 400 copies of the CAT reporter because it was expressed from an rRNA locus, then the initial molar ratio of a specific ribonucleoprotein is 1 in a million. And since there is only 2.77 *PSSA-2* mRNAs per procyclic cell, the probability to find specific regulators through this method is halved.

Table 3.1: Selected proteins associated with the *PSSA-2* 3'-UTR and the *actin* 3'-UTR

The protein ID, the annotation, the number of unique peptides identified with *PSSA-2* 3'-UTR and *actin* 3'-UTR purifications, and the molecular weight (MW) of the protein are highlighted.

Protein ID	**Proteins associated with translation**	**Peptides in *PSSA2* 3'-UTR**	**Peptides in *actin* 3'-UTR**	**MW**
Tb10.70.2650	Elongation factor 2	22	35	94
Tb09.160.3270	Probable eukaryotic initiation factor 4A	5	6	45
Tb10.26.0370	40S ribosomal protein S3	4	5	24
Tb11.01.2680	40S ribosomal protein SA	4	4	31
Tb11.01.4660	Elongation factor 1 gamma	7	7	39
Tb10.05.0220	Ribosomal protein	4	2	25
Tb11.46.0001	60S acidic ribosomal protein	3	7	35
Tb10.70.1730	40S ribosomal protein S18	2	3	18
Tb927.7.1040	40S ribosomal protein S16	3	2	17
Tb10.70.3360	40S ribosomal protein S3a	3	4	29
Tb10.70.3510	60S ribosomal protein L18a	1	3	21
Tb10.70.3290	ATP-dependent DEAD-box RNA helicase	1	4	46
Protein ID	**Proteins associated with RNA metabolism**	**Peptides in *PSSA-2* 3'-UTR**	**Peptides in *actin* 3'-UTR**	**MW**
Tb927.11.510	UBP2	1	2	20
Tb927.6.4440	RNA-binding protein 42 (RBP42)	0	2	38
Tb927.9.8740	DNA-RNA binding protein 3 (DRBD3)	0	2	37
Tb927.4.2040	ALBA-3	0	1	21
Tb927.9.10770	PABP2	0	3	62
Protein ID	**Others proteins**	**Peptides in *PSSA-2* 3'-UTR**	**Peptides in *actin* 3'-UTR**	**MW**
Tb927.8.7100	Acetyl-CoA carboxylase	23	64	243
Tb927.5.930	NADH-dependent fumarate reductase	29	28	124
Tb11.01.3110	Heat shock protein	41	34	75
Tb11.01.3550	Dihydrolipoamide acetyltransferase	11	23	41
Tb927.6.4280	Glyceraldehyde-3-phosphate dehydrogenase	18	17	44
Tb10.70.4740	Enolase	13	24	47
Tb927.6.3740	Heat shock protein 70 kDa	17	25	71
Tb10.26.1080	Heat shock protein 83 kDa	13	18	83
Tb09.211.3560	Glycerol kinase, glycosomal	9	15	56
Tb09.160.4310	NAD-specific glutamate dehydrogenase	6	11	112
Tb927.2.4210	Glycosomal phosphoenolpyruvate carboxykinase	6	7	59
Tb927.6.2790	L-threonine 3-dehydrogenase	6	6	37
Tb927.5.940	NADH-dependent fumarate reductase	19	18	95

3.3 Identification of proteins that regulate RBP10 expression

I was not able to validate the RaPID method using a known interaction. Therefore, for the identification of proteins that bind to *RBP10* 3'-UTR, I decided to deplete RNA-binding proteins that might affect RBP10 expression using RNA interference (RNAi). To look for candidate proteins, two types of profile were used. For stabilizers of *RBP10* mRNA, these would be proteins that are expressed and probably essential in bloodstream forms, can bind mRNA and can enhance expression when tethered to a reporter mRNA. A repressor of *RBP10* mRNA would be a protein that is expressed and perhaps essential in procyclic forms, can bind mRNA and repress expression when tethered to a reporter mRNA. I also checked existing literature on such proteins and looked for proteins that bind and decrease *RBP10* mRNA levels in bloodstream forms or increase it in procyclic forms after RNAi. Six different proteins satisfy at least one of the criteria cited above for regulation of *RBP10* mRNA in bloodstream forms. They are described in Table 3.2. In procyclics, the proteins ZC3H13, PUF1, ZC3H5, DRBD11, DRBD7, ZC3H39, RBP35, ZC3H8, RBP34, PUF3, and ZC3H22 were considered as potential regulators of the *RBP10* mRNA. Studies of ZC3H39 and ZC3H40 by Trenaman et al. 2019 did not reveal a role in control of *RBP10* mRNA. Erben et al. 2021 also did not find any implication of ZC3H22 in repression of *RBP10* mRNA in procyclics. Bajak et al. 2020 did not show any evidence that ZC3H5 could repress *RBP10* mRNA in procyclics. Due to time limit, I only focus on RNA-binding proteins that might regulate RBP10 expression in bloodstream forms. Depletion of ZC3H40 and ZC3H45 by RNAi in monomorphic Lister 427 bloodstream forms did not affect *RBP10* mRNA levels (data not shown). The candidate ZC3H44 was not investigated due to challenges in getting the stem-loop construct for RNAi.

Table 3.2: List of potential regulators of *RBP10* mRNA in bloodstream forms

tritrypdb.org/	tritrypdb.org/	tritrypdb.org/	Alsford et al. 2011	Erben et al. 2014	Lueong et al. 2016	
Protein ID	**Annotation**	**Category**	**RIT-Seq**	**Shotgun library tethering effect**	**ORFeome library tethering effect**	**mRNA binder?**
Tb927.2.3880	HNRNPFH	RNA binding	B3	No effect	Not in library	Yes
Tb927.11.14090	DRBD18	RNA binding	B3B6DP	increase	Not in library	Yes
Tb927.9.9450	ZC3H28	RNA binding	B3B6DP	increase	No effect	Yes
Tb927.10.14950	ZC3H40	RNA binding	B6D	increase	No effect	Yes
Tb927.11.8470	ZC3H45	RNA binding	D	No effect	3.3x	Yes
Tb927.11.7890	ZC3H44	RNA binding	B3	No effect	5.8x	Yes

Legend: RIT-Seq shows the effect of depletion of proteins in an RNAi screen. Large letters indicate a growth disadvantage. B3= bloodstream forms after three days; B6= bloodstream forms after six days; D= differentiation; P= survival as early procyclic forms.

In this part of the thesis, I will present results that were obtained with the candidate proteins HNRNPFH, ZC3H28 and DRBD18.

3.3.1 The role of HNRNPFH in regulating RBP10 expression

i. Depletion of HNRNPFH does not affect RBP10 expression levels

HNRNPFH (Tb927.2.3880) is 55 kDa RNA binding protein which contains 3 quasi RNA-recognition motifs (qRRM), a single zinc binding domain in the middle and a barrier autointegration factor domain (Figure 3.8 A). There is high evidence by that HNRNPFH is the homologue of the human HNRNPF.

To find out whether HNRNPFH regulates RBP10 expression levels in bloodstream forms, RNAi-mediated depletion of HNRNPFH was carried out using a stem-loop construct targeting HNRNPFH coding sequence. The RNAi was done in stable pleomorphic cell lines in which one allele of HNRNPFH was tagged *in situ* with an N-terminal tandem affinity purification tag and the second allele is replaced with a resistance marker encoding blasticidin S-deaminase (BSD). A schematic diagram of the cell lines used for this experiment is shown in Figure 3.8 B. Depletion of the TAP-tagged HNRNPFH protein was monitored for 6 days in four independent clones and analyzed by Western blotting (Figure 3.8 C-F). The protein level decreased by 50-70% in all the four clones after one day of tetracycline induction and the protein was not detectable in the following days, revealing successful knock-down. However, in day 4, there was about 20-30% increase in HNRNPFH protein levels after tetracycline induction of the RNAi. This implies that as days go, the regulation of the RNAi was being lost. Surprisingly, depletion of HNRNPFH led to an increase of about 1.8-fold in RBP10 protein levels (Figure 3.8 G) after one day of tetracycline induction. At later days (day 2-6), RBP10 protein levels was rather unchanged. In addition, although HNRNPFH was reported to be essential in bloodstream forms, the depletion of TAP-HNRNPFH did not affect cell growth (Figure 3.8 H): the cells without induction of the RNAi grew relatively similar to those with induction. However, only in the second clone (Figure 3.8 H: orange and green), after induction of the HNRNPFH RNAi, the cells grew slower than the control. Taken together, it appears that the small amount of HNRNPFH protein left after RNAi induction is enough to allow normal growth of bloodstream forms and normal expression of the RBP10 protein. It is also possible that the cells have adapted to the loss of HNRNPFH during culturing. In regard of the above results from the stem-loop RNAi construct, an RNAi construct using opposing T7 promoters and targeting a bigger region of the HNRNPFH coding sequence was generated (pHD 3244). However, even with this construct, the growth of bloodstream form trypanosomes was not affected although HNRNPFH protein was completely depleted (data not shown).

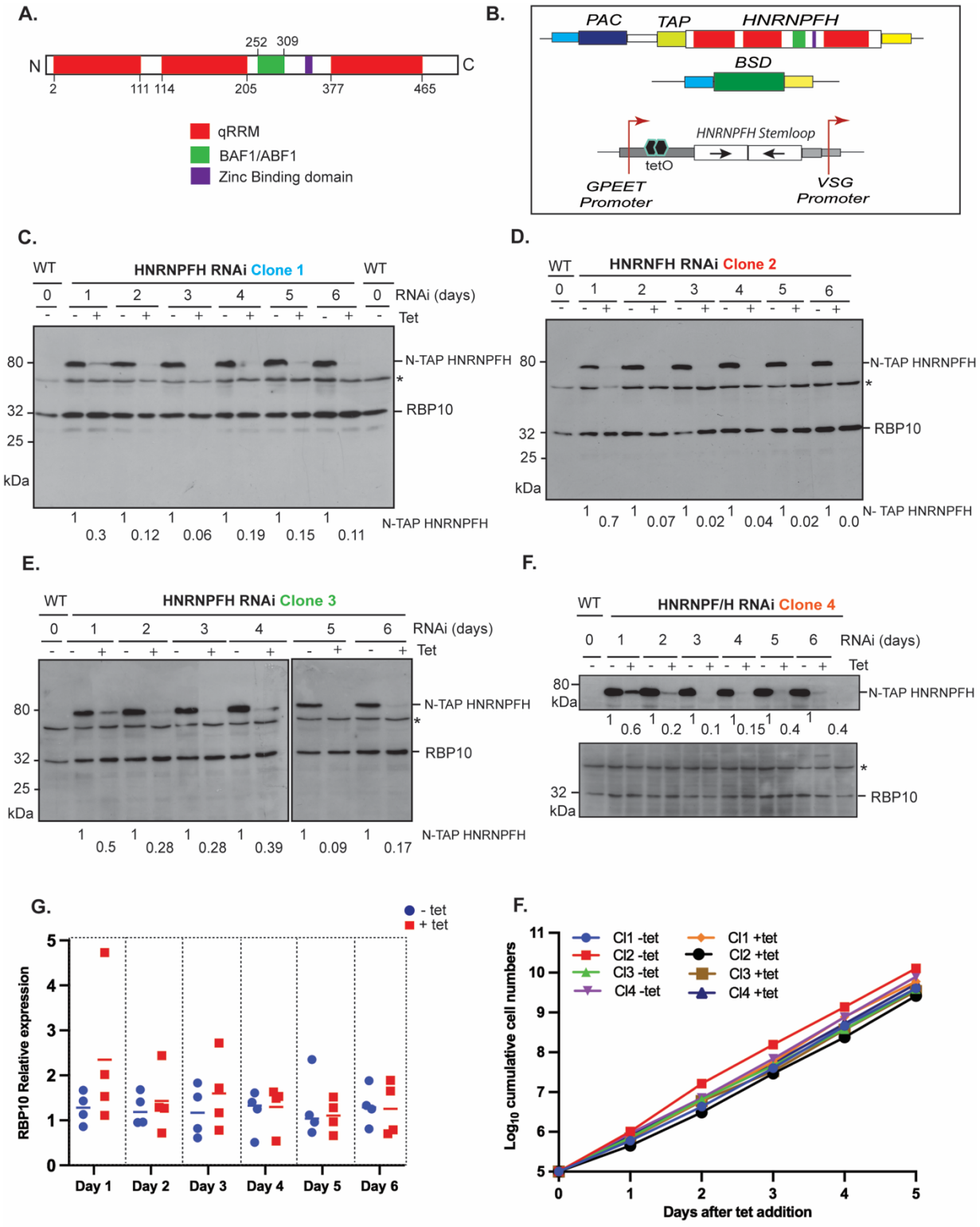


Figure 3.8: Knockdown of HNRNPFH does not affect RBP10 expression levels and cell growth. (Legend continues on the next page) **(A)** Schematic representation of the HNRNPFH protein sequence including the domains. Red indicates the quasi RNA recognition motif domain, green represents the barrier autointegration factor domain and purple represents the zinc binding domain. **(B)** Schematic representation of the cell used for

this experiment: one allele of HNRNPFH encodes an in-situ N-TAP tagged HNRNPFH and the other allele encodes blasticidin S-deaminase (BSD). The stem loop construct is expressed from an rRNA locus. **(C-F)** Western blots showing TAP-HNRNPFH and RBP10 protein levels before and after RNAi induction over six days. A wild-type cell line EATRO 1125 bloodstream form is used as a control. Peroxidase-labelled anti-peroxidase (PAP) antibody was used for detection of the TAP-tagged HNRNPFH. The polyclonal antibody anti-RBP10 was used for detection of RBP10 and a non-specific band (*), which is used as the loading control. **(G)** RBP10 expression levels were measured from the western blots (C-F) in four independent clones before and after induction of HNRNFH RNAi. Values are normalized to those obtained before induction of the HNRNPFH RNAi (- tet). **(H)** Cumulative growth curves of bloodstream forms with and without HNRNPFH RNAi in four independent clones.

I then decided to check whether depletion of HNRNPFH by RNAi affects *RBP10* mRNA levels. For that, RNA from cells with and without induction of HNRNPFH RNAi were analyzed by Northern blotting. The blots were hybridized to RBP10 probe from the coding sequence (Figure 3.9 C). Successful knock-down of HNRNPFH was confirmed by Western blotting, which indicated the loss of the protein after 24 to 72 hours of RNAi induction (Figure 3.9 A). Northern blots revealed that *RBP10* mRNA levels were not drastically affected by depletion of HNRNPFH in bloodstream forms (Figure 3.9 C). This result, however does not exclude a potential role of HNRNPFH in regulating RBP10. Since *RBP10* 3'-UTR contains several regulatory sequences (Figure 3.5 B) which might involve several RNA-binding proteins; depleting one protein might not necessarily affect the expression levels of *RBP10* mRNA and protein.

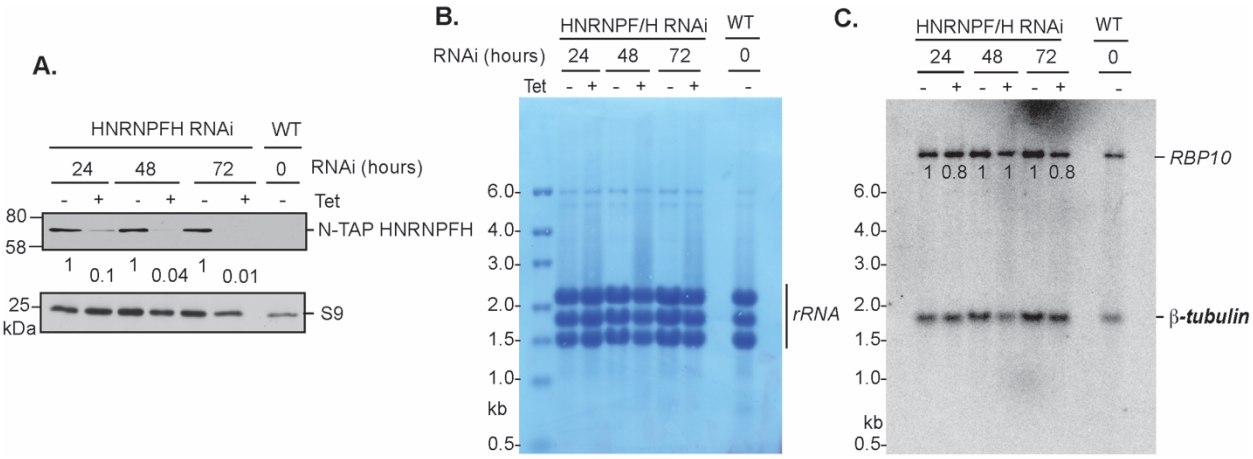


Figure 3.9: Knockdown of HNRNPFH does not affect *RBP10* mRNA levels. (A) Western blot showing TAP-HNRNPFH protein levels with and without induction of HNRNPFH RNAi for 24, 48 and 72 hours. Peroxidase-labelled anti-peroxidase (PAP) antibody was used for detection of TAP-HNRNPFH; and ribosomal S9 protein is the loading control. WT is a wild-type cell line with no stem-loop construct for HNRNPFH RNAi. **(B)** RNA from the samples used in (A) were used to examine the effects of HNRNPFH RNAi on *RBP10* mRNA. The methylene blue staining is shown to reveal proper transfer and loading. **(C)** Northern blot was hybridized with RBP10 and β-tubulin probes.

ii. HNRNPFH associates with mRNAs encoding various proteins

The depletion of HNRNPFH does not affect drastically *RBP10* mRNA and protein levels. But does HNRNPFH directly bind *RBP10* mRNA? Does it also bind other mRNAs in trypanosome bloodstream forms? To identify mRNAs bound by HNRNPFH, RNA immunoprecipitation was performed from cells expressing functional *in situ* N-TAP tagged HNRNPFH. The protein was purified using magnetic IgG beads and it was released from the beads with the tobacco etch virus (TEV) protease. Western blot of aliquots from each step of the purification showed that about 98% of the protein was bound to the beads and after cleavage, the protein was released leaving only the protein A attached to beads (Figure 3.10 A). The bound RNAs and the RNAs from unbound fraction were purified and analyzed by RNA sequencing. The experiment was done in triplicates under the same conditions.

The principal component analysis showed reproducibility between the replicates from the bound (eluate) and the flow-through (unbound) fractions (Figure 3.10 B). About 423 transcripts were more than 2-fold enriched in all the three purifications compared to the unbound fractions (Figure 3.10 C). Strikingly, the majority of the bound mRNAs encoded proteins with unknown precise functions as well as proteins with various functions (noted as "others" in the functional category, Figure 3.10 C). The mRNA encoding HNRNPFH was highly enriched in the three purifications, as well as the mRNAs encoding RBP5, DHH1, ZPF2 and Complex III cytochrome b/c complex oxidase (Table 3.2). When the affinity tag of the protein being purified is at the N-terminus, it is common to see that the protein purifies its own mRNA, probably through the purification of polysomal RNA via the nascent polypeptide (Erben et al. 2021). Selected mRNAs that were more than 10-fold enriched in the bound fraction in all the three purifications are shown in Table 3.2. The complete raw data with the corresponding read counts are deposited in the Array express with the accession number E-MTAB-10708. The *RBP10* mRNA (Table 3.3, last row in red) was only 2-fold enriched in all the three purifications, which suggests that binding of HNRNPFH is probably not specific.

Next, the 3'-UTRs of the 423 bound mRNAs were analyzed using the motif-based sequence analysis tool (MEME) to identify enriched motifs. The mRNAs that were not bound by HNRNPFH (< 1-fold) with the same sized 3'-UTRs as the bound mRNAs were used as background control. A purine rich motif AAAA(G)A(G/C)A was highly enriched ($E = 6.2\ e^{-023}$) in the bound fraction. Previous results from (Gupta et al. 2013) revealed that a purine rich motif AAGAA was highly enriched in both untranslated regions flanking the 5'-splice site and the poly (A) sites of the mRNAs that were affected by HNRNPFH depletion in trypanosomes.

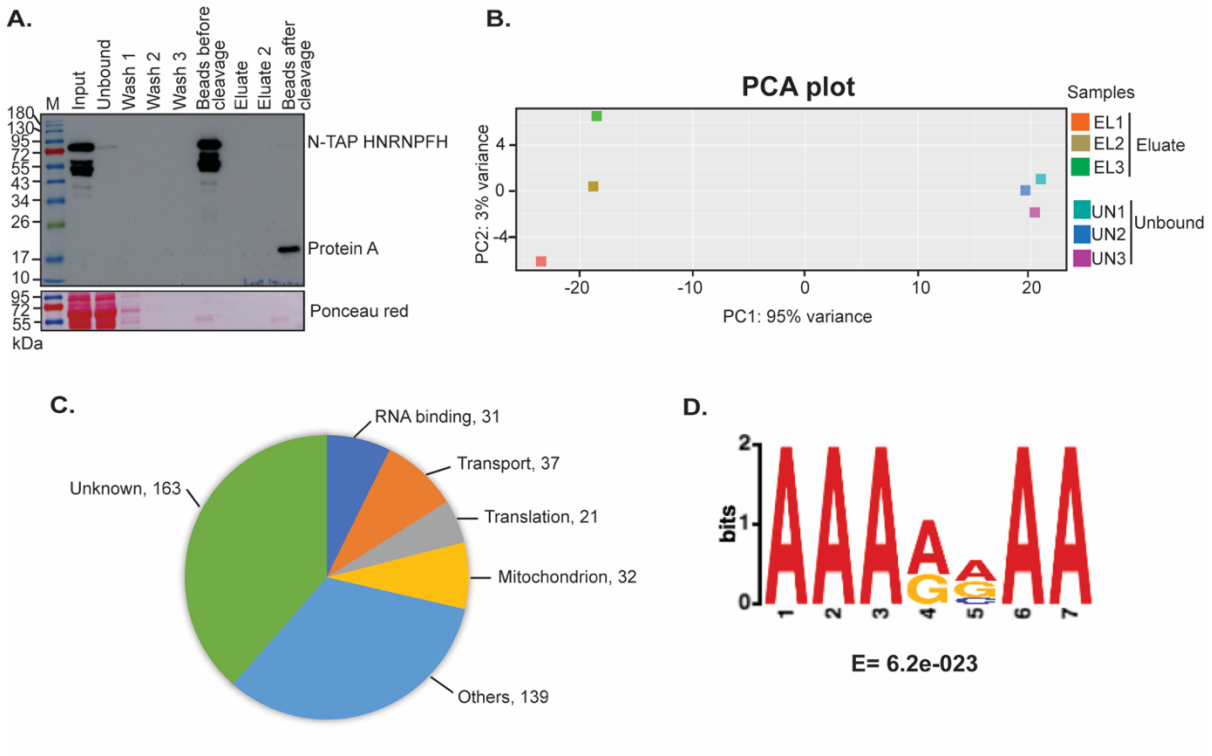


Figure 3.10: HNRNPFH binds mRNAs which encode a wide variety of proteins. Affinity purification from a bloodstream form cell line HNRNPFH$^{TAP/-}$ was carried out in order to identify the mRNAs bound by HNRNPFH. The RNAs from bound and unbound fraction were analyzed by RNA sequencing. **(A)** For each step of the TAP purification, 20 µl sample was taken and mixed with 2x Laemmli buffer for analysis by Western blotting. Peroxidase-labelled anti-peroxidase (PAP) antibody was used for the detection of TAP-HNRNPFH and the red ponceau staining was used for loading control. M is the marker. **(B)** Principal component analysis (PCA) plot of the Eluate (EL) and the Unbound (UN) RNAs in triplicates. **(C)** Functional categories enriched in HNRNPFH RIP-Seq data. **(D)** The motif enriched in the 3'-UTRs of mRNAs strongly bound by HNRNPFH was searched using the MEME tool.

Since the depletion of HNRNPFH by RNAi did not affect growth of the trypanosome bloodstream forms (Figure 3.8 H), I was not able to identify mRNAs that are affected by HNRNPFH depletion using deep RNA sequencing. However, results obtained by Gupta et al. 2013 using microarrays revealed that the growth was affected after depletion of HNRNPFH and different transcripts were up- and down-regulated. Of the 423 target mRNAs, only 32 were downregulated and 63 were upregulated. It is therefore not evident that HNRNPFH might stabilize its target mRNAs. In addition, *RBP10* mRNA was not affected.

Overall, the results obtained in this section revealed that HNRNPFH binds various mRNAs encoding proteins of numerous functions and containing a purine rich motif AAAA(G)A(G/C)A in their 3'-UTRs.

Table 3.3: Selected mRNA targets of HNRNPFH identified by RNA Sequencing

Gene ID, Annotation and category of the mRNAs are indicated. The enrichment from the three purifications is shown as a fraction of the reads per million from the eluate fraction (B) divided by that from the unbound fraction (UN).

Gene ID	Annotation	Category	Enrichment B/UN (1)	Enrichment B/UN (2)	Enrichment B/UN (3)
Tb927.2.3880	**HNRNPHF**	RNA processing	297	96	173
Tb927.11.12100	RBP5	RNA binding	129	102	73
Tb927.10.3990	DHH1	RNA degradation	42	33	36
Tb927.11.14950	ZFP2	RNA binding	46	39	31
Tb927.9.12300	Replication factor C	DNA	36	29	33
Tb927.10.4280	Complex III cytochrome b/c	Mitochondria transport	55	29	42
Tb927.3.740	ZC3H5	RNA binding	44	28	27
Tb927.9.15100	Hypothetical protein	Unknown	25	22	22
Tb927.7.6770	Acyl-CoA binding protein, putative	Unknown	28	22	18
Tb927.9.12510	ATP-dependent DEAD/H RNA helicase	Translation	18	17	27
Tb927.1.3950	D-alanine aminotransferase	Mitochondria pathway	20	17	19
Tb927.9.13990	DRBD2	RNA binding	27	15	19
Tb927.11.16770	Glucosamine-6-phosphate isomerase	Glucose and glycerol	15	15	18
Tb927.4.1870	Phospho-protein phosphatase	Protein phosphatase	24	14	26
Tb927.4.3430	Hypothetical protein	Unknown	22	14	15
Tb927.3.720	ZFP3	RNA binding	32	21	14
Tb927.6.5010	Hypothetical protein	Unknown	16	13	13
Tb927.10.6200	Hypothetical protein	Mitochondria pathway	23	14	12
Tb927.1.3850	Sec61 beta family and TM domain	Vesicular transport	15	12	16
Tb927.9.11720	Iron-sulfur cluster assembly protein	Mitochondria pathway	17	14	12
Tb927.7.4500	PX domain protein	Unknown	16	13	12
Tb927.10.9030	Mechanosensitive ion channel	Unknown	21	12	16
Tb927.4.1220	Leucine-rich repeat protein (LRRP)	LRRP	26	11	14
Tb927.5.1200	CSL4 exosome component	RNA degradation	16	11	13
Tb927.7.3520	MPC2 mitochondrial pyruvate carrier	Mitochondria membrane	17	13	11
Tb927.3.2930	RBP6	RNA binding	17	11	19
Tb927.11.530	RBP3	RNA binding	10	13	13
Tb927.3.1030	RBP28	RNA binding	12	11	10
Tb927.10.14140	Pyruvate kinase	Glucose and glycerol	13	11	10
Tb927.8.2780	**RBP10**	**RNA binding**	**2.8**	**2**	**2**

iii. HNRNPFH does not bind to RBP10 3'-UTR

Gupta et al. 2013 showed that the motif AAGAA was enriched in the mRNAs that were affected after depletion of HNRNPFH in trypanosomes. The RIP-Seq results from this study confirms a purine-rich motif, AAA(A/G)(A/G/C)AA as a putative binding site of HNRNPFH. Interestingly, the *RBP10* 3'-UTR fragment 1.1 contains approximately 6 putative binding sites of HNRNPFH (the "AAGAA" motif).

To investigate whether HNRNPFH regulates *RBP10* mRNA level through binding to the putative motif, the CAT reporter containing *RBP10* 3'-UTR fragment 1.1 was transfected in the cell lines in which both alleles of HNRNPFH are modified to encode an *in situ* N-TAP tagged HNRNPFH and a blasticidin S-deaminase. The cell lines also express from the rRNA locus a stem loop construct targeting HNRNPFH coding sequence for RNAi. Pleomorphic cell line EATRO 1125 bloodstream form with no inducible expression of HNRNPFH RNAi and without *in situ* N-TAP tagged HNRNPFH was used as a control.

Knock-down of HNRNPFH was induced every 24 hours for 3 days in two independent clones (Figure 3.11 A and B). Simultaneously the CAT activity and mRNA levels were measured respectively by CAT assay and Northern blotting from the cells with and without RNAi induction. Depletion of HNRNPFH protein was successful as shown in the western blots from the two independent clones (Figure 3.11 A and B (i)). After 72 hours of RNAi induction, HNRNPFH protein was not detectable. The mRNA levels were then measured and the values obtained from the cells without induction of the RNAi were set to 1. As shown in Figure 3.11 A and B (ii), the reporter mRNA levels were not affected by HNRNPFH depletion. Similar to mRNA levels, the CAT activity was also not drastically affected after HNRNPFH RNAi induction. There was an increase of about 1.2-fold in the first clone after 48 hours of RNAi induction, but the effect was not maintained after 72 hours (Figure 3.11 A (iii)). This suggests that the reporter levels were rather not affected after depletion of HNRNPFH.

Taken together, there is no clear evidence that HNRNPFH binds directly the *RBP10* mRNA and therefore a role of HNRNPFH in regulating RBP10 in bloodstream forms could not be validated.

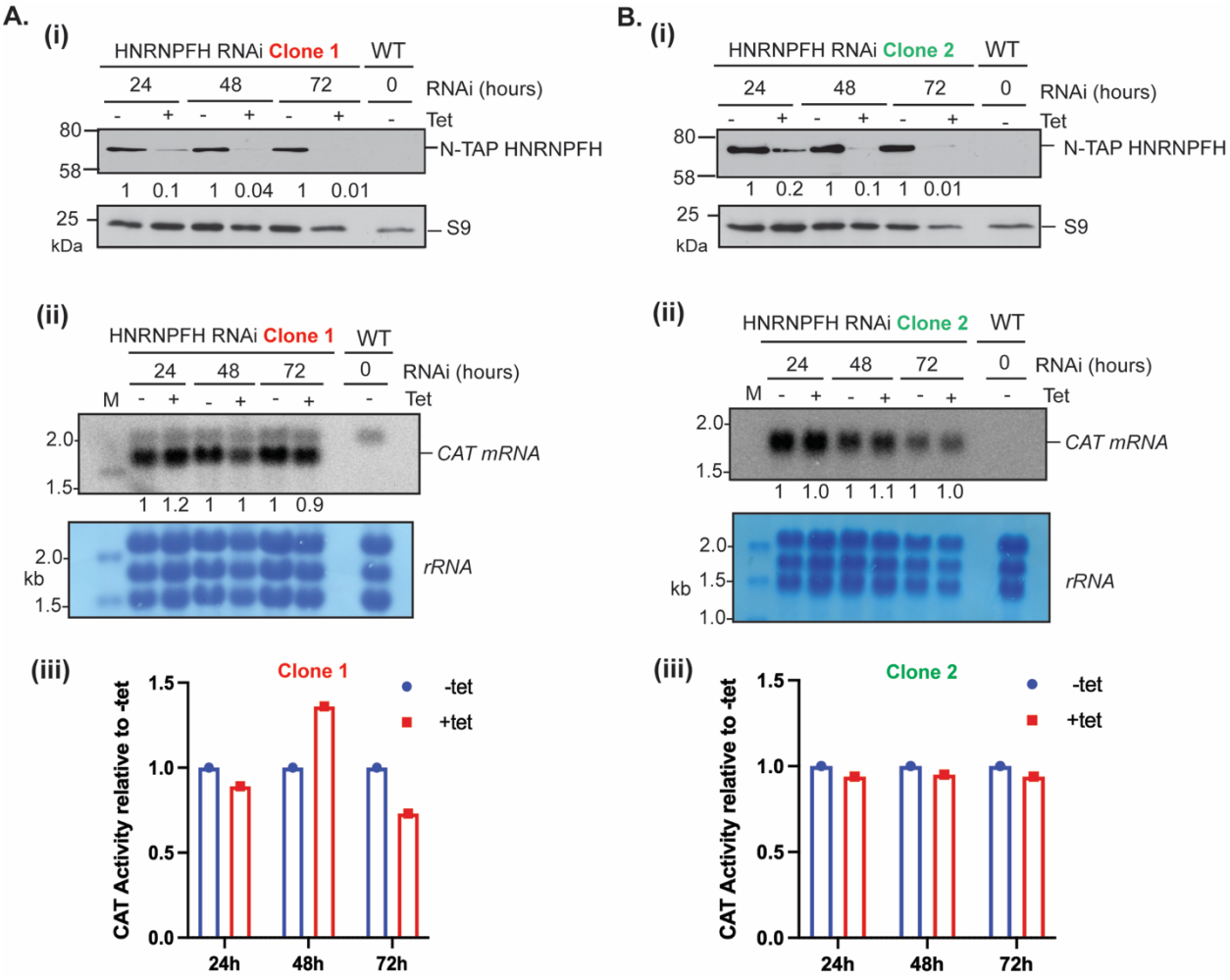


Figure 3.11: Depletion of HNRNPFH does not affect the CAT reporter bearing *RBP10* 3'-UTR fragment 1.1. **(A) (i)** Western blot showing TAP-HNRNPFH protein levels with and without induction of HNRNPFH RNAi in a permanent cell line (clone 1). The cell line used here has an in-situ N-TAP tagged HNRNPFH, an RNAi targeting HNRNPFH coding sequence and a CAT reporter bearing *RBP10* 3'-UTR fragment 1.1. Peroxidase-labelled anti-peroxidase (PAP) antibody was used for detection of TAP-HNRNPFH and ribosomal S9 protein is the loading control. **(ii)** RNA from the samples used in (i) were used for Northern blotting in order to check the effect of RNAi on the *CAT* mRNA containing *RBP10* 3'-UTR fragment 1.1. M is the marker. A section of methylene blue staining is depicted to show the loading and integrity of the ribosomal RNAs. The amount of *CAT* mRNA in presence of tetracycline was divided by the mRNA level in the absence of tetracycline. **(iii)** The CAT activity levels were measured in one permanent cell line (clone 1). The activity in absence of tetracycline is set to 1. **(B)** same as **(A)** for a second permanent cell line (clone 2).

3.3.2 The role of ZC3H28 in trypanosome bloodstream forms

Since ZC3H28 is an RNA-binding protein that was never characterized before, I decided not only to investigate its role in regulating the expression of RBP10 but also to characterize it and find out its potential function in trypanosome bloodstream forms.

i. ZC3H28 is essential for growth in bloodstream forms

ZC3H28 (Tb927.9.9450) is a 114 kDa RNA-binding protein which consists of 1030 amino acids with a single CCCH zinc finger domain spanning residues 998-1025 at the C-terminus. In its central part, ZC3H28 contains multiple histidine- and glutamine-rich regions, including $(H)_{11}$, $(H)_{17}$, $(Q)_{11}$ and $(H)_{14}$ (Figure 3.12 A). Sequence analysis reveals that ZC3H28 is conserved in Kinetoplastids; its position in the genome upstream of the gene encoding the peroxisomal proteins PEX13 is also conserved. Most sequence identity is concentrated at the C-terminus around the zinc finger domain and also towards the N terminus (Supplementary information/ section 5). All sequences examined have the C-terminal zinc finger domain. They all also have numerous histidine- and glutamine-rich regions (Supplementary information/ section 5). TrypTag database showed that C-terminally GFP tagged ZC3H28 is localized in cytoplasmic aggregates in slightly-stressed procyclic forms (Dean et al. 2017), but the N-terminally tagged version was uniformly distributed in the cytosol. Fritz et al. 2015 revealed that ZC3H28 was not enriched in granules formed after prolonged starvation. High-throughput RNAi target sequencing in trypanosomes suggested that ZC3H28 is essential in bloodstream forms, during differentiation but also in procyclic forms (Alsford et al. 2011).

To investigate whether ZC3H28 is essential for growth in bloodstream forms and to confirm the high-throughput results, RNAi-mediated depletion of ZC3H28 was induced in stable cell line Lister 427 in which one allele of ZC3H28 was tagged *in situ* with an N-terminal TAP tag (N-TAP ZC3H28). The experiment was carried out in two independent clones, and protein samples were collected after 24 and 48 hours to monitor the depletion of ZC3H28 (Figure 3.12 B). A wild-type cell line Lister 427 bloodstream form without the stem-loop RNAi construct and with intact ZC3H28 alleles was used as a control. Depletion of ZC3H28 by RNAi induction was successful in both clones as shown in Figure 3.12 B. As expected, cell growth was inhibited after 24 hours in both clones and cells died until the fourth day (Figure 3.12 C). However, after 4 days, some cells survived and resumed growth, probably because ZC3H28 was re-expressed and the RNAi effect was lost. I did not check whether this hypothesis is true but loss of RNAi regulation in trypanosomes is very common.

In conclusion, the results show that depletion of ZC3H28 affects trypanosome bloodstream forms growth, suggesting that the protein is essential at this life stage.

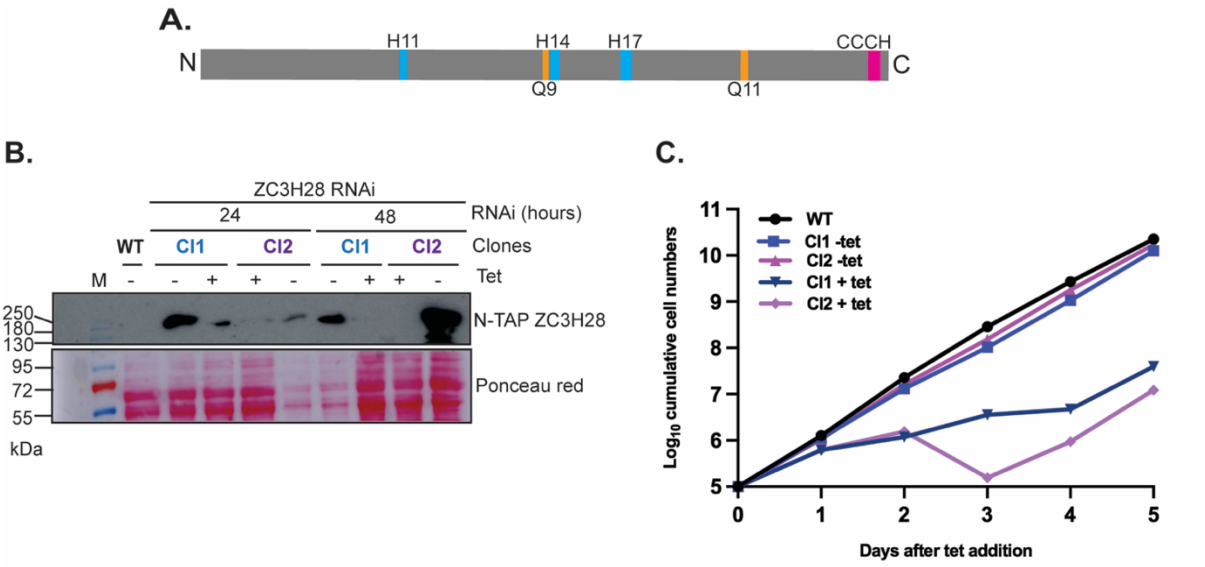


Figure 3.12: Depletion of ZC3H28 affects cell growth in trypanosome bloodstream forms. **(A)** Schematic representation of the ZC3H28 structure. The zinc finger domain CCCH is indicated, as well as the poly-histidine and poly-glutamine sequences. **(B)** Western blot showing the N-TAP tagged ZC3H28 in two independent clones with and without induction of RNAi for 24 and 48 hours. Expression of the stem-loop RNAi construct was induced with tetracycline (Tet). Peroxidase-labelled anti-peroxidase (PAP) antibody was used for detection of TAP-HNRNPFH and the red ponceau staining is used as loading control. M is the marker. As expected, the TAP-tagged protein is not detectable in the wild-type cell line. **(C)** Effect of ZC3H28 RNAi on cell numbers over 5 days. The cell density of the two independent clones used in (B) were monitored for 5 days with or without tetracycline. The cells with no RNAi construct are included for comparison.

ii. ZC3H28 increases the abundance of an attached reporter mRNA

In a tethering assay in trypanosomes, a protein of interest is fused to the lambda N (λN) peptide at the N-terminus and a myc tag at the C-terminus. The protein is expressed together with a reporter mRNA containing the "Box B" sequences. The lambda N (λN) peptide binds with high affinity to the Box B sequences (Erben et al. 2014). In high-throughput tethering screen, λN-ZC3H28-myc was found to activate the expression of the Box-B containing reporter (Erben et al. 2014). To find out whether this is true, expression of λN-ZC3H28-myc was induced for 24 hours in cell lines constitutively expressing the CAT reporter construct with 5 copies of "BoxB" sequences (Figure 3.13 A). The expression of λN-ZC3H28-myc was induced with tetracycline for 24 hours in four independent clones and this was checked by Western blotting (Figure 3.13 B). The *CAT* mRNA levels before and after induction of λN-ZC3H28-myc were analyzed by Northern blotting (Figure 3.13 C). CAT activities were measured enzymatically. A cell line with no expression of the CAT reporter and no inducible expression of λN-ZC3H28-myc and a cell line with CAT reporter but no inducible expression of λN-ZC3H28-myc were used as controls. Results from the four independent clones

confirmed that ZC3H28 increases the CAT reporter protein levels by 1.5-2.7-fold while the increase in the mRNA abundance was about 2-3.5-fold (Figure 3.13 D).

Overall, I herein report that tethering ZC3H28 increases the mRNA and expression levels of the attached reporter. The increase in mRNA level is higher than that of the expression level.

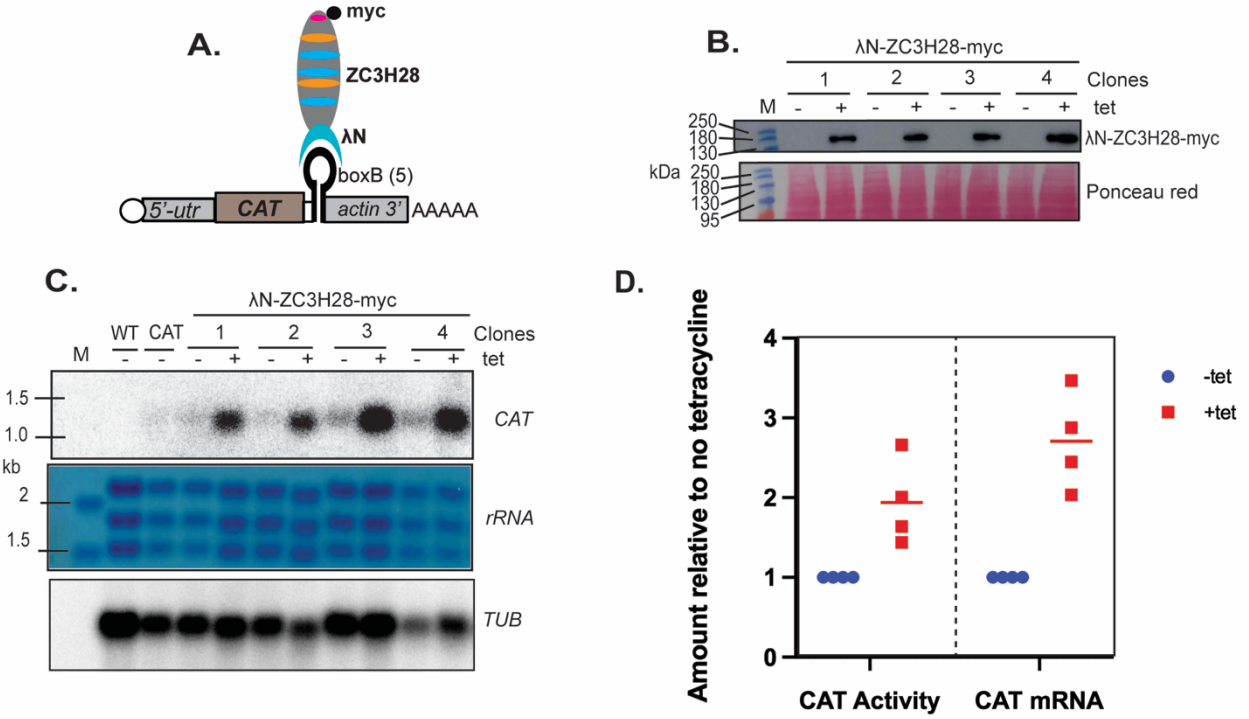


Figure 3.13: ZC3H28 increases expression levels of an attached CAT reporter mRNA. (A) The λN-BoxB tethering system. The ZC3H28 is fused at the N-terminus with the lambda N (λN) peptide and at the C terminus with the myc tag. The protein is tethered to the CAT reporter containing 5 copies of Box B sequences (for simplicity, only one copy is shown) upstream of the actin 3'-UTR. The lambda N (λN) peptide has high affinity to the Box B sequences. **(B)** Expression of λN-ZC3H28-myc was induced for 24 hours with tetracycline in four different independent clones. The protein was detected with anti-myc antibody and the red ponceau staining was used as loading control. M is the marker. **(C)** Effect of tethering ZC3H28 on CAT mRNA. The RNAs obtained from the samples used in (B) were purified and analyzed by Northern blotting. Cells with no CAT reporter and cells with CAT reporter but no inducible expression of the ZC3H28 (CAT) were used as negative controls. Methylene blue staining and beta-tubulin were used as loading controls. M is the marker. **(D)** Effect of tethering ZC3H28 on CAT activity and mRNA levels. For each independent clone, the amount of CAT mRNA or CAT activity in presence of tetracycline was divided by the mRNA level or CAT activity in the absence of tetracycline.

iii. ZC3H28 interacts with the MKT1 complex and several ribosomal proteins

To understand the mechanisms through which ZC3H28 could increase the reporter mRNA and expression levels, proteins associated with ZC3H28 were identified. For that, a Lister 427 bloodstream form cell line in which one allele of ZC3H28 was modified to encode a N-terminally TAP-tagged ZC3H28 was used. To find out whether the tag was fully functional, I attempted to delete the unmodified ZC3H28 gene, which should result in cells expressing only the endogenous N-TAP tagged ZC3H28. Unfortunately, it was not possible to select such cells. This could mean that the N-TAP tagged ZC3H28 is not fully functional; or alternatively the trypanosomes need both copies of ZC3H28 gene for normal growth so that a deletion creates a haploid insufficiency. Nevertheless, the N-TAP tagged ZC3H28 was purified using IgG beads. Only one step purification was done in order to avoid loss of material. The protein was released from the beads using the TEV protease (Figure 3.14 A). The experiment was done in three replicates under the same conditions. The cell lines inducibly expressing C-TAP tagged GFP and the cells with endogenous C-TAP tagged DRBD18 (described later) were used as controls. Western blot showing aliquots from each step of the affinity purification revealed that there was about 3-5% of the proteins bound to the beads after TEV cleavage, suggesting that most of the purified proteins were successfully released from the beads after digestion with the TEV protease (Figure 3.14 A and B). The eluate was run on SDS-PAGE gel and stained with Coomassie. Samples in each eluate replicate were divided into two slices (numbered boxes in Figure 3.14 C) and analyzed by quantitative liquid chromatography-tandem mass spectrometry (LC/MS). Proteins were denoted as being associated with ZC3H28 if they fulfil one of these two criteria: being at least 4-fold enriched with P-values < 0.01 (this criterion is obtained with the Perseus algorithm); and present in all three ZC3H28 purifications but absent in all the GFP/DRBD18 purifications.

