

# The role of phosphoinositides in the control of variant surface glycoprotein gene expression in trypanosomes.

RAJESH, Rishi

Parasitology

McGill University, Montreal

December 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of M.Sc. Parasitology.

© Rishi Rajesh, 2020

## Table of Contents

1. Acknowledgement .....	iv
2. Contribution of authors .....	v
3. List of abbreviations .....	vi
4. List of figures .....	ix
5. List of Tables .....	x
6. Abstract .....	xi
7. Resume .....	13
8. Introduction .....	1
8.1.1. Etiology and epidemiology .....	1
8.1.3. Course of infection and associated symptoms .....	5
8.1.4. Diagnosis and treatment .....	6
8.2. Genetic features of <i>Trypanosoma brucei</i> .....	8
8.2.1. Genome organisation .....	8
8.2.2. Transcription of housekeeping and parasite specific protein coding genes .....	11
8.2.3. Transcription by RNA polymerase I .....	13
8.2.4. RNA processing, transport, and decay .....	14
8.3. Antigenic variation in <i>T. brucei</i> .....	16
8.3.1. What is antigenic variation? .....	16
8.3.2. Mechanisms of antigenic variation .....	17
8.3.3. Structure of the ES chromatin and Pol I transcription .....	18
8.3.4. Factors regulating ES transcription .....	19
8.3.5. Epigenetic modifications of the ES chromatin .....	23
8.3.6. Mechanisms involved in the switching of VSGs. ....	27

8.3.7.	Telomeric position effect and silencing of ES transcription.....	28
8.4.	The role of Repressor activator protein 1 (RAP1) in VSG gene expression .....	29
8.5.	Phosphoinositides and the regulation of VSG expression .....	30
8.6.	Concluding remarks .....	34
<i>9.</i>	<i>Project overview</i> .....	<i>36</i>
9.1.	Rationale and hypothesis.....	36
9.2.	Specific aims .....	36
<i>10.</i>	<i>Methods</i> .....	<i>38</i>
10.1.	<i>T. brucei</i> cell culture and growth curve analysis.....	38
10.2.	Immunofluorescence analysis.....	38
10.3.	Western blot analysis.....	39
10.4.	RNA extraction and qPCR analysis.....	40
10.4.1.	RNA isolation .....	40
10.4.2.	DNase treatment and cDNA synthesis.....	41
10.4.3.	Quantitative real-time PCR analysis.....	41
10.5.	Recombinant protein expression and purification.....	44
10.6.	Phosphatidylinositol binding assays.....	45
10.7.	Electrophoretic mobility shift (EMSA) or gel shift assays.....	46
10.8.	Microscale thermophoresis (MST).....	47
10.8.1.	DNA acquisition and preparation.....	47
10.8.2.	Binding checks to confirm binding.....	48
10.8.3.	Equilibrium dissociation constant ( $k_d$ ) measurements.....	48
10.9.	Statistical analysis.....	49
<i>11.</i>	<i>Results:</i> .....	<i>50</i>
11.1.	PIP5Pase is important for the developmental regulation of VSGs.....	50
11.1.1.	PIP5Pase localizes in the nucleus in PFs.....	50

11.1.2. PIP5Pase is essential for the growth of PFs .....	51
11.1.3. Knockdown of PIP5Pase results in the expression of silent VSGs in PFs .....	52
11.2.1. Purification of RAP1 and its domains .....	53
11.2.2. BRCT domain of RAP1 binds to PI(3,4,5)P3 .....	55
11.2.3. BRCT domain of RAP1 binds specifically to PI(3,4,5)P3 .....	57
11.3. Myb and Mybl domains of RAP1 bind ES chromatin DNA. ....	58
11.3.1. RAP1 binds directly to 70 bp repeats and telomeric repeats <i>in vitro</i> . ....	58
11.3.2. Myb and Mybl domains of RAP1 bind to the ES chromatin DNA <i>in vitro</i> . ....	62
11.3.3. PI(3,4,5)P3 disrupts the binding of RAP1 to telomeric repeats and 70 bp repeats. ....	63
<b>12. Discussion</b> .....	66
12.1. PIP5Pase plays a key role in regulating VSG expression in different developmental stages. ....	66
12.2. Characterization of RAP1 and its domains .....	68
12.3. PI(3,4,5)P3 regulates RAP1 binding to the ES chromatin. ....	71
<b>13. Conclusions</b> .....	72
<b>14. References</b> .....	75

# 1. Acknowledgement

First and foremost, I would like to thank Dr. Igor Cestari, my supervisor for enabling me to work on an interesting project with constant support and guidance throughout this experience. I would also like to thank my advisory committee member, Dr. Reza Salavati for his feedback, advice, and insights throughout my project. I want to thank McGill University and the Institute of Parasitology for the Graduate Excellence Award that enabled me to pursue my project and for allowing me to be a part of a wonderful group of students and researchers. I would like to thank past and present members of Dr. Cestari's laboratory for their help and support throughout the duration of my project. I would like to single out Fernando Altamura for his friendship and academic support throughout this experience. I would also like to extend my gratitude to members of Dr. Salavati's laboratory, specially Nisha Ramamurthy, Akshaya Srikanth and Vaibhav Mehta for their friendship and support. I want to thank all the members of the staff at the Institute of Parasitology. I would like to thank Mr. Serghei Dernovici and Dr. Norma Bautista-Lopez for their technical support and for enabling the access to various state of the art equipment. I would like to extend my appreciation and gratitude to Ms. Shirley Mongeau for her administrative support. I am also grateful for the various student representatives at the Institute of Parasitology for ensuring the well-being of students. I would like to thank students, past and present, at the Institute of Parasitology, specially, Jennifer Noonan, Tosin Opadokun, Georgia Limniatis, Mark Kaji and Maude Dagenais for their support throughout my time at the Institute. I cannot thank enough, my family and friends back in India for being extremely supportive during this important phase of my career.