With these criteria, 92 proteins were identified as being putative interactors of ZC3H28. The complete list of the proteins have been deposited in ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD027792. The proteins of the MKT1 complex (labelled in black in Figure 3.15 A and B) were significantly enriched. In the ZC3H28 purifications, there were 33-44 peptides identified for ZC3H28, 19-23 peptides for MKT1, 3-6 peptides for LSM12, 1-2 peptides for XAC1, 16-27 peptides for PABP1 and 40 peptides for PABP1. In the GFP purifications, no peptide was identified for either ZC3H28, LSM12, XAC1 or PABP1, only 1 peptide for MKT1, and 2-8 for PABP2. Oddly, PBP1 was only detected in two purifications out of the three ZC3H28 replicates, but in none of the controls. Since TAP-ZC3H28 is not fully functional, I suspect that the tag might have interfered with some interactions. When comparing these results with TAP-DRBD18 purifications, the MKT1 complex was still highly enriched (Figure 3.15 B): no peptide was identified for either ZC3H28, XAC1, LSM12 or PBP1; there was only 2-4 peptides for MKT1.

However, PABP1 and PABP2 were also associated with DRBD18 (19-24 peptides for PABP1 and 43-48 peptides for PABP2).

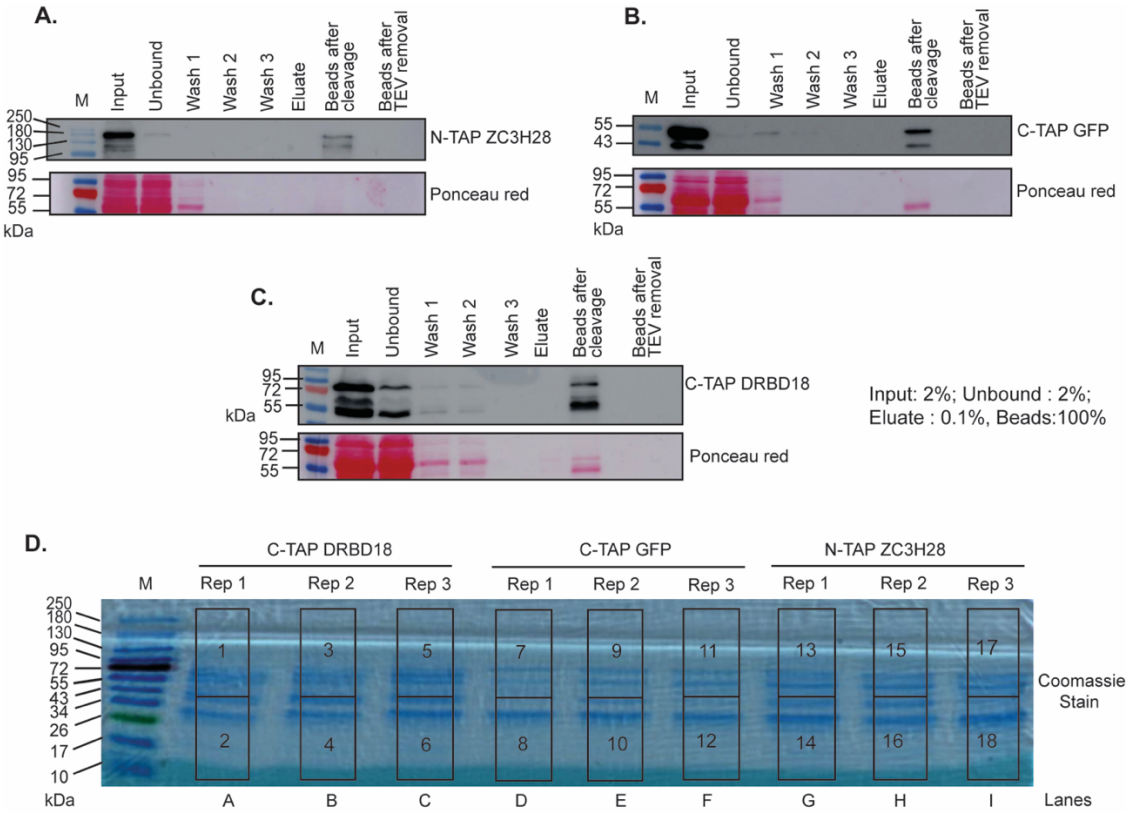


Figure 3.14: Tandem affinity purification of N-TAP ZC3H28, C-TAP GFP and C-TAP DRBD18. (A) N-TAP ZC3H28 was purified using IgG beads; 2% of the input, unbound, wash fraction, all of the beads aliquot and 0.1% of the eluate were collected and applied for Western blotting. About 3% of the protein remains bound to the beads after TEV protease cleavage. Peroxidase-labelled anti-peroxidase (PAP) antibody is used for the detection of the N-TAP ZC3H28, the antibody recognizes the protein A part of the TAP tag. The staining with red ponceau is used as loading control. M is the marker. (B) C-TAP GFP was purified using IgG beads similar to (A). M is the marker. About 5% of the protein remains bound to the beads after TEV protease cleavage. (C) C-TAP DRBD18 was purified with IgG beads similar to (A). M is the marker. About 5% of the protein remains bound to the beads after TEV protease cleavage. (D) Coomassie stain gel of the eluates obtained from one-step purifications of C-TAP DRBD18, C-TAP GFP and N-TAP ZC3H28 is shown. Each protein was purified in triplicates. Lanes A-C: triplicates of C-TAP DRBD18, D-F: triplicates of C-TAP GFP, G-I: triplicates of N-TAP ZC3H28. Each sample was sliced in two parts and later analyzed by quantitative mass spectrometry. M is the marker.

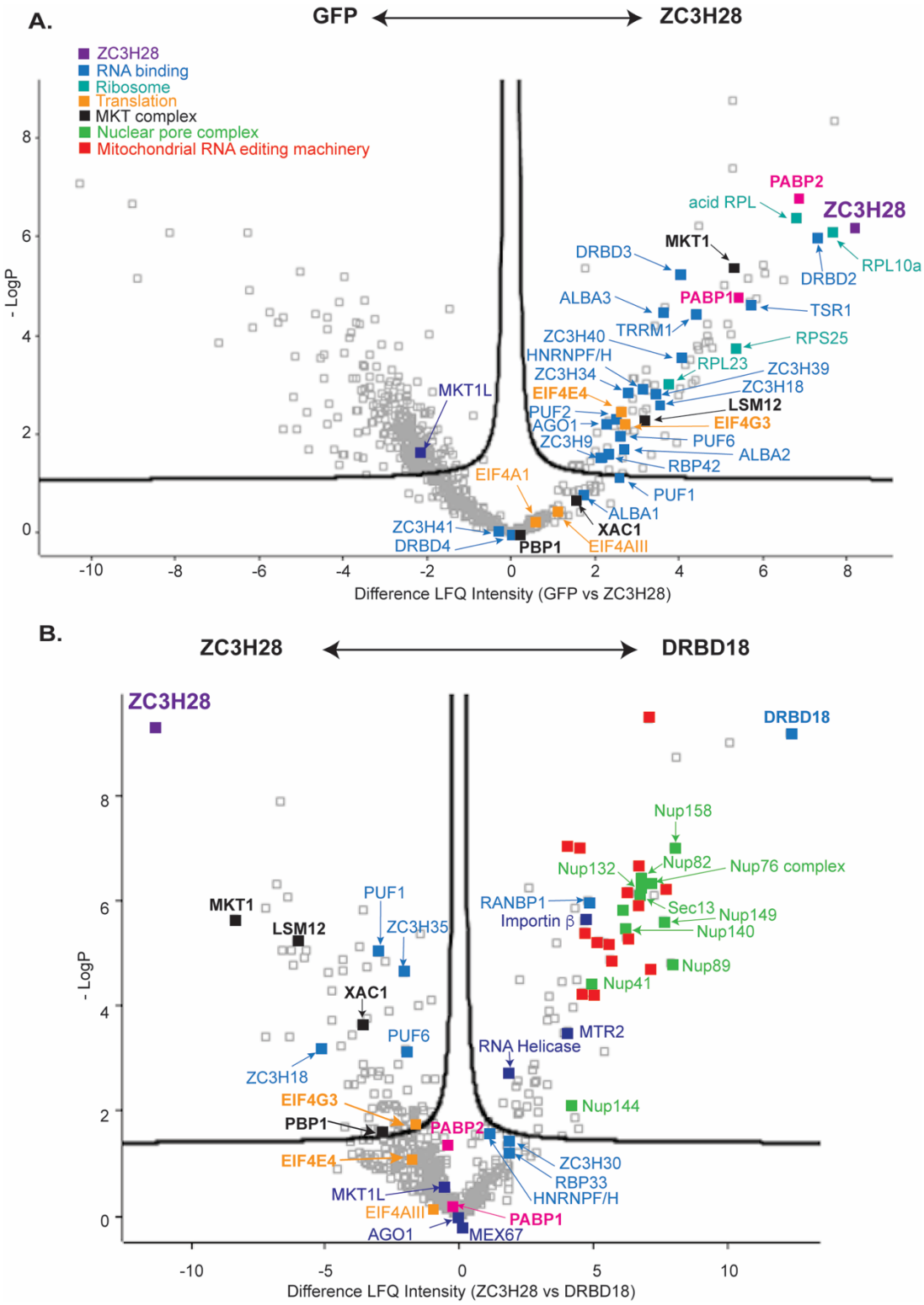


Figure 3.15: ZC3H28 associates with the MKT1 complex and several ribosomal proteins. (legend continues on the next page) (A) Volcano plot showing proteins interacting with TAP-ZC3H28 compared with GFP-TAP which serves as negative control. The figure was generated using Perseus software (Tyanova et al. 2016) the y-axis represents – $\log_{10}$ of the p-value for significant enrichment and the x-axis shows the differential

signal intensities drawn on a $\log_2$ scale. Selected enriched proteins are indicated but the full list of proteins is deposited to the ProteomeXchange consortium via the PRIDE partner repository with the dataset identifier PXD027792. The color code of the selected proteins is shown on the top left of the figure: purple denotes ZC3H28, blue represents RNA-binding proteins, cyan indicates ribosomal proteins, orange shows proteins involved in translation, black is proteins of the MKT1 complex, polyA binding proteins are indicated in pink. (B) Volcano plot showing proteins that are significantly enriched with ZC3H28-TAP and DRBD18-TAP (described later). Color code is similar to (A).

A total of 27 ribosomal proteins co-purified with TAP-ZC3H28: 14 out of 27 were proteins of the small subunit and 13 from the large subunits including the four acidic subunits and the Kinetoplastids-specific ribosomal protein NRBD2. Only one of the five known EIF4E-EIF4G translation initiation complexes, the EIF4E4-EIF4G3 specifically associates with TAP-ZC3H28 and not with TAP-GFP, although the complex also co-purified with TAP-DRBD18 (Figure 3.15 A and B). Both of the poly(A) binding proteins, PABP1 and PABP2 were identified in TAP-ZC3H28 purification, although a study by Zoltner et al. 2018 showed that ZC3H28 was preferentially associated with PABP2. EIF4E4 and EIF4G3 were associated with both PABPs. Many other RNA-binding proteins co-purified with ZC3H28 including DRBD2, ZC3H34 and ZC3H41 (which were associated with both PABP1 and PABP2 in Zoltner et al. 2018), HNRNPFH, TSR1, TRRM1, ZC3H39 and ZC3H40 (which were preferentially associated with PABP2), ALBA3 (which preferentially co-purifies with PABP1), PUF2, RBP42, ZC3H18 and ZC3H9 (which are not associated with PABPs). A comparison with TAP-DRBD18 showed that the proteins PUF1, PUF6, ZC3H35 and ZC3H18 were specific to ZC3H28. The N-TAP tagged ZC3H28 also pulled down the RNA interference effector AGO1 (Shi et al. 2007). This protein is involved in mRNA degradation and is not associated with PABPs.

Since results from previous quantitative mass spectrometry using the activator expression protein XAC1 as bait suggested that ZC3H28 is an interactor partner of the XAC complex (Melo do Nascimento et al. 2020), I decided to examine the interactions of ZC3H28 with the different components of the MKT1-complex using the pairwise yeast 2-hybrid assay (Figure 3.16 A). The expression of the constructs was first assessed by growth on SD medium containing Adenine and Histidine but lacking tryptophan and leucine (double dropout) and also by Western blotting (data not shown). The constructs of XAC1 as bait and the N-terminus of ZC3H28 as prey failed to be expressed in yeast as shown in the double dropout plates. The interactions between the proteins were assessed by growth on quadruple dropout SD plates, lacking Tryptophan, Leucin, Histidine and Adenine. In this assay, ZC3H28 interacted with itself and with PBP1 but it failed to interact with either MKT1, XAC1 or LSM12. The N-terminus of ZC3H28 consisting of 527 residues, containing two poly-histidine and one poly-glutamine, also interacted with itself, with PBP1 and with the full-length ZC3H28 but only when ZC3H28 was in a “bait” configuration.

To also confirm the results from the yeast two-hybrid assay, a pulldown of TAP-ZC3H28 using IgG beads from bloodstream form cell lines expressing also either *in situ* V5-MKT1 or V5-PBP1 was performed. 2% of the input (In), unbound fraction (Un) and first wash (W1) were loaded together with 100% of the eluate from the beads (El) and analyzed by Western blotting. The eluates represent about 48 times more cell equivalents than the input and the unbound fractions. Pulldown of TAP-ZC3H28 from both cells was successful because the protein was clearly enriched in the elution fraction (Figure 3.16 B and C). MKT1 was not associated with TAP-ZC3H28 (Figure 3.16 B) but detection of PBP1 depended on the presence of N-TAP ZC3H28 (Figure 3.16 C). The result from the co-immunoprecipitation of TAP-ZC3H28 and PBP1 is however to be taken with precaution since an unspecific band of the same size as V5-PBP1 was also found in V5-MKT1 (Figure 3.16 B).

Taken together, ZC3H28 is shown here to associate with the MKT1-complex, probably through direct interaction with PBP1. It also interacts not only with several RNA-binding proteins that were previously reported to be activators of gene expression but also with the cap-binding initiation complex EIF4E4/EIF4G3. This is consistent with an association of ZC3H28 with several ribonucleoprotein complex or polysomes.

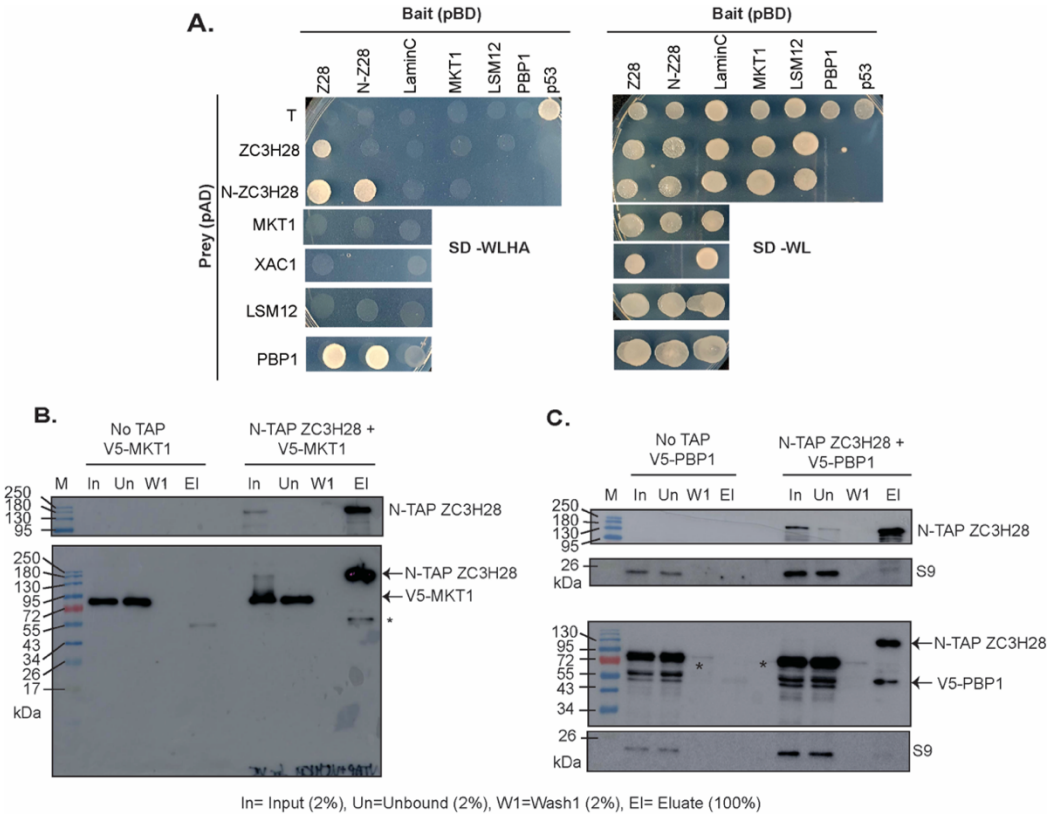


Figure 3.16: Interactions of ZC3H28 with the MKT1-complex. (legend continues on the next page)

(A) In the yeast two-hybrid assay, *S. cerevisiae* AH109 was transformed with combination of plasmids used as bait (pBD, DNA binding domain) and prey (pAD, transcription activation domain). The constructs containing ZC3H28, the N-terminus region of ZC3H28 and the proteins of the MKT1 complex were used. Expression of bait and prey was assessed by growth on media lacking tryptophan and leucin (SD-WL). Proteins interactions were then evaluated by growth in media lacking tryptophan, leucin, histidine and adenine (SD-WLHA). A positive control is the interaction between the SV40 large T antigen (T) and p53 (first row, last column) and the negative control is the combination of LaminC and the SV40 large T antigen (first row, third column). PBP1 and XAC1 fused to the DNA-binding domain are self-activators (Nascimento et al., 2020); therefore, they are used here only as activation domain fusions. **(B)** N-TAP tagged ZC3H28 was purified from cells expressing also *in situ* V5-MKT1. The peroxidase anti-peroxidase (PAP) antibody was used for the detection of TAP-ZC3H28 and anti-V5 was used to detect V5-MKT1. V5-MKT1 has a molecular weight of approximately 85 kDa and the asterisk (*) is of unknown origin. The cells expressing only *in situ* N-TAP ZC3H28 were used as controls. V5-MKT1 failed to interact with TAP-ZC3H28. **(C)** N-TAP tagged ZC3H28 was purified from cells expressing also *in situ* V5-PBP1. V5-PBP1 has a molecular weight of approximately 60 kDa; the additional band above the 72 kDa marker is of unknown origin but it was seen before in (Melo do Nascimento et al. 2020). The cells expressing only *in situ* N-TAP ZC3H28 were used as controls. Detection of PBP1 depended on the presence of N-TAP ZC3H28.

iv. ZC3H28 associates with long mRNAs that have low ribosome occupancy

To find out whether ZC3H28 binds the *RBP10* mRNA and which other mRNAs are preferentially bound by ZC3H28, the N-TAP tagged ZC3H28 was purified using IgG beads. The tag was cleaved with the TEV protease and the co-purified RNAs were identified by RNA-sequencing. The RNAs from the flow-through served as controls. Samples from each step of the affinity purification were loaded and applied for Western blotting (Figure 3.17 A). About 3% of the protein is detected in the unbound fraction suggesting that approximately 97% of the protein was bound to the beads and the cleavage of the tag with the TEV protease was successful, leaving only the protein A part of the tag attached to the beads (last lane in Figure 3.17 A). The experiment was done in triplicates under the same conditions. The complete raw data with read count has been deposited at Array express with the accession number E-MTAB-10674. The mRNA encoding ZC3H28 was 2.3-, 2.8- and 4.5-fold enriched in all the three pull-downs. The *RBP10* mRNA was not enriched in the ZC3H28 bound fraction (enrichment was 1.38-, 1.55- and 1.88-fold). A total of 180 mRNAs were more than 3-fold enriched in all the three purifications (Figure 3.17 B). These included 12 mRNAs encoding RNA-binding proteins, 14 encoding protein kinases and 89 encoding proteins of various functions (mitochondria, amino acids transporter, sterols, glycans, helicases, etc.). When comparing ZC3H28 target mRNAs with all other mRNAs, the median length of ZC3H28 bound mRNAs is twice that of all other mRNAs, suggesting the ZC3H28 preferentially binds long mRNAs. Interestingly, the median of their coding sequences was also twice that of others while their 5'-UTRs were 1.5 times longer than that of other mRNAs. The median for the 3'-UTRs of the ZC3H28 bound mRNAs was 3.7 times longer than that of all other mRNAs (Figure 3.17 C).

Although examination of the mRNAs bound by ZC3H38 showed that they are relatively long, correlation established between the mRNA binding to ZC3H28 and mRNA length was partial (Figure 3.17 D), suggesting that there is a degree of sequence specificity.

Another interesting finding is that ZC3H28 binding to mRNA was negatively correlated with number of ribosomes per kilobase of coding region (Figure 3.17 E). The median of the ribosome densities of ZC3H28 bound mRNAs was half that of other mRNAs (Figure 3.17 F), this clearly suggests that ZC3H28 preferentially binds mRNA with low ribosomal densities. One might think that ribosome density is caused only by the lengths of the bound mRNAs, this was not the reason because there was no correlation between ribosome density on the coding region and either mRNA length or 3'-UTR length (data not shown here but available in Tshitenge and Clayton 2021.

Finally, the 180 ZC3H28 target mRNAs were searched for enriched motifs using the MEME motif search. The mRNAs that were less than 1-fold enriched in the bound fraction served as controls. The motif search showed weak enrichment of poly (U) and poly (AU) sequences (Figure 3.17 G). When using annotated 3'-UTRs only and size-matched control mRNAs, there was significant enrichment of poly (AU) and polypurine motifs in the 3'-UTRs of the bound mRNAs (Figure 3.17 G). However, these motifs were not exclusive of all the bound mRNAs. The sequence $(AU)_{10}$ was present at least once in 49 of the 106 3'-UTRs of the bound mRNAs while it was also present in 36 of the 103 3'-UTRs of the unbound mRNAs.

Taken together, ZC3H28 is an RNA-binding protein, which binds long mRNAs that are slightly stable and have low ribosome occupancies on their coding sequence. Although tethering and proteomics data are consistent with a role of ZC3H28 in enhancing mRNA stability and probably translation, RIP-Sequencing revealed that ZC3H28 does not bind the *RBP10* mRNA. Therefore, a precise role of ZC3H28 in regulating the *RBP10* mRNA in bloodstream forms is not resolved.

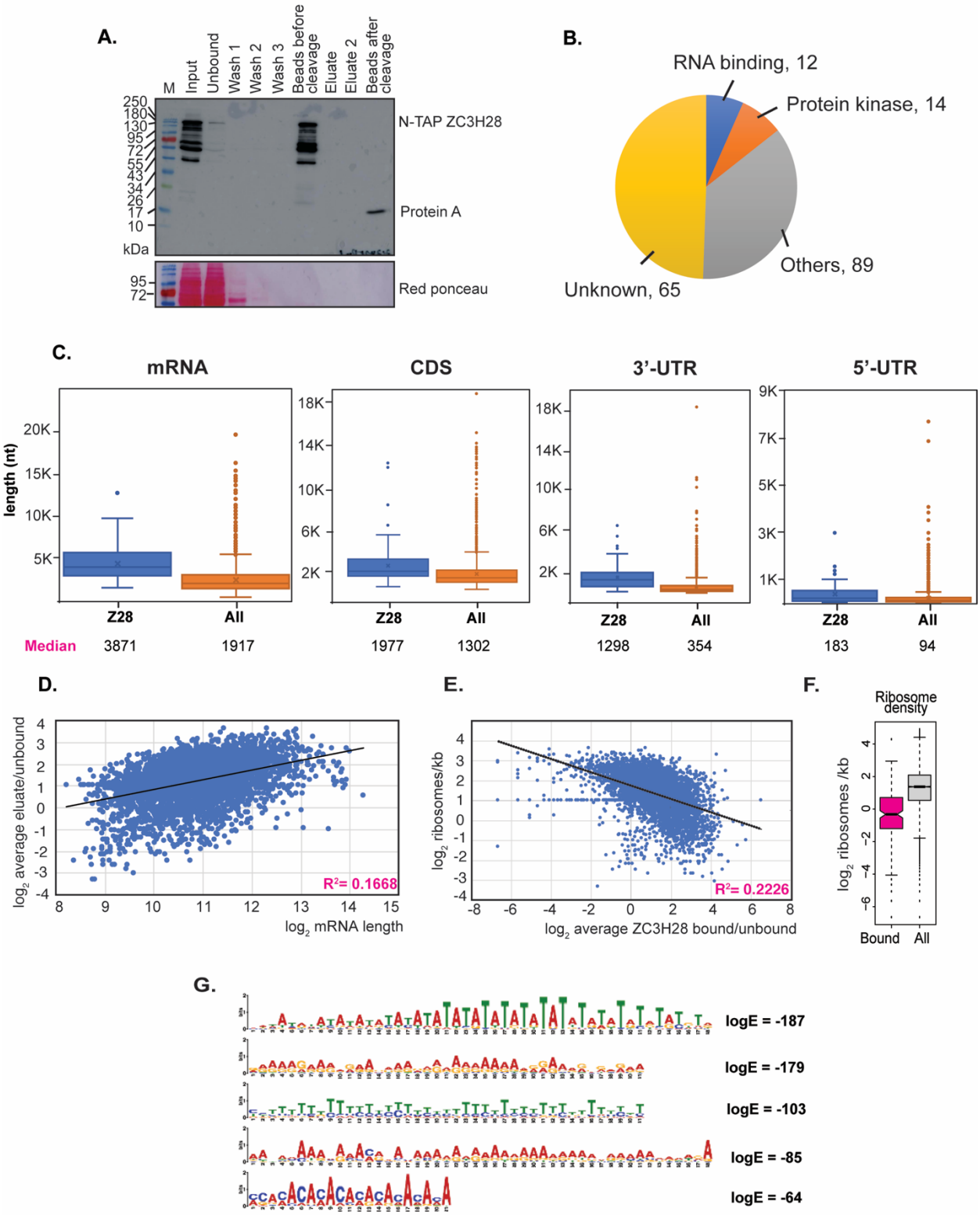


Figure 3.17: Interactions of ZC3H28 with mRNAs. (Legend continues on the next page)

(A) N-TAP ZC3H28 was purified using IgG beads and the tag was cleaved with TEV protease. For each step of the TAP purification, 20 μl sample was loaded and analyzed by Western blotting. M is the marker. Peroxidase-labelled anti-peroxidase (PAP) antibody was used for the detection of TAP-HNRNPFH and the red ponceau staining was used for loading control. No protein was detected after cleavage by the TEV protease. **(B)** Functional categories of proteins encoded by mRNAs bound by ZC3H28. **(C)** The lengths of mRNAs that were 3-fold enriched with TAP-ZC3H28 were compared with the unbound mRNAs. The coding sequences (CDS), the 3'-UTRs and the 5'-UTRs were also compared for length. The boxes extend from the 25th to 75th percentiles, the dots are outliers and the notches are indicators of variability. Median values are shown under the labels. This dataset was done with the unique gene list from (Siegel et al. 2010). **(D)** The enrichment of ZC3H28 binding to mRNAs (y-axis) was plotted against the mRNA length (x-axis). **(E)** The number of ribosomes per kilobase of coding region (taken from Antwi et al. 2016) on the y-axis is plotted against the binding of the mRNAs to ZC3H28 which is on x-axis. **(F)** The 180 mRNAs that were bound by ZC3H28 were compared with all other mRNAs for ribosome density (Fadda et al. 2014). This plot is from Prof. C. Clayton. **(G)** Motifs enriched in the 3'-UTRs of ZC3H28-bound mRNAs relative to the controls mRNAs showing less than 1-fold average enrichment in the bound fraction and same sizes with the bound mRNAs

v. ZC3H28 depletion affects trypanosome transcriptome

Since tethering of ZC3H28 showed that the protein increased the abundance of an attached mRNA reporter and proteomics data suggested that ZC3H28 might stabilize its target mRNAs through interaction with the MKT1 complex, I then checked whether the depletion of ZC3H28 affects trypanosome transcriptome. I was also interested to find out whether ZC3H28 mRNA targets are impaired after depletion of ZC3H28.

A time course study was carried out in two independent clones in order to examine how depletion of ZC3H28 affects cell growth. Starting from 12 hours after induction of ZC3H28 RNAi, the tagged protein was not detectable and remains at this state throughout (Figure 3.18 A). The cells grew normally from 0 to 12 hours after induction of RNAi. However, from 12 hours, the cells grew slightly slower than the controls, and growth was inhibited only after 24 hours (Figure 3.18 B). First, I looked at the transcriptomics from triplicate samples of the first clone (A) without tetracycline and after induction of ZC3H28 RNAi for 10 hours. Very few differences were found. Therefore, the samples from the two independent clones (A and B) were examined after 14, 16 hours of RNAi induction, and after 24 hours with and without tetracycline induction. A principal component analysis (Figure 3.19 A) showed that the first set of controls (triplicates of clone A without tetracycline, labelled A0-1, A0-2 and A0-3, colored grey squares) were well clustered and separated from the second set of controls (clone A and B after 24 hours of incubation without tetracycline, labelled A0-4 and B0, black squares).

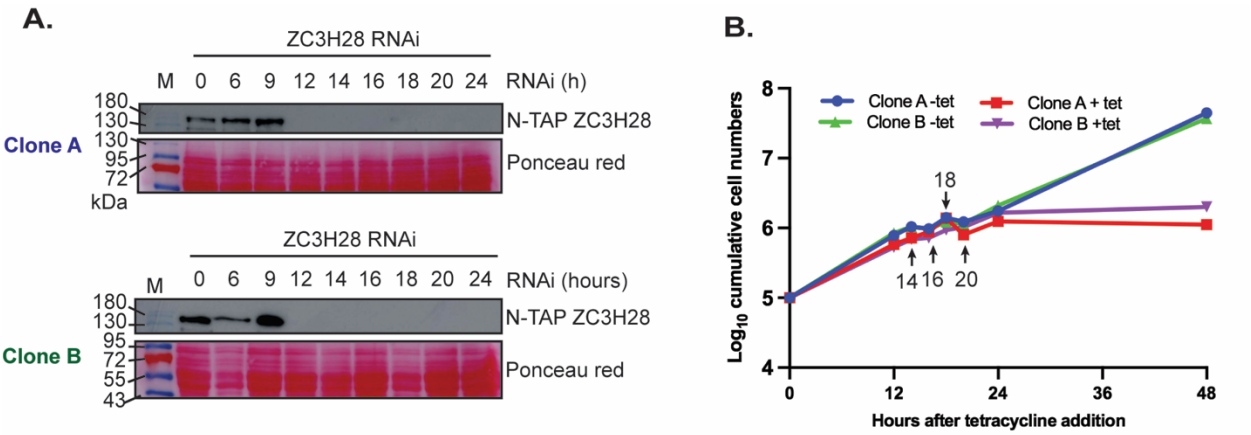


Figure 3.18: Time course study for ZC3H28 RNAi. (A) Western blots of samples taken at different time points after ZC3H28 RNAi in bloodstream form trypanosomes in two independent clones (A and B) are shown. Expression of the stem-loop RNAi construct was induced with tetracycline. Peroxidase-labelled anti-peroxidase (PAP) antibody was used for detection of N-TAP tagged ZC3H28 and the red ponceau staining is used as loading control. M is the marker. **(B)** Cumulative growth curves of cells used in (A) with and without ZC3H28 RNAi in two independent clones.

Interestingly, the triplicates samples collected with and without tetracycline induction for 10 hours clustered altogether (Figure 3.19 A) and the samples collected after 24 hours of RNAi induction (A24, B24) clustered with their controls collected without tetracycline induction of the RNAi (A0-4, B0). In contrast, there was clear effect on transcriptome after 14 hours (A14, B14) and 16 hours (A16, B16). The cell densities used for the two set of controls were different: the triplicate samples collected after 10 hours without tetracycline induction were 5-6x10^5/ml while the duplicates samples from the two independent clones (A0-4, B0) were collected after 24 hours with density of 1x10^6/ml. After 24 hours of RNAi induction, the densities of cells were 1.3-1.6x10^6 while those at 14h and 16h were lower (Figure 3.19 A). It seems that the high density of the cells affected the RNAi effect, which might explain the reason why the cells with and without tetracycline after 24 hours looked the same (Figure 3.19 A). So, the effect of cell density was assessed by looking at the differences between the two set of controls. In summary, 180 mRNAs were 1.5-fold increased at higher density, including mRNAs encoding procyclin and proteins of mitochondrial metabolism, which suggest changes with early differentiation (Silvester et al. 2018) or a stress response (Quintana et al. 2021). In contrast, 82 mRNAs were significantly decreased, including mRNAs that encode RNA polymerase I subunits, nucleotide transporters and translation elongation factor, consistent with slowing growth. While checking the effect of ZC3H28 RNAi on protein synthesis, result showed that translation was not impaired by ZC3H28 RNAi but trypanosomes at cell density over 10^6/ml showed 20-60% less translation when compared with those at density below 5x10^5/m (Figure 3.19 B, lane 1 and 7 compared to other lanes).

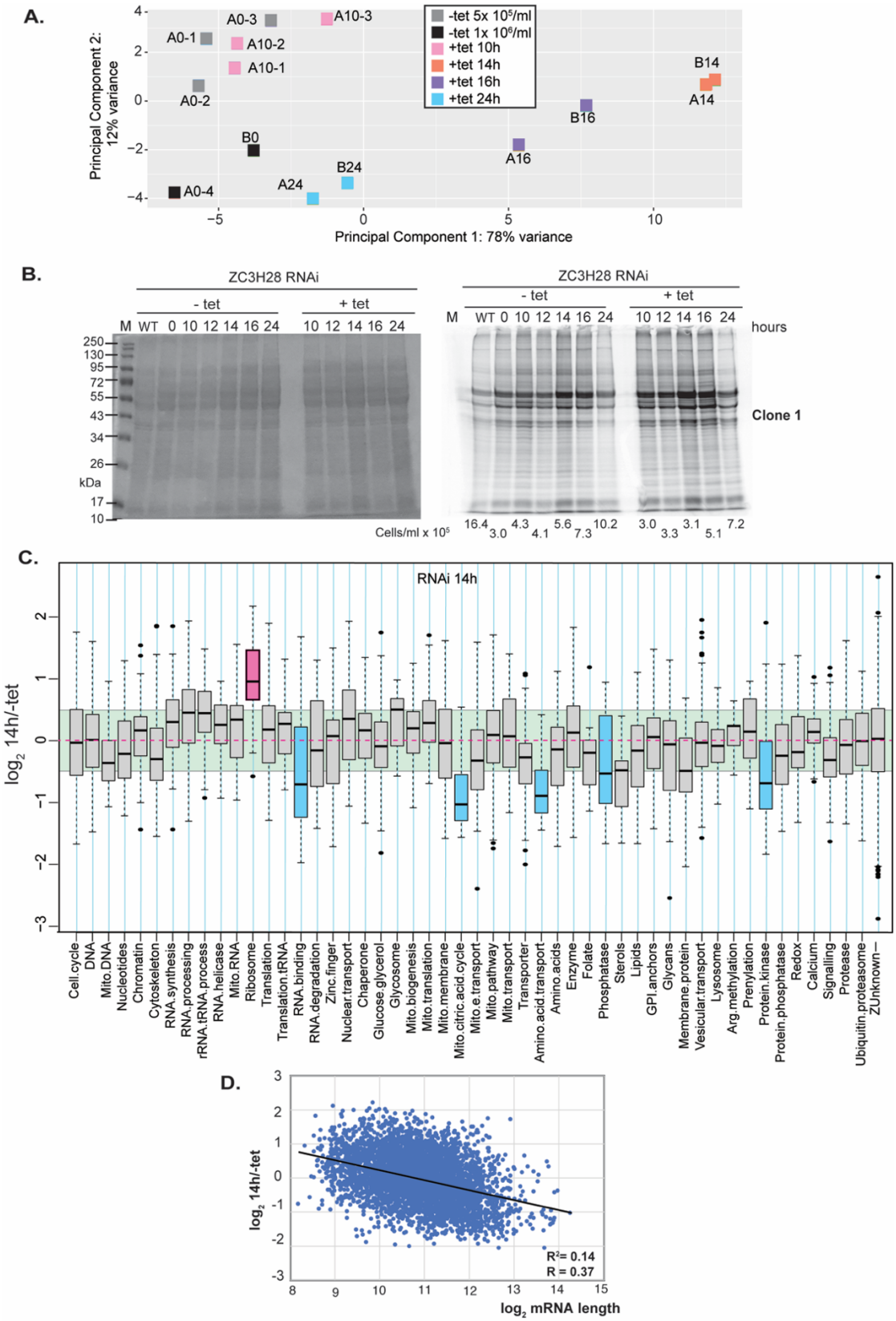


Figure 3.19: Effect of ZC3H28 depletion on the transcriptome. (Legend continues on the next page)

(A) Principal component analysis of cells used with and without depletion of ZC3H28. The control samples without induction of ZC3H28 RNAi collected after 10 hours are colored in grey and after 24 hours are colored in black. Samples collected after ZC3H28 RNAi induction with tetracycline at different time points are color-coded. Results in this figure are for the set of unique genes, to avoid over-counting repeated genes. **(B)** Effect of cell density and ZC3H28 RNAi on protein translation in clone A is displayed. The Coomassie stained membrane is on the left and the [^{35}S]-Met+Cys incorporation on the right. **(C)** The set of unique genes was placed in functional categories and the effects of ZC3H28 RNAi (14h) were plotted for each category. The green background shows the range between 25th and 75th percentiles for the total dataset and the magenta dotted line is the median. The category with a median above the 75th percentile is colored in pink while those with medians below the 25th percentile are in blue. **(D)** The effect of RNAi (y-axis) was plotted against the annotated mRNA length (x-axis). Correlation coefficients were calculated in Microsoft Excel.

It is possible that cells with density above 10^6/ml were already stressed and the effects of ZC3H28 depletion on mRNA abundance can only be seen when the cells are growing optimally at low densities. Therefore, the focus was on 14 hours after addition of tetracycline: at this time, 561 mRNAs were significantly increased by two-fold, with 48% of mRNAs encoding the ribosomal proteins and 673 mRNAs were significantly decreased by at least two-fold, including mRNAs encoding RNA-binding proteins, protein kinases, citric acid cycle enzymes, and amino acid transporters (Figure 3.19 C). The depletion of ZC3H28 RNAi was only weakly correlated with mRNA length (Figure 3.19 D). The *RBP10* mRNA was significantly decreased after ZC3H28 RNAi, but since it was not bound by ZC3H28, I believe this is a secondary effect and not a direct effect.

Moreover, from the 180 ZC3H28 target mRNAs, 56 were decreased after depletion of ZC3H28, these included mRNAs encoding 7 protein kinases and 5 RNA-binding proteins. Only two bound mRNAs were significantly increased.

Overall, this result reveals that the depletion of ZC3H28 in bloodstream forms affects significantly the transcriptome with a slight bias toward longer and ZC3H28 bound mRNAs. Since many RNA-binding proteins were bound and affected by ZC3H28, transcriptomic changes are also a result of several regulatory pathways being impaired.

3.3.3 The role of DRBD18 in regulating processing and nuclear export of the *RBP10* mRNA

The last candidate protein that was investigating for playing a role in the regulation of RBP10 expression in bloodstream forms is the double RNA-binding protein, DRBD18. As previously mentioned, DRBD18 is expressed in bloodstream forms as well as in procyclic forms. Studies by Lott et al. 2015 showed that the protein is essential in procyclics and is regulated by arginine methylation. Depending on the degree of methylation, DRBD18 was associated with different ribonucleoprotein complexes. Depletion of DRBD18 in procyclics led to a growth arrest after 48 hours and an increase in *RBP10* mRNA levels (Lott et al. 2015).

i. DRBD18 is essential in bloodstream forms and affects RBP10 mRNA processing

To investigate whether DRBD18 (Tb927.11.14090) regulates the RBP10 expression levels in bloodstream forms, RNAi-mediated depletion of DRBD18 was performed using a stem-loop construct targeting DRBD18 coding sequence. The RNAi construct was inducibly expressed in Lister 427 strain trypanosomes with tetracycline. DRBD18 was detected by Western blotting with anti-DRBD18 antibody points after induction of DRBD18 RNAi revealed loss of the protein from 12 hours post-induction (Figure 3.20 A). However, the amount of DRBD18 in the sample collected without tetracycline (0h, Figure 3.20 A) was slightly less than the level of the protein in the wild-type suggesting a small degree of leakiness from the RNAi construct. The depletion of DRBD18 caused clear growth inhibition after 24 hours of RNAi induction (Figure 3.20 B). This reveals that DRBD18 is essential for growth of bloodstream form trypanosomes. Remarkably, depletion of DRBD18 correlated with a decrease in RBP10 expression levels of about 60-70% (Figure 3.20 B). At the same time, Northern blots showed that *RBP10* mRNA longer species decreased after DRBD18 RNAi induction and shorter variants of different sizes accumulated (Figure 3.20 C). Incorrect processing of the *RBP10* mRNA implies that either *trans*-splicing or polyadenylation is impaired. Since the probe used for northern blots was from RBP10 coding sequence, *RBP10* mRNA isoforms generated after depletion of DRBD18 might be products of alternative polyadenylation. To find out whether DRBD18 RNAi affects processing on a global level, the northern blots hybridized with RBP10 probe were stripped and hybridized with beta-tubulin (β-tubulin) probe. In fact, tubulin genes are arranged in alternating alpha-beta tandem repeats; processing inhibition through heat shock, sinefungin or drug treatment leads to accumulation of tubulin RNA dimers and multimers (Begolo et al. 2018; Muhich and Boothroyd 1988). Notably, depletion of DRBD18 did not affect processing of *β-tubulin* mRNA. The upper bands seen on Northern blots (indicated as asterisk) are *RBP10* mRNA isoforms which remained due to an inefficient stripping (Figure 3.20 C).

Since the different variants of *RBP10* mRNA might be products of alternative polyadenylation, it was important to delineate the poly (A) sites of these mRNA variants by RT-PCR. However, analysis of the 3'-ends by RT-PCR (3'-RACE) using oligo $(dT)_{18}$ was not successful. This is probably because the sequences in the 3'-UTR are of low complexity and the forward primers are probably similar to the sequences found in other mRNAs.

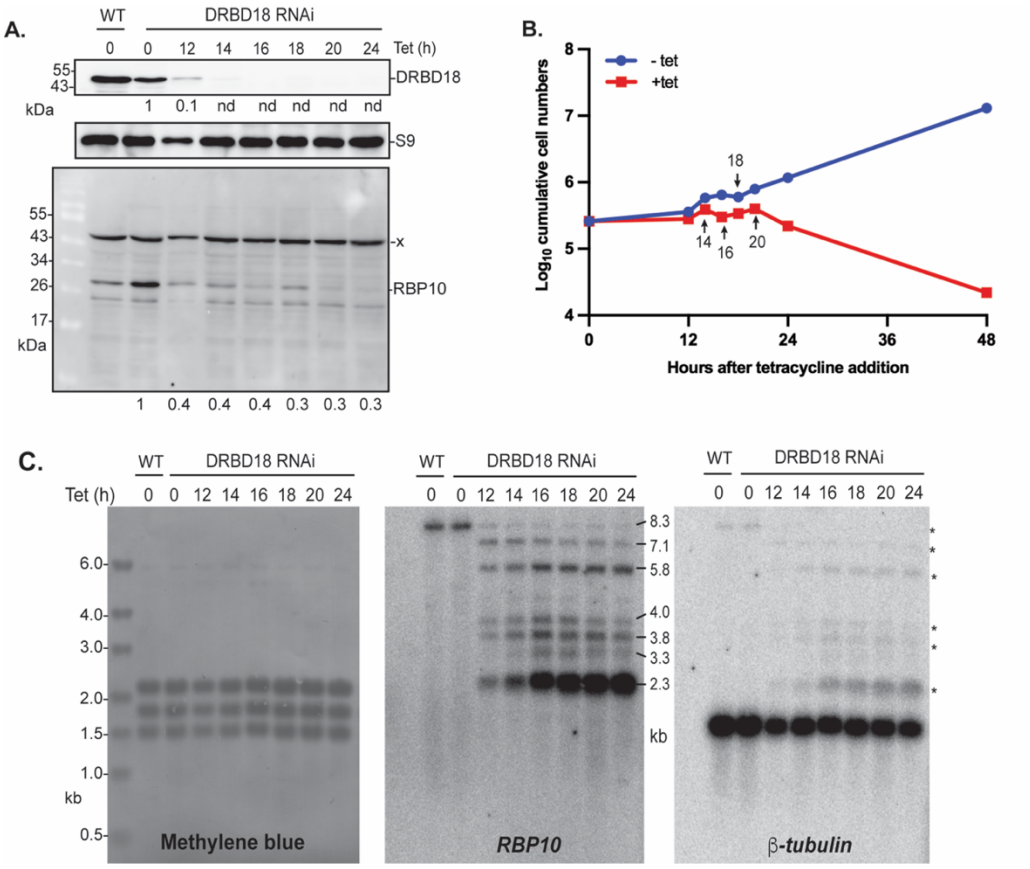


Figure 3.20: DRBD18 depletion by RNAi affects *RBP10* mRNA processing. (A) Western blots of samples collected after DRBD18 RNAi induction at different time points are shown. The wild-type Lister 427 strain bloodstream forms and the sample without induction of DRBD18 RNAi (0) were used as controls. DRBD18 was detected with anti-DRBD18 antibody and the ribosomal protein S9 is the loading control. The polyclonal anti-RBP10 antibodies were used for detection of RBP10 and a non-specific band (x). **(B)** Cumulative growth curves of bloodstream forms with and without induction of DRBD18 RNAi. The samples are the same with those used in (A). **(C)** The samples used in (A) were also subjected to RNA extraction and the effects on RBP10 mRNA were examined. The methylene blue-stained membrane is on the left, the membrane hybridized with RBP10 probe from the coding region is in the middle. The membrane was then stripped and hybridized with a β-tubulin probe as a further control; this is on the right. * are RBP10 signal that remained after stripping.

To confirm the Northern blot results, I decided to examine transcriptomes of bloodstream forms after induction of DRBD18 RNAi. A time course study was carried out to assess the

point at which depletion of DRBD18 did not affect cell growth. Western blot and cumulative growth curves showed that at 12 hours after induction of DRBD18 RNAi, the protein was depleted and cell growth was not affected (Figure 3.21 A and B). Therefore, trypanosome transcriptomes were examined after 12 hours with and without addition of tetracycline. A principal component analysis showed reproducibility between the triplicates without tetracycline induction of the RNAi and also between those with tetracycline induction. In addition, there was clear separation between the two conditions (Figure 3.21 C). When considering read changes on coding sequences of the transcripts, more than 3000 mRNAs were significantly affected by DRBD18 RNAi (padj<0.0.5). The transcriptomics raw data are deposited in Array express with accession number E-MTAB-9783. When considering those with at least 1.5-fold change, 357 mRNAs were increased (Figure 3.21 D) including mRNAs encoding enzymes involved in citric acid cycle, proteins of the electron transport chain, procyclic-specific membrane proteins and RNA-binding proteins that are more abundant in the procyclics. This suggests an early differentiation towards procyclic forms, but the effect might also be secondary. Approximately 654 mRNAs were 1.5-fold decreased (Figure 3.21 D), including mRNAs that encode proteins involved in glycerol metabolism and translation factors.

The *RBP10* mRNA was not affected by DRBD18 depletion. However, visualization of the RBP10 locus in the Integrated Genome Viewer revealed that in presence or absence of tetracycline, the read density over the *RBP10* coding region was 2-3 times higher than the density over the upstream gene, Tb927.8.2770 (Figure 3.21 D). Over the last 2 kb of *RBP10* 3'-UTR, there were fewer reads after DRBD18 depletion (cyan dotted line) as compared to the reads without induction of DRBD18 (Figure 3.21 E). This suggests that the long *RBP10* mRNA variants are lost, consistent with the Northern blot results (Figure 3.20 C). This also revealed that multiple alternative polyadenylation sites within the *RBP10* 3'-UTR might be used upon DRBD18 depletion. Consistent with this, the *RBP10* 3'-UTR harbors several polypyrimidine tracts (highlighted and underlined in blue, Supplementary 1) as well as several poly (A) tracts that might be preferentially used by the splicing and polyadenylation machinery after DRBD18 depletion.

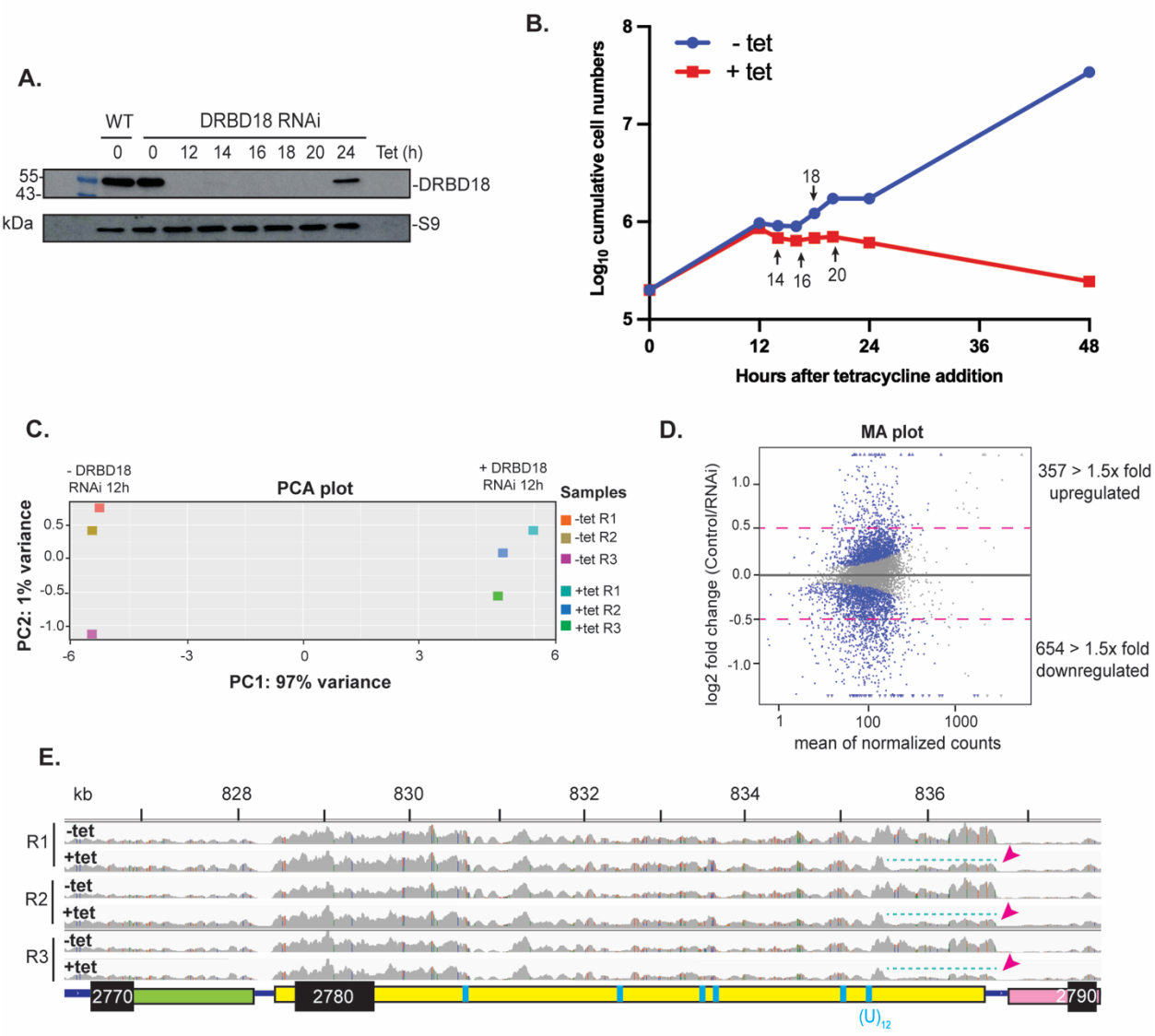


Figure 3.21: Transcriptomics data showed that DRBD18 RNAi influences RBP10 alternative polyadenylation. (Legend continues on the next page) **(A)** Time course study to determine the time point to be used for transcriptome analysis. Western blots of samples collected after DRBD18 RNAi induction at different time points are shown. Induction of RNAi was done with tetracycline (Tet). The wild-type Lister 427 strain bloodstream forms and the sample without induction of DRBD18 RNAi (0) were used as controls. DRBD18 was detected with anti-DRBD18 antibody and the ribosomal protein S9 is the loading control. **(B)** Cumulative growth curves of bloodstream forms with and without induction of DRBD18 RNAi from the samples used in (A). At 12 hours, growth of trypanosomes is not affected by DRBD18 RNAi. **(C)** Principal component analysis (PCA) plot of samples collected at 12 hours with and without induction of DRBD18 RNAi. Triplicates (R1, R2 and R3) samples were used for each condition. **(D)** mRNA changes after DRBD18 RNAi for 12 hours in bloodstream forms. The MA plot highlights the fold changes relative to the cells without induction of DRB18 RNAi. The blue dots indicate mRNAs that showed differential expression (padj<0.05); the blue dots below (-0.5) and above (0.5) the two pink dashed lines represent the subset of mRNAs with at least 1.5-fold change. **(E)** Visualization of the transcriptomics data obtained 12 hours after tetracycline-mediated induction of DRBD18 RNAi, using the Integrated Genomics Viewer. Three replicates are shown on a linear scale. The RBP10 mRNA (Tb927.8.2780) full length is shown below the tracks, together with the upstream gene (Tb927.8.2770) and the

downstream gene (Tb927.8.2790). Positions of polypyrimidine tracts $(U)_{12}$ or $(U)_6C(U)_6$ sequences are indicated in cyan, but there are numerous other polypyrimidine tracts present in the RBP10 3'-UTR (Supplementary text 1). Loss of reads at the end of the RBP10 3'-UTR after 12 hours of DRBD18 RNAi induction in the three replicates samples are highlighted in cyan dotted lines. The pink arrows also highlight the loss in the reads. Colored lines indicate read mismatches with the TREU927 reference genome.

To find out whether the downstream mRNA (Tb927.8.2790) is also alternatively *trans*-spliced, RNA samples collected with and without DRBD18 RNAi were loaded on a polyacrylamide urea gel and blotted on a positively charged membrane. The blot was hybridized with probes from the coding region of the downstream gene (Figure 3.22 B). Depletion of DRBD18 by RNAi was successful as shown in the Western blot although a small degree of protein re-expression was seen after 16 hours of DRBD18 RNAi induction (Figure 3.22 A). It is important to note that the downstream transcript has been annotated as an acetyl-coA synthetase pseudogene with a length of about 0.37 kb. Oddly, the blot gave bands that migrated slower than expected (~0.8 kb and > 1.0 kb), probably because the probe has also hybridized with the mRNA sequence present elsewhere in the genome (Tb927.8.2520) that encodes the full-length protein. I also tried to use a probe from the intergenic region but the RNAs detected were again larger than expected (data not shown).

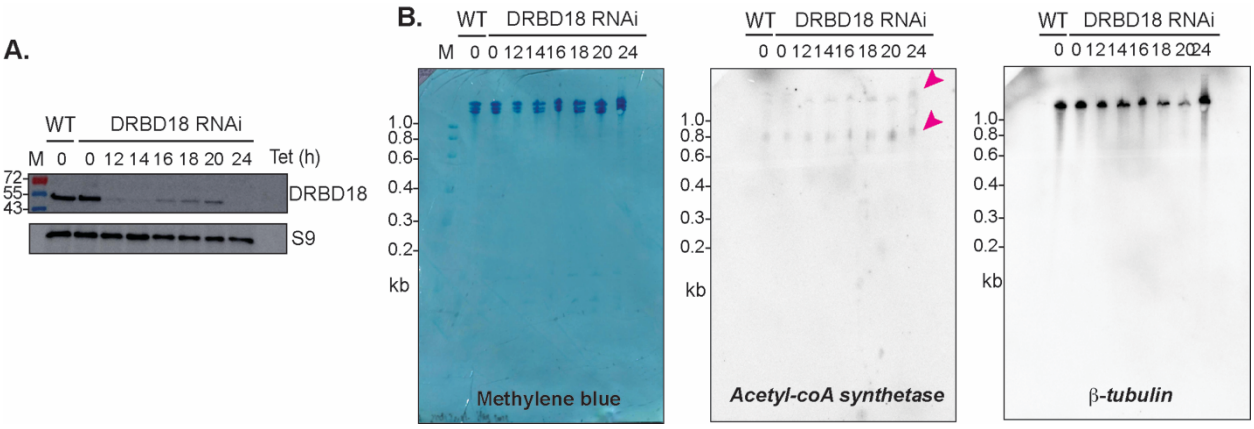


Figure 3.22: Effect of DRBD18 depletion by RNAi on the acetyl-coA synthetase mRNA. (A) Western blots of samples collected after DRBD18 RNAi induction at different time points are shown. Induction of DRBD18 RNAi was done with tetracycline (Tet). The wild-type Lister 427 strain bloodstream forms and the sample without induction of DRBD18 RNAi (0) were used as controls. DRBD18 was detected with anti-DRBD18 antibody and the ribosomal protein S9 is the loading control. **(B)** The samples used in (A) were loaded on a urea gel and the effect of DRBD18 RNAi on the acetyl-coA synthetase mRNA was examined. The methylene blue-stained membrane is on the left, the membrane hybridized with acetyl-coA synthetase probe in the middle and the membrane hybridized with a beta-tubulin probe as a further control on the right. The arrows in pink represent the detected bands which are higher than expected.

ii. DRBD18 affects processing of other mRNAs

Since transcriptomics data together with Northern blots showed that DRBD18 affected *RBP10* mRNA processing, I also checked manually whether processing of other mRNAs with long 3'-UTRs were affected. Using the Integrative Genome Viewer, results showed that the mRNA encoding DRBD12 was also alternatively polyadenylated, with a strong evidence of accumulation of shorter variants of *DRBD12* mRNA at around 4 kb after the coding region (Figure 3.23 C). Northern blotting of RNAs collected at different time points after induction of DRBD18 RNAi was performed to confirm these results. Western blotting revealed the loss of DRBD18 after induction of RNAi by tetracycline (Figure 3.23 A). As expected, from 12 hours after DRBD18 RNAi, *DRBD12* mRNA longer species got lost but smaller intermediates of about 4.1 kb and 2.6 kb long accumulated (Figure 3.23 B). Analysis of RNA-Seq data also showed that for about 154 mRNAs, the change in reads in their 3'-UTRs was at least 1.5-fold lower that changes in their coding sequence reads, suggesting that their processing was impaired after DRBD18 depletion.

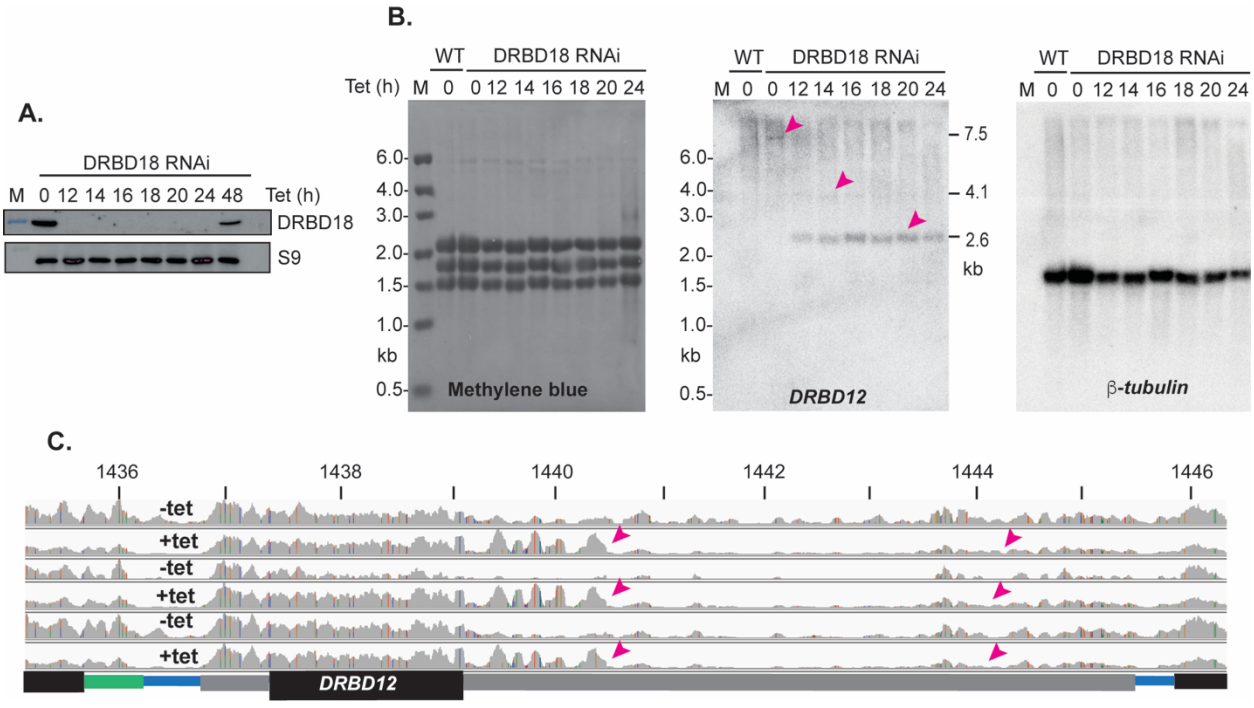


Figure 3.23: Effect of DRBD18 depletion on DRBD12 mRNA. (legend continues on the next page) **(A)** Western blots of samples collected after DRBD18 RNAi induction at different time points are shown. The sample without induction of DRBD18 RNAi (0) was used as control. DRBD18 was detected with anti-DRBD18 antibody and the ribosomal protein S9 is the loading control. **(B)** The samples used in (A) were also used for RNA extraction and loaded on agarose gel. The methylene blue-stained membrane is on the left, the membrane was then hybridized with DRBD12 probes in the middle and then with a beta-tubulin probe as a further control on the right. The arrows in pink represent the bands obtained for DRBD12 mRNA. **(C)** Integrated Genomics viewer was used to visualize the RNA-Seq results 12 hours with and without tetracycline-mediated induction

of DRBD18 RNAi. Three replicates are shown. The pink arrows show gain of reads on the DRBD12 3'-UTR after 12h of DRBD18 RNAi induction. The read densities in the middle part of the 3'-UTR are lower than the densities on the rest of the sequence, presumably because these sequences are not unique in the TREU927 reference genome. Colored lines indicate read mismatches with the TREU927 reference genome.

iii. DRBD18 mRNA binding does not correlate with RNAi effects

So far, the Northern blots together with transcriptomics data are strong evidence that DRBD18 depletion affects processing of the *RBP10* mRNA. It seems that DRBD18 binds and represses suboptimal processing sites present in the *RBP10* 3'-UTR. In absence of DRBD18, the alternative processing signals might be used by the splicing and polyadenylation complex generating alternatively polyadenylated mRNAs. If this is true, then DRBD18 might bind the *RBP10* mRNA.

To find mRNAs that are bound by DRBD18, I used a Lister 427 strain bloodstream forms in which one allele of DRBD18 was altered to encode a C-terminally TAP tagged DRBD18. To check whether the tag is fully functional, I attempted to replace the second allele with a gene encoding the blasticidin resistance marker (blasticidin S-deaminase). However, I failed to select such clones. Nevertheless, the C-terminally TAP-tagged DRBD18 was purified using IgG beads and the protein was released with TEV protease. Western blotting of samples collected at each step of the protein purification showed that about 5% of the protein was left in the unbound fraction (Figure 3.24 A), but successful release of the protein from the beads is clear since only the protein A remained in the beads after cleavage (last lane, Figure 3.24 A). The bound RNAs and the RNAs from the unbound fraction were purified and analyzed by sequencing. Three replicates were done under the same conditions. The RIP-Seq identifies 226 mRNAs that were at least 2-fold enriched (bound versus unbound) with read counts higher than 10 in all the three purifications (Figure 3.24 B). These included mRNAs encoding long RNA-binding proteins such as DRBD12, DRBD13, ZC3H48; protein kinases and mitochondrial proteins among others (Figure 3.24 B). The complete raw data set have been deposited at array express with accession number E-MTAB-10735. There was no correlation between DRBD18 binding to mRNA and the effects of DRBD18 depletion on transcript abundance (Figure 3.24 C). In addition, there was no correlation with mRNA length (data not shown). The *RBP10* mRNA was 1.6-, 2.4- and 5-fold enriched in all the three purifications. However, the mRNA encoding DRBD12 was 2.7-, 2.7- and 11.7-fold enriched. Among the 226 DRBD18 targets, 35 were at least 1.5-fold affected by DRBD18 depletion, 26 mRNAs were upregulated while 9 were downregulated (Figure 3.24 D).

The 3'-UTRs of these 35 mRNAs were analyzed for motifs search using the MEME motif suite and the size-matched mRNAs that were not bound by DRBD18 (<1x-fold) and not changed by DRBD18 depletion were used as background controls. Two 8-nt motifs that consist of polypyrimidine tracts interspaced with purine bases were enriched in the bound fraction

(Figure 3.24 E). This was a fascinating result, since splicing reactions require polypyrimidine tracts. The *RBP10* mRNA as well as the mRNAs that were bound and affected by DRBD18 depletion contain these motifs. However, these sequences are also found in many other 3'-UTRs and the binding of DRBD18 to the mRNAs whose processing were affected is questionable. Of the 154 mRNAs that showed differences in their 3'-UTRs and coding regions reads, only 11 were 2-fold enriched in the DRBD18 pulldown (Figure 3.24 F). It is possible that DRBD18 affects mRNA processing in the nucleus, presumably by binding to processing sites. This binding cannot be seen in the RIP-Seq. After export of the mRNA in the cytosol, the binding might be lost or just partial.

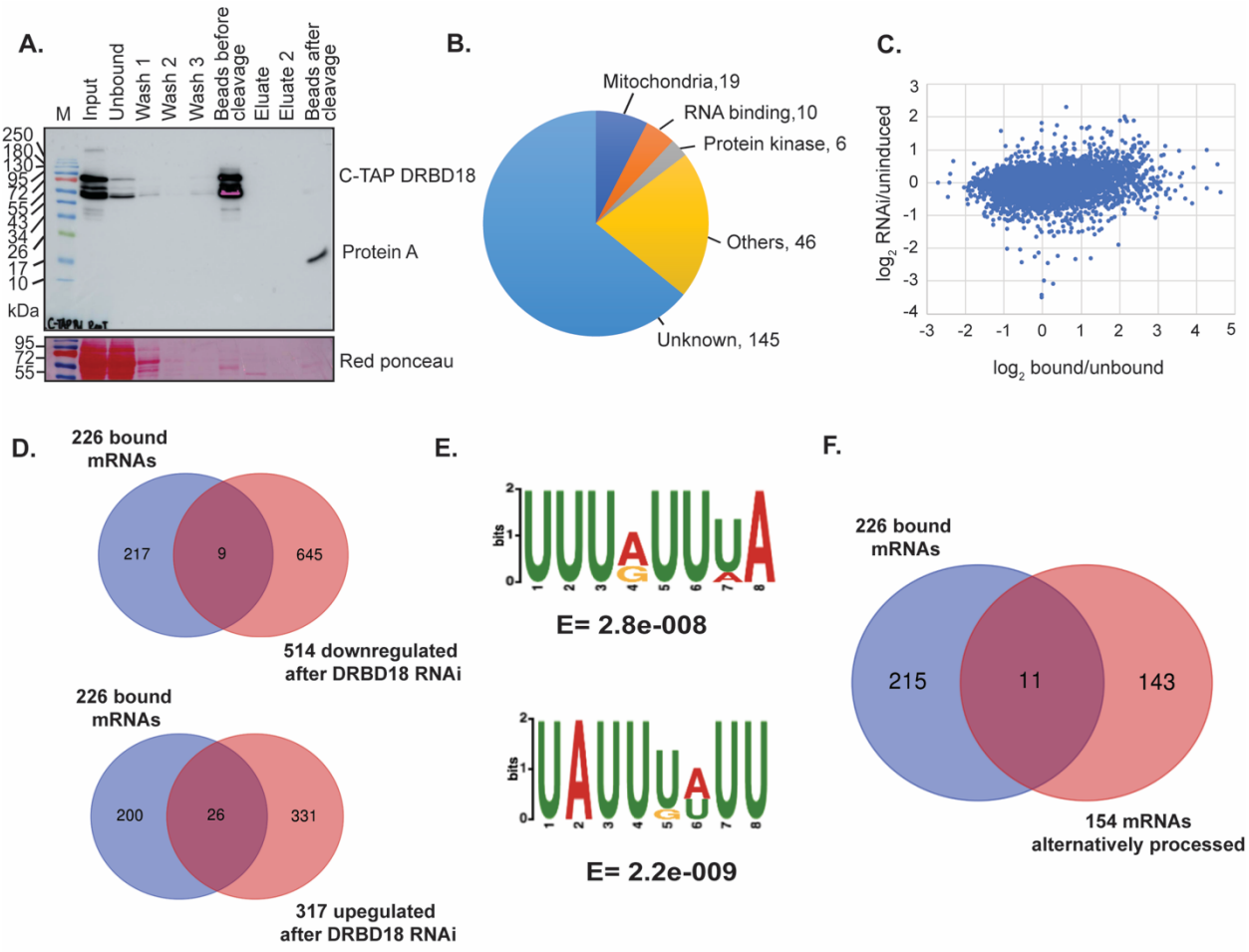


Figure 3.24: Interactions of DRBD18 with mRNAs. (legend continues on the next page) **(A)** C-TAP DRBD18 was purified using IgG beads and the tag was cleaved with TEV protease. Aliquots from each step of the purification was loaded and analyzed by Western blotting. M is the marker. Peroxidase-labelled anti-peroxidase (PAP) antibody was used for the detection of C-TAP DRBD18 and the red ponceau staining was used for loading control. No protein was detected after cleavage by the TEV protease. **(B)** Functional categories of proteins encoded by mRNAs bound by DRBD18. **(C)** The enrichment of DRBD18 binding to mRNAs (x-axis) was plotted against the RNAi effects (y-axis). **(D)** Venn-diagrams highlighting mRNAs that were 2-fold enriched in DRBD18 bound fraction and either 1.5-fold upregulated or downregulated after DRBD18 RNAi. **(E)** Motifs

enriched in the 3'-UTRs of mRNAs that were bound and affected by DRBD18 using the MEME motif search tool. **(F)** Venn diagram highlighting mRNAs that were bound by DRBD18 and whose changes in the 3'-UTR were not reflected in their coding sequences.

iv. The nuclear export factor MEX67 does not affect RBP10 mRNA processing

In procyclics, DRBD18 depletion caused partial accumulation of selected mRNAs in the nucleus (Mishra et al. 2021); therefore, it was important to check whether processing of the *RBP10* mRNA was altered by inhibition of nuclear export. The nuclear export factor MEX67 is required for general mRNA export and its depletion causes accumulation of mRNAs in the nucleus (Mishra et al. 2021). RNAi-mediated depletion of MEX67 was therefore carried out in bloodstream forms. After 24 hours of RNAi induction in two independent clones, there was clear inhibition of parasite growth (Figure 3.25 A). Northern blot hybridized with probe from the coding region of MEX67 showed reduction of the *MEX67* mRNA level (Figure 3.25 B) but *RBP10* mRNA processing was not affected (Figure 3.25 B). Overall, the alteration of *RBP10* mRNA processing after DRBD18 depletion is not just caused by retention of the mRNA in the nucleus. It seems to depend specifically on the effect of DRBD18 depletion.

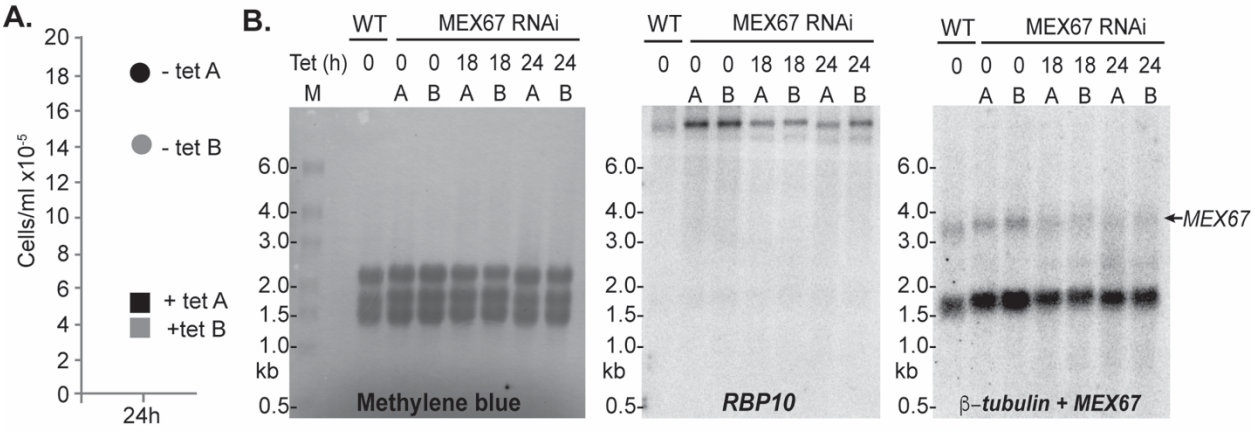


Figure 3.25: MEX67 depletion does not affect *RBP10* mRNA processing. (A) Cell densities of two independent cultures (A and B) after 24h induction of MEX67 RNAi. **(B)** RNAs from the two independent clones (A and B) were loaded on agarose gel and transferred to a positive membrane. The methylene-blue stained membrane is on the left. The blot hybridized to RBP10 probe is in middle and the blot hybridized to MEX67 and β-tubulin probes on the right.

v. DRBD18 is localized in the nucleus and in the cytosol

The above results suggested that DRBD18 affects processing of over 100 mRNAs in bloodstream forms but binding to the targets is somehow only partial, speculating that DRBD18 binds the splicing sites of its targets in the nucleus. However, subcellular fractionation of procyclic cells by Lott et al., 2014 showed that DRBD18 is localized only in the cytosol. On the other hand, C-terminally GFP-tagged DRBD18 was found in the nucleus (Dean et al. 2017). Since these results were contradictory, I decided to check the localization of DRBD18 in bloodstream forms. For that, extracts from Lister 427 strain wild-type cells and from cells depleted of DRBD18 were subjected to subcellular fractionation as described in (Biton et al. 2006). Western blotting of the nuclear and cytosol fractions showed that DRBD18 was associated with both compartments (Figure 3.26 A). This is consistent with a role in processing and nuclear export. As controls, the cytoplasmic protein RBP10 and the nuclear exoribonuclease D (XRND) were used and showed expected localizations (Figure 3.26 A and B).

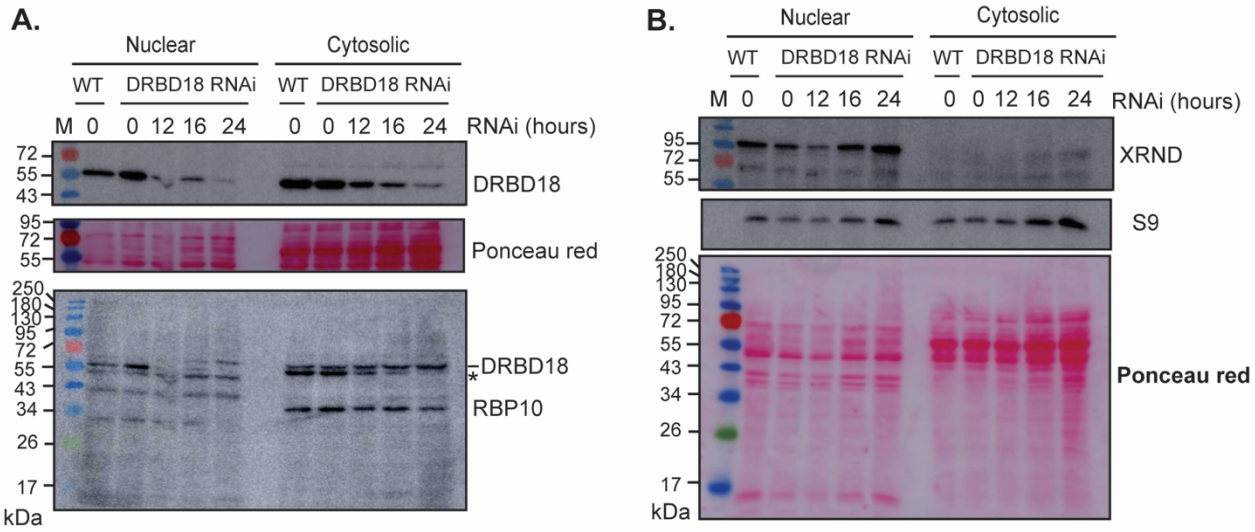


Figure 3.26: Subcellular fractionation in bloodstream forms reveals localization of DRBD18. (A) Western blots of Lister 427 bloodstream-form trypanosomes separated into nuclear and cytoplasmic fractions at various times after DRBD18 RNAi. DRBD18 was detected with anti-DRBD18 antibody. The protein is distributed in both fractions. The same blot was incubated with polyclonal antibody anti-RBP10, which detects RBP10 and a non-specific band (*). As expected, RBP10 is only associated with the cytoplasmic compartment. The DRBD18 signal is still visible and other non-specific bands. Ponceau staining is used for loading control. **(B)** Western blots of the same samples loaded in (A) are displayed for detection of a soluble nuclear protein, the exoribonuclease D (XRND) using the anti-XRND antibody. The ribosomal protein S9 (in both the nucleus and cytoplasm) is shown as a control and the ponceau staining served as a loading control.

vi. DRBD18 associates with proteins of the outer ring of the nuclear pore

To understand the mechanisms through which DRBD18 depletion alters processing of the *RBP10* mRNA as well as of the others, it was necessary to get a detailed picture of DRBD18 interactome in bloodstream forms trypanosomes. In procyclics, DRBD18 interacts directly with MTR2, and pulls down MEX67, which are both required for general nuclear export of mRNAs. Previous non-quantitative mass spectrometry results suggested that DRBD18 co-purifies some nuclear pore components and various RNA-binding proteins (Lott et al. 2015). To identify proteins that associate with DRBD18 in bloodstream forms, I made use of the Lister 427 strain encoding an *in-situ* C-terminal TAP-tagged DRBD18. The functionality of the tag is questionable since the other copy of DRBD18 (untagged) could not be deleted. Attempts to also obtain N-terminal TAP-tagged DRBD18 were not successful. Nevertheless, the C-TAP tagged protein was purified using IgG beads. The protein was released from the beads by cleavage with the TEV protease. Western blotting previously described in section 3.3.2 (Figure 3.14 C) showed successful purification of the protein. The eluate was analyzed by quantitative liquid chromatography-tandem mass spectrometry (LC/MS). The C-TAP tagged GFP and the N-TAP tagged ZC3H28 served as controls.

As mentioned earlier, proteins were identified as being associated with DRBD18 if they fulfill at least one of the two criteria, being 4-fold enriched in the DRBD18 with a p-value of less than 0.01 and being detected with at least one peptide in all the three purifications of DRBD18 but absent in the GFP and ZC3H28 controls. With these criteria, a total of 151 proteins were predominantly associated with DRBD18. It was interesting to see that all the components of the outer ring of the nuclear pore significantly co-purified with DRBD18 (Figure 3.27). They were not detected in either ZC3H28 or GFP control and their enrichment was highly significant (Figure 3.27 and 3.15 B). These included TbNup 109 outer ring, TbNup132 outer ring, TbNup140 Nup76 complex, TbNup149 Nup76 complex, TbNup152 outer ring, TbNup158 outer ring, TbNup41 outer ring, TbNup76 complex+149 which interacts with Mex67, TbNup82 outer ring, TbNup89 outer ring and TbSec13 outer ring nucleoporin. DRBD18 was also associated with MTR2, the Ran binding protein, a transportin-2 like protein, a putative Ran-GFP and the putative nuclear RNA helicase. The nuclear factor MEX67 was found in only two of the three DRBD8 purifications but in none of the controls. Oddly, multiple component of the mitochondrial editing machinery were specifically associated with DRBD18. I do not know what this specifically means. The components of the exon junction complex, Magoh and NTF2-domain protein also co-purified with DRBD18 although there were also found in ZC3H28 purifications but not in the GFP control. DRBD18 co-purifies several RNA-binding proteins, but only ZC3H30, RBP33 and ZPF2 were specifically associated with DRBD18.

The other RBPs including the two poly(A) binding proteins, DRBD2, DRBD3, ZC3H9, ZC3H34, ZC3H39, ZC3H40, UPF1, HNRNPFH, TRRM1, ALBA2 and ALBA3 were also associated with ZC3H28. The translation initiation factor EIF4E3/EIF4G4, some splicing factors, the RNA interference Argonaut protein AGO1 were also found in both DRBD18 and ZC3H28 purifications. However, members of the MKT1 complex (MKT1-LSM12-XAC1-PBP1) were not found in DRBD18 purifications, consistent with their role in mRNA stability and translation in the cytosol.

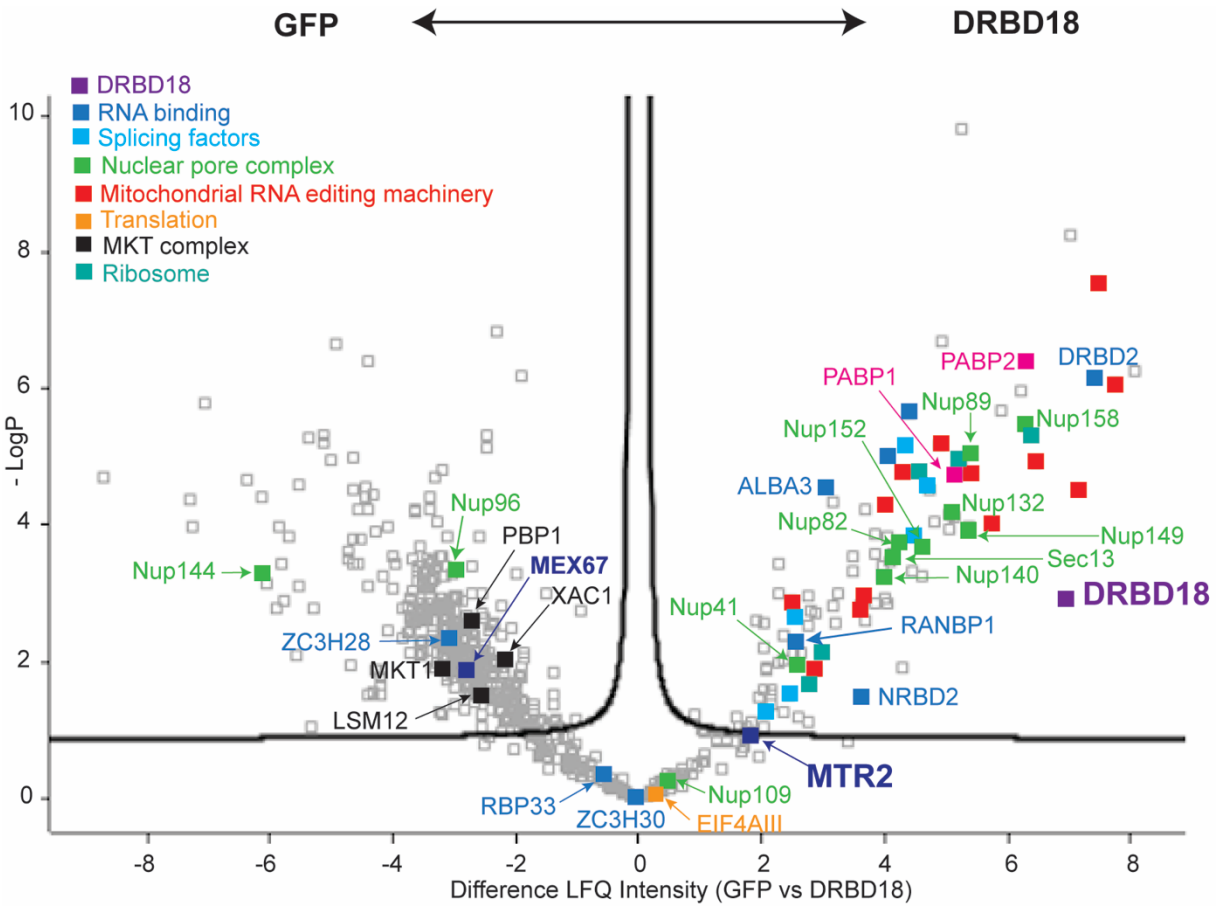


Figure 3.27: DRBD18 associates with proteins of the outer ring of the nuclear pore. Volcano plot showing proteins that are significantly enriched with C-TAP tagged DRBD18 relative to C-TAP tagged GFP. The y-axis is the -$\log_{10}$ of the p-value and the x-axis indicates $\log_2$ enrichment, generated by the Perseus algorithm. Proteins that are outside the curve are significantly enriched. Selected proteins are highlighted and colored. Purple is DRBD18, green indicates proteins of the nuclear pore complex, red highlights proteins of the mitochondrial RNA editing machinery, orange shows proteins involved in translation, black represents proteins of the MKT1 complex, in blue are the RNA binding proteins and in magenta the poly(A) binding proteins. Other proteins such as MEX67 and MTR2 are in dark blue.

Taken together, the mass spectrometry results reveal specific association of DRBD18 with proteins of the outer ring of the nuclear pore; this provides strong evidence of a potential role of DRBD18 in nuclear export of mRNAs.

vii. RBP10 long transcripts are trapped in the nucleus after DRBD18 depletion

Inhibition of nuclear export alone (Figure 3.25) was not sufficient to alter processing of *RBP10* mRNA. However, the mass spectrometry data strongly support the previous findings of Mishra et al. 2021 implicating DRBD18 in mRNA export. Therefore, it might be that DRBD18 binds and represses the suboptimal processing sites in the *RBP10* 3'-UTR, promoting efficient export of the long *RBP10* mRNA that is correctly polyadenylated. This means that, in the absence of DRBD18, the alternative signals are used by the splicing machinery generating alternatively polyadenylated *RBP10* mRNAs. If these hypotheses were true, then depletion of DRBD18 should trap long RBP10 transcripts in the nucleus.

To test how DRBD18 depletion affects *RBP10* mRNA localization, the single-molecule fluorescent *in-situ* hybridization (FISH) was performed. For that, four different probes were used to distinguish the different mRNAs (the long, the intermediates and the short): the first probe hybridizes to the coding region (pink), the second targets a region at about 2 kb after the stop coding (red), the third one hybridizes right in the middle of the 3'-UTR (red) and the last one, in green targets the terminal region of the 3'-UTR (Figure 3.28 A). In the first experiment, I use the combinations of probes 1, 2 and 4 and in the second experiment, probes 1, 3 and 4 were used (Figure 3.28 A). Details of the single *RBP10* mRNAs detected for each experiment and their relative proportions are tabulated in Table 3.5. In both experiments, about 5 to 24% of the mRNA species had signals from the coding region only (pink probe 1). These shortest species were mostly nuclear, they are probably precursors RNA that are undergoing transcription (Table 3.4). I also detected mRNAs that hybridized only to probes 1 (coding region, pink) and 2 (3'-UTR region, red). These mRNAs are thought to be 2-7 kb long and those hybridizing to 1 and 3 only were also detected and might be 5-7 kb long (Figure 3.28 A). Unexpectedly, there were about 5% of the mRNA species hybridizing to probes 1 and 4. This is very odd and might indicate that there was a small degree of failure rate for hybridization of probes 2 or 3. Also, the proportions of mRNAs hybridizing to either probes 2 or 3 only were very high, on average 17% (Table 3.4). Although these probes are very specific to *RBP10* 3'-UTR, they do contain many low complexity sequences, which might result in cross-hybridization. Very few mRNAs, on average 5%, hybridize to the green probe (probe 4) only. These might be products from the degradation process. To be certain that only specific signals are considered, for further discussion, I only focused on the long, intermediate and short mRNAs, which hybridize to the coding sequence probe and one more probe from the 3'-UTR regions.