## 2. Contribution of authors

All parts of this manuscript were written by Rishi Rajesh. Revisions and corrections were provided by supervisor Dr. Igor Cestari. The experiments were designed by Dr. Igor Cestari. The experiments presented in this manuscript were performed by Rishi Rajesh, except those indicated below.

- 1) The *Trypanosoma brucei* procyclic form conditional null cell line (C-215 was generated by Dr. Igor Cestari, and it was a gift from Dr. Ken Stuart Laboratory, Center for Global Infectious Disease Research, Seattle Children's Research Institute).
- 2) Growth curves and growth defects were determined by Dr. Cestari, and experiments repeated and data reproduced in this thesis.
- 3) VSG gene expression analysis by real-time PCR after knockdown of PIP5Pase in procyclic forms were determined by Dr. Cestari. The experiments were repeated and the data were reproduced in this thesis.
- 4) The plasmids pET-RAP1 (241), pET29-BRCT (243), pET29-Myb (245), and pET29-MybL (247) were generated by Dr. Igor Cestari.

I declare no conflict of interest.

### 3. List of abbreviations

<b>HAT</b>	Human African Trypanosomiasis
<b>GPCR</b>	G protein coupled receptor
<b>VSG</b>	Variant surface glycoprotein
<b>BES</b>	Bloodstream form expression sites
<b>MES</b>	Metacyclic form expression sites
<b>ATP</b>	Adenosine triphosphate
<b>BFs</b>	Bloodstream forms
<b>PFs</b>	Procyclic forms
<b>MFs</b>	Metacyclic forms
<b>CATT</b>	Card agglutination trypanosomiasis test
<b>ELISA</b>	Enzyme linked immunosorbent assay
<b>CSF</b>	Cerebrospinal fluid
<b>IL</b>	Interleukin
<b>VCAM</b>	Vascular cellular adhesion molecule
<b>IgM</b>	Immunoglobulin M
<b>PTRE</b>	Post treatment reactive encephalitis
<b>Mb</b>	Megabasepair
<b>kb</b>	Kilobasepair
<b>RNA</b>	Ribonucleic acid
<b>Pol II</b>	RNA Polymerase II
<b>PTU</b>	Polycistronic transcription units
<b>SSRs</b>	Strand switch regions

<b>BDF</b>	Bromodomain factors
<b>ChIP-seq</b>	Chromatin immunoprecipitation followed by sequencing
<b>Pol I</b>	RNA polymerase I
<b>ESB</b>	Expression site body
<b>CITFA</b>	Class I transcription factor A
<b>GFP</b>	Green fluorescent protein
<b>ChIP-qPCR</b>	Chromatin immunoprecipitation followed by quantitative polymerase chain reaction
<b>RNAi</b>	RNA interference
<b>mRNA</b>	Messenger RNA
<b>SL</b>	Splice leader
<b>m7G</b>	7-methyl guanosine
<b>GPI</b>	Glycosylphosphatidylinositol
<b>PI</b>	Phosphatidylinositol
<b>ESAGs</b>	Expression site associated genes
<b>RT-PCR</b>	Real time polymerase chain reaction
<b>VEX-1</b>	VSG exclusion 1
<b>VEX-2</b>	VSG exclusion 2
<b>TRF2</b>	Telomeric repeat binding factor 2
<b>NPT</b>	Neomycin phosphotransferase
<b>CAF-1</b>	Chromatin associated factor 1
<b>T7RNAP</b>	T7 RNA polymerase
<b>MNase</b>	Micrococcal nuclease
<b>FACT</b>	Facilitates chromatin transcription

<b>DSB</b>	Double strand breaks
<b>TPE</b>	Telomeric positioning effect
<b>TRF</b>	Telomeric repeat binding factor
<b>EMSA</b>	Electrophoretic mobility shift assays
<b>ITC</b>	Isothermal titration chemistry
<b>TIF2</b>	TRF interacting factor 2
<b>RAP1</b>	Repressor activator protein 1
<b>BRCT</b>	Breast Cancer type 1 C-terminus
<b>Mybl</b>	Myb-like
<b>IPMK</b>	Inositol polyphosphate multikinase
<b>PIP5K</b>	Phosphatidylinositol phosphate 5-kinase
<b>PLC</b>	Phospholipase C
<b>PIP5Pase</b>	Phosphatidylinositol phosphate 5-phosphatase
<b>CNs</b>	Conditional nulls
<b>FBS</b>	Fetal bovine serum
<b>PBS</b>	Phosphate buffered saline
<b>DAPI</b>	4',6-diamidino-2-phenylindole
<b>SDS-PAGE</b>	Sodium dodecylsulphate polyacrylamide gel electrophoresis
<b>RT</b>	Room temperature
<b>HSP-70</b>	Heat shock protein 70
<b>IPTG</b>	Isopropyl $\beta$ -D-1-thiogalactopyranoside