When considering all mRNAs hybridizing to probe 1 of the coding region, the wild-type cells had 3-3.5 *RBP10* mRNAs per cell. Fadda et al. 2014 suggested that there is about 4 *RBP10* mRNAs per cell, consistent with my results.

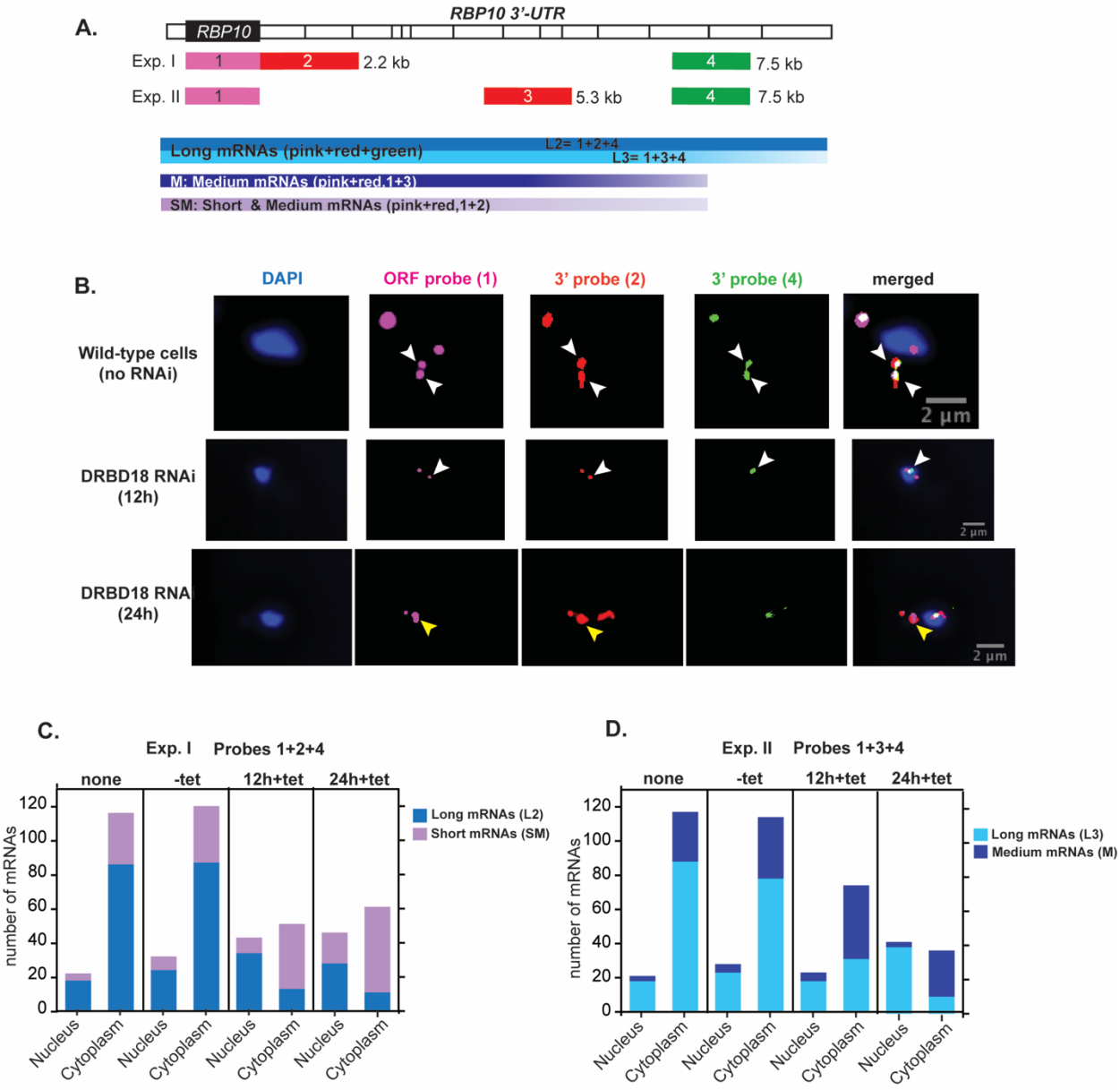


Figure 3.28: Long RBP10 mRNAs are trapped in the nucleus after DRBD18 depletion. (Legend continues on the next page). Single-molecule fluorescent in situ hybridization was used to detect individual *RBP10* mRNAs with and without DRBD18 depletion for 12 or 24 hours. For each slide used, 50 cells were analyzed. **(A)** Graphical map showing the *RBP10* mRNA and the positions of the four probes used. The different divisions on the *RBP10* 3'-UTR correspond to the ones used in Figure 3.1.5. The first experiment (Exp. I) used probes 1, 2

and 4. The second experiment (Exp. II) used probes 1, 3 and 4. The distances next to the bars (probes) indicate the position of the probe end on the mRNA. RNAs hybridizing to the different probes (1,2,3,4) could have various lengths and they are represented as dotted lines. **(B)** Examples images of *RBP10* mRNAs in a wild-type cell, in a cell depleted of DRBD18 after 12h and 24h, detected by single molecule fluorescent in situ hybridization using probes 1, 2 and 4. DNA was stained with DAPI. The white arrows indicate an mRNA that hybridizes to probes 1,2 and 4; yellow arrows point to mRNA hybridizing only to probes 1 and 2. **(C)** Total numbers of long, and short mRNAs obtained without and with DRBD18 depletion for 12 and 24 hours (Exp. I) are shown in the nucleus and in the cytosol. The data obtained are from 50 cells. None: are the wild-type cells with no RNAi plasmid; "-tet": Lister cells expressing DRBD18 RNAi but cultured without tetracycline; "12h +tet": Lister cells expressing DRBD18 RNAi after 12 hours of tetracycline induction; "24h +tet": cells treated after 24 hours of DRBDB18 RNAi induction. All slides were hybridized with the probes 1, 2 and 4 (pink, red and green probes). The long mRNAs are L2 (blue) and short are SM (purple). **(D)** Same as D for Exp. II. All slides were hybridized with the probes 1, 3 and 4 (pink, red and green probes). The long mRNAs are L3 (light blue) and medium are M (dark blue). Numbers of mRNAs hybridizing to only one probe, or red and green only, are in Table 3.5.

In the wild-type cell, on average 70% of *RBP10* mRNAs were long species (L3 and L2) of which 17% were in the nucleus. Of the 30% of the short and medium mRNAs (M and SM) detected in the wild-type cells, about 9-12% were in the nucleus (Figure 3.29 C and D, Table 3.4). Cells expressing *DRBD18* RNAi plasmid, but with no tetracycline induction, showed similar values as with wild-type, except that the proportion of mRNAs that were in the nucleus was slightly elevated (~23%), suggesting that the RNAi was very slightly leaky (Figure 3.28 C and D, Table 3.4).

After 12h after DRBD18 RNAi induction, the total number of full-length *RBP10* mRNAs was roughly halved and most of those that remained were retained in the nucleus (Table 3.4, Figure 3.28 C and D). After 24h of RNAi induction, the effect was even more pronounced with more than 70% of longer RNA species trapped in the nucleus (Figure 3.29 C and D, Table 3.4). Both the intermediate (medium) and shorter *RBP10* mRNAs were reproducibly mostly exported to the cytoplasm at all time-points. However, the total number of the shorter *RBP10* mRNA species after 24 hours of RNAi induction increased to about 64% (Figure 3.28 C, Table 3.4) while the intermediates were about 36% (Figure 3.28 D, Table 3.5). These results were very consistent with those from Northern blotting (Figure 3.20 C): the medium-length mRNAs had increased after 12h and decreased again at 24h (Figure 3.28 D), while the numbers of shorter mRNAs progressively increased (Figure 3.28 C).

Taken together, the FISH data support the hypothesis that DRBD18 is required for export of long *RBP10* mRNAs from the nucleus. However, after further processing, the shorter mRNAs can still be exported without DRBD18. From the reporter analysis, it was clear that shorter *RBP10* 3'-UTRs do not prevent translation of the mRNA (Figure 3.5 C), so the reduction in RBP10 protein level after DRBD18 RNAi might be caused by the decrease in total cytosolic *RBP10* mRNA level (Figure 3.28 C and D).

Table 3.4: Details FISH results

Total numbers of *RBP10* mRNA detected with different combinations of the probes used in the FISH experiments (probe 1 hybridizes to the *RBP10* coding sequence, probe 2 to approximately 2kb after the *RBP10* stop codon, probe 3 hybridizes to the middle region of the *RBP10* 3'-UTR and probe 4 hybridizes to the 3'-end of the *RBP10* 3'-UTR ~7.5 kb.

Probes detected and % of mRNAs based on the probes	**Localization**	**1+3+4 long and medium (Exp. II)**					**1+2+4 Long and short (Exp. I)**			
		WT	**0H**	**12H**	**24H**		**WT**	**0H**	**12H**	**24H**
PINK RED GREEN	Cytoplasm	88	78	31	10		86	87	13	11
	Nucleus	18	23	18	39		18	24	34	28
PINK RED (short or medium)	Cytoplasm	29	36	43	27		30	33	38	50
	Nucleus	3	5	5	3		4	8	9	18
PINK GREEN	Cytoplasm	4	4	3	1		6	1	5	4
	Nucleus	2	5	7	1		1	1		2
RED GREEN	Cytoplasm	13	5	2	7		8	8	5	1
	Nucleus			7	4				3	
RED only	Cytoplasm	27	54	36	26		22	34	26	34
	Nucleus			3	3				1	1
PINK only	Cytoplasm	5	4	11	3		1	7	5	5
	Nucleus	26	28	39	16		8	4	18	13
GREEN only	Cytoplasm	6		2	1		1		1	11
	Nucleus			5	2				3	3
Total long and short/medium		138	142	97	79		138	152	94	107
Average long and short/medium per cell		2.76	2.84	1.94	1.58		2.76	3.04	1.88	2.14
Total long mRNAs		106	101	49	49		104	111	47	39
Total short or medium mRNA		32	41	48	30		34	41	47	68
% long mRNA		77%	71%	51%	62%		75%	73%	50%	36%
% short or medium mRNA		23%	29%	49%	38%		25%	27%	50%	64%
% long and short/medium in nucleus		15%	20%	24%	53%		16%	21%	46%	43%
% long in nucleus		17%	23%	37%	80%		17%	22%	72%	72%
% short or medium mRNA in nucleus		9%	12%	10%	10%		12%	20%	19%	26%
Total RNAs		221	242	207	141		185	207	158	178
PINK RED GREEN %		48%	42%	24%	35%		56%	54%	30%	22%
PINK RED (short or medium) %		14%	17%	23%	21%		18%	20%	30%	38%
PINK GREEN %		3%	4%	5%	1%		4%	1%	3%	3%
RED GREEN %		6%	2%	1%	5%		4%	4%	3%	1%
RED only %		12%	22%	17%	18%		12%	16%	16%	19%
PINK only %		14%	13%	24%	13%		5%	5%	15%	10%
GREEN only %		3%	0%	3%	2%		1%	0%	3%	8%
Average of *RBP10* mRNA per cell		3.5	3.66	3.14	2		3.08	3.3	2.44	2.62

4. Discussion

RBP10 is a cytosolic protein that is exclusively expressed in bloodstream form trypanosomes (Wurst et al. 2012). Previous studies by Mugo and Clayton 2017 revealed the role of RBP10 in maintaining the bloodstream-form trypanosome life stage. The protein negatively regulates a significant number of procyclic-specific mRNAs by recognizing the motif $UA(U)_6$ present in their 3'-UTRs, targeting them for degradation and translation repression. In addition, bloodstream-form cells depleted of RBP10 grow as procyclic forms while the overexpression of RBP10 in procyclics promote their differentiation into bloodstream forms. During my PhD studies, I was interested to understand how the developmental expression of RBP10 is regulated in trypanosomes. I investigated the sequences as well as the proteins that are involved in the regulation.

4.1 Developmental regulation of RBP10 is mediated by the 3'-UTR of its mRNA

When examining RNAs from bloodstream and procyclic forms by Northern blotting, the levels of *RBP10* mRNAs were found to be developmentally regulated. This corroborates the published transcriptomics data (Fadda et al. 2014). Northern blots clearly also showed that the mRNA is approximately 8.5 kb long which gives the 3'-UTR the length of ~7.3 kb. Again, our results were in line with the RNA-Seq and the ribosome profiling data (Vasquez et al. 2014; Jensen et al. 2014) which suggest similar length. However, identification of the appropriate polyadenylation site of *RBP10* mRNA using 3'-RACE with oligo $d(T)_{20}$ was not evident, giving shorter size of the mRNA than expected. Since *RBP10* 3'-UTR contains several poly(A) tracts, it is possible that the oligos might have hybridized to the closest poly(A) tail. Our decay experiment revealed that *RBP10* mRNA is relatively stable in bloodstream forms as reported by Fadda et al. 2014. The mean half-life of *RBP10* mRNA is nearly 2 hours, quite similar to the result obtained from RNA-sequencing (Fadda et al. 2014).

But why does *RBP10* mRNA has such a long 3'-UTR? We consider that two reasons are possible: the long 3'-UTR might ensure low mRNA abundance and the second possibility is that it might contain several regulatory elements which are bound by various regulatory proteins to maintain a tight regulation. Even in trypanosomes, there is evidence that long mRNAs are generally less abundant than the short ones (Fadda et al. 2014). Indeed, there are approximately only 4 copies of *RBP10* mRNA per cell in bloodstream forms and less than 1 per cell in procyclics (Fadda et al. 2014), which reveal the low abundance of the mRNA. We speculated that like other mRNAs (Ridewood et al. 2017; Berberof et al. 1995), the 3'-UTR of the *RBP10* mRNA contains regulatory sequences responsible for the developmental expression.

We obtained evidence that the *RBP10* 3'-UTR mediates the regulation of RBP10 expression. By progressively testing different fragments of the *RBP10* 3'-UTR using the CAT reporter assays and the Northern blotting, we also found that different sequences in the *RBP10* 3'-UTR were independently promoting the stability and translation of the reporter mRNA in bloodstream forms, as well as its degradation and translation repression in procyclics.

By using Northern blotting for quantification of the mRNA levels, we also had the chance to measure the sizes of the different mRNAs. The expected sizes were measured from sequences with addition of 39-nt splice leader and 61-nt poly(A) tails. However, this calculation is not so adequate since poly(A) tails can also be extended to about 200 nt (Schwede et al. 2009). Nevertheless, in most cases, the sizes of the reporter mRNAs were as expected but in other cases, we obtained mRNAs that were shorter than expected. Since polyadenylation is coupled to *trans*-splicing of the downstream mRNA, this means that the polyadenylation site for these mRNAs was determined within their 3'-UTRs rather than within the site defined in the plasmid backbone (pHD 2164). The discrepancy in the sizes was not very high, ranging from 4 to 700-nt; making it possible to still draw conclusions from the results obtained.

The *RBP10* 3'-UTR fragments 1.1, 1.2, 1.4 and 2.2 mediated developmental regulation of the reporter levels, suggesting that both activators and suppressors of gene expression bind to these fragments (Figure 4.1). Fragments 1.3, 2.1, 2.2.4, 3.2 and 4 were able to suppress the reporter levels in procyclics with no increase in bloodstream forms, implicating their role only in mRNA degradation and translation suppression in procyclic forms. Lastly, fragments 2.2.3 and 3.1 increased the expression levels in bloodstream forms with no suppression in procyclics, meaning that only activators of gene expression bind to these regions (Figure 4.1).

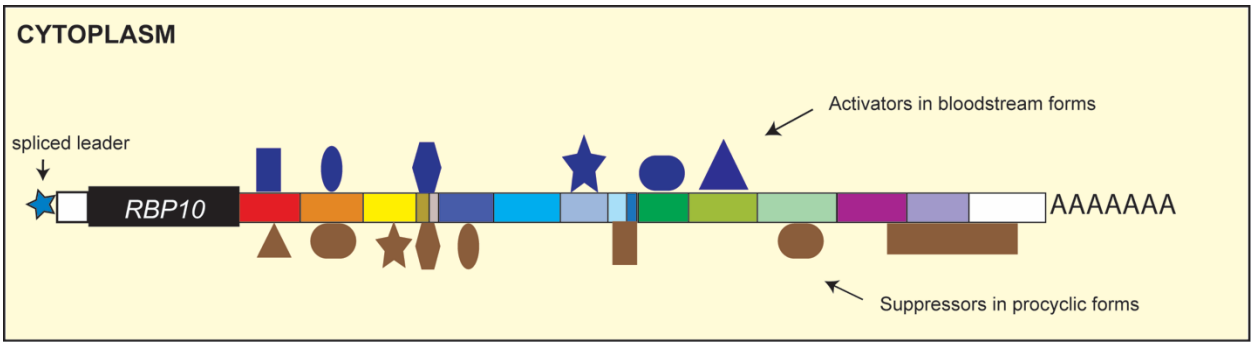


Figure 4.1: Proposed model of regulation by the *RBP10* 3'-UTR. The *RBP10* 3'-UTR contains several regulatory sequences, which are bound by proteins that stabilize the mRNA in bloodstream forms (activators, in dark blue) and destabilize the mRNA in procyclic forms (suppressors, in brown).

We speculated that a certain motif might be repeated throughout the *RBP10* 3'-UTR to enable the developmental regulation by the various fragments. By searching for motifs using the MEME tool, we did not find any *cis*-element that was repeated in the *RBP10* 3'-UTR. We also did not find any conserved motifs between *RBP10* mRNA and other co-regulated genes. It is therefore possible that several different motifs are involved in the regulation of RBP10 expression; and finding them through deletion and fragmentation of the 3'-UTR requires more time and work.

It is also likely that there are numerous *trans*-acting factors binding to these sequences and they independently ensure the stage-specific expression of RBP10 in a coordinated manner. Khong and Parker 2020 have estimated that in Opisthokont, mRNAs are bound by at least 4-18 proteins per kb. If this estimation are the same in trypanosomes, the *RBP10* 3'-UTR would be predicted to bind roughly 80 proteins. Finding these proteins responsible for RBP10 developmental regulation would require to focus on the individual regulatory fragments of the *RBP10* 3'-UTR that have been elucidated. More work is needed to narrow down the regulatory sequences to at least 30 to 50-nt. The shortest regulatory sequence (F1.4) from our study is 234-nt long. This is still a long sequence that would not be adequate for any RNA affinity purification study. However, deletion and fragmentation analysis of the 3'-UTR are also prone to confusing results, as it was the case for our study. Sometimes a deletion destroys a regulation that was seen before and sometimes it creates a regulation that was not previously there. This is due to possible disruption of the RNA secondary structure, resulting in unusual mRNP formation (Khong and Parker 2020). It is also possible that several proteins are indeed competing for binding to nearby regions within the 3'-UTR. This explains why some deletions create or disrupt the regulation.

Another approach for finding the regulatory sequences is to rely on known *cis*-elements shown in previous studies and investigate whether deletions of these sequences from the *RBP10* 3'-UTR will influence the regulation of the mRNA. Using this approach, we found that deletion of the $(AU)_{10}$ element from the *RBP10* 3'-UTR fragment 1.2 in bloodstream forms resulted in translation repression. AU-rich elements are common regulatory elements found in short-lived eukarytotic mRNAs (Barreau et al. 2005). They have long been shown to mediate degradation of the mRNAs (Chen and Shyu 1995) but also stabilization through binding of the HuR RNA-binding proteins (Katsanou et al. 2005). In trypanosomes, the expression of the human RNA-binding protein was also shown to increase the abundance of mRNAs containing AU-rich regulatory elements (Quijada et al. 2002). Although we do not know whether the effect of AU-rich element on *RBP10* mRNA is stage-specific, unpublished data from our lab showed evidence that the AU-rich element increases the abundance of the reporter mRNA levels bearing actin 3'-UTR (Igor Minia, unpublished). Droll et al. 2013 also showed that AUU-repeats were implicated in stability of the mRNA encoding the heat shock protein HSP70.

More work on this direction is ongoing in our lab. Lena Reichert, a bachelor student, has been investigating the role of the 3'-UTR in regulating the developmental expression of the isoform C of the phosphoglycerate kinase (PGKC). She found several regulatory sequences responsible for the developmental expression of the protein and we predicted that a CU-rich element is necessary for translation repression of the protein in procyclics. This sequence is also present in *RBP10* 3'-UTR fragment 1.4.2. We are currently investigating whether this element is sufficient to mediate translation suppression of the *RBP10* mRNA in procyclics. This will provide evidence of a *cis*-element shared between two bloodstream-form specific mRNAs and mediating their negative regulation.

Finding proteins that bind to the *RBP10* 3'-UTR was the second major goal of our study. Several methods are available to identify proteins that bind to a specific mRNA (Theil et al. 2019; Ramanathan et al. 2018). Using biotinylation of proximal proteins with inducible expression of the biotin ligase (Ramanathan et al. 2018), we could not validate the interaction between ZC3H20 and the *PSSA-2* 3'-UTR in procyclic-form trypanosomes. This is probably due to the low abundance of the *PSSA-2* mRNA. Because the mRNA encoding RBP10 is also not very abundant (Fadda et al. 2014), this method would not be appropriate for identifying the proteins binding to its 3'-UTR. Therefore, we decided to deplete some candidate proteins by RNAi and monitor their effect on *RBP10* mRNA and protein levels.

4.2 HNRNPFH does not regulate RBP10 expression in bloodstream forms

When HNRNPFH was depleted by RNAi in bloodstream-form trypanosomes, we did not see any effect on cell growth, although Alsford et al. 2011 revealed that the protein is essential for growth in this stage. Again, in contrast with our results, Gupta et al. 2013 silenced HNRNPFH in bloodstream forms by expressing a dsRNA from a T7 opposing promoting construct, and they showed that HNRNPFH was essential for growth. In our study, the protein was not completely depleted by RNAi, leaving 2-30% of protein; it is possible that the remaining amount of the protein is sufficient to mediate its function. The depletion of HNRNPFH by RNAi also did not affect *RBP10* mRNA and protein levels. This was consistent with the microarrays data obtained by Gupta et al. 2013 which shows that the *RBP10* mRNA was not significantly affected after depletion of HNRNPFH for 48 hours.

Nevertheless, we obtained a list of mRNAs that were preferentially bound by HNRNPFH using RIP-Seq. We found that HNRNPFH was associated with approximately 423 mRNAs containing a purine-rich motif in their 3'-UTRs. The mRNA encoding RBP10 was only 2-fold enriched in all the three purifications, which suggests that the binding by HNRNPFH was not very strong, as compared to mRNAs that were more than 10-fold enriched in the bound

fraction. The depletion of HNRNPFH by RNAi also did not affect a reporter mRNA bearing the region of *RBP10* 3'-UTR which contains the putative binding site of HNRNPFH. Therefore, we did not obtain any evidence that HNRNPFH could stabilize *RBP10* mRNA. Also, there is no evidence that HNRNPFH stabilizes its bound mRNAs in bloodstream forms, because from the 423 mRNA targets, only 32 were downregulated. Mass spectrometry analysis of HNRNPFH in bloodstream forms revealed that the protein is associated with the RNA-binding proteins, UBP1, ZC3H44 and DRBD2 (Melo do Nascimento, unpublished). These proteins are mainly found in the cytosol. UBP1 is a very abundant protein (Hartmann et al. 2007); ZC3H44 was identified in a tethering screen as a post-transcriptional activator of gene expression (Erben et al. 2014) and DRBD2 was implicated in translation control in *Trypanosoma cruzi* (Wippel et al. 2019). HNRNPFH was also associated with many ribosomal proteins. It is therefore possible that the protein is involved in stabilization and translation control of some target mRNAs, but such evidence for *RBP10* mRNA was not found.

4.3 ZC3H28 might stabilize long mRNAs through its interaction with MKT1-complex

ZC3H28 is an RNA-binding protein containing a single C3H1-type zinc finger domain with multiple histidine and glutamine-rich regions. The protein is conserved in Kinetoplastids and was reported to be essential in bloodstream forms, during differentiation and in early procyclic forms (Alsford et al. 2011). By depleting ZC3H28 using a stem-loop construct, bloodstream-form trypanosomes were killed after 48 hours, providing evidence that the protein is essential for normal growth of trypanosomes at this life stage. When ZC3H28 was tethered to a reporter mRNA, the protein increased both the mRNA abundance and the protein levels of the attached reporter. The increase in the protein level was however less than that of the mRNA. The discrepancy was very small; we think that it could be caused by the difference in the methods used for measurements. Such differences were also found in previous tethering results from our lab: tethering of MKT1 or LSM12 for example increases the reporter protein levels 3 times more than it increases the mRNA abundance (Singh et al. 2014); tethering of ZC3H11 increases the mRNA abundance of the attached reporter more than the protein levels (Droll et al. 2013). In regard of the tethering results, we speculated that ZC3H28 might stabilize its bound mRNAs. After purification of the TAP-tagged ZC3H28, our mass spectrometry revealed association of ZC3H28 with the proteins of the MKT1 complex (MKT1-PBP1-XAC1-LSM12), the translation initiation complex EIF4E4/EIF4G3, the poly(A) binding proteins (PABP1 and PABP2) and several ribosomal proteins. The MKT1 complex has been implicated in stabilization and translation activation of the target mRNAs in previous studies (Singh et al. 2014; Melo do Nascimento et al. 2020; Melo do Nascimento et al. 2021). It is believed that MKT1 is recruited to mRNA by sequence-specific RNA-binding proteins containing the interaction motif HNPY or a polyglutamine motif (Singh et al. 2014). The MKT1, then interacts with PBP1, which recruits XAC1, LSM12, and preferentially PABP2

(Melo do Nascimento et al. 2020). MKT1 also directs recruitment of the translation initiation complex EIF4E6/EIF4G5 (Melo do Nascimento et al. 2021).

ZC3H28 provides a new mechanism through which the MKT1 complex can be recruited to the mRNA. Using yeast two-hybrids and co-immunoprecipitation, we did not find any evidence that ZC3H28 directly interacts with MKT1 protein, bearing in mind that ZC3H28 does not contain the HNPY interaction motif. However, we found that ZC3H28 specifically interacts with PBP1. This is the first evidence that the MKT1 can be recruited to an mRNA via direct interaction between PBP1 and the RNA-binding protein (Figure 4.2). However, the motifs implicated in this interaction are not known. Follow-up studies could include yeast two-hybrids assays with truncated versions and mutated versions of the protein sequences to unravel the motifs involved in this interaction. In addition, the mass spectrometry data also reveals that the translation initiation complex EIF4E4/EIF4G3 and both PABPs were specifically associated with ZC3H28; unlike other RBPs that associate with the MKT1 complex, recruiting the translation initiation complex EIF4E6/EIF4G5 and preferentially only PABP2 (Melo do Nascimento et al. 2020; Melo do Nascimento et al. 2021). Having found that both PABPs associated with ZC3H28 is interesting because PBP1 is known to interact with both of them (Singh et al. 2014). In addition, Zoltner et al. 2018 showed that both PABPs were associated with the translation initiation complex EIF4E4/EIF4G3, which was found enriched in ZC3H28 pulldown. Therefore, ZC3H28 provides here new insights on recruitment of the MKT1 complex to mRNAs and evidence that other translation initiation complex together with many other proteins might play a role (Figure 4.2). The EIF4F/EIF4G complex are known to bind the mRNA cap and they play a critical role in translation initiation (Shirokikh and Preiss 2018). The complex EIF4E4/EIF4G3 is essential for normal growth of bloodstream forms; it might also be active in translation but its precise role is unknown (Freire et al. 2017).

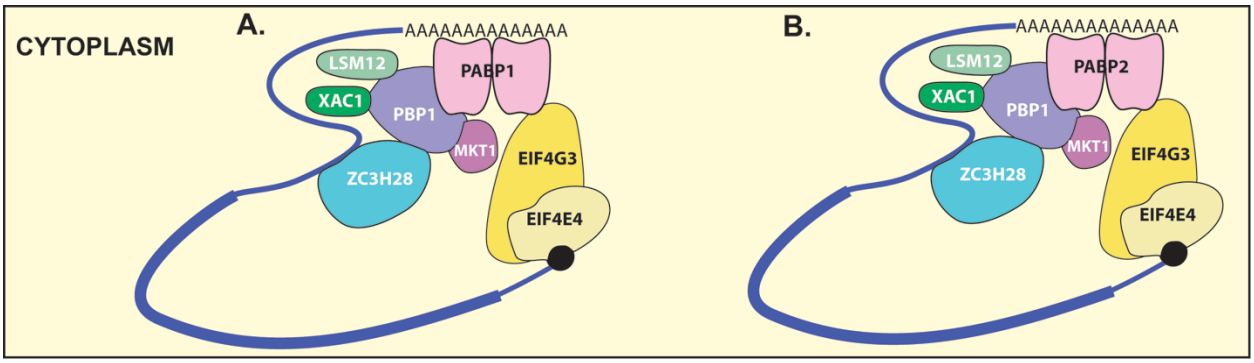


Figure 4.2: Proposed model for ZC3H28 function. ZC3H28 might stabilize its bound mRNAs by recruitment of the MKT1 complex (taken from (Melo do Nascimento et al. 2020)). The protein directly interacts with PBP1 which in turn recruits the MKT1-XAC1-LSM12 proteins together with either PABP1 (A) or PABP2 (B). The PABPs interact directly with the cap-binding protein EIF4G3, which associates with EIF4E4.

To obtain more information about ZC3H28 association with mRNAs, RIP-Sequencing was performed using the TAP-tagged ZC3H28. We found that ZC3H28 was associated with ~180 mRNAs, which tend to be relatively long in their 3'-UTRs. Among those mRNAs, we found those encoding for protein kinases and RNA-binding proteins. These mRNAs were reported to have longer-than-average 3'-UTRs (Clayton 2019). Erben et al. 2021 indicated that an RNA-binding protein that selects long UTRs might preferentially binds to AU-rich sequences. This was the case of ZC3H28: poly(AU) and polypurine motifs were enriched in the 3'-UTRs of ZC3H28 target mRNAs, although such sequences are also found in other 3'-UTRs. In addition, we did not find evidence that ZC3H28 binds to a specific short motif enriched in the 3'-UTRs of these mRNAs. This is probably because ZC3H28 has only one zinc finger domain, therefore high specificity is not expected. To understand whether ZC3H28 stabilizes its target mRNAs, we also look at the transcriptomes after depletion of ZC3H28 by RNAi. The greatest effect was seen after 14 hours of ZC3H28 RNAi induction where cell density was less than $1x10^6$ cells/ml. Above this cell density, depletion of ZC3H28 did not have any major effect on trypanosome transcriptome. We speculated that ZC3H28 might be essential in logarithmic phase when cells are growing optimally. Looking at the effect of ZC3H28 depletion after 14 hours together with the RIP-sequencing data, we realized that one third of ZC3H28 target mRNAs were decreased after depletion of ZC3H28, consistent with an mRNA stabilizing function of ZC3H28 as shown in the tethering assay and the mass spectrometry analysis. However, despite having a potential role in stabilizing and activating translation of its target mRNAs, ZC3H28 was associated with mRNAs that have very low ribosome occupancy in their coding regions. This suggests that these mRNAs might have either a low rate of translation initiation or rapid translation elongation.

In conclusion, we believe that ZC3H28 acts to preserve poorly translated mRNAs by stabilizing them.

4.4 DRBD18 affects processing and nuclear export of *RBP10* mRNA

DRBD18 is an RNA-binding protein with two RNA-recognition motifs. The protein was reported to be essential in procyclic forms where it is regulated by arginine methylation and it was found to associate with proteins of the nuclear pore complex (Lott et al. 2015). Depletion of DRBD18 in procyclics forms led to partial retention of mRNAs in the nucleus (Mishra et al. 2021).

When we depleted DRBD18 by RNAi using a stem-loop construct in bloodstream forms, we found that trypanosomes stop growing after 24 hours and they were rapidly killed after 48 hours. Downregulation of DRBD18 led to a decrease of RBP10 protein levels, and accumulation of shorter variants of *RBP10* mRNA with loss of the longer variants. When we analyzed the transcriptomes of bloodstream-form trypanosomes after depletion of DRBD18

by RNA sequencing, we confirm that *RBP10* mRNA processing was impaired, generating variants of alternatively polyadenylated mRNAs. Interestingly, this effect was also shown in over 100 mRNAs including the mRNAs encoding DRBD12 and ZC3H48.

We speculated that in normal cells, DRBD18 affects *RBP10* mRNA processing (and also other target mRNAs), by binding to suboptimal processing sites present in *RBP10* 3'-UTR and preventing binding of the *trans* spliceosome/polyadenylation complex to these sites. As a consequence, the *RBP10* mRNA is correctly processed resulting predominantly in the very long *RBP10* mRNA, which is then exported in the nucleus, through interaction of DRBD18 with the export machinery (Figure 4.3 A). This hypothesis is supported by several facts. First, *RBP10* 3'-UTR harbors several polypyrimidine tracts of different compositions and lengths as well as various poly(A) tracts (Supplementary information/ text 1). Second, the RIP-Seq performed with TAP-tagged DRBD18 identifies over 200 mRNAs that were preferentially bound by DRBD18. Analysis of mRNAs that were bound by DRBD18 and affected by DRBD18 showed enrichment of polypyrimidine tracts in their 3'-UTRs. This is consistent with the assumption that DRBD18 binds processing signals, although only few mRNAs whose processing was impaired by DRBD18 were bound by DRBD18. It is therefore possible that DRBD18 binds the processing signals in the nucleus and once in the cytoplasm, the binding is lost; hence it is not possible to see it. Third, DRBD18 is localized in the nucleus and in the cytoplasm, as shown in the subcellular fractionation. We also found specific association of DRBD18 with proteins of the outer ring of the nuclear pore, consistent with the role in nuclear export. Lastly, depletion of the nuclear export factor MEX67 did not affect *RBP10* mRNA processing, consistent with a specific role of DRBD18 in mediating both processing and export of the mRNA. It is also possible that when DRBD18 fails to bind *RBP10* mRNA under normal circumstances, some few mRNAs will be shorter than expected.

This was confirmed in the FISH experiment, which reveals presence of shorter variants in the wild-type cells. However, based on the analysis of the *RBP10* 3'-UTR, these shorter variants can still ensure developmental regulation of the *RBP10* mRNA, because the first 2kb of the *RBP10* 3'-UTR (Fragment 1) was sufficient to suppress translation of the mRNA in procyclics. After binding *RBP10* mRNA in the nucleus and mediating export of the *RBP10* mRNA, DRBD18 might still be partially bound to *RBP10* mRNA in the cytosol (Figure 4.3 A) because RIP-Seq showed a weak enrichment of *RBP10* mRNA in the bound fraction.

In cells depleted of DRBD18, we speculated that the suboptimal processing signals are bound by the splicing machinery, generating different variants of *RBP10* mRNAs. The long *RBP10* mRNA is made but it is retained in the nucleus due to the absence of DRBD18 and impairment of nuclear export (Figure 4.3 B). We also thought that the different variants of *RBP10* mRNAs are stuck in the nucleus and could not be exported in the cytosol. We found evidence looking at the transcriptomic data on integrative genome viewer that depletion of DRBD18 indeed

causes loss of the long variants. Using the FISH method, we could also clearly see accumulation of longer variants of *RBP10* mRNA in the nucleus while in normal cells, these variants were mostly in the cytosol. Also, the FISH experiments showed that the number of shorter variants was doubled after depletion of DRBD18 while the intermediate ones were slightly increased, consistent with results obtained in Northern blotting. Oddly, our FISH experiments also reveal that the shorter and intermediate variants of *RBP10* mRNA are exported in the cytosol, in contrast with what we thought. (Goos et al. 2019) showed that mRNA export can be initiated before polyadenylation has occurred, so it seems that the export happens from the 5'-end. In that case, perhaps the exon-junction complex is recognized by some proteins involved in the nuclear export, mediating export of the shorter variants in the cytosol (Figure 4.3 B)

But how does depletion of DRBD18 caused reduction in the protein levels if these variants can be exported in the cytosol? We think that although shorter variants are exported in the cytosol and might be translated, the general decrease in total *RBP10* mRNA levels after depletion of DRBD18 is responsible for decrease in protein levels. Despite all these findings, we still do not know why DRBD18 choose some mRNA precursors to influence their processing and their nuclear export. Specific sequences and secondary structure might play a role, but we are uncertain.

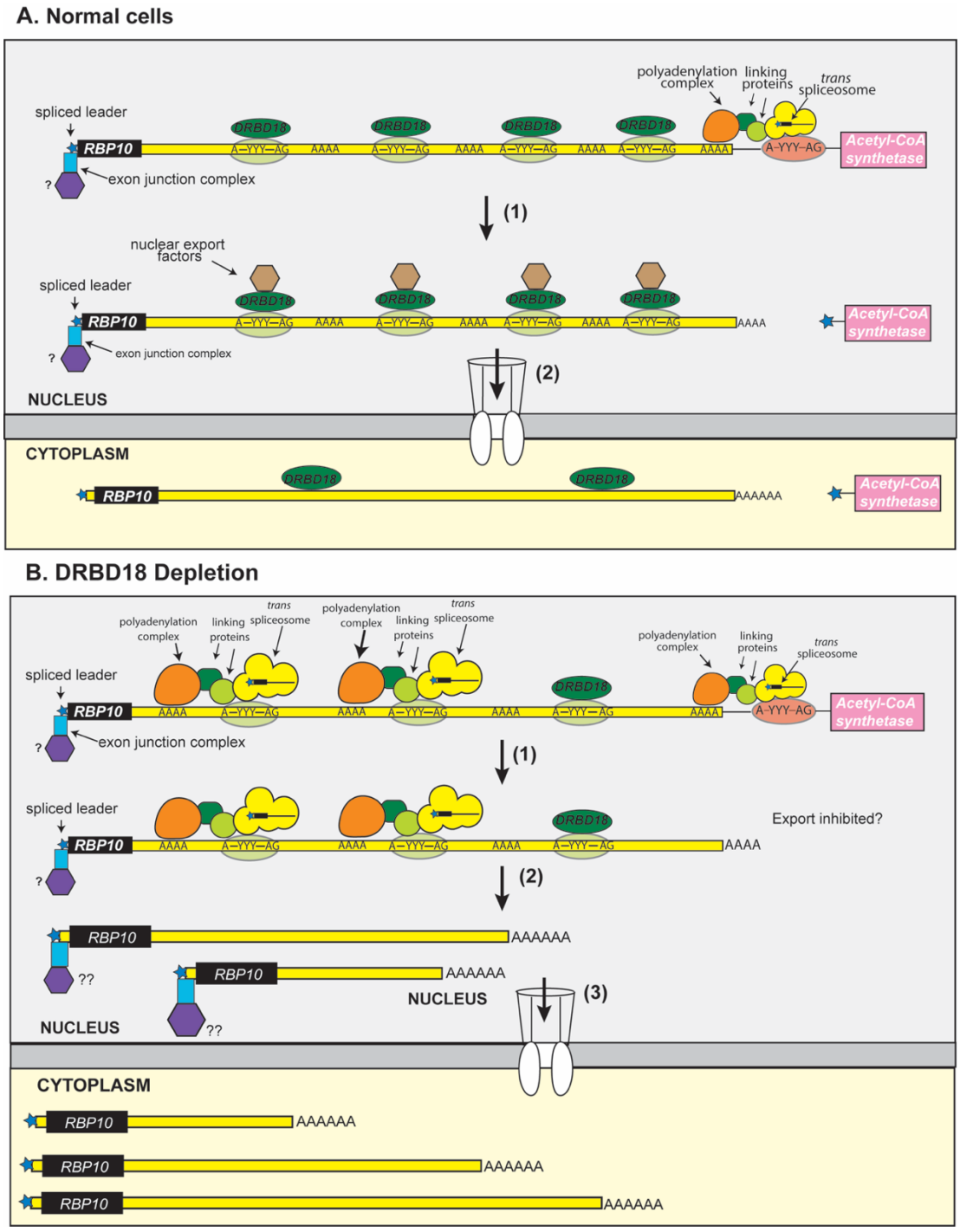


Figure 4.3: Proposed model of regulation of *RBP10* mRNA by DRBD18. (A) DRBD18 binds to suboptimal processing signals in *RBP10 3'-UTR* (1), and together with nuclear export proteins mediates processing and export of the long *RBP10 mRNA* variant. **(B)** In absence of DRBD18, the splicing machinery binds to the suboptimal signals in the *RBP10* 3'-UTR, delaying export of the long RBP10 mRNA (1), allowing re-processing of the mRNA and generation of alternatively polyadenylated mRNAs which are exported in the cytosol (3) probably through recognition of their 5'-cap. [Splicing machinery taken from (Clayton 2019).

5. References

Van Den Abbeele, J., Y. Claes, D. Van Bockstaele, D. Le Ray, and M. Coosemans. **1999**. Trypanosoma brucei spp. development in the tsetse fly: Characterization of the post-mesocyclic stages in the foregut and proboscis. Parasitology. 118: 469–478.

Ahmed, Y., A. Hagos, B. Merga, A. Van Soom, L. Duchateau, B. M. Goddeeris, and J. Govaere. **2018**. Trypanosoma equiperdum in the horse - A neglected threat? Vlaams Diergeneeskd. Tijdschr. 87: 66–75.

Alibu, V. P., L. Storm, S. Haile, C. Clayton, and D. Horn. **2005**. A doubly inducible system for RNA interference and rapid RNAi plasmid construction in Trypanosoma brucei. Mol. Biochem. Parasitol. 139: 75–82.

Alsford, S., D. J. Turner, S. O. Obado, A. Sanchez-flores, L. Glover, M. Berriman, C. Hertz-fowler, and D. Horn. **2011**. High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. Method. 21: 915–924.

Antwi, E. B., J. R. Haanstra, G. Ramasamy, B. Jensen, D. Droll, F. Rojas, I. Minia, M. Terrao, C. Mercé, K. Matthews, P. J. Myler, M. Parsons, and C. Clayton. **2016**. Integrative analysis of the Trypanosoma brucei gene expression cascade predicts differential regulation of mRNA processing and unusual control of ribosomal protein expression. BMC Genomics. 17: 1–16.

Archer, S. K., V. D. Luu, R. A. De Queiroz, S. Brems, and C. Clayton. **2009**. Trypanosoma brucei PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle. PLoS Pathog. 5: 1–15.

Auty, H. K., S. J. Torr, T. M, S. Jayaraman, and L. J. Morrison. **2015**. Cattle trypanosomosis : the diversity of trypanosomes and implications for disease epidemiology and control Trypanosome species of relevance to cattle. Rev. sci. tech. Off. int. Epiz. 4: 587–598.

Bailey, T. L., J. Johnson, C. E. Grant, and W. S. Noble. **2015**. The MEME Suite. Nucleic Acids Res. 43: W39–W49.

Bajak, K., K. Leiss, C. Clayton, and E. Erben. **2020**. A potential role for a novel ZC3H5 complex in regulating mRNA translation in Trypanosoma brucei. J. Biol. Chem. 295: 14291–14304.

Barreau, C., L. Paillard, and H. B. Osborne. **2005**. AU-rich elements and associated factors: Are there unifying principles? Nucleic Acids Res. 33: 7138–7150.

Begolo, D., I. M. Vincent, F. Giordani, I. Pöhner, M. J. Witty, T. G. Rowan, Z. Bengaly, K. Gillingwater, Y. Freund, R. C. Wade, M. P. Barrett, and C. Clayton. **2018**. The trypanocidal benzoxaborole AN7973 inhibits trypanosome mRNA processing, PLoS Pathog. 14(9): 1-33.

Beisang, D., and P. Bohjanen. **2009**. Perspectives on the ARE as it turns 25 years old Daniel. Wiley Interdiscip. Rev. RNA. 23: 1–7.

Benz, C., D. Nilsson, B. Andersson, C. Clayton, and D. L. Guilbride. **2005**. Messenger RNA processing sites in Trypanosoma brucei. Mol. Biochem. Parasitol. 143: 125–134.

Berberof, M., L. Vanhamme, P. Tebabi, A. Pays, D. Jefferies, S. Welburn, and E. Pays. **1995**. The 3'-terminal region of the mRNAs for VSG and procyclin can confer stage specificity to gene expression in Trypanosoma brucei. EMBO J. 14: 2925–2934.

Bevkal, S., A. Naguleswaran, R. Rehmann, M. Kaiser, M. Heller, and I. Roditi. **2021**. An Alba-domain protein required for proteome remodelling during trypanosome differentiation and host transition. PLoS Pathog. 17: 1–30.

Biton, M., M. Mandelboim, G. Arvatz, and S. Michaeli. **2006**. RNAi interference of XPO1 and Sm genes and their effect on the spliced leader RNA in Trypanosoma brucei. Mol. Biochem. Parasitol. 150: 132–143.

Boshart, M., and M. Engstler. **2004**. Cold shock and regulation of surface protein trafficking convey sensitization to inducers of stage differentiation in Trypanosoma brucei. Genes Dev. 18: 2798–2811.