## 4. List of figures

- Figure 1** Life cycle of *T. brucei*.
- Figure 2A** Structure and organisation of chromosomes in *T. brucei*.
- Figure 2B** Structural organisation of BF and MF expression sites in *T. brucei*.
- Figure 3** Two mechanisms of antigenic variation in *T. brucei*.
- Figure 4** ES associated proteins and their regulation of VSG gene expression.
- Figure 5A** Immunofluorescence analysis of V5-tagged PIP5Pase in *T. brucei* PFs.
- Figure 5B** Control: Immunofluorescence analysis of V5-tagged PIP5Pase in *T. brucei* PFs in the absence of tetracycline.
- Figure 6A** Growth curve analysis after PIP5Pase knockdown after PIP5Pase knockdown In *T. brucei* PFs.
- Figure 6B** Western blot analysis of V5-tagged PIP5Pase in *T. brucei* PFs.
- Figure 7** VSG gene expression analysis by qPCR after knockdown of PIP5Pase in *T. brucei* PFs.
- Figure 8A** Diagram of recombinantly purified *T. brucei* His-tagged RAP1 (rRAP1-His) and its domains- rBRCT-His, rMyb-His, rMybl-His.
- Figure 8B** rRAP1-His and its domains were purified to homogeneity.
- Figure 9A** Diagram of PI binding assays used to study the binding of rRAP1-His, rBRCT-His, rMyb-His, rMybl-His to PIPs/IPs.
- Figure 9B** Phosphoinositide binding assay of rRAP1-His and PIPs/IPs.
- Figure 9C** PI(3,4,5)P3 binding with rRAP1-His and its domains.
- Figure 10A** PI(3,4,5)P3 competition assay.
- Figure 10B** Binding assays with rBRCT-His and PIPs/IPs.
- Figure 11A** Diagram of EMSA used to study the binding of rRAP1-His and its domains to 70 bp repeats and the telomeric repeats.
- Figure 11B** EMSA of rRAP1 with biotin-labelled 70 bp or telomeric repeat DNA sequences.
- Figure 11C** Binding check by MST to confirm the binding of rRAP1-His to the telomeric repeats.
- Figure 11D** Binding check by MST to confirm the binding of rRAP1-His to the 70 bp repeats.
- Figure 11E** Binding check by MST to rule out non-specific binding of rRAP1-His using the scrambled sequence.

- Figure 12A** Equilibration dissociation constant ( $k_d$ ) calculated for rRAP1-telomeric repeats interaction from the dose response curve obtained from MST.
- Figure 12B** Equilibration dissociation constant ( $k_d$ ) calculated for rRAP1-70 bp repeats interactions from the dose response curve obtained from MST.
- Figure 13** EMSA with rMyb-His, rMybl-His and rBRCT-His with telomeric repeats and 70 bp repeats.
- Figure 14A** Binding of rRAP1-His with telomeric repeats in the presence of PI(3,4,5)P3.
- Figure 14B** Binding of rRAP1-His with 70 bp repeats in the presence of PI(3,4,5)P3.
- Figure 15A** Binding of rMyb-His with telomeric repeats in the presence of PI(3,4,5)P3.
- Figure 15B** Binding of rMybl-His with telomeric repeats in the presence of PI(3,4,5)P3.
- Figure 15C** Binding of rMyb-His with 70 bp repeats in the presence of PI(3,4,5)P3.
- Figure 15D** Binding of rMybl-His with 70 bp repeats in the presence of PI(3,4,5)P3.
- Figure 16** Diagram of a silent ES.
- Figure 17** Diagram of an active ES.

## 5. List of Tables

- Table 1.** List of drug candidates used to treat HAT.
- Table 2.** List of qPCR primers, DNA sequences for EMSA and MST.
- Table 3.** List of oligonucleotide sequences for EMSA.
- Table 4.** List of oligonucleotide sequences for MST.
- Table 5.** Conditions tested to optimise the purification of proteins- rRAP1, rBRCT, rMyb and rMybl.
- Table 6.** Yield of proteins (rRAP1, rBRCT, rMyb, rMybl) after affinity purification.

## 6. Abstract

*Trypanosoma brucei* bloodstream forms (BFs) evade the host immune response by switching their variant surface glycoprotein coat (VSG) by a process called antigenic variation. The genome of *T. brucei* encodes for about 2,500 VSG genes. However, only one is expressed from one out of 20 telomeric expression sites (ES). VSG genes are silent in the procyclic forms (PF) of the parasite which express another surface protein known as procyclins. The regulation of VSG expression is multifactorial as it involves many proteins that regulate the transcription of VSGs at different levels. This includes proteins that affect promoter recognition by RNA polymerase I, proteins involved in chromatin remodelling and proteins that interact with the telomeres. A multiprotein complex found at the telomeres functions in regulating the transcription of VSGs. This complex includes proteins involved in RNA processing, nuclear lamina proteins, DNA binding proteins as well as regulatory proteins. Two members of this multiprotein complex: phosphatidylinositol phosphate 5-phosphatase (PIP5Pase) and repressor activator protein 1 (RAP1) have been implicated in the regulation of VSG transcription. These proteins regulate the silencing of VSG genes through protein-DNA interactions with the telomeric ES chromatin. These proteins have been shown to affect the silencing of VSGs in BFs of the parasite. How is the transcription of VSGs regulated by PIP5Pase and RAP1? Interestingly, RAP1 binds to PI(3,4,5)P<sub>3</sub> which is a substrate of PIP5Pase. This interaction influences the binding of RAP1 to the ES chromatin which in turn affects the silencing of VSGs in BFs. Knockdown of PIP5Pase was lethal in PFs and led to the expression of silent VSGs from the ESs. Immunofluorescence analysis showed that PIP5Pase localised in the nucleus of PFs similar to what was seen in BFs. Recombinant RAP1 as well as its truncations were purified. They were used in binding assays and it was found that the N - terminal BRCT domain of RAP1 binds to PI(3,4,5)P<sub>3</sub> while the central Myb and the C - terminal Myb1 did not. Additionally, the binding of the BRCT domain to PI(3,4,5)P<sub>3</sub> was competed out by molar excess of unlabelled PI(3,4,5)P<sub>3</sub> and not PI(4,5)P<sub>2</sub>. The recombinant proteins were also used in gel shift assays and microscale thermophoresis (MST) with synthetic ES chromatin DNA sequences and it was determined that RAP1 bound directly to the ES chromatin DNA sequences in vitro. MST revealed that RAP1 binds with different  $K_{ds}$  to the telomeric repeats and the 70 bp repeats. Gel shift assays with the Myb and Myb1 domains showed that they bound to the ES chromatin DNA while the BRCT domain did not. Interestingly, the presence of PI(3,4,5)P<sub>3</sub>

affected the binding of RAP1 to these sequences in these assays. The data shows that the binding of RAP1 to the ES chromatin DNA is affected by PI(3,4,5)P3 which is a substrate of PIP5Pase. These data suggest that the regulation of VSG gene transcription occurs by the regulation of RAP1 at the telomeres by PIP5Pase. In the presence of PIP5Pase, PI(3,4,5)P3 is dephosphorylated and this prevents the interaction of the metabolite with RAP1. As a consequence RAP1 binds to the ES chromatin leading to the silencing of the associated VSG gene. Upon the depletion of PIP5Pase, RAP1 is able to bind to PI(3,4,5)P3 and does not bind the ES chromatin, allowing the transcription of the VSG gene. Therefore, PIP5Pase regulates VSG gene expression by regulating the activity of RAP1.

## 7. Resume

Les formes sanguines de *Trypanosoma brucei* (BFs) échappent à la réponse immunitaire de l'hôte en changeant leur variante du revêtement glycoprotéique de surface (VSG) par un processus appelé variation antigénique. Le génome de *T. brucei* code pour environ 2 500 gènes VSG. Cependant, un seul est exprimé à partir d'un site d'expression télomérique (ES) sur 20. Les gènes VSG sont silencieux dans les formes procycliques (PF) du parasite qui expriment une autre protéine de surface connue sous le nom de procyclines. La régulation de l'expression des VSG est multifactorielle car elle implique de nombreuses protéines qui régulent la transcription des VSG à différents niveaux. Cela inclut des protéines qui affectent la reconnaissance du promoteur par l'ARN polymérase I, des protéines impliquées dans le remodelage de la chromatine et des protéines qui interagissent avec les télomères. Un complexe multiprotéique présent au niveau des télomères fonctionne en régulant la transcription des VSG. Ce complexe comprend des protéines impliquées dans le traitement de l'ARN, des protéines de lamina nucléaire, des protéines de liaison à l'ADN ainsi que des protéines régulatrices. Deux membres de ce complexe multiprotéique : la phosphatidylinositol phosphate 5-phosphatase (PIP5Pase) et la protéine d'activation du répresseur 1 (RAP1) ont été impliqués dans la régulation de la transcription des VSG. Ces protéines régulent le silence des gènes VSG par des interactions protéine-ADN avec la chromatine ES télomérique. Il a été démontré que ces protéines affectent la réduction au silence des VSG dans les BF du parasite. Comment la transcription des VSG est-elle régulée par la PIP5Pase et la RAP1? Il est intéressant de noter que RAP1 se lie à PI(3,4,5)P3 qui est un substrat de la PIP5Pase. Cette interaction influence la liaison du RAP1 à la chromatine ES qui, à son tour, affecte le silence des VSG dans les BFs. L'élimination de la PIP5Pase est létale dans les PF et conduit à l'expression de VSG silencieux dans les ES. L'analyse par immunofluorescence a montré que la PIP5Pase se localisait dans le noyau des PF de manière similaire à ce qui a été observé dans les BFs. La RAP1 recombinante ainsi que ses tronçatures ont été purifiées. Ils ont été utilisés dans des tests de liaison et il a été constaté que le domaine BRCT N-terminal de RAP1 se lie à PI(3,4,5)P3 alors que le Myb central et le Myb C-terminal ne le font pas. En outre, la liaison du domaine BRCT au PI(3,4,5)P3 était concurrencée par un excès molaire de PI(3,4,5)P3 non marqué et non de PI(4,5)P2. Les protéines recombinantes ont également été utilisées dans des tests de déplacement de gel et de thermophorèse à micro-échelle (MST) avec des séquences d'ADN de chromatine ES synthétique et il a été déterminé que RAP1 se liait directement aux séquences d'ADN de chromatine ES in vitro. La MST a révélé que RAP1 se lie avec différents Kds aux répétitions télomériques et aux répétitions de 70 bp. Les tests de déplacement de gel avec les domaines Myb et Myb1 ont montré qu'ils se lient à l'ADN de la chromatine ES, alors que le domaine BRCT ne le fait pas. Il est intéressant de noter que la présence de PI(3,4,5)P3 a affecté la liaison de RAP1 à ces séquences dans ces tests. Les données montrent que la liaison de RAP1 à l'ADN