Bradford, M. M. 1976. A Rapid and Sensitive Method for the Quantification of Microgram Quantities of Protein Utilizing the Priniple of Protein-Dye Binding. Anal. Biochem. 72: 248–254.

Brun, R., R. Don, R. T. Jacobs, M. Z. Wang, and M. P. Barrett. **2011**. Development of novel drugs for human African trypanosomiasis. Future Microbiol. 6: 677–691.

Bühlmann, M., P. Walrad, E. Rico, A. Ivens, P. Capewell, A. Naguleswaran, I. Roditi, and K. R. Matthews. **2015**. NMD3 regulates both mRNA and rRNA nuclear export in African trypanosomes via an XPOI-linked pathway. Nucleic Acids Res. 43: 4491–4504.

Burkard, G., C. M. Fragoso, and I. Roditi. **2007**. Highly efficient stable transformation of bloodstream forms of Trypanosoma brucei. Mol. Biochem. Parasitol. 153: 220–223.

Büscher, P. P., G. Cecchi, V. Jamonneau, and G. Priotto. **2017**. Human African trypanosomiasis. Lancet. 390: 2397–2409.

Caro, F., N. Bercovich, C. Atorrasagasti, M. J. Levin, and M. P. Vázquez. **2006**. Trypanosoma cruzi: Analysis of the complete PUF RNA-binding protein family. Exp. Parasitol. 113: 112–124.

Cassola, A. 2012. RNA Granules Living a Post-Transcriptional Life: the Trypanosome's Case. Curr. Chem. Biol. 5: 108–117.

Cassola, A., J. G. De Gaudenzi, and A. C. Frasch. **2007**. Recruitment of mRNAs to cytoplasmic ribonucleoprotein granules in trypanosomes. Mol. Microbiol. 65: 655–670.

Cayla, M., L. McDonald, P. Macgregor, and K. R. Matthews. **2020**. An atypical DYRK kinase connects quorum-sensing with posttranscriptional gene regulation in Trypanosoma brucei. Elife. 9: 1–52.

Cecchi, G., M. Paone, J. R. Franco, E. M. Fèvre, A. Diarra, J. A. Ruiz, R. C. Mattioli, and P. P. Simarro. **2009**. Towards the Atlas of human African trypanosomiasis. Int. J. Health Geogr. 8: 1–12.

Chen, C. Y. A., and A. Bin Shyu. **1995**. AU-rich elements: characterization and importance in mRNA degradation. Trends Biochem. Sci. 20: 465–470.

Clayton, C. 2019. Regulation of gene expression in trypanosomatids: Living with polycistronic transcription. Open Biol. 9:1-24.

Clayton, C. E. 2002. Life without transcriptional control? From fly to man and back again. EMBO J. 21: 1881–1888.

Clayton, C. E. 2014. Networks of gene expression regulation in Trypanosoma brucei. Mol. Biochem. Parasitol. 2: 96–106.

Clayton, C., and S. Michaeli. **2011**. 3′ Processing in Protists. Wiley Interdiscip. Rev. RNA. 2: 247–255.

Clayton, C., and M. Shapira. **2007**. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. Mol. Biochem. Parasitol. 156: 93–101.

Cléry, A., M. Blatter, and F. H. T. Allain. **2008**. RNA recognition motifs: boring? Not quite. Curr. Opin. Struct. Biol. 18: 290–298.

Czichos, J., C. Nonnengaesser, and P. Overath. **1986**. Trypanosoma brucei: cis-Aconitate and temperature reduction as triggers of synchronous transformation of bloodstream to procyclic trypomastigotes in vitro. Exp. Parasitol. 62: 283–291.

Das, A., R. Morales, M. Banday, S. Garcia, L. Hao, G. A. M. Cross, A. M. Estevez, and V. Bellofatto. **2012**. The essential polysome-associated RNA-binding protein RBP42 targets mRNAs involved in Trypanosoma brucei energy metabolism. Rna. 18: 1968–1983.

Dean, S., J. D. Sunter, and R. J. Wheeler. **2017**. TrypTag.org: A Trypanosome Genome-wide Protein Localisation Resource. Trends Parasitol. 33: 80–82.

Decker, C. J., and R. Parker. **2012**. P-Bodies and Stress Granules: Possible Roles in the Control of Translation and mRNA Degradation Carolyn. Cold Spring Harb. Lab. Press. 4: 1–16.

Deeks, E. D. 2019. Fexinidazole: First Global Approval. Drugs. 79: 215–220.

Dejung, M., I. Subota, F. Bucerius, G. Dindar, A. Freiwald, M. Engstler, M. Boshart, F. Butter, and C. J. Janzen. **2016**. Quantitative Proteomics Uncovers Novel Factors Involved in Developmental Differentiation of Trypanosoma brucei. PLoS Pathog. 12: 1–20.

Desquesnes, M., F. Biteau-Coroller, J. Bouyer, M. L. Dia, and L. Foil. **2009**. Development of a mathematical model for mechanical transmission of trypanosomes and other pathogens of cattle transmitted by tabanids. Int. J. Parasitol. 39: 333–346.

Desquesnes, M., and M. L. Dia. **2003**. Trypanosoma vivax: mechanical transmission in cattle by one of the most common African tabanids, Atylotus agrestis. Exp. Parasitol. 103: 35–43.

Desquesnes, M., P. Holzmuller, D.-H. Lai, A. Dargantes, Z.-R. Lun, and S. Jittaplapong. **2013**. Trypanosoma evansi and surra: a review and perspectives on origin, history, distribution, taxonomy, morphology, hosts, and pathogenic effects. Biomed Res. Int. 2013: 194176.

Dhalia, R., N. Marinsek, C. R. S. Reis, R. Katz, J. R. C. Muniz, N. Standart, M. Carrington, and O. P. de Melo Neto. **2006**. The two eIF4A helicases in Trypanosoma brucei are functionally distinct. Nucleic Acids Res. 34: 2495–2507.

Dostalova, A., S. Käser, M. Cristodero, and B. Schimanski. **2013**. The nuclear mRNA export receptor Mex67-Mtr2 of Trypanosoma brucei contains a unique and essential zinc finger motif. Mol. Microbiol. 88: 728–739.

Droll, D., S. Archer, K. Fenn, P. Delhi, K. Matthews, and C. Clayton. **2010**. The trypanosome Pumilio-domain protein PUF7 associates with a nuclear cyclophilin and is involved in ribosomal RNA maturation. FEBS Lett. 584: 1156–1162.

Droll, D., I. Minia, A. Fadda, A. Singh, M. Stewart, R. Queiroz, and C. Clayton. **2013**. Post-Transcriptional Regulation of the Trypanosome Heat Shock Response by a Zinc Finger Protein. PLoS Pathog. 9:1-17.

Engstler, M., T. Pfohl, S. Herminghaus, M. Boshart, G. Wiegertjes, N. Heddergott, and P. Overath. **2007**. Hydrodynamic Flow-Mediated Protein Sorting on the Cell Surface of Trypanosomes. Cell. 131: 505–515.

Erben, E., C. Chakraborty, and C. Clayton. **2013**. The CAF1-NOT complex of trypanosomes. Front. Genet. 4: 2–5.

Erben, E. D., A. Fadda, S. Lueong, J. D. Hoheisel, and C. Clayton. **2014**. A Genome-Wide Tethering Screen Reveals Novel Potential Post-Transcriptional Regulators in Trypanosoma brucei. PLoS Pathog. 100(3):457-471.

Erben, E., K. Leiss, B. Liu, D. I. Gil, C. Helbig, and C. Clayton. **2021**. Insights into the functions and RNA binding of Trypanosoma brucei ZC3H22, RBP9 and DRBD7. Parasitology. 148: 1186–1195.

Estévez, A. M. 2008. The RNA-binding protein TbDRBD3 regulates the stability of a specific subset of mRNAs in trypanosomes. Nucleic Acids Res. 36: 4573–4586.

Estévez, A. M., T. Kempf, and C. Clayton. **2001**. The exosome of Trypanosoma brucei. EMBO J. 20: 3831–3839.

Etheridge, R. D., D. M. Clemens, P. D. Gershon, and R. Aphasizhev. **2009**. Identification and characterization of nuclear non-canonical poly(A) polymerases from Trypanosoma brucei. Mol. Biochem. Parasitol. 164: 66–73.

Fadda, A., V. Färber, D. Droll, and C. Clayton. **2013**. The roles of 3′-exoribonucleases and the exosome in trypanosome mRNA degradation. Rna. 19: 937–947.

Fadda, A., M. Ryten, D. Droll, F. Rojas, V. Färber, J. R. Haanstra, C. Merce, B. M. Bakker, K. Matthews, and C. Clayton. **2014**. 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. 94: 307–326.

Falk, F., K. K. Marucha, and C. Clayton. **2021**. Mechanisms of translation repression by the EIF4E1-4EIP cap-binding complex of Trypanosoma brucei: potential roles of the NOT complex and a terminal uridylyl transferase. bioRxiv 10.1101/2021.05.19.444837.

Färber, V., E. Erben, S. Sharma, G. Stoecklin, and C. Clayton. **2013**. Trypanosome CNOT10 is essential for the integrity of the NOT deadenylase complex and for degradation of many mRNAs. Nucleic Acids Res. 41: 1211–1222.

Ferguson, M. A. J., and A. F. Williams. **1988**. CELL-SURFACE ANCHORING OF PROTEINS VIA GLYCOSYL-PHOSPHATIDYLINOSITOL STRUCTURES. Annu. Rev. Biochem. 57: 285–320.

Fevre, E. M., B. Wissmann, S. C. Welburn, and P. Lutumba. **2008**. The Burden of Human African Trypanosomiasis. 2 (12):1-7.

Field, M. C., and M. Carrington. **2009**. The trypanosome flagellar pocket. Nat. Rev. Microbiol. 7: 775–786.

Florini, F., A. Naguleswaran, W. H. Gharib, F. Bringaud, and I. Roditi. **2019**. Unexpected diversity in eukaryotic transcription revealed by the retrotransposon hotspot family of Trypanosoma brucei. Nucleic Acids Res. 47: 1725–1739.

Franco, J. R., G. Cecchi, G. Priotto, M. Paone, A. Diarra, L. Grout, P. P. Simarro, W. Zhao, and D. Argaw. **2018**. Monitoring the elimination of human African trypanosomiasis: Update to 2016. PLoS Negl. Trop. Dis. 12: 1–26.

Franco, J. R., P. P. Simarro, A. Diarra, and J. G. Jannin. **2014**. Epidemiology of human African trypanosomiasis. Clin. Epidemiol. 6: 257–275.

Freire, E. R., A. M. Malvezzi, A. A. Vashisht, J. Zuberek, E. A. Saada, G. Langousis, J. D. F. Nascimento, D. Moura, E. Darzynkiewicz, K. Hill, O. P. de Melo Neto, J. A. Wohlschlegel, N. R. Sturm, and D. A. Campbell. **2014**. Trypanosoma brucei translation initiation factor homolog EIF4E6 forms a tripartite cytosolic complex with EIF4G5 and a capping enzyme homolog. Eukaryot. Cell. 13: 896–908.

Freire, E. R., D. M. N. Moura, M. J. R. Bezerra, C. C. Xavier, M. C. Morais-Sobral, A. A. Vashisht, A. M. Rezende, J. A. Wohlschlegel, N. R. Sturm, O. P. de Melo Neto, and D. A. Campbell. **2018**. Trypanosoma brucei EIF4E2 cap-binding protein binds a homolog of the histone-mRNA stem-loop-binding protein. Curr. Genet. 64: 821–839.

Freire, E. R., N. R. Sturm, D. A. Campbell, and O. P. de Melo Neto. **2017**. The role of cytoplasmic mRNA cap-binding protein complexes in Trypanosoma brucei and other trypanosomatids. Pathogens. 6 (55):1-25.

Freire, E. R., A. A. Vashisht, A. M. Malvezzi, J. Zuberek, G. Langousis, E. A. Saada, J. D. F. Nascimento, J. Stepinski, E. Darzynkiewicz, K. Hill, O. P. De Melo Neto, J. A. Wohlschlegel, N. R. Sturm, and D. A. Campbell. **2014**. eIF4F-like complexes formed by cap-binding homolog TbEIF4E5 with TbEIF4G1 or TbEIF4G2 are implicated in post-transcriptional regulation in Trypanosoma brucei. Rna. 20: 1272–1286.

Fritz, M., J. Vanselow, N. Sauer, S. Lamer, C. Goos, T. N. Siegel, I. Subota, A. Schlosser, M. Carrington, and S. Kramer. **2015**. Novel insights into RNP granules by employing the trypanosome's microtubule skeleton as a molecular sieve. Nucleic Acids Res. 43: 8013–8032.

De Gaudenzi, J., A. C. Frasch, and C. Clayton. **2005**. RNA-binding domain proteins in kinetoplastids: A comparative analysis. Eukaryot. Cell. 4: 2106–2114.

Gibson, W., and M. Bailey. **2003**. The development of Trypanosoma brucei within the tsetse fly midgut observed using green fluorescent trypanosomes. Kinetoplastid Biol. Dis. 2: 1-13.

Gingras, A.-C., B. Raught, and N. Sonenberg. **1999**. eIF4 Initiation Factors: Effectors of mRNA Recruitment to Ribosomes and Regulators of Translation. Annu. Rev. Biochem. 68: 913–963.

Giordani, F., L. J. Morrison, T. G. Rowan, H. P. De Koning, and M. P. Barrett. **2016**. The animal trypanosomiases and their chemotherapy: A review. Parasitology. 143: 1862–1889.

Goos, C., M. Dejung, A. M. Wehman, E. M-Natus, J. Schmidt, J. Sunter, M. Engstler, F. Butter, and S. Kramer. **2019**. Trypanosomes can initiate nuclear export co-transcriptionally. Nucleic Acids Res. 47: 266–282.

Graham, S. V., B. Wymer, and J. D. Barry. **1998**. Activity of a Trypanosome Metacyclic Variant Surface Glycoprotein Gene Promoter Is Dependent upon Life Cycle Stage and Chromosomal Context. Mol. Cell. Biol. 18: 1137–1146.

Gupta, S. K., V. Chikne, D. Eliaz, I. D. Tkacz, I. Naboishchikov, S. Carmi, H. W. Ben-Asher, and S. Michaeli. **2014**. Two splicing factors carrying serine-arginine motifs, TSR1 and TSR1IP, regulate splicing, mRNA stability, and rRNA processing in Trypanosoma brucei. RNA Biol. 11.

Gupta, S. K., I. Kosti, G. Plaut, A. Pivko, I. D. Tkacz, S. Cohen-Chalamish, D. K. Biswas, C. Wachtel, H. Waldman Ben-Asher, S. Carmi, F. Glaser, Y. Mandel-Gutfreund, and S. Michaeli. **2013a**. 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. 41: 6577–6594.

Gupta, S. K., I. Kosti, G. Plaut, A. Pivko, I. D. Tkacz, S. Cohen-Chalamish, D. K. Biswas, C. Wachtel, H. Waldman Ben-Asher, S. Carmi, F. Glaser, Y. Mandel-Gutfreund, and S. Michaeli. **2013b**. 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. 41: 6577–6594.

Haile, S., A. M. Estévez, and C. Clayton. **2003**. A role for the exosome in the in vivo degradation of unstable mRNAs. Rna. 9: 1491–1501.

Hartmann, C., C. Benz, S. Brems, L. Ellis, V. D. Luu, M. Stewart, I. D'Orso, C. Busold, K. Fellenberg, A. C. C. Frasch, M. Carrington, J. Hoheisel, and C. E. Clayton. **2007**. Small trypanosome RNA-binding proteins TbUBP1 and TbUBP2 influence expression of F-box protein mRNAs in bloodstream trypanosomes. Eukaryot. Cell. 6: 1964–1978.

Hartmann, C., H. R. Hotz, M. McAndrew, and C. Clayton. **1998**. Effect of multiple downstream splice sites on polyadenylation in Trypanosoma brucei. Mol. Biochem. Parasitol. 93: 149–152.

Hehl, A., E. Vassella, R. Braun, and I. Roditi. **1994**. A conserved stem-loop structure in the 3' untranslated region of procyclin mRNAs regulates expression in Trypanosoma brucei. Proc. Natl. Acad. Sci. U. S. A. 91: 370–374.

Hendriks, E. F., and K. R. Matthews. **2005**. Disruption of the developmental programme of Trypanosoma brucei by genetic ablation of TbZFP1, a differentiation-enriched CCCH protein. Mol. Microbiol. 57: 706–716.

Hendriks, E. F., D. R. Robinson, M. Hinkins, and K. R. Matthews. **2001**. A novel CCCH protein which modulates differentiation of Trypanosoma brucei to its procyclic form. EMBO J. 20: 6700–6711.

Hinnebusch, A. G., and J. R. Lorsch. **2012**. The mechanism of eukaryotic translation initiation: New insights and challenges. Cold Spring Harb. Perspect. Biol. 4: 1–25.

Hirumi, H., and K. Hirumi. **1989**. Continuous Cultivation of Trypanosoma brucei Blood Stream Forms in a Medium Containing a Low Concentration of Serum Protein without Feeder Cell Layers Author (s): Hiroyuki Hirumi and Kazuko Hirumi Published by : Allen Press on behalf of The American Soc. J. Parasitol. 75: 985–989.

Hoek, M., T. Zanders, and G. A. M. Cross. **2002**. Trypanosoma brucei expression-site-associated-gene-8 protein interacts with a Pumilio family protein. Mol. Biochem. Parasitol. 120: 269–283.

Horn, D. 2014. Antigenic variation in African trypanosomes. Mol. Biochem. Parasitol. 195: 123–129.

Hotz, H. R., C. Hartmann, K. Huober, M. Hug, and C. Clayton. **1997**. Mechanisms of developmental regulation in Trypanosoma brucei: A polypyrimidine tract in the 3'-untranslated region of a surface protein mRNA affects RNA abundance and translation. Nucleic Acids Res. 25: 3017–3025.

Hug, M., H. R. Hotz, C. Hartmann, and C. Clayton. **1994**. Hierarchies of RNA-processing signals in a trypanosome surface antigen mRNA precursor. Mol. Cell. Biol. 14: 7428–7435.

Hummel, H. S., R. D. Gillespie, and J. Swindle. **2000**. Mutational analysis of 3' splice site selection during trans-splicing. J. Biol. Chem. 275: 35522–35531.

Isaac, C., M. Ciosi, A. Hamilton, K. M. Scullion, P. Dede, I. B. Igbinosa, O. P. G. Nmorsi, D. Masiga, and C. M. R. Turner. **2016**. Molecular identification of different trypanosome species and subspecies in tsetse flies of northern Nigeria. Parasites and Vectors. 9: 1–7.

Jacobs, R. T., B. Nare, S. A. Wring, M. D. Orr, D. Chen, J. M. Sligar, M. X. Jenks, R. A. Noe, T. S. Bowling, L. T. Mercer, C. Rewerts, E. Gaukel, J. Owens, R. Parham, R. Randolph, B. Beaudet, C. J. Bacchi, N. Yarlett, J. J. Plattner, Y. Freund, C. Ding, T. Akama, Y. K. Zhang, R. Brun, M. Kaiser, I. Scandale, and R. Don. **2011**. Scyx-7158, an orally-active benzoxaborole for the treatment of stage 2 human african trypanosomiasis. PLoS Negl. Trop. Dis. 5 (6):1-11.

Jaé, N., P. Wang, T. Gu, M. Hühn, Z. Palfi, H. Urlaub, and A. Bindereif. **2010**. Essential role of a trypanosome U4-specific sm core protein in small nuclear ribonucleoprotein assembly and splicing. Eukaryot. Cell. 9: 379–386.

Jensen, B. C., G. Ramasamy, E. J. R. Vasconcelos, N. T. Ingolia, P. J. Myler, and M. Parsons. **2014**. Extensive stage-regulation of translation revealed by ribosome profiling of trypanosoma brucei. BMC Genomics. 15: 1–21.

Jha, B. A., S. K. Archer, and C. E. Clayton. **2013**. The Trypanosome Pumilio Domain Protein PUF5. PLoS One. 8: 4–9.

Jha, B. A., A. Fadda, C. Merce, E. Mugo, D. Droll, and C. Clayton. **2014**. Depletion of the trypanosome pumilio domain protein PUF2 or of some other essential proteins causes transcriptome changes related to coding region length. Eukaryot. Cell. 13: 664–674.

Jones, T. W., and A. M. R. Dávila. **2001**. Trypanosoma vivax – out of Africa. Trends Parasitol. 17: 99–101.

Katsanou, V., O. Papadaki, S. Milatos, P. J. Blackshear, P. Anderson, G. Kollias, and D. L. Kontoyiannis. **2005**. HuR as a negative posttranscriptional modulator in inflammation. Mol. Cell. 19: 777–789.

Khong, A., and R. Parker. **2020**. The landscape of eukaryotic mRNPs. Rna. 26: 229–239.

Kocabas, A., T. Duarte, S. Kumar, and M. A. Hynes. **2015**. Widespread Differential Expression of Coding Region and 3' UTR Sequences in Neurons and Other Tissues. Neuron. 88: 1149–1156.

Kolev, N. G., J. B. Franklin, S. Carmi, H. Shi, S. Michaeli, and C. Tschudi. **2010**. The transcriptome of the human pathogen Trypanosoma brucei at single-nucleotide resolution. PLoS Pathog. 6: 1–15.

Kolev, N. G., K. Ramey-Butler, G. A. M. Cross, E. Ullu, and C. Tschudi. **2012**. Developmental progression to infectivity in Trypanosoma brucei triggered by an RNA-binding protein. Science (80-.). 338: 1352–1353.

Kramer, S. 2014. RNA in development: How ribonucleoprotein granules regulate the life cycles of pathogenic protozoa. Wiley Interdiscip. Rev. RNA. 5: 263–284.

Kramer, S. 2017a. The ApaH-like phosphatase TbALPH1 is the major mRNA decapping enzyme of trypanosomes. PLoS Pathog. 13: 1–20.

Kramer, S. 2017b. Simultaneous detection of mRNA transcription and decay intermediates by dual colour single mRNA FISH on subcellular resolution. Nucleic Acids Res. 45 (7):1-12.

Kramer, S., and M. Carrington. **2011**. Trans-acting proteins regulating mRNA maturation, stability and translation in trypanosomatids. Trends Parasitol. 27: 23–30.

Kramer, S., N. C. Kimblin, and M. Carrington. **2010**. Genome-wide in silico screen for CCCH-type zinc finger proteins of Trypanosoma brucei, Trypanosoma cruzi and Leishmania major. BMC Genomics. 11 (1):1-13.

Kramer, S., S. Piper, A. Estevez, and M. Carrington. **2016**. Polycistronic trypanosome mRNAs are a target for the exosome. Mol. Biochem. Parasitol. 205: 1–5.

Krüger, T., M. Hofweber, and S. Kramer. **2013**. SCD6 induces ribonucleoprotein granule formation in trypanosomes in a translation-independent manner, regulated by its Lsm and RGG domains. Mol. Biol. Cell. 24: 2098–2111.

Kuepfer, I., E. P. Hhary, M. Allan, A. Edielu, C. Burri, and J. A. Blum. **2011**. Clinical presentation of T.b. rhodesiense sleeping sickness in second stage patients from Tanzania and Uganda. PLoS Negl. Trop. Dis. 5 (3):1-9.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg. **2009**. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10 (3):1-14.

Langousis, G., and K. L. Hill. **2014**. Motility and more: the flagellum of Trypanosoma brucei. Nat. Rev. Microbiol. 12: 505–18.

Lee, J. H., T. N. Nguyen, B. Schimanski, and A. Günzl. **2007**. Spliced leader RNA gene transcription in Trypanosoma brucei requires transcription factor TFIIH. Eukaryot. Cell. 6: 641–649.

Leppek, K., and G. Stoecklin. **2014**. An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. Nucleic Acids Res. 42 (2):1-15.

Li, C. H., H. Irmer, D. Gudjonsdottir-Planck, S. Freese, H. Salm, S. Haile, A. M. Estévez, and C. Clayton. **2006**. Roles of a Trypanosoma brucei 5′/3′ exoribonuclease homolog in mRNA degradation. Rna. 12: 2171–2186.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. **2009**. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25: 2078–2079.

Li, H., and C. Tschudi. **2005**. Novel and Essential Subunits in the 300-Kilodalton Nuclear Cap Binding Complex of Trypanosoma brucei . Mol. Cell. Biol. 25: 2216–2226.

Ling, A. S., J. R. Trotter, and E. F. Hendriks. **2011**. A Zinc Finger Protein , TbZC3H20 , Stabilizes Two Developmentally Regulated mRNAs in Trypanosomes * □. J. Biol. Chem. 286: 20152–20162.

Liu, B., K. Kamanyi Marucha, and C. Clayton. **2019**. The zinc finger proteins ZC3H20 and ZC3H21 stabilise mRNAs encoding membrane proteins and mitochondrial proteins in insect-form Trypanosoma brucei. Mol. Microbiol. 430–451.

López-Estraño, C., C. Tschudi, and E. Ullu. **1998**. Exonic Sequences in the 5′ Untranslated Region of α-Tubulin mRNA Modulate trans Splicing in Trypanosoma brucei . Mol. Cell. Biol. 18: 4620–4628.

Lott, K., S. Mukhopadhyay, J. Li, J. Wang, J. Yao, Y. Sun, J. Qu, and L. K. Read. **2015**. Arginine methylation of DRBD18 differentially impacts its opposing effects on the trypanosome transcriptome. Nucleic Acids Res. 43: 5501–5523.

Love, M. I., W. Huber, and S. Anders. **2014**. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15: 1–21.

Lueong, S., C. Merce, B. Fischer, J. D. Hoheisel, and E. D. Erben. **2016**. Gene expression regulatory networks in Trypanosoma brucei: Insights into the role of the mRNA-binding proteome. Mol. Microbiol. 100: 457–471.

Lukes, J., D. L. Guilbride, J. Votýpka, A. Zíková, R. Benne, P. T. Englund, J. Voty, and A. Zíkova. **2002**. Kinetoplast DNA Network : Evolution of an Improbable Structure MINIREVIEW Kinetoplast DNA Network : Evolution of an Improbable Structure. Society. 1: 495–502.

Lunde, B. M., C. Moore, and G. Varani. **2007**. RNA-binding proteins: Modular design for efficient function. Nat. Rev. Mol. Cell Biol. 8: 479–490.

Luu, V. D., S. Brems, J. D. Hoheisel, R. Burchmore, D. L. Guilbride, and C. Clayton. **2006**. Functional analysis of Trypanosoma brucei PUF1. Mol. Biochem. Parasitol. 150: 340–349.

Mair, G., H. Shi, H. Li, A. Djikeng, H. O. Aviles, J. R. Bishop, F. H. Falcone, C. Gavrilescu, J. L. Montgomery, M. I. Santori, L. S. Stern, Z. Wang, E. Ullu, and C. Tschudi. **2000**. A new twist in trypanosome RNA metabolism: Cis-splicing of pre-mRNA. Rna. 6: 163–169.

Mani, J., A. Güttinger, B. Schimanski, M. Heller, A. Acosta-Serrano, P. Pescher, G. Späth, and I. Roditi. **2011**. Alba-domain proteins of trypanosoma brucei are cytoplasmic RNA-Binding proteins that interact with the translation machinery. PLoS One. 6 (7):1-16.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. Tech. Notes EMBnet.journal 17.1. 7: 10–12.

Marucha, K. K., and C. Clayton. **2020**. Roles of the Pumilio domain protein PUF3 in Trypanosoma brucei growth and differentiation. Parasitology. 147: 1171–1183.

Matthews, K. R. 2005. The developmental cell biology of Trypanosoma brucei. J. Cell Sci. 118: 283–290.

Matthews, K. R., C. Tschadi, and E. Ullu. **1994**. A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. Genes Dev. 8: 491–501.

Mayho, M., K. Fenn, P. Craddy, S. Crosthwaite, and K. Matthews. **2006**. Post-transcriptional control of nuclear-encoded cytochrome oxidase subunits in Trypanosoma brucei: Evidence for genome-wide conservation of life-cycle stage-specific regulatory elements. Nucleic Acids Res. 34: 5312–5324.

Mayr, C. 2017. Regulation by 3′-Untranslated Regions. Annu. Rev. Genet. 51: 171–194.

McCulloch, R., C. A. Cobbold, L. Figueiredo, A. Jackson, L. J. Morrison, M. R. Mugnier, N. Papavasiliou, A. Schnaufer, and K. Matthews. **2017**. Emerging challenges in understanding trypanosome antigenic variation. Emerg. Top. Life Sci. 1: 585–592.

Melo do Nascimento, L., F. Egler, K. Arnold, N. Papavasiliou, C. Clayton, and E. Erben. **2021**. Functional insights from a surface antigen mRNA-bound proteome. Elife. 10: 1–26.

Melo do Nascimento, L., M. Terrao, K. K. Marucha, B. Liu, F. Egler, and C. Clayton. **2020**. The RNA-associated proteins MKT1 and MKT1L form alternative PBP1-containing complexes in Trypanosoma brucei. J. Biol. Chem. 295: 10940–10955.

Meyer, A., H. R. Holt, F. Oumarou, K. Chilongo, W. Gilbert, A. Fauron, C. Mumba, and J. Guitian. **2018**. Integrated cost-benefit analysis of tsetse control and herd productivity to inform control programs for animal African trypanosomiasis. Parasites and Vectors. 11: 1–14.

Meyer, A., H. R. Holt, R. Selby, and J. Guitian. **2016**. Past and Ongoing Tsetse and Animal Trypanosomiasis Control Operations in Five African Countries: A Systematic Review. PLoS Negl. Trop. Dis. 10: 1–29.

Michaeli, S. 2011. Trans-splicing in trypanosomes: Machinery and its impact on the parasite transcriptome. Future Microbiol. 6: 459–474.

Minia, I., and C. Clayton. **2016**. Regulating a Post-Transcriptional Regulator: Protein Phosphorylation, Degradation and Translational Blockage in Control of the Trypanosome Stress-Response RNA-Binding Protein ZC3H11. PLoS Pathog. 12: 1–31.

Mishra, A., J. N. Kaur, D. I. McSkimming, E. Hegedűsová, A. P. Dubey, M. Ciganda, Z. Paris, and L. K. Read. **2021**. Selective nuclear export of mRNAs is promoted by DRBD18 in Trypanosoma brucei. Mol. Microbiol. 116: 827–840.

Monica, M., C. George, and P. Nina. **2015**. The in vivo dynamics of antigenic variation in Trypanosoma brucei. Science (80-.). 347: 1470–1473.

Moreira, D., P. López-García, and K. Vickerman. **2004**. An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: Proposal for a new classification of the class Kinetoplastea. Int. J. Syst. Evol. Microbiol. 54: 1861–1875.

Morriswood, B., K. Havlicek, L. Demmel, S. Yavuz, M. Sealey-Cardona, K. Vidilaseris, D. Anrather, J. Kostan, K. Djinović-Carugo, K. J. Roux, and G. Warren. **2013**. Novel bilobe components in Trypanosoma brucei identified using proximity-dependent biotinylation. Eukaryot. Cell. 12: 356–367.

Mugnier, M. R., C. E. Stebbins, and F. N. Papavasiliou. **2016**. Masters of Disguise: Antigenic Variation and the VSG Coat in Trypanosoma brucei. PLoS Pathog. 12: 1–6.

Mugo, E., and C. Clayton. **2017**. Expression of the RNA-binding protein RBP10 promotes the bloodstream-form differentiation state in Trypanosoma brucei. PLoS Pathog. 13: 1–37.

Muhich, M. L., and J. C. Boothroyd. **1988**. Polycistronic transcripts in trypanosomes and their accumulation during heat shock: evidence for a precursor role in mRNA synthesis. Mol. Cell. Biol. 8: 3837–3846.

Mukherjee, N., H. H. Wessels, S. Lebedeva, M. Sajek, M. Ghanbari, A. Garzia, A. Munteanu, D. Yusuf, T. Farazi, J. I. Hoell, K. M. Akat, A. Akalin, T. Tuschl, and U. Ohler. **2019**. Deciphering human ribonucleoprotein regulatory networks. Nucleic Acids Res. 47: 570–581.

Müller, L. S. M., R. O. Cosentino, K. U. Förstner, J. Guizetti, C. Wedel, N. Kaplan, C. J. Janzen, P. Arampatzi, J. Vogel, S. Steinbiss, T. D. Otto, A. E. Saliba, R. P. Sebra, and T. N. Siegel. **2018**. Genome organization and DNA accessibility control antigenic variation in trypanosomes. Nature. 563: 121–125.

Nagai, K., C. Oubridge, T. H. Jessen, J. Li, and P. R. Evans. **1990**. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. Nature. 348: 515–520.

Nagle, A. S., S. Khare, A. B. Kumar, F. Supek, A. Buchynskyy, C. J. N. Mathison, N. K. Chennamaneni, N. Pendem, F. S. Buckner, M. H. Gelb, and V. Molteni. **2014**. Recent developments in drug discovery for leishmaniasis and human african trypanosomiasis. Chem. Rev. 114: 11305–11347.

Niednery, A., F. T. Edelmanny, and D. Niessing. **2014**. Of social molecules: The interactive assembly of ASH1 mRNA-transport complexes in yeast. RNA Biol. 11: 998–1009.

Obado, S. O., M. Brillantes, K. Uryu, W. Zhang, N. E. Ketaren, B. T. Chait, M. C. Field, and M. P. Rout. **2016**. Interactome Mapping Reveals the Evolutionary History of the Nuclear Pore Complex. PLoS Biol. 14: 1–30.

Obado, S. O., M. C. Field, and M. P. Rout. **2017**. Comparative interactomics provides evidence for functional specialization of the nuclear pore complex. Nucleus. 8: 340–352.

Opperdoes, F. R., and P. Borst. **1977**. Localization of nine glycolytic enzymes in a microbody-like organelle in Trypanosoma brucei: the glycosome. FEBS Lett. 80: 360–364.

Overath, P., and M. Engstler. **2004**. Endocytosis, membrane recycling and sorting of GPI-anchored proteins: Trypanosoma brucei as a model system. Mol. Microbiol. 53: 735–744.

Palenchar, J. B., and V. Bellofatto. **2006**. Gene transcription in trypanosomes. Mol. Biochem. Parasitol. 146: 135–141.

Palfi, Z., N. Jaé, C. Preußer, K. H. Kaminska, J. M. Bujnicki, H. L. Ju, A. Günzl, C. Kambach, H. Urlaub, and A. Bindereif. **2009**. SMN-assisted assembly of snRNP-specific Sm cores in trypanosomes. Genes Dev. 23: 1650–1664.

Parker, R., and H. Song. **2004**. The enzymes and control of eukaryotic mRNA turnover. Nat. Struct. Mol. Biol. 11: 121–127.

Paterou, A., P. Walrad, P. Craddy, K. Fenn, and K. Matthews. **2006**. Identification and stage-specific association with the translational apparatus of TbZFP3, a CCCH protein that promotes trypanosome life-cycle development. J. Biol. Chem. 281: 39002–39013.

Patzelt, E., K. L. Perry, and N. Agabian. **1989**. Mapping of branch sites in trans-spliced pre-mRNAs of Trypanosoma brucei. Mol. Cell. Biol. 9: 4291–4297.

Perez-Riverol, Y., A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D. J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Pérez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, Ş. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A. F. Jarnuczak, T. Ternent, A. Brazma, and J. A. Vizcaíno. **2019**. The PRIDE database and related tools and resources in 2019: Improving support for quantification data. Nucleic Acids Res. 47: D442–D450.

Pinger, J., S. Chowdhury, and F. N. Papavasiliou. **2017**. Variant surface glycoprotein density defines an immune evasion threshold for African trypanosomes undergoing antigenic variation. Nat. Commun. 8 (1):1-8.

Queiroz, R., C. Benz, K. Fellenberg, J. D. Hoheisel, and C. Clayton. **2009**. Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons. BMC Genomics. 10: 1-19.

Quijada, L., C. Guerra-Giraldez, M. Drozdz, C. Hartmann, H. Irmer, C. Ben-Dov, M. Cristodero, M. Ding, and C. Clayton. **2002**. Expression of the human RNA-binding protein HuR in Trypanosoma brucei increases the abundance of mRNAs containing AU-rich regulatory elements. Nucleic Acids Res. 30: 4414–4424.

Quintana, J. F., M. Zoltner, and M. C. Field. **2021**. Evolving Differentiation in African Trypanosomes. Trends Parasitol. 37: 296–303.

Ramanathan, M., K. Majzoub, D. S. Rao, P. H. Neela, S. Mondal, J. G. Roth, H. Gai, J. R. Kovalski, T. D. Palmer, J. E. Carette, P. A. Khavari, and P. Alto. **2018**. RNA–protein interaction detection in living cells. Nat. Methods. 15: 207–212.

Rettig, J., Y. Wang, A. Schneider, and T. Ochsenreiter. **2012**. Dual targeting of isoleucyl-tRNA synthetase in Trypanosoma brucei is mediated through alternative trans-splicing. Nucleic Acids Res. 40: 1299–1306.

Reuner, B., E. Vassella, B. Yutzy, and M. Boshart. **1997**. Cell density triggers slender to stumpy differentiation of Trypanosoma brucei bloodstream forms in culture. Mol. Biochem. Parasitol. 90: 269–280.

Rico-Jiménez, M., G. Ceballos-Pérez, C. Gómez-Linãn, and A. M. Estévez. **2021**. An RNA-binding protein complex regulates the purine-dependent expression of a nucleobase transporter in trypanosomes. Nucleic Acids Res. 49: 3814–3825.

Ridewood, S., C. P. Ooi, B. Hall, A. Trenaman, N. V. Wand, G. Sioutas, I. Scherwitzl, and G. Rudenko. **2017**. The role of genomic location and flanking 3′UTR in the generation of functional levels of variant surface glycoprotein in Trypanosoma brucei. Mol. Microbiol. 106: 614–634.

Robinson, D. R., T. Sherwin, A. Ploubidou, E. H. Byard, and K. Gull. **1995**. Microtubule polarity and dynamics in the control of organelle positioning, segregation, and cytokinesis in the trypanosome cell cycle. J. Cell Biol. 128: 1163–1172.

Rodrigues, C. M. F., H. A. Garcia, A. C. Rodrigues, D. L. Pereira, C. L. Pereira, L. B. Viola, L. Neves, E. P. Camargo, W. Gibson, and M. M. G. Teixeira. **2020**. Expanding our knowledge on African trypanosomes of the subgenus Pycnomonas: A novel Trypanosoma suis-like in tsetse flies, livestock and wild ruminants sympatric with Trypanosoma suis in Mozambique. Infect. Genet. Evol. 78: 1-10.

Rotureau, B., and J. Van Den Abbeele. **2013**. Through the dark continent: African trypanosome development in the tsetse fly. Front. Cell. Infect. Microbiol. 4: 1–7.

Roux, K. J., D. I. Kim, M. Raida, and B. Burke. **2012**. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J. Cell Biol. 196: 801–810.

SANOFI. **2021**. US FDA approves fexinidazole as the first all-oral treatment for sleeping sickness. Media Updat. 1–3.

Schumann Burkard, G., S. Käser, P. R. de Araújo, B. Schimanski, A. Naguleswaran, S. Knüsel, M. Heller, and I. Roditi. **2013**. Nucleolar proteins regulate stage-specific gene expression and ribosomal RNA maturation in Trypanosoma brucei. Mol. Microbiol. 88: 827–840.

Schürch, N., A. Furger, U. Kurath, and I. Roditi. **1997**. Contributions of the procyclin 3' untranslated region and coding region to the regulation of expression in bloodstream forms of Trypanosoma brucei. Mol. Biochem. Parasitol. 89: 109–121.

Schwede, A., L. Ellis, J. Luther, M. Carrington, G. Stoecklin, and C. Clayton. **2008**. A role for Caf1 in mRNA deadenylation and decay in trypanosomes and human cells. Nucleic Acids Res. 36: 3374–3388.

Schwede, A., T. Manful, B. A. Jha, C. Helbig, N. Bercovich, M. Stewart, and C. Clayton. **2009**. The role of deadenylation in the degradation of unstable mRNAs in trypanosomes. Nucleic Acids Res. 37: 5511–5528.

Serpeloni, M., C. B. Moraes, J. R. C. MunizJoã, M. C. M. Motta, A. S. P. Ramos, R. L. Kessler, A. H. Inoue, W. D. daRocha, S. F. Yamada-Ogatta, S. P. Fragoso, S. Goldenberg, L. H. Freitas-Junior, and A. R. ÁvilaAndré. **2011**. An essential nuclear protein in trypanosomes is a component of mrna transcription/export pathway. PLoS One. 6 (6):1-14.

Sharma, R., L. Peacock, E. Gluenz, K. Gull, W. Gibson, and M. Carrington. **2008**. Asymmetric Cell Division as a Route to Reduction in Cell Length and Change in Cell Morphology in Trypanosomes. Protist. 159: 137–151.

Shi, H., N. Haven, K. Butler, N. Haven, C. Tschudi, and N. Haven. **2019**. 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. 6: 50–56.

Shi, H., C. Tschudi, and E. Ullu. **2007**. Depletion of newly synthesized Argonaute1 impairs the RNAi response in Trypanosoma brucei. Rna. 13: 1132–1139.

Shirokikh, N. E., and T. Preiss. **2018**. Translation initiation by cap-dependent ribosome recruitment: Recent insights and open questions. Wiley Interdiscip. Rev. RNA. 9: 1–45.

Siegel, T. N., D. R. Hekstra, L. E. Kemp, L. M. Figueiredo, J. E. Lowell, D. Fenyo, X. Wang, S. Dewell, and G. A. M. Cross. **2009**. Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei. Genes Dev. 23: 1063–1076.

Siegel, T. N., D. R. Hekstra, X. Wang, S. Dewell, and G. A. M. Cross. **2010**. Genome-wide analysis of mRNA abundance in two life-cycle stages of Trypanosoma brucei and identification of splicing and polyadenylation sites. Nucleic Acids Res. 38: 4946–4957.

Siegel, T. N., K. S. W. Tan, and G. A. M. Cross. **2005**. Systematic Study of Sequence Motifs for RNA trans Splicing in Trypanosoma brucei. Mol. Cell. Biol. 25: 9586–9594.

Silvester, E., A. Ivens, and K. R. Matthews. **2018**. A gene expression comparison of Trypanosoma brucei and Trypanosoma congolense in the bloodstream of the mammalian host reveals species-specific adaptations to density-dependent development. PLoS Negl. Trop. Dis. 12: 1–25.

Simarro, P. P., G. Cecchi, M. Paone, J. R. Franco, A. Diarra, J. A. Ruiz, E. M. Fèvre, F. Courtin, R. C. Mattioli, and J. G. Jannin. **2010**. The Atlas of human African trypanosomiasis: A contribution to global mapping of neglected tropical diseases. Int. J. Health Geogr. 9: 1–18.

Simpson, A. G. B., J. R. Stevens, and J. Lukeš. **2006**. The evolution and diversity of kinetoplastid flagellates. Trends Parasitol. 22: 168–174.

Simpson, L., N. Neckelmann, V. F. de la Cruz, A. M. Simpson, J. E. Feagin, D. P. Jasmer, and J. E. Stuart. **1987**. Comparison of the maxicircle (mitochondrial) genomes of Leishmania tarentolae and Trypanosoma brucei at the level of nucleotide sequence. J. Biol. Chem. 262: 6182–6196.

Singh, A., I. Minia, D. Droll, A. Fadda, C. Clayton, and E. Erben. **2014**. Trypanosome MKT1 and the RNA-binding protein ZC3H11: Interactions and potential roles in post-transcriptional regulatory networks. Nucleic Acids Res. 42: 4652–4668.

Smale, S. T. 2010. Chloramphenicol acetyltransferase assay. Cold Spring Harb. Protoc. 5: 1–5.

Sonenberg, N., and A. G. Hinnebusch. **2009**. Regulation of translation initiation in eukaryotes. Cell. 136: 731–745.

Stern, M. Z., S. K. Gupta, M. Salmon-Divon, T. Haham, O. Barda, S. Levi, C. Wachtel, T. W. Nilsen, and S. Michaeli. **2009**. Multiple roles for polypyrimidine tract binding (PTB) proteins in trypanosome RNA metabolism. Rna. 15: 648–665.

Stockdale, C., M. R. Swiderski, J. D. Barry, and R. McCulloch. **2008**. Antigenic variation in Trypanosoma brucei: Joining the DOTs. PLoS Biol. 6: 1386–1391.

Stuart, K. 1983. Kinetoplast DNA, mitochondria DNA with a difference. Mol. Biochem. Parasitol. 9: 93–104.

Stuart, K., R. Brun, S. Croft, A. Fairlamb, R. E. Gürtler, J. McKerrow, S. Reed, and R. Tarleton. **2008**. Kinetoplastids: Related protozoan pathogens, different diseases. J. Clin. Invest. 118: 1301–1310.

Subota, I., B. Rotureau, T. Blisnick, S. Ngwabyt, M. Durand-Dubief, M. Engstler, and P. Bastin. **2011**. ALBA proteins are stage regulated during trypanosome development in the tsetse fly and participate in differentiation. Mol. Biol. Cell. 22: 4205–4219.