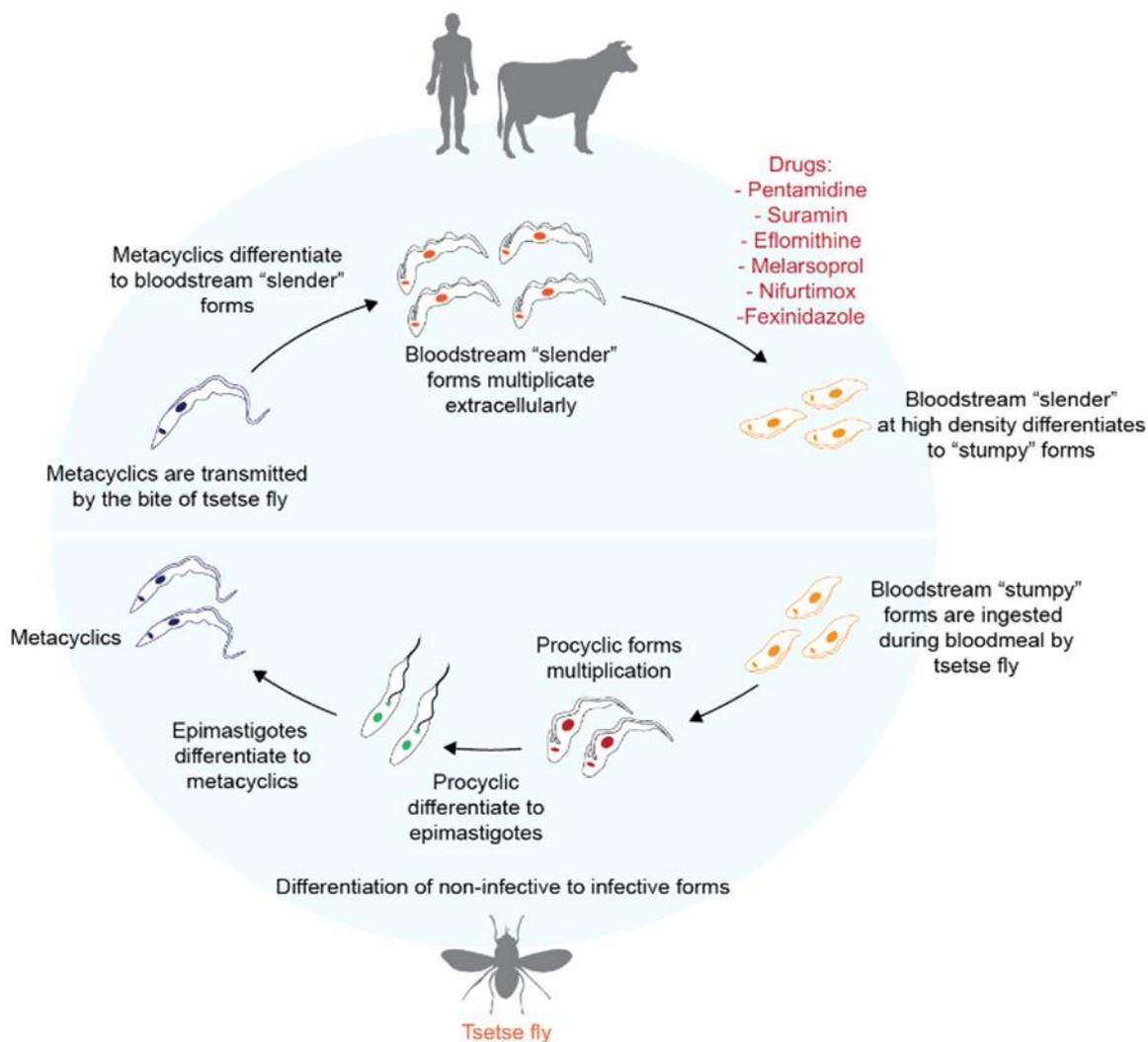
de la chromatine ES est affectée par la PI(3,4,5)P3 qui est un substrat de la PIP5Pase. Ces données suggèrent que la régulation de la transcription du gène VSG se fait par la régulation de RAP1 au niveau des télomères par la PIP5Pase. En présence de la PIP5Pase, le PI(3,4,5)P3 est déphosphorylé et cela empêche l'interaction du métabolite avec le RAP1. En conséquence, RAP1 se lie à la chromatine ES, ce qui entraîne la réduction au silence du gène VSG associé. Lors de l'épuisement de la PIP5Pase, RAP1 est capable de se lier à PI(3,4,5)P3 et ne se lie pas à la chromatine ES, permettant la transcription du gène VSG. Par conséquent, la PIP5Pase régule l'expression du gène VSG en régulant l'activité de RAP1.

## 8. Introduction

### 8.1.1. Etiology and epidemiology

African trypanosomiasis are neglected tropical diseases caused by the flagellated protist *Trypanosoma brucei* of the phylum Euglenozoa and affect both humans and animals. The presence of a characteristic kinetoplast, a network of circular DNA found in the mitochondrion distinguishes this protozoan group from other protozoan such as sarcodines, ciliates, and sporozoans, and are therefore grouped under the class Kinetoplastea (1). They are further classified under the family Trypanosomatidae which includes other monoflagellated protozoans of medical importance, namely *Trypanosoma cruzi* and *Leishmania spp.* (2). Humans can be infected by two morphologically similar subspecies of *T. brucei*: *T. brucei gambiense* and *T. brucei rhodesiense*. Despite their similarities, about 98% of the reported cases are caused by the anthroponotic *T. brucei gambiense* while 2% are caused by the zoonotic *T. brucei rhodesiense* (3). In the western and central sub-Saharan, *T. brucei gambiense* is known to cause the chronic form of the disease, referred to as West African sleeping sickness which primarily affects the Democratic Republic of Congo, Angola, Sudan, Central African Republic, Chad, and northern Uganda. In contrast, *T. brucei rhodesiense* is known to cause the more acute East African sleeping sickness, primarily affecting Uganda, Tanzania, Malawi, and Zambia. (3). On the other hand Animal African Trypanosomiasis is caused predominantly by *T. vivax*, *T. congolense* and less commonly by *T. brucei brucei* affecting both domestic and wild animals. Endemic in 37 countries, animal African trypanosomiasis has led to an estimated annual economic loss of about 4.5 billion USD.

### 8.1.2. Transmission and Life Cycle



**Fig. 1. Life cycle of *T. brucei* (4).** *T. brucei* is transmitted by tsetse flies to mammals. In the insect, *T. brucei* metacyclic trypomastigotes are released into the vertebrate host by the bite of the insect. Then, they develop into proliferative, slender bloodstream forms in the bloodstream of the mammal. Bloodstream forms circulates extracellularly in the bloodstream, adipose tissue, skin or brain. At high densities, bloodstream trypomastigotes slender forms differentiate into non-proliferative stumpy forms that can be transmitted to the tsetse fly where they develop into replicative procyclic forms, followed by differentiation into epimastigotes, and then to infective metacyclic forms that can be transmitted to mammals. This figure has been reproduced here from the journal Drug Development Research with the permission from John Wiley and Sons from the article The current drug discovery landscape for trypanosomiasis and leishmaniasis: Challenges and strategies to identify drug targets, by Altamura F, Rajesh R, Catta-Preta CMC, Moretti NS, Cestari I, 2020 (4).

*T. brucei* is primarily transmitted by the bite of a tsetse fly during a blood meal (5). *T. brucei gambiense*, is transmitted via the bite of the riverine species *G. palpalis* and *G. fuscipes* whereas *T. brucei rhodesiense* is transmitted via the bite of the savannah group species *G. mortisans* and *G. pallidipes* (6, 7). Other routes of transmission have also been reported, including sexual contact, congenital transmission, transmission during blood transfusions, organ transplantation and iatrogenic transmission (medical or laboratory accidents) (8-12).

The development of the parasite is complex with various physiological and morphological changes. Within the mammalian host, bloodstream trypomastigotes exist as long, slender, proliferative forms (Fig. 1) (13). Proliferation continues until they receive a signal to develop into the transmissible stumpy forms. The signals that mediate this transmission are known as stumpy inducing factors (SIF) (14). A recent study that looked further into this signalling indicates that oligopeptide signalling via a receptor called TbGPR89 drives the development into stumpy forms (15). The stumpy forms are well adapted to changes in conditions within the tsetse fly such as high protease levels, alkaline pH conditions, and changes in the available nutrients. (16). Shortly upon entering the midgut of the fly, they encounter a cold shock induced by a sudden drop in temperature from 37°C in the human or animal body to about 27°C in the tsetse fly (17). The exposure to tricarboxylic acid cycle metabolites: citric acid and cis-aconitate have been identified as contributing factors towards the development of PFs from stumpy forms (18). Another study on protein tyrosine phosphatases, specifically TbPTP1 in these kinetoplastids suggests a mechanism of differentiation in which TbPTP1 negatively regulates the development of BFs to PFs (19). However, the presence of CCA, which is transported via the he carboxylate transporter proteins associated with differentiation (PAD), inhibits TbPTP1 and promotes the differentiation of BFs to PFs (20, 21). A recent study has implicated the inositol polyphosphate multikinase (IPMK) enzyme, a member of the inositol phosphate (IP) pathway in the development of BFs to PFs (22). IPMK is a kinase that phosphorylates Ins(1,4,5)P<sub>3</sub> and Ins(1,3,4,5)P<sub>4</sub> and its knockdown led to the development of BFs to PFs with a loss in VSG gene expression and expression of procyclins accompanied by activation of a functional mitochondrial oxidative phosphorylation system (22). The BF to PF differentiation was enhanced in the presence of CCA (22). This suggests that the various factors described above work together in regulating the process of development in *T. brucei*. In summary, the development of the parasite from BFs to PFs involves a cascade of

signalling events accompanied by morphological, physiological and metabolic changes. After colonization of the midgut, the stumpy forms develop into replicative procyclic forms (PFs) that eventually develop into the epimastigotes in the foregut (Fig. 1) (23). The epimastigotes then migrate to the salivary gland where they transform into the mammal-infective metacyclic forms (MFs) (Fig. 1) (24).

The surface coat of the bloodstream form (BF) of the parasite is almost entirely occupied by the highly immunogenic variant surface glycoproteins (VSGs) which the parasite periodically switches to evade the host antibody response (25). However, the PFs exhibit an entirely different surface coat consisting of procyclins that enable them resist proteolytic cleavage within the midgut of the tsetse fly (26). After the epimastigotes migrate to the salivary gland they differentiate into MFs, which express a subset of VSGs from metacyclic expression sites (MES) (discussed below). In addition, *T. brucei* undergoes numerous biochemical changes across the different developmental stages. The proliferative BFs are surrounded by the glucose-rich environment of their mammalian host. They completely depend on glycolysis for the production of ATP in addition to the excretion of the pyruvate product into the host bloodstream (27, 28). The stumpy forms of the parasite, although present in a glucose rich environment, undergo changes at the gene expression level in order to adapt to the environment within the tsetse fly (29). The differences in gene expression include differences in regulation of transcripts associated with processes such as metabolism, cell structure, protein transport, proteolysis and translation (30). Transcripts associated with cell division, maintenance of cytoskeletal structure, VSGs were more abundant in the slender BFs whereas they were down regulated in stumpy forms. Transcripts associated with processes such as procyclin expression, proteins associated with differentiation (PAD) family, lipid biosynthesis, lysosome activity were upregulated in stumpy forms (30). Within the midgut of the tsetse fly, in a glucose-depleted environment, the PFs utilize L-proline which is catabolized in the mitochondrion (28). This results in the production of alanine as the end product along with many reduced cofactors which can then be reoxidised in the respiratory chain for ATP production via oxidative phosphorylation (OXPHOS) (28). They also catabolize other amino acids such as leucine, and threonine (31). Leucine and threonine are primary carbon sources for lipid production by the fatty acid biosynthesis pathway or the mevalonate pathway to generate isoprenoid and sterols (32, 33).

However, if glucose is present, the PFs utilize the glucose via glycolysis as the primary energy source despite the presence of proline or other amino acids (28).