Sutherland, C. S., C. M. Stone, P. Steinmann, M. Tanner, and F. Tediosi. **2017**. Seeing beyond 2020: an economic evaluation of contemporary and emerging strategies for elimination of Trypanosoma brucei gambiense. Lancet Glob. Heal. 5: e69–e79.

Tarral, A., S. Blesson, O. V. Mordt, E. Torreele, D. Sassella, M. A. Bray, L. Hovsepian, E. Evène, V. Gualano, M. Felices, and N. Strub-Wourgaft. **2014**. Determination of an optimal dosing regimen for fexinidazole, a novel oral drug for the treatment of human African trypanosomiasis: First-in-human studies. Clin. Pharmacokinet. 53: 565–580.

Telleria, E. L., J. B. Benoit, X. Zhao, A. F. Savage, S. Regmi, T. L. A. e Silva, M. O'Neill, and S. Aksoy. **2014**. Insights into the Trypanosome-Host Interactions Revealed through Transcriptomic Analysis of Parasitized Tsetse Fly Salivary Glands. PLoS Negl. Trop. Dis. 8 (4):1-16.

Terrao, M., K. K. Marucha, E. Mugo, D. Droll, I. Minia, F. Egler, J. Braun, and C. Clayton. **2018**. The suppressive cap-binding complex factor 4EIP is required for normal differentiation. Nucleic Acids Res. 46: 8993–9010.

Tetley, L., C. M. R. Turner, J. D. Barry, J. S. Crowe, and K. Vickerman. **1987**. Onset of expression of the variant surface glycoproteins of Trypanosoma brucei in the tsetse fly studied using immunoelectron microscopy. J. Cell Sci. 87: 363–372.

Theil, K., K. Imami, and N. Rajewsky. **2019**. Identification of proteins and miRNAs that specifically bind an mRNA in vivo. Nat. Commun. 10: 1–13.

Trenaman, A., L. Glover, S. Hutchinson, and D. Horn. **2019**. A post-transcriptional respiratome regulon in trypanosomes. Nucleic Acids Res. 47: 7063–7077.

Tshitenge, T. B., and C. Clayton. **2021**. Interactions of the Trypanosoma brucei brucei zinc-finger-domain protein ZC3H28. bioRxiv (preprint). 10.1101/2021.08.09.455650.

Turner, C. M. R., K. Vickerman, J. D. Barry, and I. Maudlin. **1988**. An estimate of the size of the metacyclic variable antigen repertoire of Trypanosoma brucei rhodesiense. Parasitology. 97: 269–276.

Tyanova, S., T. Temu, P. Sinitcyn, A. Carlson, M. Y. Hein, T. Geiger, M. Mann, and J. Cox. **2016**. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat. Methods. 13: 731–740.

Ullu, E., K. R. Matthews, and C. Tschudi. **1993**. Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts. Mol. Cell. Biol. 13: 720–725.

Urbaniak, M. D., M. L. S. Guther, and M. A. J. Ferguson. **2012**. Comparative SILAC proteomic analysis of trypanosoma brucei bloodstream and procyclic lifecycle stages. PLoS One. 7.

Urwyler, S., E. Vassella, J. Van Den Abbeele, C. K. Renggli, P. Blundell, J. D. Barry, and I. Roditi. **2005**. Expression of procyclin mRNAs during cyclical transmission of Trypanosoma brucei. PLoS Pathog. 1: 0225–0231.

Utter, C. J., S. A. Garcia, J. Milone, and V. Bellofatto. **2011**. Poly(A)-specific ribonuclease (PARN-1) function in stage-specific mRNA turnover in Trypanosoma brucei. Eukaryot. Cell. 10: 1230–1240.

Vanhamme, L., and E. Pays. **1995**. Control of gene expression in trypanosomes. Microbiol. Rev. 59: 223–240.

Vasquez, J. J., C. C. Hon, J. T. Vanselow, A. Schlosser, and T. N. Siegel. **2014**. Comparative ribosome profiling reveals extensive translational complexity in different Trypanosoma brucei life cycle stages. Nucleic Acids Res. 42: 3623–3637.

Vassella, E., J. Van Den Abbeele, P. Bütikofer, C. K. Renggli, A. Furger, R. Brun, and I. Roditi. **2000**. A major surface glycoprotein of Trypanosoma brucei is expressed transiently during development and can be regulated post-transcriptionally by glycerol or hypoxia. Genes Dev. 14: 615–626.

Vassella, E., M. Probst, A. Schneider, E. Studer, C. K. Renggli, and I. Roditi. **2004**. Expression of a Major Surface Protein of Trypanosoma brucei Insect Forms Is Controlled by the Activity of Mitochondrial Enzymes. Mol Biol Cell. 15: 3986–3993.

Vaughan, S., and K. Gull. **2003**. The trypanosome flagellum. J. Cell Sci. 116: 757–759.

Vazquez-Pianzola, P., H. Urlaub, and B. Suter. **2011**. Pabp binds to the osk 3'UTR and specifically contributes to osk mRNA stability and oocyte accumulation. Dev. Biol. 357: 404–418.

Vickerman, K. 1985. Developmental Cycles and Biology of Pathogenic Trypanosomes. Br. Med. Bull. 41: 105–114.

Viegas, I. J., J. P. De Macedo, M. De Niz, and J. A. Rodrigues. **2020**. N 6 -methyladenosine in poly (A) tails stabilize VSG transcripts. 1–25.

Wedel, C., K. U. Förstner, R. Derr, and T. N. Siegel. **2017**. GT -rich promoters can drive RNA pol II transcription and deposition of H2A.Z in African trypanosomes . EMBO J. 36: 2581–2594.

Wippel, H. H., J. S. Malgarin, A. H. Inoue, F. D. V. Leprevost, P. C. Carvalho, S. Goldenberg, and L. R. Alves. **2019**. Unveiling the partners of the DRBD2-mRNP complex, an RBP in Trypanosoma cruzi and ortholog to the yeast SR-protein Gbp2. BMC Microbiol. 19: 1–12.

World Health Organization. **2020**. Trypanosomiasis, Human African (sleeping sickness). World Heal. Organ. Fact Sheets 259.

Wurst, M., A. Robles, J. Po, V. D. Luu, S. Brems, M. Marentije, S. Stoitsova, L. Quijada, J. Hoheisel, M. Stewart, C. Hartmann, and C. Clayton. **2009**. An RNAi screen of the RRM-domain proteins of Trypanosoma brucei. Mol. Biochem. Parasitol. 163: 61–65.

Wurst, M., B. Seliger, B. A. Jha, C. Klein, R. Queiroz, and C. Clayton. **2012**. Expression of the RNA recognition motif protein RBP10 promotes a bloodstream-form transcript pattern in Trypanosoma brucei. Mol. Microbiol. 83: 1048–1063.

Yamashita, A., T. C. Chang, Y. Yamashita, W. Zhu, Z. Zhong, C. Y. A. Chen, and A. Bin Shyu. **2005**. Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. Nat. Struct. Mol. Biol. 12: 1054–1063.

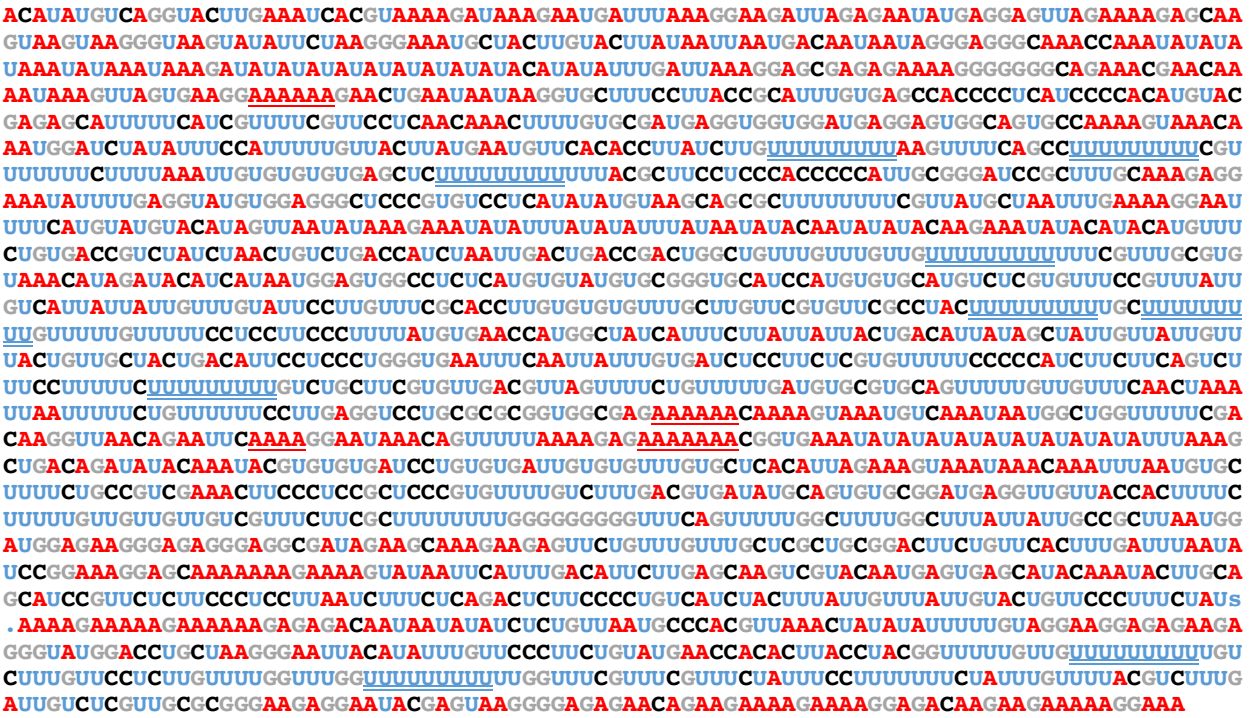
Zamudio, J. R., B. Mittra, D. A. Campbell, and N. R. Sturm. **2009**. Hypermethylated cap 4 maximizes Trypanosoma brucei translation. Mol. Microbiol. 72: 1100–1110.

Zoltner, M., N. Krienitz, M. C. Field, and S. Kramer. **2018**. Comparative proteomics of the two T . brucei PABPs suggests that PABP2 controls bulk mRNA. PLoS Negl. Trop. Dis. 12: 1–18.

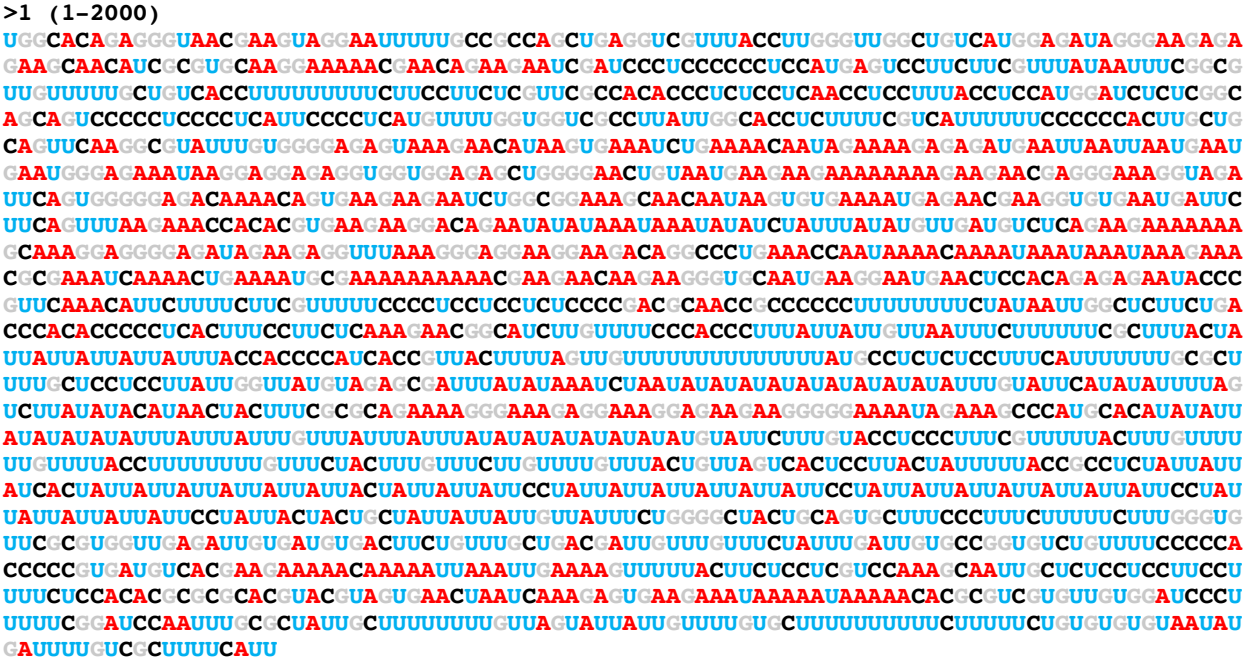
6. Supplementary information

Supplementary 1: Sequences of *RBP10* 3'-UTR and *RBP10* 3'-UTR fragments

>Full sequence with UUUAUUUA and (U)$_9$ underlined as well as poly(A) tracts



>1 (1-2000)



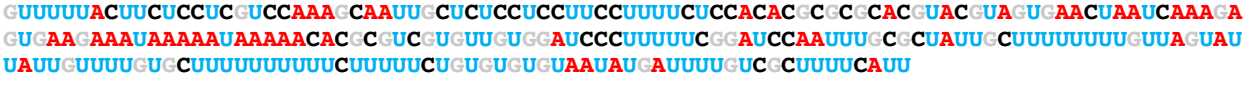
>1.1 (1-618)

>1.2 (598-1262)

>1.2 ΔAU

> 1.3 (1239-1772)

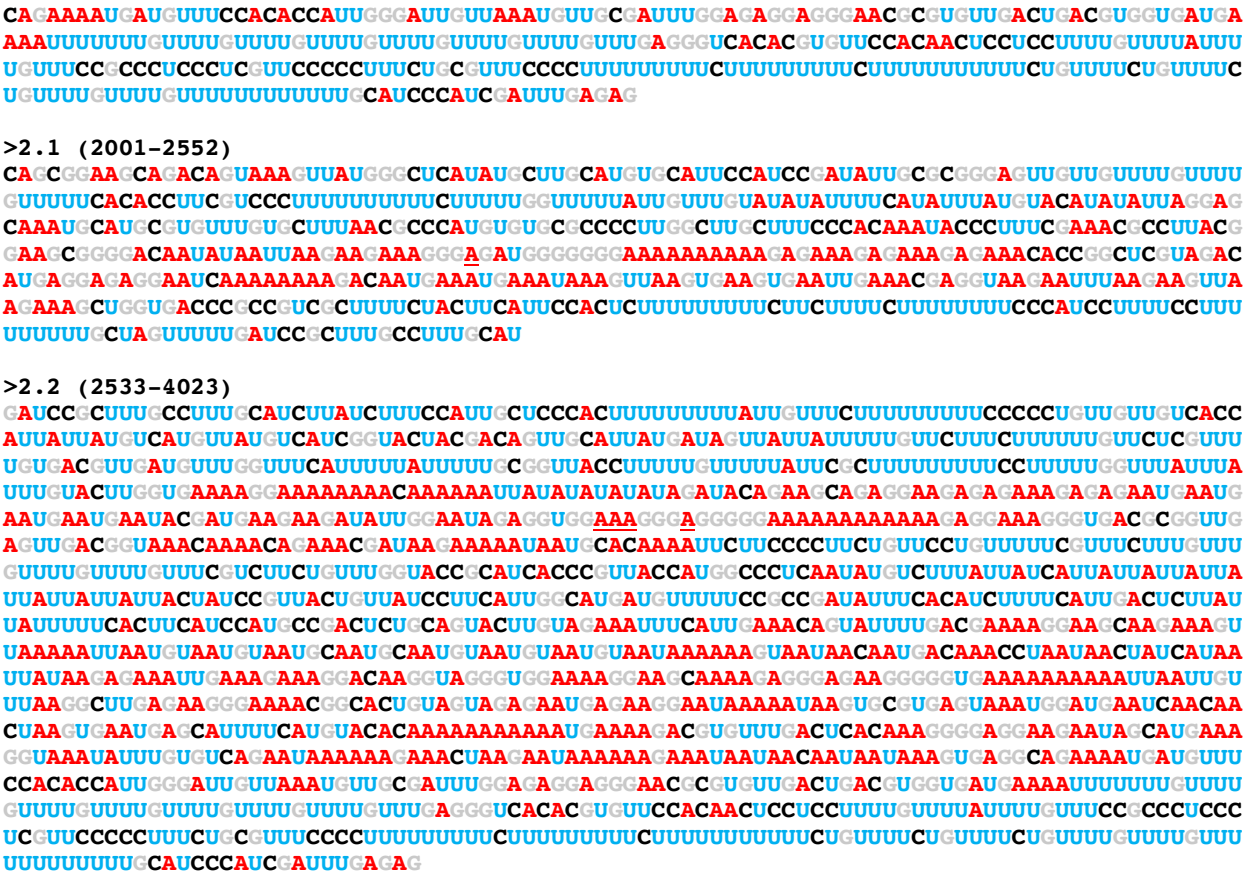
> 1.4 (1766-2000)



> 1.4.1 (1766-1879)

> 1.4.2 (1874-2000)

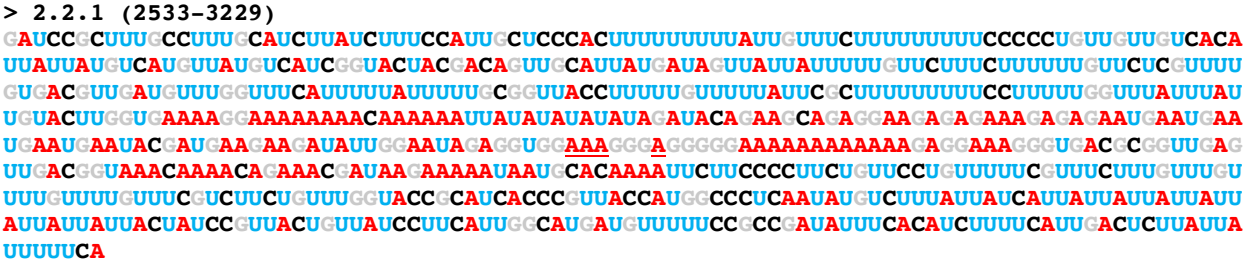
>2 (2001-4023)



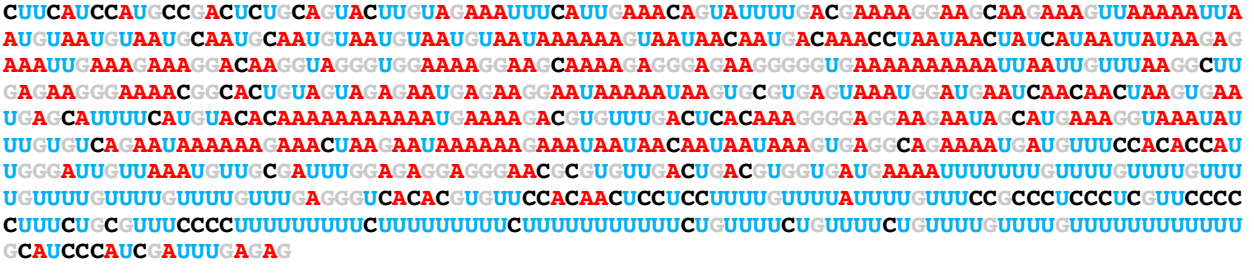
>2.1 (2001-2552)

>2.2 (2533-4023)

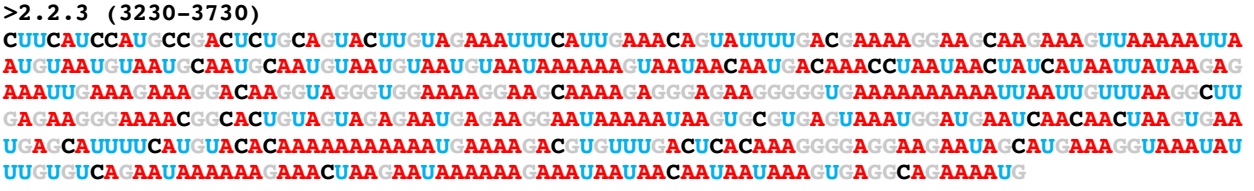
> 2.2.1 (2533-3229)



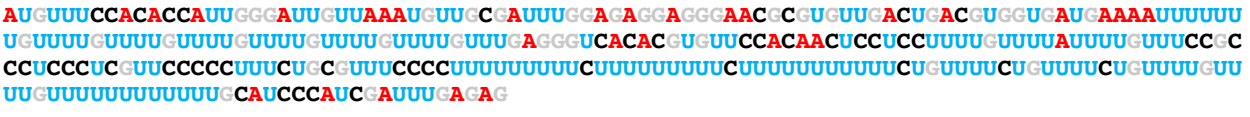
> 2.2.2 (3230-4023)



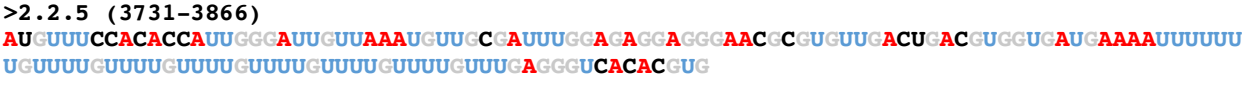
>2.2.3 (3230-3730)



> 2.2.4 (3731-4023)



>2.2.5 (3731-3866)

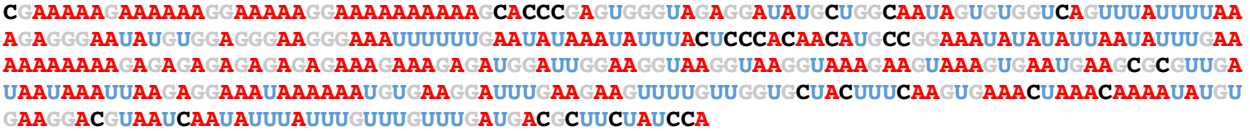


>2.2.6 (3859-4023)

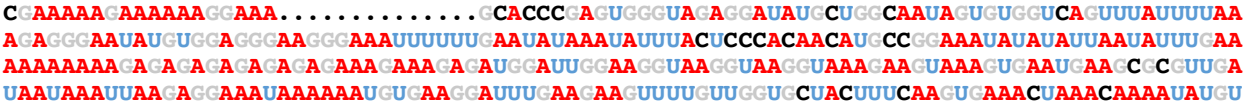
>3 (4024-5999)

> 3.1 (4024-4531)

>3.1.2 (4139-4531)

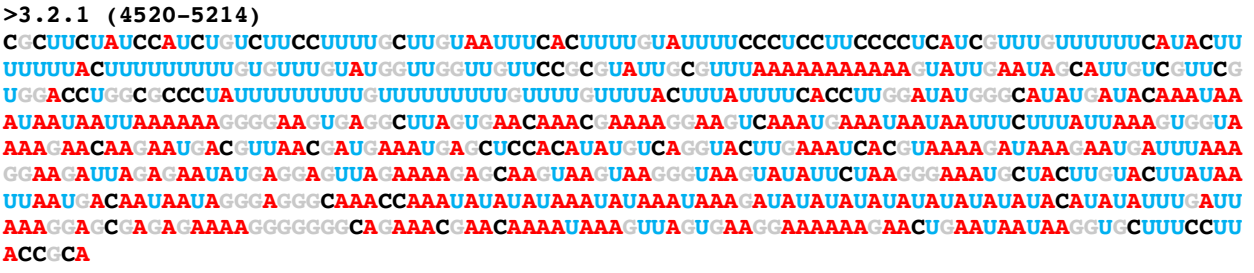


>3.1.2 ΔAAGG(A10)



>3.2 (4520-5999)

>3.2.1 (4520-5214)



>3.2.2 (5208-5999)

>4 (6000-8081)

AAAAGGAAA: possible polyadenylation site

AG: possible splice acceptor site for downstream pseudogene RNA

Lower case - possible intergenic region

GCUUACCGUCAAUAUCAGUAACCAUGUCACGCGGUUCACAAAUGUUAAAUUUAAUAACAGUAAUAAAACGAAUAUACAAUUGAAUA
GACGUGCAGAAGUUUCGAAGUUAAGUCGAAAGAGCAAAAAAGUUAAAGAAAAAAUAUGUAGAGGGUGGAGCCAACAAUAACGUAU
GGGAGGAAUGAAUGUGCUUUAUGUCCGUGACGGAGGCGUAAAAUGAUGAUGAACUGAAGUUAUGGAAACUAACAAAAUAAGGGAAA
GGACAUAUAUAUAUAUAUAUAUAUGCGUGCAUAUAUUGUGCCGGAAGGAAGCGGAAAUAUACACUGAAGCGUACCGGUCAAAAAUU
GAAAGUGCGGAACCCCGCGUCUCCGCGUAGUGUCACCUGUUAUCCGCGCAUUUUCUUAGUGUCUCUGCUUUUAGCGUUGCACAUCA
CCUUUGCAACAACCGAGUUAUCGCCGUUUCACACCUGUGGCACAGAUUUUCGUGGCUCCUUCAUGACACUUAUGCUCACUUUACAU
CUCACCUGCCGUUUUU

>ACT
UAACACCGGGUUGUGGCCAAAAUUGUUCUGUAGUCGCUGUGAGUUGACACGGCUAGUGCUUAUGAUUUUCCUCGCGUGUGGUGC
CUGUACUCAGCCCUAUGCCUUAUUUGCAACACAUUUACGUACAGCGCACAAGAGAAGAGAAGAUCACUUGAAGAUAAUAAAUAUAG
GGUUGUAGGCAUCUUGUUUAACUCAAAUUUUCUCGUCUUGGUGU

Supplementary 2

Table 6.1: Affymetrix probes used for the FISH experiment

Name	Assay ID	Description	Type	Target sequence
AF0078	**VPMFWYW**	***RBP10* codon region**	**6**	ATGGGAGACTCGATATCACCTTCGGAATTACTTCCAGGGCACGCTGG GGAAGAACATTTCTTATCACTTGAAGAAGGAACGACTCCGCAGGTCC ACCGGCAGGAAGTGACTACAGATGCAATGACAAACCTTCAACGGCGT AACGTTTATGTCTCTGGCCTCCCGGAAACGTACCGTGCTTCCGAGTTT CGTGATCTGTGTCAGTCGTTCGGTCGGGTGGAGGCGTCGAAACTCTG CGTTGACAGTAAGTGCCGGCCGGCGAAGGGTTATGGTTTTGCACTTT TCTTCGAGGAGGAGGACGCATTGAAGTGCATCGAAGGACTTAACGGC CGAGTGCTTATGGGCGGCCGACCGTTGCAAGCACGAATAGCAGACGC CGCTGCAGCGCCCGCGCCACTGGACCCATCAATAGCCCACCCGCCCAT TAGCCGCACGCGTCAGGGAATGAAAAGTCACCACCGACAGCTTGGCG CAAGCATCAGCAGCGGAAGCGGTCCTCTCGACAGCTCCATCGTTCTCG GGCGGAGCATGTCCTTTTCCTCTCGAGATGATCTTACAGGCTCACAAC TCGGGCTCGCAACAGCACTCATCCCGGGAACCGGAACCCCATCGGCCC CGACAGCTCTGCACGCACCGATGCAAGCGCCAATTCACGGTGCAGGGG TCACACAACCACCAGGGGCCGTTGGTTACTCACTCGCGCCGTTCCCTG TGCCTGGTGGCTACATGCCTGTTGGGCACTCCTTCATTCCGCAACAAA CTGTTGGCATGCCTCTATTTGTGACTATGCCACAAACCGCGCCCGTCT CCACTACTGTTGCACAGCAGGTCGTCGTCCCGACGACAGCAATTGTCG CACCGAACGGCTATACACCACAGCTTTGTTCATCGCTGGTGTATCCTC CGGTTCGAATGGAGTGA
AF0079	**VPNKRJU**	***RBP10* 3′-UTR 1-1280**	**1**	TGGCACAGAGGGTAACGAAGTAGGAATTTTTGCCGCCAGCTGAGGTC GTTTACCTTGGGTTGGCTGTCATGGAGATAGGGAAGAGAGAAGCAAC ATCGCGTGCAAGGAAAAACGAACAGAAGAATCGATCCCTCCCCCCTCC ATGAGTCCTTCTTCGTTTATAATTTCGGCTGTTGTTTTTGCTGTCACC TTTTTTTTTCTTCCTTCTCGTTCGCCACACCCTCTCCTCAACCTCCTTT ACCTCCATGGATCTCTCGGCAGCAGTCCCCCTCCCCTCATTCCCCTCA TGTTTTGGTGGTCGCCTTATTGGCACCTCTTTTCGTCATTTTTTCCCC CCACTTGCTGCAGTTCAAGGCGTATTTGTGGGGAGAGTAAAGAACAT AAGTGAAATCTGAAAACAATAGAAAAGAGAGATGAATTAATTAATG AATGAATGGGAGAAATAAGGAGGAGAGGTGGTGGAGAGCTGGGGAA CTGTAATGAAGAAGAAAAAAAAGAAGAACGAGGGAAAGGTAGATTC AGTGGGGGAGACAAAACAGTGAAGAAGAATCTGGCGGAAAGCAACAA TAAGTGTGAAAATGAGAACGAAGGTGTGAATGATTCTTCAGTTTAAG AAACCACACGTGAAGAAGGACAGAATATATAAATAAATATATCTATT TATATGTTGATGTCTCAGAAGAAAAAAAGCAAAGGAGGGGAGATAG AAGAGGTTTAAAGGGAGGAAGGAAGACAGGCCCTGAAACCAATAAAA CAAAATAAATAAATAAAGAAACGCGAAATCAAAACTGAAAATGCGA AAAAAAAAACGAAGAACAAGAAGGGTGCAATGAAGGAATGAACTCC ACAGAGAGAATACCCGTTCAAACATTCTTTTCTTCGTTTTTCCCCTCC TCCTCTCCCCGACGCAACCGCCCCCCTTTTTTTTCTATAATTGGCTCT TCTGACCCACACCCCCTCACTTTCCTTCTCAAAGAACGGCATCTTGTT TTCCCACCCTTTATTATTGTTAATTTCTTTTTTCGCTTTACTATTATT ATTATTATTTACCACCCCATCACCGTTACTTTTAGTTGTTTTTTTTTT TTTTATGCCTCTCTCCTTTCATTTTTTTGCGCTTTTGCTCCTCCTTAT TGGTTATGTAGAGCGATTTATATAAATCTAATATATATATATATATA TATATTTGTATTCATATATTTTATGTCTTATATACATAACTACTTTC GCGCAGAAAAGGGAAAGAGGAAAGGAGAAGAAGGGGGAAAATAGAA AGCCCA

AF0080	**VPPRJ4R**	***RBP10* 3′-UTR 2880-4023**	**1**	AATGAATACGATGAAGAAGATATTGGAATAGAGGTGGAAAGGGAGG GGGAAAAAAAAAAAAGAGGAAAGGGTGACGCGGTTGAGTTGACGGT AAACAAAACAGAAACGATAAGAAAAATAATGCACAAAATTCTTCCCC TTCTGTTCCTGTTTTTCGTTTCTTTGTTTGTTTTGTTTTGTTTCGTCT TCTGTTTGGTACCGCATCACCCGTTACCATGGCCCTCAATATGTCTTT ATTATCATTATTATTATTATTATTATTATTACTATCCGTTACTGTTA TCCTTCATTGGCATGATGTTTTTCCGCCGATATTTCACATCTTTTCAT TGACTCTTATTATTTTTCACTTCATCCATGCCGACTCTGCAGTACTTG TAGAAATTTCATTGAAACAGTATTTTGACGAAAAGGAAGCAAGAAAG TTAAAAATTAATGTAATGTAATGCAATGCAATGTAATGTAATGTAAT AAAAAAGTAATAACAATGACAAACCTAATAACTATCATAATTATAAG AGAAATTGAAAGAAAGGACAAGGTAGGGTGGAAAAGGAAGCAAAAG AGGGAGAAGGGGGTGAAAAAAAAAATTAATTGTTTAAGGCTTGAGA AGGGAAAACGGCACTGTAGTAGAGAATGAGAAGGAATAAAAATAAG TGCGTGAGTAAATGGATGAATCAACAACTAAGTGAATGAGCATTTTC ATGTACACAAAAAAAAAAATGAAAAGACGTGTTTGACTCACAAAGGG GAGGAAGAATAGCATGAAAGGTAAATATTTGTGTCAGAATAAAAAA GAAACTAAGAATAAAAAAGAAATAATAACAATAATAAAGTGAGGCA GAAAATGATGTTTCCACACCATTGGGATTGTTAAATGTTGCGATTTG GAGAGGAGGGAACGCGTGTTGACTGACGTGGTGATGAAAATTTTTTT GTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTGAGGGTCACACGT GTTCCACAACTCCTCCTTTTGTTTTATTTTGTTTCCGCCCTCCCTCGT TCCCCCTTTCTGCGTTTCCCCTTTTTTTTTCTTTTTTTTTCTTTTTTT TTTTCTGTTTTCTGTTTTCTGTTTTGTTTTGTTTTTTTTTTTTGCATC CCATCGATTTGAGAG
AF0081	**VPRWEPN**	***RBP10* 3′-UTR 5315-6314**	**4**	AGTGCCAAAAGTAAACAAATGGATCTATATTTCCATTTTTGTTACTT ATGAATGTTCACACCTTATCTTGTTTTTTTTTAAGTTTTCAGCCTTTT TTTTTCGTTTTTTTCTTTTAAATTGTGTGTGTGAGCTCTTTTTTTTTT TTACGCTTCCTCCCACCCCCATTGCGGGATCCGCTTTGCAAAGAGGAA ATATTTTGAGGTATGTGGAGGGCTCCCGTGTCCTCATATATGTAAGC AGCGCTTTTTTTTCGTTATGCTAATTTGAAAAGGAATTTTCATGTAT GTACATAGTTAATATAAAGAAATATATTTATATATTTATAATATACA ATATATACAAGAAATATACATACATGTTTCTGTGACCGTCTATCTAA CTGTCTGACCATCTAATTGACTGACCGACTGGCTGTTTGTTTGTTGTT TTTTTTTTTTCGTTTGCGTGTAAACATAGATACATCATAATGGAGTG GCCTCTCATGTGTATGTGCGGGTGCATCCATGTGTGCATGTCTCGTGT TTCCGTTTATTGTCATTATTATTGTTTGTATTCCTTGTTTCGCACCTT GTGTGTGTTTGCTTGTTCGTGTTCGCCTACTTTTTTTTTTGCTTTTTT TTTGTTTTTGTTTTTCCTCCTTCCCTTTTATGTGAACCATGGCTATCA TTTCTTATTATTACTGACATTATAGCTATTGTTATTGTTTACTGTTG CTACTGACATTCCTCCCTGGGTGAATTTCAATTATTTGTGATCTCCTT CTCGTGTTTTTCCCCCATCTTCTTCAGTCTTTCCTTTTTCTTTTTTTT TGTCTGCTTCGTGTTGACGTTAGTTTTCTGTTTTTGATGTGCGTGCAG TTTTTGTTGTTTCAACTAAATTAATTTTTCTGTTTTTTCCTTGAGGT CCTGCGCGCGGTGGCGAGAAAAAACAAAAGTAAATGTCAAATAATGG CTGGTTTTTCGACAAGGTTAACAGAATTCAAAAGGAATAAACAGTTT TT

Supplementary 3

Table 6.2: Measurements of CAT reporter mRNAs containing the *RBP10* 3'-UTR fragment and the *actin* 3'-UTR

Expected and measured sizes of all reporter mRNAs. The expected sizes are predicted from the sequences, with an additional 100 nt added to allow for a 39nt spliced leader and a 61nt poly(A) tail; known poly(A) lengths can be up to about 200nt. Some mRNAs were shorter than expected, suggesting that a poly(A) site within the 3'-UTR sequence was used. BS is bloodstream form and PC is procyclic form.

Fragments detailed	Clones	Measured length (nt)	Expected length (nt)	Expected length + 100 nt	Differences (meas - exp)
Actin 3'-UTR	CL1 BS	1365	1211	1311	54
	CL1 PC	1365	1211	1311	54
	CL2 BS	1307	1211	1311	-4
	CL2 PC	1446	1211	1311	135
	CL3 BS	1348	1211	1311	37
	CL3 PC	1465	1211	1311	154
F1 (1-2000)	CL1, 2, 3 BS and PS	2611	2913	3013	-402
F2 (2001-4023)	CL1, 2 BS and BS	2842	2934	3034	-192
	CL 3 BS and PC	2830	2934	3034	-204
F3 (4024-5999)	CL1, 2 BS and PC	2687	2853	2953	-266
	CL3 BS and PC	2785	2853	2953	-168
F4 (6000-8080)	CL1, 2, 3 BS and PC	2105	2994	3094	-989
F1.1 (1-618)	CL1, 2 BS	1528	1530	1630	-102
	CL1, 2 PC	1576	1530	1630	-54
F1.2 (598-1282)	CL1, 2, 3 BS and PC	1733	1578	1678	55
F1.3 (1239-1772)	CL1 BS	1224	1447	1547	-323
	CL1 PC	1259	1447	1547	-288
	CL2 BS	1211	1447	1547	-336
	CL2 PC	1191	1447	1547	-356
	CL3 BS	1224	1447	1547	-323
	CL3 PC	1238	1447	1547	-309
F1.4 (1766-2000)	CL1 BS and PC	1259	1148	1248	11
	CL2 BS and PC	1304	1148	1248	56
F1.4.1 (1766-1874)	CL1 BS and PC	1007	1027	1127	-120
	CL2 BS and PC	1029	1027	1127	-98
	CL3 BS and PC	1034	1027	1127	-93
F1.4.2 (1874-2000)	CL1 BS and PC	1063	1034	1134	-71
	CL2 BS and PC	1086	1034	1134	-48
	CL3 BS	1074	1034	1134	-60
	CL3 PC	1146	1034	1134	12
F2.1 (2001-2532)	CL1, 2, 3 BS and PC	2214	1465	1565	649
F2.2 (2533-4023)	CL1, 2, 3 BS and PC	2667	2405	2505	162
F2.2.1 (2533-3229)	CL1 BS and PC	1707	1611	1711	-4
	CL2 BS and PC	1707	1611	1711	-4
	CL2 BS (short band)	987	1611	1711	-724
	CL2 PC (short band)	1059	1611	1711	-652
	CL3 BS and PC	1707	1611	1711	-4

Fragments detailed	Clones	Measured length (nt)	Expected length (nt)	Expected length + 100 nt	Differences (meas - exp)
F2.2.1 (2533-3229)	CL3 BS (short band)	980	1611	1711	-731
	CL3 PC (short band)	980	1611	1711	-731
F2.2.2 (3230-4023)	CL1, 2, 3 BS and PC	1553	1707	1807	-254
F2.2.3 (3230-3730)	CL1 BS	1929	1414	1514	415
	CL1 PC	1956	1414	1514	442
	CL2 BS	1947	1414	1514	433
	CL2 PC	1974	1414	1514	460
	CL1 BS (shorter band)	1417	1414	1514	-97
	CL1 PC (shorter band)	1464	1414	1514	-50
	CL2 BS (shorter band)	1430	1414	1514	-84
	CL2 PC (shorter band)	1512	1414	1514	-2
F2.2.4 (3731-4023)	CL1 BS	1825	1206	1306	519
	CL1 PC	1885	1206	1306	579
	CL2 BS	1885	1206	1306	579
	CL2 PC	1868	1206	1306	562
F2.2.5 (3731-3860)	CL1, 2, 3 BS	1046	1048	1148	-102
	CL1, 2, 3 PC	1112	1048	1148	-36
F3.1 (4024-4531)	CL1 BS and BS	1332	1421	1521	-189
	CL2 BS	1406	1421	1521	-115
	CL2 PC	1258	1421	1521	-263
F3.2 (4520-5208)	CL1 BS	2725	2393	2493	232
	CL2 BS	2774	2393	2493	281
	CL3 BS	2725	2393	2493	232
F3.2.1 (4520-5208)	CL1 BS	1550	1608	1708	-158
	CL1 PC	1575	1608	1708	-133
	CL2 BS	1362	1608	1708	-346
	CL2 PC	1453	1608	1708	-255
F3.2.2 (5208-5999)	CL1 BS	1968	1705	1805	163
	CL2 BS	1942	1705	1805	137
	CL3 BS	1892	1705	1805	87

Supplementary 4

Sequences of 3'-UTR from the selected bloodstream-form specific mRNAs

>Tb927.10.5620 Fructose-biphosphate aldolase, glycosomal

AAUAAUUAUGCCCAUUUAGUUGGCUUUUCCCUUGUCUCGUGUCUUUUCCGUGGAAAGGUUCCCGGAGUAAUCUGAUGGCACAGCAG
GGAGGUGCGCCUGCAGGUUGGUUAGGAAGGGGGGAUGAUGUAAAAGAAGAAAAUGGGGGGAUAUCUUAUGUUUAAGAGGAUAAAUA
AUGUGAAGGGGCUUUAUAUGCUUGCUUGGUUGUCUCGUUGGUGCAUGGGGAAUCUGCAUGUUUGCUUUGGAGCACGCGUGGUACAU
CAAUGGGAUCAUAUCUUGCUGCCUCCCGCAGCUCACCGUGCGAGCUGCCGGGACCCCGUUUCUCAUAGGCGUGUGCACCUUCCGUC
UGAGACCUGUAAAUGGUUAAAAAGGAAUAUAUAAUUACCUUUUGAAAAGUGGUAAAAAACGAAUAUAUCUUUUUUUUUGGUUUUAA
UACGCUCUUUUGUGUGUAUGAGAGGAAUAAAUAAGUGUGUGUGUGUGUGUGUGUGUUGUUGUUGUUGAAAAAAUGAAGGACACG
AAAGCGCUCAUUAAGGGUGGGGAUACACCAAGCAUAAUUAUCCGAAAAUAUUUUGAUGCACCAAAUAAGUGAACAUACUGACGAAA
UCAAAGAGUUGGAGGAUGGAUAGGGAGGCCUCAUUGGCAGUGGUAAAUUGAUUUGACUU

>Tb927.1.700 Phosphoglycerate kinase C (PGKC, region needed)

GCUGGGAAAUGAACUCUAAAAAUGAGAAAUAAAGGGAAAGAGAAAGGAAGAGUGAUAUAUAUAUUUUUUGGAAAAAAAAACACUUU
UCUUUUGCUUGCGCUGCUGAGUGGGAGAUCAUUCUCCGUGUUAUAUGUCCUUUUUCUAGUGGUUGAGAUUGUGUUGUUGUUUUUUC
AAUUUCUUCUGUGGAUAAUCUUCCUCGUGAAGAAGACGCAGAAAGCGGGCCACACGGAGUGAAUUCAUACCUUACUUAAAAUAAUA
UAAAACGCAUUAAAAUAUGUAAUUAUAUUUAUAUAUUUUUUUCCCUUUCUUUCUUUUAAAAAAAAAUUCUCUUUUGUGCUUCUUGC
UUCUCUCGUUUUCUAAACUGGGCAAUUAAUAUGCUCGAAAGUAAAUAUUGAGGUUAUUGAAGAGGGUUGGGGUGUGAA

>UHU1_UNIQUE_REGION

UAAAGCACUGAAGCUUAAUUCAUCUGGUGAGGUAUUGUUUGUCUCGCUCGCAUGACUCAUGUGCUGGGGAGGUGUAAAGGGGGGAU
GGCGACGAAGUUGUUUCUUGCAUAUUUCUCGCGCAUCUGAUGAAUUAAAAAAAAACGAUUAUUGCAUAACAUGAUUAUCUGACCAC
AAAACGUUUUGUAGUUUGAAGGAGGUAAUUGGGUAAUGUUUUAGAGGUCGUCAAUAUUAGUGGCGUUAAUGAAAACGGAUUUUAAA
AUUUACUUCUUUUUGCUGUUUUAUGUUGUCUAUAUAUACUUUUGUUUUUCCAUCAAGUCGACUGUGCCUAUUAUUAUCUGCUCGGU
UUUGUAGCAGCGGAUGGACAGAUGGAUGAAGUGAUAUAUGAGGGCAGUAUGCUGUUAGUGUGUAUGUGCACUCUAAAGCUGCUGCU
GUGUCGGGAUAGUGAUUUACGUAGGGCAGUUGAUUUUUCUUUUUUCCUUUUUUUUUGUAAUAUUAUACAAUUGAAGACGUUUCUUA
AUAGUU

> Tb927.1.3830 Glucose-6 phosphate isomerase, glycosomal (PGI)

ACAACCGACUGAAGAAAAAUAUUUAUGUUGACUAAUAGUUCAUCAACUGACUUGGGGUUAAUUUUUUUUCCAUUUGCUCCCCCUCU
CUUUAAAUCCUAUUUUAACAAAAACGACGCAGACGAUCAGUGAGCUACUCGACGCAGCAAAAACUUCCCUACUGCUUCCGUUCUUU
CGAAGGGAAAUAUAUUGAAUGCUUGCGCUGAGUAUCUUCAUAUAUUUUCUUUUCACUUAUUUUCCUUUAUAAAAUUUCGUUCCCAC
CUCUUAUUGAUGACCUUUUCCAUCUCUGGAUGGGAUCGUUCUUGGUGUUGGGUCUGUUAGCUGUUGUUCAUGUGUGAC