### **8.1.3. Course of infection and associated symptoms**

Human African Trypanosomiasis (HAT) manifests in 2 stages: the early haemolympathic stage (stage 1) and the later meningoencephalitic stage (stage 2). Upon infection, *T. brucei* BFs proliferate and spread through the blood and the lymphatic system, establishing the haemolympathic stage of the disease. Interestingly, motile parasite populations have been observed in the skin, specifically in the dermis and subcutis (34). The initial symptoms such as arthralgia, headache, and fatigue are less indicative of the disease. The onset of clinical features such as hepatomegaly, pericarditis, congestive heart failure, splenomegaly define the haemolympathic stage of the disease (35). In the case of West African trypanosomiasis, the haemolympathic stage can last from about a few months up to several years whereas, in the case of East African trypanosomiasis, the haemolympathic stage persists for a few weeks. Once the parasite invades the blood brain barrier (meningoencephalitic stage) more symptoms such as sleepiness during the day, insomnia at night, slurred speech, persistent confusion, changes in personality, paralysis and loss of balance may arise during the later stages of infection (36). West African trypanosomiasis can persist for years, whereas East African trypanosomiasis usually persists for a few months resulting in death of the affected individual if left untreated. During later phases of the haemolympathic stages, symptoms such as myositis, cerebral ataxia are accompanied by visual and sensory complications. A characteristic clinical feature in both the haemolympathic and meningoencephalitic stages of the disease is the occurrence of daytime somnolence and insomnia at night. Eventually the patient experiences deterioration of various sensory perceptions, accompanied by features such as slurred speech, optic neuritis, motor neuropathy, psychiatric disorders and seizures (36). Eventually, the affected individual experiences a relentless loss of conscience, followed by coma and death if left untreated (37).

#### 8.1.4. Diagnosis and treatment

It is crucial to diagnose the disease as early as possible to prevent progression to the meningoencephalitic stage of the disease. The available diagnostics involves the detection of antibodies, the parasite, or the clinical manifestations of the haemolympathic and meningoencephalitic stages of the disease. Some of the commonly used methods are Card Agglutination Trypanosomiasis Test (CATT), the lymph node examination (palpation and aspiration), microscopic examination of blood and lymph, serodiagnostics such as ELISA and immune Trypanolysis, and molecular diagnostic tools such as polymerase chain reaction (PCR), proteomics, microarray chips, as well as high throughput methods (38-42). Studies are in progress to establish biomarkers of disease progression, possibly including markers for neuroinflammation such as IL-10, IL-6, IL-1 $\beta$ , ICAM-1, VCAM, IgM and neopterin (43-45).

Currently, there are five drugs available for the treatment of HAT (Table 1). Pentamidine is primarily used to treat the heamolymphatic stage of the infection caused by *T. brucei gambiense* whereas suramin is used to treat the heamolymphatic stage of the infection caused by *T. brucei rhodesiense* (4). Eflornithine monotherapy and eflornithine in combination with nifurtiox are used to treat the meningoencephalitic stage of the infection caused by *T. brucei gambiense*. Melarsoprol is effective in treating the meningoencephalitic stage of infections caused by either species. However, these drugs have various associated side-effects that cause harm to treated individuals (Table 1). More recently, in 2019, Fexinidazole has been approved for the treatment of both stages of the infection caused by *T. brucei gambiense* (46).

Table 1 below lists the drugs used to treat either stage of infections caused by both *T. brucei gambiense* as well as *T. brucei rhodesiense*. The dose regimens of each drug as well as the associated side effects have been described. This table has been partially reproduced from Altamura et. al. (4)

Table 1. List of drug candidates used to treat HAT.

Species	Stage of infection	Drug	Mode of administration and dosage unit	Toxic effects of the drug
<i>T. brucei gambiense</i>	Hemolymphatic stage	Pentamidine	Intramuscular injections (7-10 units)  4 mg/Kg/unit.	<ul style="list-style-type: none"> <li>○ Hypotension</li> <li>○ Hypoglycaemia</li> <li>○ Nephrotoxicity</li> <li>○ Abnormal pancreatic activity</li> <li>○ Stevens- Johnsons syndrome</li> </ul>
	Meningoencephalitic stage	Eflornithine	Monotherapy: intravenous injections (56 units)  100 mg/Kg (adults) 150 mg/Kg (children)  Nifurtimox Eflornithine Combination Therapy: 400 mg of eflornithine every 12 hr for 7 days with 15 mg/kg of nifurtimox administered orally thrice a day for 10 days	<ul style="list-style-type: none"> <li>○ Nausea</li> <li>○ Seizures</li> <li>○ Head and abdominal ache</li> <li>○ Insomnia</li> </ul>
	Both stages	Fexinidazole	Oral (600mg/ tablet) Adults: 3 tablets/day for 4 days followed by 2 tablets/day for 6 days.  Children:	<ul style="list-style-type: none"> <li>○ Nausea</li> <li>○ Emesis</li> <li>○ Diarrhea</li> <li>○ Paresthesia</li> <li>○ Myalgia</li> <li>○ Keratitis</li> <li>○ Asthenia</li> <li>○ Dyspepsia</li> </ul>

			2 tablets for 4 days/day followed by 1 tablet/day for 6 days.	
<i>T. brucei rhodesiense</i>	Hemolymphatic stage	Suramin	Intravenous injections of 20 mg/kg every 3-7 days for 4 weeks	<ul style="list-style-type: none"> <li>○ Neuropathy</li> <li>○ Abnormal renal behaviour</li> <li>○ Anaphylactic shock</li> </ul>
	Meningoencephalitic stage	Melarsoprol	Intravenous injections (10 units)  2.2 mg/kg/day	<ul style="list-style-type: none"> <li>○ Post treatment reactive encephalitis (mental deterioration followed by pathologies such as comas, convulsions)</li> </ul>

Acoziborole (SCYX-7158) is another drug currently in the pipeline for the treatment of trypanosomiasis. It was selected while screening boron-based compounds (benzoxaborole 6-carboxamides) (47). Demonstrating trypanocidal activity representative against *T. brucei brucei*, *T. brucei gambiense* and *T. brucei rhodesiense* strains, it successfully completed phase 1 clinical trials in 2015 and entered phase two and three in the year 2016 (by Drugs for Neglected diseases initiative) (47).