>Tb927.3.3270 ATP-dependent 6-phosphofructokinase, glycosomal (PFK)

AAUGACCCUCUGGUAACGGAGUUCUGGGAAAGGGAGAAUGAGGAGGGCACUCUCUGAUAUAUCGCCACGUCAGCCAGGCGUGAGCC
GGGAGGGCCACGAAACAGAGUUAAGAAACUAAAACAACAGCAGCAUCAUCGAUCAGACGGAAAAAAAGAAGUGUUAAAAACAGGAG
AAUAUUCACACCAAAACGAAUUUUAUGGGGUUUAAUGAGGGUAAGUGGACAAAUGAGGGAAAGGAAUAUGAUGAGAAUAAAUGAGG
AUGAAAACAGAAAGGCGCAUUAAGGCCACGUGACCCCAGCGCGUUGGACCCCUCUUAUUCUUCUUUUUGAUUUUUCCUUUUUUAGU
UUUGUCUCUUUGAUCUUCUCCACAAAAGGGGCGAGACACUUGUGUGAAUGGGGAGGCGAAUACCGACAAGCCAGACAAACAUGCAU
GAGGGGUGAGAAUAUGUGAAGAAAAUGUAGUGGAGAUCAAUGAAUUAGAGAAAUCAAUGGUGAACACAUGGACCGAAUGAUGAGGA
CGAAAGCAAUGCCAAGUAAUCUGCUUGUGUGAUAAAGCUCGGAUAGGAAUAAGAAAAGCAAACAGCUGUGAAAGGGGGGGGGGAAC
AGCAUGUAACAAAAUUUAGCAGGCUAUCACCAAGAAGGGACAGUGGAAAAAUGUACACCAACUGUUAAUGUGAAAUGAGUGGUGCU
UGUAAUAAUUGUGCGUGAUCAGCGGAGGAGUGGGGUGAGUGCUUCUGGGACAGAAGGAGCAACAAAGGGAAGGAAGUUGUUUCCUU
CCAUUCCCCCUUCCUUUGGUUUAUUUUCCAUUCAUUUCUUCGUUCAUUUUUUCCCCUUUCCCCUAUAUUUCUCUAUAUCAUUUCUC
CACUAGUGGUUUAUUCCCUCCUCUUUCCUUCAUUCCUUCCUUCCUUCCUCUGUCACUUGCUUUAUUGUUACCGUUGCCACCACCGC
ACUGCACACAACAUCAAUACAUGCCUUUGGGUUGUACGAGUAAGACUCAAAAUAGAUGGAGGAGAGGGUGAAGAAAAAAAGGAGAG
ACGCUCGUCGUCUCUCGUGCCUCACAGUAGUAUUGAUGCUGUCAACAUAUAUGUAUAUGUUUAUUUAUACAUGUAUGCGUAUGAUU
UUGCAUGUGUGUGUGUGUGUGCGCGUAGUUUUGUUUGUUUGUUCAGUCACGCAACCCAACUUUCCCAAUUCACUUCAGCCCACC
UUUAAAAGCGUACGAACUUCCGGGUGUGCUAAAGGGGGGUGAUAUAUGUGUAUGUAUAUGCAUGUAUUUGUAUGAACGUGUGUAUG
GGAAAUGGGGACAUCCUCUCACUGCGAUGAUAAGUGUAUUCCAAACCUUAAAUUAAGGGAUGCGCAUUGGAAGGGAUGGGGAAAUC
AAAUAAAGAGAAGGAAGAAGAGAGGAUCGAAGGAGUAAUAUGUAUGAAAUGGUGACUGAAGAAGCCACUAACGGCCCCUCCCCAUC
AAAAAACAAUAACAACAGCAACAAAGCAAGACUAACCGGAAGAGAAUGGAUUCAAAACAAAAUCAAAGGAAAAGGAAAAGGGGGAG
AAAAGAGGGGAACACUCUAAAGCGCAACCUGUACGUUGGAAGGUCAUUAAUGGUGGUUGGGGAGUCAAAUAAAAGAGAGGUUAAUA
GAAACAAGGUAAUUAAGAAACUAUUGGCUGACCAAAUGUUGAGCAACAACCGGUAACGCGCUGUAUAUUUGUCCCUUUUUUUCUUU

UGGAGUUUGAUUGCUAUUACUGUUGCUAGCAGACUCCCCUCUUUUUUGUUCCUUCAUGGUCAUUUUUUUGCUGAUUUUCAUAUUUU
UGUCCCUAAGUCACUGCUAUUAUAAUUAUGGAUAAACUACUCAGGUUUUUUUUUCUGUGUCGGGCUGGGGUCAAGGUACUUUUUCA
UUUUUUUGUUUAUCUCCUAAUUUUUGUUUAAUUUGCGAACUACGAAGUAGUUGACAAA

> Tb927.10.7090 Alternative oxidase, mitochondrial (AOX)

GAGCUGACAUUACCUUUGAUGGAGUUCCGCCUUUCAGAGUGAUUUGAAUGCAUGCAUUUGUGUUUCUUUUUUUUUUUGGCUCAAUA
UUGUGUUCAUGUGUGUGUGUGUGUGUGUGUGGUGCGUGUGGAUUGUGGGGAUGGGGGCGGGGGAAAUUAUGAAAUAAUCGGCAG
AGGGCAGGGUCAAAUGAGAGGGAAAGAGGAAUAUAUAAAUAUAUAUUUAUUUAUAUUUAUAUUUCUACUCGCUUAUUUGUGCUUGG
AUGCCAUCAGUGAAAAUUAUAUGCGUCCUCGAACUCACCGGAUGUAAGGAAGCGCGCAGGUUUACUCACGUUACAGUGAAAAUGUG
UAACGCAAGAAAAAACAAAAAGGCUUGUGAGCAAAAAAGAGGAGGAAGGAGGAGAAGGUUAAAAAGAGGAGAGAUGCAAAAAAAAA
AAAGGUGAAACAGGAGCAAGAAAAGGUACAGAUGUUCAAACAGAGGCAACAAAAAGAGAGACACAAAGUCAUAAAAUAUAUUUAGU
GAUCUAUUGAAAGUGAUUUGAGAUAUUUCAAAAAAAAAAAUAAGUGACAACGAUGGGAACAAUUUUUAAUAUUAAAAAGAAAAGGG
GAAAAAACUAUUUUUCCCAACGUGGAUUUUUUUUGUUUUUUCUCCCCUUUUCUUUUCCCACGUCUUUAUCCUCUAUCCCGUUGGCU
GCUUUCCUUUUUUUGUUGUUGUUUUUUUUUUACUGUUUGUUUUUUGUUCCGUCCAUUAUUGAUCGAUUACUGAAGUUUUCGUAUCA
UAUGAAAUAUUUGCUUCUGUUUUGUUGUAUAAAUAAAUAUAAAUUAUUUUUCUUUUAAUCUCUCUUUCCUACUCCUCUUUCAUUUG
ACCUCGCUGUUUUGCCGUCAUUUGCUUCCGGUUUUGUUUGGGAUGUUUUAUGUUAAGUUGUGUUGUGUUAUAUUUUCCUCCUCGAC
AUUAUGCGAGAGGGAGAAAUGACGGAGUGAAAGAAAGGGGGGAAAAAAUGAAAAAUAAAAAUAAAAAAAAGAUGGAACGGUACGGU
GCCCCCAAUGAAACCGAUAUGAUUAAAUGAGUGAACUGAAAGAUUUGGUUUGUAUUUUGUUGUGUUAUGUUGUGUUAUGUGGCGAU
GCCUUCCCCUGUGGAGGAGCAACAAUAAAAGCAGUGCCUUGGGAAUAAUAAAAAGAAGAGAGAGAGAGAGAGAAGAGUAAUGUGAAGG
GCUGCACAAACAUAGUUUUUUUUUCCCACACACACACACACACACGCACACUUUGGUUUGUUUCGUUUUGUGUUCUUUAUUGUGUG
CAUGAGACUGUCAUUAAAAAGAAAAAGAAAUAAAAAGAGAGAAAACAUAAAAGAAUACGAAAUGACGGUGACUUCAUACAAGCGUA
UACGCAUGAAUAUAAUAAAUAUAUUUAAAUCAUGGGUAUAUGUAUUUAUAUUUAUAUUUGUAUGAAGUUGCUAGUGUUCGUGUGCA
UGUGUUUGGGUGGUGGGGAGAAGCAAGUGACUGAAAUGAAAAAAAAAAAGGAAAGAGGUUAUAAAAACAAUUAUGAUAAUAAUAAU
AUACGCGUUGCUUUAGAAAUACAGAAUGUUUUUUUUUAAAAAAAAAAUAAAUAAAACAGUGGACGGGAAGUGAGAUAUGUUUUCCUU
UGUUACUAUAACAAAACAACAACAACAAUAAUAAUAAUAACAAUAGCAGUAACGGUAACGAUAACGAUAAUUAUAAGAGUAAUUUA
GAAAGUAAAAGCAACAAAUAUUUAAUUAUUCAUUUAUGUGGUAGAGAAAGAGAGAGAGAGAGUAAAAGGGGAAAAGGUAGGAAUAG
GAAUAACAUUGUUUGUGUUUGUGCUUGUUUGGCUUUUUUUUCUUUUUUUUCUUUUUUUUCAUGUGUGUGUGUUGGGGGGGGGGAAC
UUCUUUAAUUCAACCACUUGUACAACAAAACUUCCAGUAAUCGAAGGGUAAAAUCAAAUGAGAAGCGGGGGAAAAUGUUUUUUUUU
GUUUGUUUGUUUGUUUAAUGAAAGGAGAAGUAUAUAUGUAUAUUUAUAUAUACGAAUAUGUGCAUCUUUUAAAUAUUUUAGGUCAU
CUUCAAUAUCCUUUCUUCUUUUUUUUUGUGUGUGUGAGAGAGGGAGAGAGAGAUUUUUUUUUGUUCGCUUUGUGGCUCUGACGUGU
UCGGCAAUCCCAUUAGAGAAAAGAAAAAGUUUAGGUAUUUUUUUUUAUUUAAAAAAAAAAAAGCGUCGUAAGAUUUGCGGUCCCCG
UAUGCACAUGGCGGUUGCAUCUCGUAUGUUUUCUCCAUUUUUUUUCUUUUUUAUUCAGUUAUCAACAGUUUCCUUUCUAUCUUUUU
UUUCUUUUUUUUCUUUAUUAUAAUAAUUAUUGUUCCUUUUCUUUUCAUUUUACCUUUUUCUUUUCUGUUUUUGUGUCGUUUAUGCU
GUUAAAAGAAAAAAAAAUGAAAGAAGGGAAAAAAUAAAAGGAGUGAAGGAGGGACGGACUGGCGGGUGGGAAAAGAAAAGAAAAGA
AAAAAACAAAAAAAAGGAAAAGGAAAAAGGAAUAUAUUUUACGAUUCGAAGCAGCCAAAAUGAAGUUGUGGUCAGUUAGACGGGGU
GAAAGUUUUUUUUUAAUUCUUUUCUUUUUCCACUGUUUUUACUCCCUUUCCCCCUUUUCUCUCCCGUCCUUAUCAUUAUUAUUAUC
AUCAUCAUCCGCUUUGUUUUGUUAUAUUUUAUUCAUUUUUCAUUUAUAUCCUUCAGAUUCUUCUUUUGGUUGUUGUAAUGACGAUG
GGUCUGUAGUUGUUUUUUUUUUUGUUUUUUUUUGUCCCAAACUUUAACCAACCUUAAGUUUGAGGAUCAUUCAAAACGCAUCAACU
GGUGGAAAAUGCGGGGAAAGGGUUAAUGGGCGAGAAAAAGAAAAAAAUGAAACAACAGCAACAAAUAAAAAAAAGAGAGAGAGAACA
AAAAAAAAAAAAAAAUAUAUAUAUAUAAGAGAAACGGAAGGACAAAGUUGAUAGUUAUGAUGUUUUUAAAAAAAAAAAAGAAAAG
UUAACUUUCCUCCGUUUAUUUCAUAUUUACACAAAUGACCAUUAUGCACCAGUUAUACAGCAUCAGAAAGAGAAAAGGGGAAAGGA
AAUAAAUAAGGAAGGGAGGUACUCAUGUCUGUUUUAUGUUUUGUGCAUUUGUAUGUAUGACUGUGCGUAGUUGCACACAUUUUCUU
CCUCCUUUCUUUCUUUUUAUUUAUUUGUUUGUUUCUUUCUUUCUCAUCAUUUGAAUGGUGGUAAGAAGGUUAUAUGGGGGAGGAAA
ACAAAAAAUAUAUAUUUAUAUCUACAUCCAUACCUAUAUUUAUUAAAUUCCACAUGCAGACACAUAUGUAUAUAUAAAAAUGAGAGU
GUUUAGAGACGGCGCAUUUGUCACUGAUCUCUCACGCUCUGAUUUGAGAUAUGUUUUGCUUUUCAUCUUCUCGCCUCUUCGUCCUU
CUUUUCCUUGUUUUGUCUCUAUCUUCUAUCUUAAUAUCAUUGUUUUUUUUGUUUUUUUUUUUAGAAAUGAAGCCUUCAAAACAAAACA
AAAAGCAAUGAUUAUGUUAUGUUCUAACUUAGUUUCAUUUAGAUGAAUAUACGCGACAUGUUCACAACGCAUGGAAAAGUUGCUGA
AGUAAACAGCAUGGAAAUCAUGUGUAUAAUCGCGGAGGCGAAGUUGCGCGGAGGGAACAAGGGAAAUGCAUUGAAUUAAUAUUAUA
AAAGCAACAGGAGAUGUAAAAAUAUUUAAAAAAAAAAAGAAACAAAGUCACCAGUGAAAUGGUUCUGAUGCAUGACGUUAAGGUGG
AUUUCAUCACUCACAGGCUGUGUUAGUUUGUGUUAUUUAUCUUUAGAUUUCAGGAAUUAUUUUCACUGAGAAGUCGAGUAAAA

>Tb927.10.14140 Pyruvate kinase

GCAUUGGUGUGUGGGUGAGCAAUGUAAUGAGGGGUUAAUGAAAAAAAAAAAUGCGAUAUGCUGGUCACUGCAGCAAUUUAUUUUAUU
CGGGAGGAAAAGAGUAAAAUGGAGGAAGAAUUAAUAUUAGGGGAUAUUAUUGAUCUGGGAGGGAUAAACUUUAAGAAUGGGGAAUA
UAUUUGUAUGUAAGAUGAUAAACCGGGGGGGAUGCAUUGAAGCUAUAUAGCAUGUAAGUGAAGGGUAUAAGGGUGAACAAACUGAU
GUCCAAAAACUACGUGCGCCAUCCCCUCAGAUUCAUAUGUGACGAUGGGACAAGUGUCGCAGAAAGCGUGGAAAGUUAGGGAUAUG
CAAACUGAGGUGGCGCGAUGCCGUUAAAAAAAGAGCAUCCUGUGCCAAAUAAGUGGAAUAUAGGAAGGGAGAGGUGAGUAUUGCUG
AGAAAUCUCCCUUUCUUCGCCCCCUUUUUGAUUUUUCUUUUGGUUCUUGCUAUUUAAAUUUAAAGCAAUAAUAUCCCUAUUUGACU
GUUUAUUUGAUAUAUAUAUAUAUAUAUAUAUAUAUCUUCUUUUUGCACGUCUCUGUGUGGUUGUAUUAUCUAAUUUUUUUGUUC
UGCUAUCAUCUCAUGUUACUAAAAUGGAUAAAACAGCGCUUUUCUUCAUAUCUUCUGGCAAUCCUUGGUGAUGUUUCUAAGCUGUA
CUACCGAAAAGGGGAGCUGUUAAUGUUAUUAGGCUACAAGGGAAGGAAAGUGUAGUGAUGGUGGAAAGUGGUGAAUCUCUUAAUGU
UUUCAAUACGAGCAAGCACUGAAAUUUUUUUCAGCGGGUGAGUGCCAUGACGGGUGAAUUUUGUUGGUGGUGUAUCGUUGAUAGCG
CUAAGACAACGGAAAGGGGUUGCGCACUUGAUUUGGACUUCUUUUCCACGUUGAAGGGUGGAGUAGUCAUUUGAUUUGUGGUCUUG

CUGUAGUUUUGUCGCGGAGUACUCCGUGCUGUGGGACCAUUGGGGAAUAACCCUAACACGUUUCCCAUUUCCCUUUAUUCUGUUUC
CGUUCCGUACGUAAUUCACCUGUGAGCCAGUAGCUGUGCUCAUUAAGAUAUAUACGCGAACACGGAGUUUAGA

>Tb927.10.14160 Aquaglyceroporin

UCUGCAUGAAGGAAAAAGGAAGAAUUAUAAUACUAAAUGGAAUUCCAAAUAACAUUUAUGGUGCCUGGUUAAAAGCAAAGGUAAAA
CGCAAGGGGAAUAUUUUUCACCGUAUGUACUAAGUAUAUAAAAAGCGUUAUUUUAAUUCGUGUUUAAAAAAAAACGAUUCUUUUCC
UGUUUUUGAGUGCAUGAGCUGCCUUUCUGGUUGCCGUUUUGUUUCAUUGUAUUAAGUACAAUGAUCGAAGGGGACACUAAUUUUUA
GAGAUAAAUAAGGUGAACAGAUGGCGCAAUCCAUAGGUGCUGUUGCUCCCUUCAGCAGCGUAAGAAAGUCUUUCCCCAUCACUAAU
CUCAUGUUUACUUUUUUCUUAUAUAUAUAUAUAUAUAUGUGUGUGUGUGUGUGUGCGCGCGCUUGUUCAUGUGUUUUUUUUUUU
CGAGCCACCCCGAUUAUGUUCGCUCAGGGCACCAAUGGAAGAAUACACACAUAAAAUUUGUCACUGAAUAGCGCUGAACCACUUGG
AUAAUCCCUCGUUGCACCCCUAACGUCAUUUAAAUACAUAUUUUGGGAAGACGUUAAGGUGAAUCUCGGUUUACUUCACUCAUGUC
GUACAUCUCUUUGUUGUUGCACUUUACUUCUCUUCAUCCCACAGCUGUUUGCAUGUGUGUUUUAACCGCUGAUUUUGUUGUGUUGG
GAGGUGAGGGAUAGUAUGCCGGGGCAAAUUCGCACUUGAUGAUUCACUUCUGUUUUACAAAUGAGGUGGUGGUGCAUUCUAUUGGG
GAAUAUUUGAUUGUUAUUAUAGAUUUUAAUGAGCUUUCUUUAUUUUUACUUUGUUUGAUACCUGGAGCUCCUCCACACAAACAAAG
GGAGGUGGAGAGCGACUUCCGAAUUCGCGUGGGAAAAGAGAAAACGGCAGCCAGUUUAACCAGUAAAUAUUACAUUGAAGGUUUCG
UUUAUUUUCGUAAUUUUCGUCAGCACGAUGGAUAGGGACCCUGAGAGGAAGUUUCUUACCCACUCACGGAGACCAUUCACUUUGGU
GGGAAGGUGUUUUCACAUGCAAAUUAUUAUUUCUGGGUUCCACUAGUUAUUCCAUUGUUUGGAGGUAUCUUUGGACUUUUUCUGUA
CAAAUAUUUUGUGCCACACUAAUCUGCAUGAAGGAAAAGGGAAGAAUUAUAAUACUAAAUGGAGUUCCAAAUAACAUUUAUGGUGC
CUGAUUAAAAGCAAAGGUAAAAUGCAAGGGGAAUAUUUUUCACCUUAUGAAUUGUGAACUUGAUGGGAACGCUGUUCCUGGGGUGU
GGAAGGGGACGGAACAUGUUCGUUAUGUUUCCGCCACCUCCACUAAAAUUUGUUUGAGGAAGCGGAUAUAAUCACAUCUGCCCCUU
UAGAUGCCCUUUGCAGCACACUCUCAGCGAAAGAAUACGUAAGCACAUCCAUUUAUUUACAUGCUGGAGUACAUGAAGAAGAAGGA
AAAGAAUAUGUUCCCGAUUUAUGUAAUGUGGGGGCGAGGCGGGGGUGGGCAUUACCCUUUUAUUUCUGUACACAUGUCUUUUGUUU
GGGCUGCUUUGCACCGCUCCGCAUCCGGAGGUGCACACCAGUGGAUCGGUUUGGCAACUCAGGGGAAUUCUCAAAGGAAGAGGAGA
GGGGGAAGCAGGAAAACUCCACCGGCUGCGUCAUCAAGUUUAUGGUAGUCGCAUGUGUACUCGCACAACUCCUAUUGCCACCAUUU
GAAUUUAUGCGGGCCAAUACGUGGCGCAGCAACGCAAAAGCUCGCCGACGAUAGCGCCAUGUGUUUGGCAUCCUCUAGGGGUAUUG
UCUCCCU

>Tb927.10.15910 Glycerol Uptake Protein 1

GCUGCUUUAUUUGAUUGGAAACUGAGAGCUGAUUGCGCUUCAGUAGUUGCACAUUCCUUGUGUGUUCGCGCAACACAUGAACCGUA
AGAUGUCAGACGUCUUGUGUGAAGUUUUUAUACUUCAUUUCUGAAAAUCGUAUAGUCAUUUAUCCUCUAUCUCAUCAUUAUUCUUU
CCCUUCCUCUUUUGUUGCCGUUACUUAAAUGGCUUGUUUGAUUGUGUGUUUGUUUUCUUCUCGCUUAACCGUUGAAUUCUAUAUAC
UUUGUUAAUAUUUUUUAAUAAAAAAAAAAGGAAAAGGAAAGAAAAAAGGAAAACGGAAUAAAUGCACGGUGUGGGGGCAUAAUGAAU
UCCACAGAAUGAAAAAAAAGGGGGGAAAGUAACUUGUAGGCAAGAAGCGGAUCAAAAUAUCCGUAAAGGCGGUGGAAUGAAAACAA
GUAAAAUAGAGGGCGUACGAAACGCUUUAAAAAGAUAAAUAAAUAAAUAAUGGCGGUGUAGUCGAAUAUAUUUAACGAUUAUGACA
AAGAGGGAAAAAAGAAAAAGAAAAAGGUAAGAGAAGGGCAGUCGGAUUGUAGGGGUGUUUGUAUGCACUACCCGACUCCGCUUUUG
CGGCAUAGUGAUGAUGACAGAUGCGGGCAAGUAUCUAAGUAAAAUUAGUUCAGUCAUCAAUUUGCUCCUCUACUACCGCUCUUCGU
CUGAUUCCUAUUUUCGCUAUGUGUUGUCAUUUUCUUUUUUUUCUUUGUUUGCUUGUCUUUCUAUCAGUGUUUUUUGUUUCUUUUUC
UUUUUUAUUUUUGUUUUUUCUCUUGAUUUAUCUCUUUUUUUAAUCUUCACAUUUCAUUACUCGAAGUUGAUUGAAUGGAAUUAACA
CAUGGAAUUGCAGCAAAAAGCAGAACACGGAAUUGGAUAAAAAUGGAUUAGGCGAAACGCUUCACAAAUGCUAUUUGCAUUUUGGC
UCAUACCAGACCGGGUAACCACCGCUACCGUCAUCAACACUGCUGCCGCUGCUGCUGCUGAAUCCUGAGUGAAGGUGCCAAAAAAA
AAAAAAAAGAGAAAAGAGAAGAGAAGAAAAGAAAAAUCAAACAAAAGAAAAAAAUAGCACGGCAACAUUUUCUCACCGCGACGAAG
UCACCCAACAGCGCUUAAGACAAACAGCCAGUGGAAGAUGUAUUAAGUGUAAUCGUACCGCUCCAUUAUUUUACGGUGCUGUUUUA
UUCCUUUCAUUCUUUUCUUCGCCCC

>Tb427.02.6000 Glycosylphosphatidylinositol-specific phospholipase C (GPI-PLC)

AAAUGUGGGAAAAAAAACAAAUCUCUGCUGCGUUUGUGACUGAAACAAGGGUUUAAAAGUAAAAAAAAAGUUAAGUGGACGACGAA
AUAACAACAAAGGGAAGGAAGGAAAAUUAAAAUUAAAAAAAAAAAUUAAAAAAAAAAAGGGCGACAACGACGUUGGAAGAGUGUGAAC
UAUAUACAUAUAUAUAUAUGUAUUCCUCCUGUUUUUUUUUGAUUUGUCUCUUGCUUAAAAAAGAAAGAAGAAGAAGAGGAGGAGGG
GGAGAGAAAAGAAAAAAGAAAGCGAUCAAAUGAAAGGUCAGUGGAUUAUUGAGGGGGACGGUGUUUGAAAGAGAGAGUGAAUGCGA
AUACAUGUGCCAGUGGUACUGCUGCUAUAUUAAAAAAAAAAAUUAUACGGGUAUGGUGUGGAAGCAAAUGUGGGGGUUUUUAUUUU
UAAUCCUUAAAAAGCAAAAGCAAAAGCAAAAAACGGGAAACUAAUACGCUAUGCGUUUUGUGCCUUAAAUGUACGGGCCCUCACAU
GCUUUAUUUUCCAUUCUGGUUUUUACCCGUAUUCAACUUUUAUUCCAUUUUUUCUUUAUUUUUAUAUAUGCAUGCGUGCAUAUAUA
UAUAUAUAUAUAUUUAUAUGUUGAUUUAUUUAUUUAUGCAUUAUUGUGUCCCUUAUUAUGCUUUCUUUUCUUCCCAAAUUUCUUCC
AUUUUGCUUUGAAAAUUUGUGGUCAUACGGAAUUGGUAUGAAAGGAGGAGUUUUACAACAGCAACAAAAAAAAAAAGAAAAAGAAA
AAAAAGAGAAAAGAAAAGAAUAAUUUAAAGAAAAUAAAAUAAAAUAAAGAAAAGAAAAGAGAGAUGAUGUCGGUUAAAAAAAAAAG
GGAAGGAUGUUGAGUGGAGAAGUGGAGGAAAUAAAUAACAACAAUGCACCAAACAGCAGUUUAGUUUUUUUUUUUUUGCUACACAUA
UUUCUACAUGUAUAUUUACUUAUGUUUCCGUACGAAUACAAAAUGAAUACAUUUCUAUAUUUUUGUAUUUGUAAUCGCCACUCUUU
UGAAUUUAAACAGAAAAGAAAAAAAAAAUUAAAAAGAAGAGAAGGAGUGAAAAGUGCGAAGUGAAAGAGGGGGAAAAAAAAAAAGA
AAUCGGAAGGGAAUUGCACGUGUCUCCCGUGAAGGAAUGUUUUUUUUUUUUUAACGUAAAAGUAAAUAAAUAAUGCAAGGAAAAGA
GUGAAAAUCUAAUAAAAAUAAAAUGAACUUUUAGAUAUUGAUGAAACGGCUGCUGCAAAACAAAUGAUAUGUUUUUUUUUUAAAU
GCACCACCUCAUCAAAGUAGAGAAUUAAAGUUUUUUUUUUUUAGAAAAAAAAGCGGAAAAUAUACGUAUGGAAUUGGGAAUGGAAU

AAGAGGAGGAGAAAGAAGAAGAGGAAGAAGAAAAGUGGUGCUUUUUUGUUUGUUUGUUUGUUUGUUUGUUUGUUUGUUUGUUAGGU
UGCGUAAAACAUAUACGCAUCAAUAGAAAACAAAUAAAAAGUGAAGUGAAGUGAAGGGGAAGGGGAAGAAACAAAAAAAAAAAGAA
AGAAAGAAACAAAGGAAAGAGCACCUUUUUGGUUUUUUUGUUUAAAUUUUAUUUUUACUGAAAAGAAAAAAAGUGCAAGAACGAAG
UUUGGUGUGCAGAUCCUGUGAGGUUUAUAUUCCUUAUUUAACUUCCACCUAGUUUGGUGCAUGUUUAAUUUUUUUUUCUCUUUUUC
UACUUGGUAUUUAAUUUGGUUUCUGUCUAUUUUCGUUUUGAUGCACUUAUGCAUUUGUUUGCUUACGCCUUUUGCCUUCCUUUGCU
UUGUUUUGUCUCAACCUCUCUUCCAUUCGGUUGCUAUCAUUGAUUAUUAAAUAAAUAUCUAAAUAUUUAUAUUUUUAUGGUGAUUU
UGAUUAUCACGAUUAUCAUCAUCAUUAUUAUUAGUAUUAGUAUUAUUGGUGAUUUGUUUAUUUGUGGGUACAACCACCGUAUGAUU
AUUGCCGUUAUUGUUGUACUUAUUAUCCUUUAAAAAAAAUAAUAAUAAUAAUUGUGUAGGCUGUUUCUUCUCCUCUCUUCUCCCCU
CUUCUUCUUCUUUUUUUUUAAACUUCGUCUUUUCAAUUUUUUUCUUUUUGAGGAAGGGGGAGGAGAGGGGAGGGGCGUUGUCACUC
GUACACACACACACACAUACACACAUACAUAUGUAUACGUAUAUUGUAUAUACAUAUGUAUCUGUGUACGCUUAUAUAUAUAUAUA
UAUAUAUAUGUAUAUUUAUUAUAUAGUUAAAUAAUAGCUAAUAAAUGAAUAAUAUAAAUAAUAAUAAUAAUUGAUUAAUUAAAUUA
UUAAAGCGUGAAUUCCAUUGUGUUACUUCUAUAUUUACCGUGUUUUCUUGUUUGACAUGGUGUGGCGAUUAUAAUUCGUCUUCUUC
UUAUCCCUUUUUCACUCGCUUUCUCAUUUUCUCUGACUUCUCUUUCAUAUUUUUCUUUUGGUUAUUGAAAUGUUAUUGAAGUUUAA
GUAUCCGAUAGUGUGUAACCCCCUUUUCCCUCUUUUUCAUUCUCGUUGUUUUUGUUUAUUUUAAUAUUUACAUCCAAUUUCUCUUU
UACGAUUUUUCUUUUUUUUUUUUCCUUUUUUAAUCACUUCAUAAUAUUUCCCUCCUUUAAGGAAUAAUAGUAAUGUGAAUAAUAAA
AAGGAAAGGAUUAAAGGAUGGAAAUACAUACAUAUGUGUAUACAUACAUAAAUAAAUAUAUAUAUCUACACAUGCAUAUAUUAAAA
UGUGCUUCAAAUGUGUGAAGUGAUGUGACAUGCCGGUUUAAUGUCAAAUGAAAUAUUUAUUCGUUUAAUUGUGGUUGUUGCAGUCG
UAACUUCAGGAGGAAAUGAGAGAGGUAUAUUUAUACAGGUGACACAGAUCCCUUUCAAUGCUUCAUGAAAUUUAAAAUAAUUCGGG
AAACUACACACGCACAGUAGAUGAUACGAGACA

>Tb927.10.6880 Glyceraldehyde 3-phosphate dehydrogenase, cytosolic (GGAPDH)

GUGAUGUUCUCGCUCCCUACGCGCUUGGCUGUGGUGCGGGUGUAAGAGCCCUGCAUUUAGCACUAUGACCAUCCGUACUCUUGUUU
UGUAAUGCACUUUGGUUUCGAAUGACCAAUUGCUACCGACAAACUGUCUUAGCAGACCAGAUUAAGAAUAUAUGUGUGACGGACAA
GAAGGGAUGGCAAAGCCCCUAUGUGUAGAAUGGAACAUCCGCUGAGCACUUUGCGUCCUUUGAUAUUUAAGCAGUAAUGUGUGGAG
GGGGAGAGAAUGCUACAAAACGUAAGAGAUUUGAUUUUUAUGUACCUAAAGGAAUAAUAUUCUAGGGUAGUGGGAGUAAGUGAAAA
UUGGGGAAAAUAUAUGAAUUGUGUGGAAUGGUGAACACAGUUAAGGGGAAAUGGAUGCAUAAUUGAGCAAGGUUUCAUGCCUGGAU
CUUUCUUUCCUAUGCUCGGGAACCGUGUCUUCAGCUUUGUGCGUGUGCAUAUGCGCCUCUUGUGCUUUUACAGUUAUGAUGUAGGG
AUAACGUCAUCGGUUUCGAAGGUUGGAUGUGUAG

>Tb927.9.12550_Glycerol kinase, glycosomal

UAUUCUGAGGUUAUGUGUGCUUUGUGAUUUGCGUAAUGGACUAACAGGCUGAACGGUAUAUCUUUUUUCUAAAUCAUCUCGAAGAU
UCUGUUUGUUUUUAAGAGGAUUAUCUUGGGACCUACCACGUUUUGUUAUGUUUUCCCAUUCCAAUAUUUUCCCUUUUUGUUCAACA
CUUUUUUAUUGUUUCCUUCGGAUUUGGUCACUAUUGCUCCUUCUUUGUUGAUAUCCCUAUGUUUUGUUGUUUUCGCCCCCCCCCCC
ACACACACACACACCACUAAGUACGGUUGCCUUCUAUAUCAACCACGUUCCGGUCAUAUGACCUCGUGGAAUAUCACUUCUGUUUC
GUGUCCUUGAAGCAUCAGAAGUGGCUGCGGUUUCUGUUUGUAGGGAGUCGGCUUUUUGACUUGGUUCAUUCCGUCGAGGUUUGACA
CGGGGGAUGGCACCAACUGUAUUUGCUGUGGGGUUUAUACUUGGCGCGGUUUCGCACACGCGGAUGCCGUGAAAAGGAAUGCCACG
GCAAUAAGUAACACUCAAGACGAAUAUUGAUUUUUUUUUAUUUCUCCUUUUUCUCUCUUUUUUAAUGUGUGCCGAAAUAAUAACAA
UAACGCCAACGGGCCAUGAUUUGGCUGAAUUGGAGGAGUGGAAAAGAAGAAUGCAUUCACUUUGUGGUGAAGUGUGAUCCUCCAUC
GC

>Tb927.2.3270 65kDa Invariant surface glycoprotein (ISG65)

ACUCACCUUCACUCUUAGACAAAUGUGCAGUUUUUUUUUUUUAAACAGCGAGUGUAAAUAAGGAAAACCAAAGAGAAUAAAAAUGA
UAAUGAGUUGAAUAAAAGUAUAUAGAUUCAUGAGAUAAGGAAGAAGGUCUAGCAGCCUUUUCGUUGUUUUUGUUUUGUGGUAAUAC
CUUACCUAUAUUUUUUUUAAAAAUAAAGAAUUGACACAAUAGAGUAAAUAUUUUCGCUUGAAGAAGCCGUUGUUGAUUUUCUCUUC
CAGUCGUAAAUUCAGCUGAUCGUUUGAAGCAUCCACCUCUCACACUGUACGUCAUCAUGGCUGGCGUGCUUCAUGAAACUGCCACA
ACCGUAUCAUACAGGGGAGCGCUAUGCAAAUUAGCAGGUGUUGAGCUACAAGGGAUUAAUUUAUCACCACCUCAAAGGCUGUGCCA
CUUGGGGUUGCUGAUCAGUCGUGAGAUCAACUGGCUGGCAGCAGUCCUCGAGGAAUGGGAAGAAAAAUAUAUGCUUUUGACAUGUG
UGCCGUACCGUGGUGUUGUGGACGCGUUUACAACCCUGCUCAGGACAAUUAAGCGGCACCUGGAGACAGAGAACAGUUCAAAGGAC
AAUGAAAAGGAGUGGGAUGCGUUGCUGGCUUCUAGCAUUACUGAAUUUCAGGUGGAGGUUAUGGAAAUCGGAAAGGGUAAGAGUUA
CUAGCAUUUUUUUCUUUUCUUCCUAAAUGGACAUUUAACUUUUUUUCCUCUUUUUUUUACGGUGGUAUAUCCUCUUUUUUUGUCAC
CCAAUGGUUAAAUUACCAUGCGGUGACAAAACAAAGACUCGGUUUUUAAACUCAUUGAGAAUAAUAGAGUACGCAUACCUAAAAAA
AAGUGUAUGCGGAGUGGAUUAUGUAUUCUCAUGAAAUCAAUCAUUCAAGUAUCUAAUGCAAAGCGCAAACAACUGGACGACGGAAA
UUUUUUUAAUUUUUGCGGUGCCUCUAUAAAUAACUCCUUCACUAACCGUUAUUGCUCGACGUAGCAUCACUAGGUAAACCUUUACG
CGUUCGUCUGCUUUGAAUGCACUAGUUCAUACAGCUCCCUAAAGUCUAACUCCGUCUCAACUCUUUCCCCACUCUUACACAUCGAA
UAGUCUUUGGCAAAAUACAAACCUUUGGGAGUCACACCUCCACAAGAAAUUAGUUAUAGCUAGCCUUCCCCCACUGCCCACAUUAU
GCAUCCUUAUCAAUUGCCUCGGCUGAUAACUUAAAUUGACAGCGUAAUUGAACUAUCCGUCUCUGUUAUGAGUCCCUGAUGGCAGU
ACUCGUGGUUUCCAUAGGAGGGGGAACGAUCCGUCCGAAUAAUCCAAAUUUGAGGGGUUAAAGGGUAGUAAAAUACAUUACUUUCA
UCAGAUGAAGAACACACGAAAUGACUUAAGGGUGCGAGAGUGGCAUUCUGAACCUGUAAAAUGUGUACCUUGCCUGCUUCUUCAUG
CCAUGAAGUGUAAAUAUAUCCCAAACAUUCAAGUCUGCUUUGUCGAUAUAUUUGUCACAGAAGUGUAGUAGGUUACAUGGUCUGUG
UGAACUGGGGAUCGAAAGCUCUCAGGUGGCAAUGCUUGGCGGUGUAAUGUAUGCAUGGCCGCAGUGUUGAAUACCAUUCCUUAUAU
CUACUCGAUGAGUUCAGUUUUGGCGAAUUAGCCUAUGAUUGUGUCUUCGAUGUUUUAUCUCGGGGUUUAGUAGAGGUGGACACAUA
UAAAGUUAUGGGGGCGGCUGUUACCUGUGGGAGUCCUAUUAUAAAGUGGUAAGUACUGUGCCUCUUUUAUGGGUGCUGUCAGACAA

GCAUUUUGUUUUGUGUGGAGACGGACUUUCCUCCAUUUGCUUUGGAGGGACGAUUUAAGUCACUGCUGGUUUCUGUUGUGCAAUAG
AAACGCGAGAGGCAAAGUGUGUCUGUUGAAGUGUGAUUAUUCUUUUUGGGGCGUCUUCAUUACCGGGAUGUUGGGAAUGCUGAUUA
CUGCUGUUUCCACUGUAGUGAGGCGAAUAAACAGCUGCAAGUGUGUUUCGGAUCCCUUUA

>Tb927.11.7620 Major surface protein MSPA

UAAUAUUAUGCGUAAGCGGCACUCUUACUGCAAUUACUUUAUGAUGCCUUUCUUAGCUGGAUUUCACAAGCAUUUGUUGAGAGUUU
UGGUGUGCUUUAUUAACAGUUUACUAACCAAUCUUUCUUUCUUUCACAUGAUGUACAUACAUGCAUUUUCAUGAAUGGGGAGCUGG
UGUCGUUUAUCACGUCAUAUCUCUUUUACGACAGCAUCACACGACACUUGUUGCUGUUGCAGUGUAAAAGAGUAUUAAUAAAUAUA
CAUGAAUAUUGAAUGUUUGUUUUAUUGCUCUUAAAUUGAAAAUUGUGUGGAUAUAUUGAAAGGAAUAAUGGAAAACCUACAUAAUG
UUUUGGGUACCGUUUUACCUGGAGGUGACGGUAUUGUUGAUACACAUCAGCAACAUUGUUCUUCCGUCUCUUUUUGUUUCGUUUGC
GCACAUCACAAUAAUUUAUGUGUGUGAAAUAUCAAAUGGUUACCUGCUUGAUAUAGAUAGAGCAGCUUAAACUUUCAUAUCAUGAU
UUGCAUCAACGAUUAUUGUGACCUGUCAUGUUGAUUCUUAUAUUUUUCUCCCUUGCCUUCUCUCUUUUUUUUCCUCUGCGAAGGUG
GUAACAGCCAUAUGUGAAGAAUAUCUCUUUCUUCUAUCUUUCUACUUUGAUUUCUUUUCAUGUACGUAAUUCCUUGGUGUGUUGUA
UUCUCACUGGUUUCUUUCAUUCAAUGUGCAUGAACCUUUCAUUUGUUUCUUGUUAUAAGGCGAAAUCAUUCCCAUGUGUGUGAUGC
GCUGCUCUCAUUGGACCCUCAUGAUCUUUUUUUUGUAUUAUUUUCCCAACCCUCUUUACUUGAGCAUGAGAUUUUCACACCUUUAU
CAUGAUUCAUA

>Tb927.3.2610 Glycosylphosphatidylinositol deacylase 2

GAAAAUAAAAGCGUGUAUAUUAACUUUGCUUUUGGAAUCACGGCUCUCGUGUUCCUUUCCUUUCCUUUUUGUUUUUUUCGUUUUCU
UCCUCCUCUCUCCCUGUCUUUUUUUUUGUUGCUGCUUGGUGGACUGGCGGCACACUGGGAGACAAACGACGCUGCAGUGGUGUCGU
ACGAGACUUUCGUAGUCGACCUCCAACCUGCGAUGAAAAUCCACACUUGUCUCGGCAGGAAGGCGUGAGGGCCAUAACUUUUUAUA
AUAUUGCGGCUACCCAAUCCAUGGUUGACGUUUGUCGUCUUCUGUUUUUGUCGGUGUUGUUUCAUUUUUCUUUUCUUGCCGCUCCU
UUGCAUUUCGCACGCAGUUGUGUAUGCUGCGACGGUCUCUCCUUCCCUGUGUUCGUGUGUUUGCUUAUCCACGGCAUACUUG

>Tb927.10.4780 Glycosylphosphatidylinositol deacylase 2 PRECURSOR

UGCGUUAGAGUACAUCGGGCAAACCUGGUGGCGGUUCCUUCUGCACCUGUGAGGUUCCCCUUCUCUUUAGCGCACUUUUGCAAAAA
UAAAAAAUAAAAAUCGCUGCCUUUUCGUUAUCUCUUCUUGAGGUACCGAUUGAGCGUUGAAAAACGGUGCAAGUGAGUCUGAUAUG
CAUUGCCUUUUUUUUUUUCGGAUUGUUGUUUCCCUUUGCCCGCGCCGGAAUCCAUAGCCGCGAGGGGGUAUGUCACAUACCUGGCG
UGUCUUUCUGUUUUUGUUGUGGGCGUGGAACCGCAUAUUUUUCAAUUCGGACGUCGUUUUACUAAUCUUGUUUGACGCGCUUUCGG
UGCAGUGCAGUCUAUUUGCGGUAAAUAUUUUCUUUCCUUCACCUUUAUUGCAACCUACGCGGACCGCACAGGAUUUUUUACUUUGU
CUUCGGGUUCAUUUCACCACGUCCAAAUUUUUUCCGCUUCUUCUUUUUCCUUUUUUUUCUUUCCCCCCUCUUGUGUGUGUGCUAUC
AGUUAUUGAUCUUUUACCCGGGUGCUACAUCUAUCUCUUCUUUAUGGACUCGACUGUUGAUUUUACUGAUUUUUCUUUCCCCAUAU
GGUGGUGACGGUGGCACUUCUGCUGUUGGUGGUUGGACCGGGAAAAAAACGUACCGAGCGUUAGAAGGAGUUCACUAAAUUGUUCC
AAUGAGUUGAGCCGGAGCCUUCACAUGAGUUUUUCAGAUAUCUUUUCUUGCCUUUUCGCAUAUGUGCAAGGUACUCAGUGUUGCGU
AAGCUGAGCCGUUAGCUUCACCCUCACGAUUGGUUCUUUAUAUUUGUUGAAGUUUCCAGCCGUCUGGUCAUGGUUACCGCCAGUUU
UGAAUGUUGCGGUUGUGGGCAAUUGAUAAAGUCGGUGGCAUAUAUUGUUUUUGCUUUUCCUGCUGACUGUGGGUUGAAAUGAUGAG
AAAGUUUGGGGAGGGGAUUAAAGUCACGUAAGUUUUGCGUUGUCAGACAACCACCCACAUUUACGUUGCAUAUAGUCUCUA

>Tb927.1.4650 Cyclin-like F-box protein 2 (CFB2)

UGACUUCUCUUUUUGUUGAUUCGCUGCGAAAAUUUUCAUAUAUUAGUGUUUCUUAUGCUUCGGUAUGAAAGUUUUUUCUCUGUCGU
AAUGAUUCGUUACGACCAUUGCAGCUGAGUCAUUGAGUUUGUUCUCGGUGCGCGUGCCUUUGUGUGUUUUAUAUCUUCUUGUUUUA
CUUAUGCUCGGAAGUGCAUCUUUCUUUCAUGUAUGACGGAGUGUUCGUCUCUCUGUACGGUACUUUCUUGCUGGAUGGAAUGCAGU
CUUUCCCUCCUUUUCCUUCUUUUUUUUUAAUUUUUUUUUCUCUUCAUUUAUAUUUUGUGUGUGUGUGUUUUUUUUUUAUUAUUU
AUAUUAUUUGAUUGGAAUGAAAGAGGCGCUCGCUUGGAUGGGAUAGUGUUGGGAGAGGUGUUUUGCUGUUCCCUUUCUCUUCUUUU
UUUUUUGUUCGUGGGGCUGUUUGGAAGGCGGGAAAUAAUAUAAUGAUGAAAAUUGUAUAAUGAUGCAGUUUUAUACCAGUUUAUGA
AAAAAAAAAAGAGGAAGAAAAGUCGGGGAAUGGAAGUGAAAAAGAAAAAAUAUAACAAAAGCAAAAGCAUGCGCACAUUUGUGUAU
GUAUGAUGAUGUCUUAUUAUUGCUAUUAUUUUGUGUCUACCAUCUUUGCUCGAUUUUCUUUUCUCAUCUUUGCUUUCGCUACAUUU
ACUGAACUGUUUUUCUCUUAUACACAAAUAUAAGUAAAUAUUUAAAUAUACCUUCAUGAGAAAAGAGAAAAGAGAAAAGGGAAAAG
AGAAAAAAGAAAAAGAAGCGAAAGAAACCGAAGCGAAAAAAUGUUUCGGCUCCGCUUUCAUUUUUCUGCAUCCUCCUUUCUCUCUG
UUUUUAUAUCUUUUUGAUAUAUAUUUAUUUAUUUAUUUAUGUGUGUACGAUGCAGCAGUCGGAUACGUUUUGCUAAUGUAGAACUG
AGAGAGCAGGUUCCAUUAUUAUUAUUAUUAAUAUUCUUAUUAUUAAUGAUUUUAGUGGGAAACUCUUCUUUCUUUUUUUUUUUGUU
CCCGUCGCAAGAGAAACAGGGAAAUAAUUGUGUGAAAAGGAAAAAUAAAUAAAUAAAUGAAUAGUUGGAAGUGCAUGGCCUUCUGU
UUUCCUCACUGGAAUUCAUUUGUUCACCUAAUGCAAAACUCAAAAAAAAAAAAAAGCAAGAAAUGAACAGAAGAGAGGGGGAAAAAA
AUGAGGGAAAAAUAACUUGUUUUUGGGAGCAAUUGGGACGUGUGCGACGGGGAGCUGAGCUGUUGGAAUUUCAUUUUAUAUUUGAU
UUUUCCAAUUCACUUGACUUCAUGUUAAGUCGGAUUUCGGUUGCAGCAGUUUUAUUUCAUUUCAUCUUUUUUGUGUGCUUUUUUUU
UUCUUCUCCCCUCCUUUUCACCCUUCCAUUCUUUCAUCUUUUAUUAUUUUCAUUAUUACUUUGAGUGUAAUUCUGCAUCAUAUGUA
UAUAUAUAUAUAUAAAUAUAUUUAUAUAUAUAAAUAUAUUUACAUACAUAUAUUUGUUAUUUCAUUUUAUUCUCCCGCUUUCCAAA
CAGUUUUUUUUGUUUGCGCGUGUUCUUGCGAGUUUUAUUAUCAGUUUUAUUAUUAUUAUUGAUGGUGUUUUCGUUUUUUUGUA
CUAAGUUUAUAGAGUUAUGGGAGGGAAAAUGAAGGCACAAAAAGCAGAAAGAAAGAGGAAAAAAAAAGAAAAAAAAUGAAGCUGAA
AUGAUUUCGAGUGAAAUUUUUUCUUUUCUUUUCUUUUUUUGCUCGUUUUUAUUUCGUUUUGUUGGGUAUUAAUAAAAUAUUUGUUG
CUAUUAUCUACAGAGAAGCAAACAAAUAUUUAUUAUUAAUCUGUUCUUUCUUCUUUUUUUUCUUUUUGUGCCUUUUGAGUGGGCUU

AUAUUAUUCUUAAUAUCACUUCUCCCGAUUAUUAUUAUUAUUAAUAUAUUAUUAUUAUCCGCCAGUGAGUCGCCAUCAUU
AUUCAUUUUUCAUCCAUCUCAAGAAGAUAAAACAAAACAAGACAACAGUAAUAAAAUCAAUUUAGCAAAAAAAAAAGUGAAAAUCG
AAAAAAAAUCGAAACAUAAAAGAAUAUGGAAAAAAAAAAGAAGAGAAAAUAAUUAUUAAAAAAAAAAGAAGGAAGGGAGAGGAGGU
AUUUGCUCCUUUUUGUUUUUGUUUUUGUUUUUUUUUCCACCCACACACGCUUAAAAGUUUGUAUGACGUGCAGUCGUAUGGCUGUA
AUUUUUCUGUUGUGCAUUCUUUUGGUUUGUCUUUUUUUUUUUUUCCUCCUCCCCCCCCCCAAUAAUCUUCGUAUUCUAGUUGUUUCUA
AUUAUUCUCAUAUAAAAUAUAUAAAUAUAUAAAUGAAUAUAUAAAUGAAAGGUUAAAUAUAUUUAUAUAUAUAAAUAUAUAUUUAA
UCAGAUGUACAGAUGUAUGAUUGCGUGUUGCUGAUGAGUAGCCGCAUAUUUAAUGUGUGGCGGGAGAGGUGAUGAAUCACUAUUUG
GCACGUGCGAAAGCGAAAUGACGGUUUUCUUUUUAUUAUUAUUAUUAUUGAUGUUUUUUGUUUUGUUGCCUCUCUGAUGUAUUUCA
CCGUCAUACCACCUUGUUCAUUUUUUAUAUUUGUUUAUUUAUUAGUGUUUGUAUAUGCGUUUGUUUGUUUGUGUUUGUUUUUGUUU
UUGUGUUUGUAUUUGUGUUUGUGUAGUGAAGAAGAUUAUGAUGAUGAUGAUGAUUAUUAUUAUUAUUAUUAUGAAAAUGAUGA
UGAUAGCAAAUGAAUUAGAUAAUUAUAUGGAUAUGAAAGUGUGUGAAACGGUACACGUGAUUUGAAAUGAAUCGAUGUAAAUGAAU
UGCUAGUUUCCUUCUACAUGAGGGUUCAAUUUAUUUCAUUUCAUUUAUGCAUUUAAAUAUCUAUUUGUUUAGUUAAAAACUUGCAA
AAAAGAAAAAGAAAAAAAAAGGAAAGAAAAAAAAAAGGAAAAGAAAAAGAAAAAUUAAAAAUUAAAAAAAAUGUGAAAAAAAAAAG
AGUCAGACAGGCACAAAAACACAAACACGCAAACACACGCAUAUGCGCACACAUAUAAAUACACAUAUAGAUAAAUACAAAUAAUG
AUGGGAAAAAAAAUGUAAAUAGCUAUUACUAUUAUUACUAUUAUUAUUAUUAUUACUAUUAUUAUUAUUGUUAUUAGUUCGUCAUU
AUUAUGAGUUUUCGCUUUUUGUUCAUUCCCCAUUUGCACAAAACAUAUGAAACACACACAUGUAUGCAUAUUCGCGUCCAUUCGCU
CAUAGAAAGCAGAAAAGAAGAAACAAAAACAACAACAACAGCAACAACAACAGAAAAAAGAUUCAAAUGACAAAAAUCAUCAGAAG
AAAAAAAAAGAAAAAAGAGAAAACGUUUAUGUAAAUAUAAUUAAGAAAAGUUACGUUAGCGUUUGGAUGCGGGCGCAUAUUUAUGU
GUUUUGUGUGUUUUAUGUGUGUGGAAUAAAUGUUUGAGGAAUGAAAUAAUGAGCUGGCAGAAGCUGACGGUACGAAAUGACAGCAA
AAAACAAUAACAACACCCCUCAUCUUUUUCUUUUUCUUCUUUUUUUUUUUUUCAUCAUUUUCUGAAUGGGAAAUUUCCCUUCUUUUU
CUUUUUUUUCUUUUUUUCUUCUUUUUUUUCCACUCA

>Tb927.4.1230 Hypothetical protein

ACGGCAGUUUAAUCGACCAAUUUGCUCUAAUUUUUUCUUCCGCGAGCGUGAGCGCGCACAUACACACAUGCACACACACACACACA
AUAUAUAUACAUAUAUAUAGUUUGCACCCACCUUGUGCCGAUGUUGUGCUUGUGCAUGCUCC

>Tb927.11.7550 Hypothetical conserved

AAUGAACAACAACGACAGCAACAGCGACAGCGAAGUGAGUGGAAGGAAAGAGGGAAAAGAAAGAAAGAGGAGCCGAAAUAACGGCA
AGAAAAUCGAAUGAAAAGGGCAUGUUUGAGUUAGUAUACGCUAACACGCGUAUGAAUGUGUCUGUGGAAGUGACAUAAAUUCAGUG
ACUACGGGCGGUAACGGGAAGGAUAGAGGAGGGGAAAGACUAAAAGAAAAAAGUGACACAAGGGAGCGGUUGAAAAGAAAAGAGUU
GUGGAAUUAGUUAUUGGGACUUGUUUUAACGUGCACCAGCAGGUGUGGACAUCCUGAGGGAAGAUUUAAAAAAAAAAAACAGCGAUC
ACGAUGAAAAGAACUAAUUAAAUAAUAGUGCAUGUCGGCACCGCGUUUGCUUGGAUCCCCUUCUCAUUACCCCUUCGUAUAGCCUG
UUGGUUACUACAUAUGCUUUUGUUUGCUGUGCUUGCACUUGUUCCCCAUUAAGCCAAAGAC

>Tb927.7.4650 J biosynthesis protein 2

UAUACCCUUUCAUUUUUUAAUAUUUGUUCCAUGUGAUCACCCCUACAGCUGUGCAGAUCAAAUGUGUUUUAUUUCUCUUAUAUUUC
UCGUACAACUGCCGUGGUAUUGAUAAUUUUGAUAUUGGGCUGAGUUGUCUGUGGAAAUCUGAAGUAAGUACUAUUCUGUGCGCUUU
UCCGUAAGUGUUGUUCCUUGAUUUGUUGUUUCGGAUCCCCUACACUUUGUUAUCUUCUUGGGUUAUGGCUGUUCUUAUUGCCAACG
UCUUUCCUCUCUCGUACGAAUCUUGUCCGACUGUGUAUGUGUUGGCGGACCCAUUUUGUGUGUUUGUGUUCGUGUUUGUGUAUGUU
UUUUUUUUUUAACUUGCUCCGCUAGCUUGCCCAGACUACGCAACUGUACUCGAUCCUACGUGUAGUUCCUUCUCCCUUGAGCAUAU
UUGAGCGCUAUUUGCGGGUAGCGCAAAUGAGAGAAACAUGUGCGGUGCUUGAGGUGGGUCCUGGGGCGUCUUUGUUGUUUUUUUUC
GAUGUUCAGCUUUGAUUGGCCUUGCUCGUUUGCUUGCGUAAAUUGCUGGGCCGUGACACGGAGCGUUUACCACAAUAUCCAAGACA
CUCCGAACAAACGCUCCGGGGUAGAGUCGCUGGCAGUCGGUGUAGACUGCGGGGCUUUCGGGGUAGUUUAGGAUGUAUUACGUGAG
UGUAUGGUUUUAUAUGUAGGUUUGUUGCAUGCAUAUCCGUCCUCUUUCCUGUUGUUUGCUCGUUUUUAUUGUAAAAAUUAUUAAUC
UGUUGAUGUUCAGUGGGUUG

>Tb927.8.3720 Sumo-interacting motif-containing protein

UCGGCCGAAUGCUAUAUAAGGGCGGAUAAUUUCCGUUUUGCUUUAUCCAAACAGCGUUUUAUAACUUUCUAAGUUUCUUCUUACAA
UUACAAUCGUUUCCUUACCGCCUUCGUGGUUUUUUCCCCCCUCCUUGAGAGAACUCUGAAGAAGAAAAGAAAAAAAACGAUGCUUU
CCUUUGAAUGCCUUCGAACUUCAUUUUUUUUCUUUUUUGCACUUCAAA

>TB927.7.3730 RBP26

ACCAUCAUUGGUUGCAAAAUGAGGAGUGGCGGGCUGCUCGUGUCCACCUGUGAGUACUUAUGGUGUAGGGAAGACUUGUCUUUGAA
AGGAGAUGACCACCGUUUAAUGGAGUUGGUGUAUCUGUUGUUGCUGUCGUCUUUUGUUUUUACCACCAACGUCGUCUUUGUUAGGG
AUUGUGAGAUGAUGAUCGUCAAUUUAUCCUUAGGACUGGGAGGUGUGUGCGUGUAUAUUUGUGUCCGUGCGUGGGGAGUGAAACUG
ACGAAGUAAAUGCGAGAAUGAUGGAAAUAAAAGGGUUUGGAGUGGAGAAAAGAAAAGAAAAGGAGUGACGCGCAGCGUACACCAUA
CCAUUUUGGAUUACCGGAGUCGUUUGCCCUUUUUCUCCUCCAUUUUCUUUCGUGUAUUCCCCUUUGUUUGCUCCCCGCGUCCUGGA
AGGCCCCCACAUGUUUAUGGCAUUUUUCUUUGUUUCAUGUGUUCUUUUUGUCGUACUUCCUCUCAUUUCGUAUGUUUUCGUGUGCC
GUUUCUGGUGACCCUCUCAGUGGGAAAGGCACAAGCGGGUUUUUGCAUUGGUUUAUUUACUAUGUGCCCGCUGCCGUCUUCUGUCU
GUGCGCAGACUCAUUGUCUUUCGUAGCUCUGCGUUCUUCUGUUUGUCUUCCUUUCUCCUUUUUUUUUCUUUCUAUUCUACCAGAUU

UUAUGUCUUUCCCCCUUUGCUUUGCUUGUUUUGUUCAAAUCCCCCUUCCCCCAAGCUUUUGUUUUGGGUGGCUUUUGUUGUCGCUU
UUUGUUUUCCCCCCUUUAAAUCAUUUUUUUUUUUUCCUCCCCCCUCUCCGCCAAAUUAUCACGCGCUGGUACACAUUACUUGUGAAG
AAGUGAACAGUGCCCUCGGUUUUUUUUUGUUACCAUCAUUAUCUUUUUUGCCAGUUUUUUUCGUCGUUGACCUUUUCCUUUUUCUU
UUCUUCCGUUCCCUUGUCCAUUUUUUUUUUUUUCGUUUCUCCUUGCGUGGACACGAUCAUUAGUAUUAGUAACAUUAUUCCCGUCAU
UAUCAUCAUUUUUUUUAUGAUCCUUUUUUUUUCUUUUCAUUAAUCUGCUUGGUUCAUCGCAGUUAUUUUCCUUUGUUAUCUCAGUGG
GUCUUCAGGUGCCAAAGAGAAAGAAAAGAAGAAUU

>Tb927.7.470 Enriched in surface-labelled proteome protein 14

GUAAAUGCGUCAGGAAUGUUACAAUGAUAUGAAAUAACAAAACAACUUCAGUGUAAUGAGAGGAGGGAUGGGAUACGAUAGCAGGG
CAGAGAAAAAGAAAAGCGAAGAAAAAAAAACGUAAGUGGAAGCGGUGUCGUCCACUUUGUGCAUAAACCUUCCCCGGCGUUUCUGU
UUUGUCGGAUGACCCGCUGUUUAAAAUAAGGUGACGAGUAACGUAUUUGCGGUUCUGCAUGGAUAUCUUUUAAAUGUUCUUUUAAU
GUGCCGUUGUGUGGAAAAUAAAUACUAGGUUGCAGGAGAUGAGGUAUGGGAGAACUGGCGUUCAAGAGGGCUAAGAGGGCAAAAGG
AAGGAGACGCCAUCAGUCCAAUUGCUACAUUUGCUGUGUGCGUGUAAUUUAGGGGGAGGGGUUGUGUUUAUUUCUCCCCAAUUAUG
CGCCUUUCACCGAUUUCCUCUGUGGAGUGAAAGUCAUUAUUUAGUAACCGUUUUGUUCAGGAAGAAUCUCAAUGCGUUUAUUCUGA
AGGUGCACAUUAAGUAUUUCCUAUCAUAGGUUGCUUUUUUUUUUUUGCUUCCUGCGCCUUCCCCAUGACGCCUUAACUUUGCUCGG
CGUAUUGAACUUUGUGCUUCUAGGGUAAGUGAUUAGCUGUGUGGUGAAUACGCUGUUCUUUUGGAAGUUGGGAAGGGGAGAACAUG
ACGAAACCGAUUGAAAUGAUGGUUGCUCUCCCUUAUUAGUUAUAUAUAAAACUAUGUUUAUAUAUCAGGGGAGAGAGACACGUAAA
ACUUAGUUUGACCCUUUCGCUUACCCGUGGAGUUGCAUUGGUUGUAU

>Tb927.8.7110 Serine-threonine protein kinase NEK12.1

GUGGUUCCCUCUUUUUUGUUUUUUUUUUUCUUGUCCUUCUUCUUCCUUUUUUUUCCCCCAUCUUUUUUCCCUCUUUCAUUUCUUCU
UUCAGUUUUGUGGCAUUUGGUGUGUAAUGAUUGGAACGUUGGAACACAAGACGAGUUGAAAUGACAAGAAAAUAUGUCUUUAGUUU
UAGCCCCAUUUCCCCUCAGUAAAAAAAAAAAAAGUCUAUCAAAAGCAGCGUUACCGGUGGAUGAGAGGUGCUUUCUUUUUCUUUUUU
UUCUUUUUUUUCUUUUCUUUCUUUUUUUUUUCUUUUUUUUUUUUUUACUGGAAACACAAGGGUAAUAAUUAGCAAAAAAAAAAAAGAUG
ACAAAUAAUAUUGGAAACAUUUGUUUUAAAUGUAUACAUGACCAGAAUUUUCUUCUGUUUUUUUUUCUUUUUACCGCGUGCCCUUA
UGCAUGGAAAUGGGAAAUUACCAAUUACACUUACUGCACACAUGUUAAUCGGAUAAAGGAAGGGUUGAGGAAAAGAAACAAAUAAA
UAGCACUCGAGCGAAAGAGAAAAAGUAAAAUAAAAUAAAAUAAAAUAGAAAGAAAAUAAAUAAAUAAAAGAGAUGCAAGAAUUAAA
AUGAAAACGAUAAUAACACUUACAGUUCCGAUGAAACGGAGGGAAUGUAUAAUAAAUAGCGAAAACAAAAGAUGAGGGGACAAAAA
GUAAUGUUACAGAUUACAAAUCAAACAUUACUUGAAGGAUGAGGAAAAUUUGUUGUUGUUGUUUUUGUUGUUUCGCAUUGCUGGAA
GAAGAGGGGAAAAAAAGCAAUGGGAAAUUAGAAAGUAGGAAUGAAAUGAAAUGAAAUGAAAUGAAAUAAAAGCAAAAGGGAAACCG
UAUACGUGUUUAACCGGUGUAUUUCAAGGGAUUUAACAAGGAGCAAAAUUACAUCCAUAAAAGAAAACGGACAAACAAAUCAUAAG
GAAUAGGGGAAAAGAAAAAAGGAAGAAGAAAAAAGAAAAAGGAAGCAGCAUUUAACGGAUUAAUGCUUACAUUCAUAUUUAUAUAU
AUUUAUUUCUUUUUUUUUUUUGAUGUUGGGUUGGGUGACUGUAUUUCCUUAAUUUUCUUCUUUUUUUUUUUGUUUGUUUGGACAAA
ACCCAUCAGGUUAUAUGCUCAGUACUGCUACCGUCAUUACACUGGUGGUUUCUUUUGUUUCUUGUCUCAUUUUCACUCAACCCCCG
AUGUUCCAUAUUUCACCUUUUGUUGUUUUUGCUGCUGUUUUGUUUUUUUUUUCCUUCUUUUUUUUAAAAAAAUUUUAUCUUACCUU
AUCUUAUCUUAUCUUUUUUUCUCUUUAUUUGACUUGUCUUAUUUAUGCUUAUUUGCUGCGACUGUUUCUUGUUUGUUUUGUUUUGC
AUUUUUUUUUUCACCUCUUUAUUCUUCUUUAGCAGGCGUUGUUUAUUAUUUCAUUUGGUCUCAUUCCUAAGCGUAUAACUUGUGCA
UGUGUGUAGAGGCGUGGCGGACAAUCUUUUUCUCUUUUUUUUUUCUUAUUAUUCUGUGUUGUAUUUUCACUGUUGUUUUAUAUCUU
CUUGUUUAUUAAUGAAUUUGACUAGAUGGGGGAAAUAAACCGAAACUUAUGAAUGAAUAAAUGGAGUGAAUUUCCCAACAACGACG
AAAGCGAUCAUGUAAAUAUAUUUGUGCACUUGUAUGUAAAAAAAUAAAUAAAUAAAUAAUAAUAAUAGUAUAUAUAAAAGGGA
GGAGAAAAGGAAACAAUUGUUGUAUUUAUAUUGCUUCCUCAACAUGAGAAGGAGGUGUUGACUGAAAAAAAUUCCAAAAAUAAAAA
UAAAAAGAUAAAAUAAAAUAGAAGAGGGAAGUGUGCAAUGGUAAAGGGUGUAAAGGGAAAAUAGCGAUAGAAAGUCGUAGUGUUCG
GUGCCACCGUAAUUCUCCUUUUGCUUAUUUGCUUUUUAAUUUUUACUUUUUACUUUUUUGACUUGUUUUGUUUUGUAUUGUUGUUG
UUGUUGUUGUUGUUGUUGUUGUUGUUGUUGUUUUUUGGUCUAGCGUAUACGUACAUAUGUUAUAUGUAUCGCCAUAUGGA
GAAGAAGAAAAAAAGGGGGGAAAAGGAACAUUCCGUAUAACUCAAUUUCUGCUCGUACAGGCAUUUGUGUUUUCUUUUUUUAUUGU
UUUUAUUGUUUUGUUGUUUGUGUUUUGCGUGUGUCCGAAUUCGGCUCAUCGUACCUUUGGUUACUGUGGGCUUUUGUUGUGUAUUC
UUCGCGUUGCGGAUUCCUUUUUUUUUUUGUUUUGUUUUGUUUCUUGCUGCGGUUUCGUCUUUGGCUGACUUGUGACCUCCCCCCCCC
CCCCACCCCCUCACCAUUUGUUUCUUGGAUGCUUUUAAUCCUUCAUUUUUUUCUUUCUUUUUUUUUUUCCUUUCAUUCCCCAUGAU
GGAUGUGAUGUGAUAUCUUUCCAUCACUUACACUCUCCCAACCAAAUCGUCACCGCGCUCAUGAAAGAACUGCUUCGUGUUCUCCU
CAGAGUGUUGUUGUUUUUUAUUUUUGUUGUUUUUUUUUGUUUUUGCUGUUGCUUAAUCCUUAGUGUUUUAUCUUUACAUAAGUUUU
UUUUGUGUGUGUGUAUAUGCGUGGAUUUCAAAGAAAAUGGAAAACAAAUACAUGAUUGAAGAGAUGCUAAGACGAUGUGUCUUCCG
CCCUUUCAUUUCACUUCAUGUUUGUUUGCUGUGUUGUUGUUGUUGUUGUUUUCUUCUUCGUCCUCGCCCGUCUGUUGUUUGUU
UGUUUUUUUUUCUUUUGGUUGUUGUUGUUAUACGUCUCUCUUUCUUCCCACCGGUACACUUGAAAUGAAUGGAAUAGUUCCGG
UGUGCUC

>Tb927.5.630 Acidic phosphatase

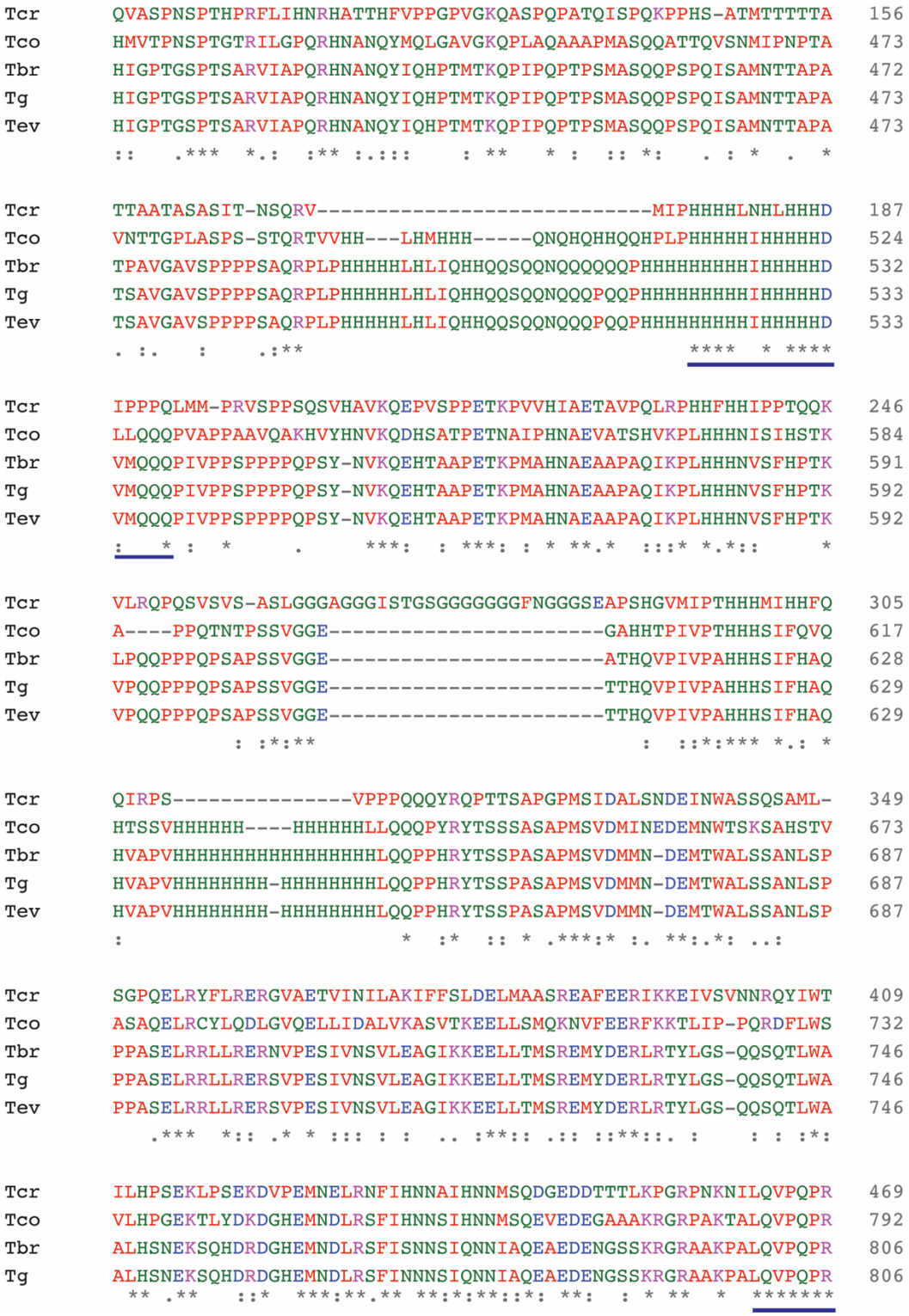
ACGGAUGGUUGAGACAAAACAUAGGUGGAAGUUGAAUUAAAAUUAAUGUAGGGGACCUUCCGCCUCGUGCGAAGUACGUGAGUGAA
GCCAAAAGAAUCCAGAAAAACCAGAAAGAAUGCGUGAAAGAAUGAAUAAUCUAUAUAAGCGAGAGGAGGAGGGUUGUUUUUUGUGU
GUAUCGGACAAUGUUACAAUACUAUUGCUUACUUAUUGGCAACGACGGGCAAAAAGACUUUGGUGCAGCUGCCUCUCAUUGAAGGA
GGGCAGUCUGUAUUCGGAAGGCACACUAUUCACGCUUAAGUUCGCAUUUCCGUCUCGUUUUCUGCUUUUUUUUUUUGCUUUUUCGU
UACCUUUCUAUGUGUUGUUUUUGUUUUGCUUUUUUUUUUUUUUGAACAGCGCCUACCAUCAUUAUACGUAUACCUUUAUUCAAGUGG
CCAGUGGGAGGCCACAUAAAGGUAAUAUUUUUGCUUGCUGAUCUACAGAGAACGGGAGCUGCGCUUCCAUCUAAAAGCGGAGGGGA

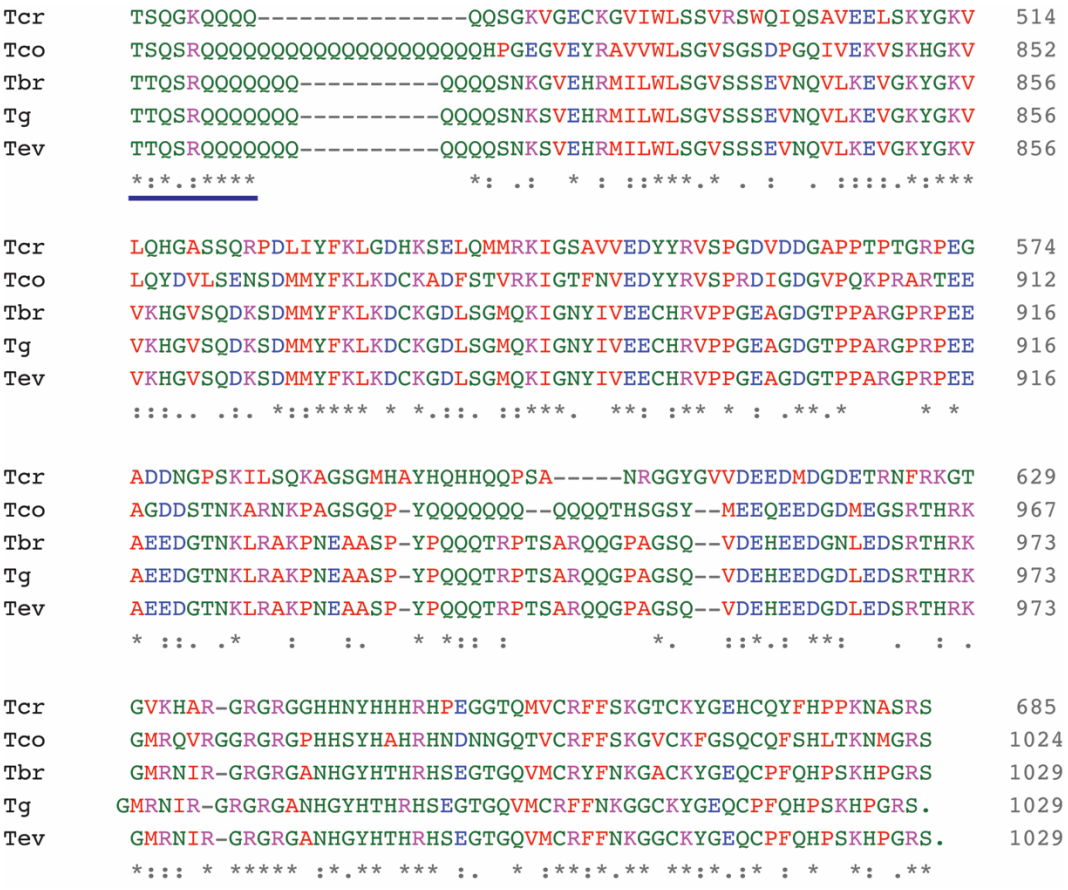
GGAAUAUUUUUUAUCCUUCGGUUUCUUCUGGAAAUUCCUUUCAGAAUUAAAUCACGGUGCUCUGCGAAAGGAAGCUUCGUCUUUGU
CUAGAAAAAAAAAUUGCUAGUGAUCUGAUUCAUUCGACCGAGCUGCAACAAGCUCUGGAUGUGGAAGCGCCUCGCCGCUACAUCCCG
CCUACUCUCGUCACCUCUUUCCUUAUGACAUGCUGUGGAGGCAAUAUUUCUUCGCAACACCCUUUUCUCCUCCAUCUUUUUUAUGC
UUAGCCUCUAGCUUCUGGCAUAAAUGUCGUUAUGUUACACCUUUUUUUCGGUACGGAAACUGACAACAGAUUAGAGGCAGACGACU
CAACACUCAUCGGAAAAUGAACACAAAAGUGGUCUGUUUCGCAGCUGGGGGCGCAGCACACAUGUAUGAAUAUUCACAUCCUCCCA
UUGUUCAGCAAACGUCAUUAGAGGUGAUGUAUAUUCCGUUGCAUGUUGAUGCUUUUCAUAAUGUGUAUGGGUAUAACCUGAGAGGA
AGACAGUGGUAAAAACGCUGAGUCGGCCCCUCACAUUUUGUU

Supplementary 5

Sequence analysis of ZC3H28 and its homologues from selected Kinetoplastids using Clustal Omega with default settings. The asterisk (*) indicates a fully conserved residue and colon (:) marks conservation between amino acids groups with similar properties. Period (.) highlights conservation between amino acid groups of weakly similar properties.

```
Tcr      ------------------------------------------------------------  0
Tco      MRS--SEKETPGAHILLPASVGEQQVEQDEGQLQAGAPARGEEEGKSNENVPPVLAHLQQ  58
Tbr      MYSSEKEKEAPETLTSLPANAEEQQRERDDNHS---ASAQ-GEEDKNQDNTPVALLHLRQ  56
Tg       MYSSEKEKEAPETLTSLPANAEEQQRERDDNHS---ASAQ-GEEDKNQDNTPVALLHLRQ  56
Tev      MYSSEKEKEAPETLTSLPANAEEQQRERDDNHS---ASAQ-GEEDKNQDNTPVALLHLRQ  56

Tcr      ------------------------------------------------------------  0
Tco      NFGVMSRVLEREVYRRGLVQSTVEEMCQVFSTASMVLVDILRSEIEHRLLRSQMISMIYT  118
Tbr      NFGVLTRVLEREVFRRGLINSTVEDMYSAFSGASVVLVDILRSEVEPRLLRSQLGDMLYS  116
Tg       NFGVLTRVLEREVFRRGLINSTVEDMYSAFSGASVVLVDILRSEVEPRLLRSQLGDMLYS  116
Tev      NFGVLTRVLEREVFRRGLINSTVEDMYSAFSGASVVLVDILRSEVEPRLLRSQLGDMLYS  116

Tcr      ------------------------------------------------------------  0
Tco      ALEASYGPDHARMLAMVIDNSFDDLYLFNAVLSPTTLDSLVQQTQEWMNTSSSRAGSSST  178
Tbr      NLELSYGAEHARMLATVIGSSLDDVYLFHAICSPSALEALVHQAQELINPPSNRTGGSNT  176
Tg       NLELSYGAEHARMLATVIGSSLDDVYLFHAICSPSALEALVHQAQELINPPSNRTGGSNT  176
Tev      NLELSYGAEHARMLATVIGSSLDDVYLFHAICSPSALEALVHQAQELINPPSNRTGGSNT  176

Tcr      ------------------------------------------------------------  0
Tco      YNDSVRGGAQAQSYGNTSGTARPSFSEFGKDRRVDIQHQTGVDGMLPRGPAAYGDSVPAG  238
Tbr      YSDGARA-VQSQSYGNAVGLARPSFTEFGKERRGEVQP--TVEAKLPRVSNTYADGAAPA  233
Tg       YSDGARGTVQSQSYGNAVGLARPSFTEFGKERRGEVQP--TVEAKLPRVSNTYADGAAPA  234
Tev      YSDGARGTVQSQSYGNAVGLARPSFTEFGKERRGEVQP--TVEAKLPRVSNTYADGAAPA  234

Tcr      ------------------------------------------------------------  0
Tco      GINQEGDAMLPA-VSGGWTNPQRRHVEPEREVPLHHGHNHQLRSRPPMNQFTVPQRTHAP  297
Tbr      GMNQPEEEVIPSVTTGGWSNAQRKHVEAEREVLPHHAHHHQLRARLPVNQFTAPQRPQVP  293
Tg       GMNQPEEEVIPSVTTGGWSNAQRKHVEAEREVLPHHAHHHQLRARLPVNQFTAPQRPQVP  294
Tev      GMNQPEEEVIPSVTTGGWSNAQRKHVEAEREVLPHHAHHHQLRARLPVNQFTAPQRPQVP  294

Tcr      ----------------LVRWMQQQQQQQQQQQQQESPQTDSFPRAPVPQVQQKRMVERLP  44
Tco      SVG----HHHHHHQVVVGTGSLGFQQQDRQQQQQDIVQTDTFPRAPVPHGLSKRPMDRLQ  353
Tbr      SVPHHHHHHHHHHHVVVGTGPVGLQQQDRHHQQLENPQNDAFPRIPVATGLPKRPVDRVL  353
Tg       SVPHHHHHHHHHHHVVVGTGPVGLQQQDRHHQQLENPQNDAFPRIPVATGLPKRPVDRVL  354
Tev      SVPHHHHHHHHHHHVVVGTGPVGLQQQDRHHQQLENPQNDAFPRIPVATGLPKRPVDRVL  354
                         :       ***::::** :  *.*:*** **     ** ::*:

Tcr      PGT---QESDEPHVRGPFWRVPPAPSGTQQETVTVPQPSYHAQNPSSTPTPHSN----QH  97
Tco      HQQAMQGGEDAMHARGVWRGASSSPAPQQQGAPGIPQASYNVRNNTSPLPPHATHGQPLH  413
Tbr      HQQSLQGHDEASHQRGAWRGMGPSPTPQ-QGAAGVSQPNYNVRNNSPPVPSHVNHNQHPH  412
Tg       HQQSLQGHDEASHQRGAWRGMGPSPTPQ-QGAAGVSQPNYNVRNNSPPVPSHVNHNQHPH  413
Tev      HQQSLQGHDEASHQRGAWRGMGPSPTPQ-QGAAGVSQPNYNVRNNSPPVPSHVNHNQHPH  413
```


Tcr QVASPNSPTHPRFLIHNRHATTHFVPPGPVGKQASPQPATQISPQKPPHS-ATMTTTTTA 156
Tco HMVTPNSPTGTRILGPQRHNANQYMQLGAVGKQPLAQAAAPMASQQATTQVSNMIPNPTA 473
Tbr HIGPTGSPTSARVIAPQRHNANQYIQHPTMTKQPIPQPTPSMASQQPSPQISAMNTTAPA 472
Tg HIGPTGSPTSARVIAPQRHNANQYIQHPTMTKQPIPQPTPSMASQQPSPQISAMNTTAPA 473
Tev HIGPTGSPTSARVIAPQRHNANQYIQHPTMTKQPIPQPTPSMASQQPSPQISAMNTTAPA 473
Tcr TTAATASASIT-NSQRV----------MIPHHHHLNHLHHHD 187
Tco VNTTGPLASPS-STQRTVVHH---LHMHHH-----QNQHQHHQQHPLPHHHHHIHHHHHD 524
Tbr TPAVGAVSPPPPSAQRPLPHHHHHLHLIQHHQQSQQNQQQQQQPHHHHHHHHHIHHHHHD 532
Tg TSAVGAVSPPPPSAQRPLPHHHHHLHLIQHHQQSQQNQQQPQQPHHHHHHHHHIHHHHHD 533
Tev TSAVGAVSPPPPSAQRPLPHHHHHLHLIQHHQQSQQNQQQPQQPHHHHHHHHHIHHHHHD 533
Tcr IPPPQLMM-PRVSPPSQSVHAVKQEPVSPPETKPVVHIAETAVPQLRPHHFHHIPPTQQK 246
Tco LLQQQPVAPPAAVQAKHVYHNVKQDHSATPETNAIPHNAEVATSHVKPLHHHNISIHSTK 584
Tbr VMQQQPIVPPSPPPPQPSY-NVKQEHTAAPETKPMAHNAEAAPAQIKPLHHHNVSFHPTK 591
Tg VMQQQPIVPPSPPPPQPSY-NVKQEHTAAPETKPMAHNAEAAPAQIKPLHHHNVSFHPTK 592
Tev VMQQQPIVPPSPPPPQPSY-NVKQEHTAAPETKPMAHNAEAAPAQIKPLHHHNVSFHPTK 592
Tcr VLRQPQSVSVS-ASLGGGAGGGISTGSGGGGGGGGFNGGGSEAPSHGVMIPTHHHMIHHFQ 305
Tco A----PPQTNTPSSVGGE----------GAHHTPIVPTHHHSIFQVQ 617
Tbr LPQQPPPQPSAPSSVGGE----------ATHQVPIVPAHHHSIFHAQ 628
Tg VPQQPPPQPSAPSSVGGE----------TTHQVPIVPAHHHSIFHAQ 629
Tev VPQQPPPQPSAPSSVGGE----------TTHQVPIVPAHHHSIFHAQ 629
Tcr QIRPS---------------VPPPQQQYRQPTTSAPGPMSIDALSNDEINWASSQSAML- 349
Tco HTSSVHHHHHH----HHHHHHLLQQQPYRYTSSSASAPMSVDMINEDEMNWTSKSAHSTV 673
Tbr HVAPVHHHHHHHHHHHHHHHHLQQPPHRYTSSPASAPMSVDMMN-DEMTWALSSANLSP 687
Tg HVAPVHHHHHHHH-HHHHHHHHLQQPPHRYTSSPASAPMSVDMMN-DEMTWALSSANLSP 687
Tev HVAPVHHHHHHHH-HHHHHHHHLQQPPHRYTSSPASAPMSVDMMN-DEMTWALSSANLSP 687
Tcr SGPQELRYFLRERGVAETVINILAKIFFSLDELMAASREAFEERIKKEIVSVNNRQYIWT 409
Tco ASAQELRCYLQDLGVQELLIDALVKASVTKEELLSMQKNVFEERFKKTLIP-PQRDFLWS 732
Tbr PPASELRRLLRERNVPESIVNSVLEAGIKKEELLTMSREMYDERLRTYLGS-QQSQTLWA 746
Tg PPASELRRLLRERSVPESIVNSVLEAGIKKEELLTMSREMYDERLRTYLGS-QQSQTLWA 746
Tev PPASELRRLLRERSVPESIVNSVLEAGIKKEELLTMSREMYDERLRTYLGS-QQSQTLWA 746
Tcr ILHPSEKLPSEKDVPEMNELRNFIHNNAIHNNMSQDGEDDTTTLKPGRPNKNILQVPQPR 469
Tco VLHPGEKTLYDKDGHEMNDLRSFIHNNSIHNNMSQEVEDEGAAAKRGRPAKTALQVPQPR 792
Tbr ALHSNEKSQHDRDGHEMNDLRSFISNNSIQNNIAQEAEDENGSSKRGRAAKPALQVPQPR 806
Tg ALHSNEKSQHDRDGHEMNDLRSFINNNSIQNNIAQEAEDENGSSKRGRAAKPALQVPQPR 806



Legend: Tcr (*Trypanosoma cruzi*); Tco (*Trypanosoma congolense*); Tbr (*Trypanosome brucei*); Tg (*Trypanosoma gambiense*) and Tev (*Trypanosoma evansi*)

Parts of this thesis have been submitted to scientific journals for publications:

Tania Bishola Tshitenge and Christine Clayton. "Interactions of the *Trypanosoma brucei brucei* zinc-finger domain protein ZC3H28 (preprint 10.1101/2021.08.09.455650). Manuscript in review, accepted in *Parasitology*."

Tania Bishola Tshitenge, Bin Liu and Christine Clayton. The RNA-binding protein DRBD18 regulates processing and export of the mRNA encoding the *Trypanosoma brucei* RNA-binding protein 10. (Manuscript submitted to *Nucleic Acid Research*).

Tania Bishola Tshitenge, Lena Reichert and Christine Clayton. Developmental regulation of the glycosomal phosphoglycerate kinase in *Trypanosoma brucei brucei*. (Manuscript in preparation, to be submitted to *Plos One*).

Parts of this thesis have been presented in international conferences and seminars:

Talks

Tania Bishola "Regulating a post-transcriptional regulator of gene expression, the RNA-binding protein 10, in *Trypanosoma brucei*"
Invited speaker at the Chair of Cell and Developmental Biology of the University of Würzburg (Germany), 4th August 2021

Posters

Tania Bishola, Monica Terrao, Bin Liu and Christine Clayton. "Roles of a monster 3'-UTR in regulating the expression of the master regulator, the RNA-binding protein 10, in *Trypanosoma brucei*: stabilization, degradation and alternative processing"
26th Annual meeting of the RNA Society, Virtual meeting, May 25-June 5, 2021.

Tania Bishola, Monica Terrao, and Christine Clayton. "Regulating the master regulator RBP10 in *T. brucei"*
Trypanosomiasis and Leishmaniasis Seminar 2020 organized by the "British Society of Parasitology" in Granada (Spain), March 9-11, 2020 (**Best Poster Award**)

Tania Bishola, Monica Terrao, Bin Liu and Christine Clayton. "Regulation of the expression of the RNA-binding protein RBP10 in *T. brucei"* 8th Kinetoplastid Molecular Cell Biology Meeting in Woods Hole, April 27-May 1, 2019.