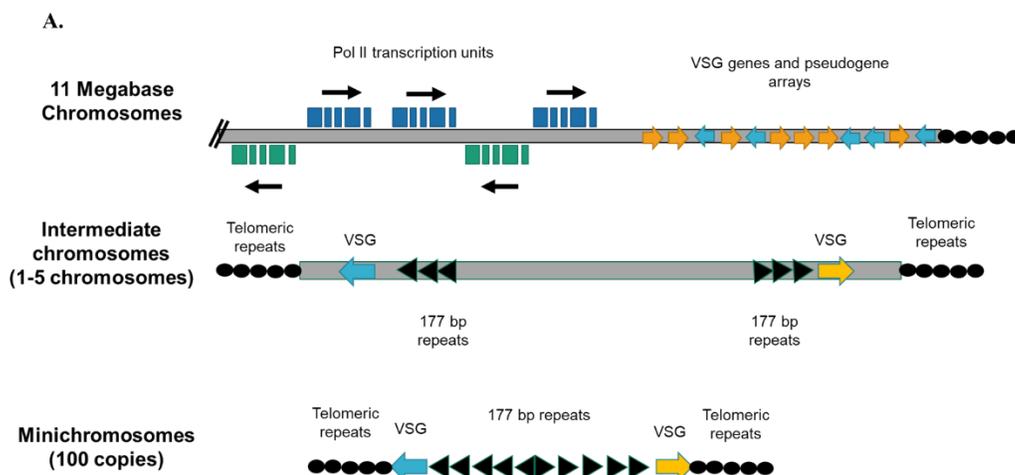
## 8.2. Genetic features of *Trypanosoma brucei*

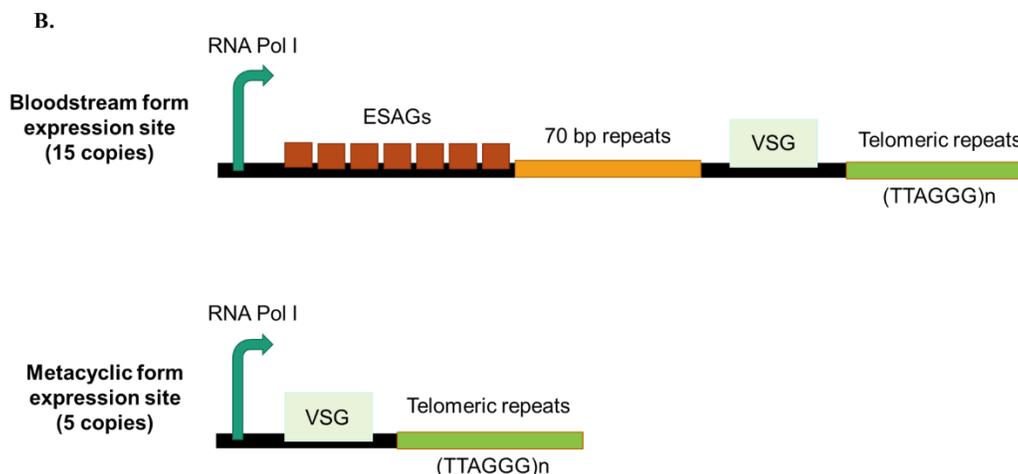
### 8.2.1. Genome organisation

The first complete genome of *T. brucei* was published in 2005 (48). The diploid genome of *T. brucei* is about 25 Mb and includes 11 megabase chromosomes (1-6 Mb), about 100 minichromosomes (50-150 Kb) and 5 intermediate size chromosomes (200-900 Kb per haploid genome) (48). The megabase chromosomes have long gene clusters (Fig 2) on either strand which are transcribed as polycistronic units, and are post-transcriptionally processed by trans-splicing and polyadenylation (48). About 50% of the genome contains coding sequences with a GC content of 46.4% (48). There are 9068 protein coding genes among which about 1700 genes are predicted to be *T. brucei* specific and 904 are predicted to be pseudogenes (48). RNA genes include 65 transfer RNA (tRNA) genes, 56 ribosomal RNA (rRNA) genes, fewer than 28 small interfering

RNA (siRNA) genes, five small nuclear and 353 small nucleolar RNA genes (48). Notably, a large portion of the genome (20%) is dedicated to VSG protein coding genes and pseudogenes that capacitate the parasite to evade the host antibody response (48). The genome sequencing also accelerated the discovery of pathways or enzymes that have the potential to serve as drug targets. This includes genes coding for proteins involved in carbohydrate metabolism, electron transport and oxidative phosphorylation, GPI anchor synthesis, amino acid metabolism, trypanothione biosynthesis, purine and pyrimidine salvage pathways which are potential drug targets. For instance, phosphonopyruvate decarboxylase was found to be present in all major kinetoplastids: *T. brucei*, *T. cruzi* and *L. major* with no homologs in other eukaryotes, making it a potential drug target (48).

The protein-coding genes in *T. brucei* are located on megabase chromosomes. They are organised in long gene arrays as polycistronic transcription units (PTU) (48-50) (Fig. 2A, top). The intermediate chromosomes contain 177 bp repeats and VSG genes that act as a source of VSG sequences for recombination. The minichromosomes consist of large palindromic region of 177 bp which are also flanked by VSG genes (51) (Fig. 2A, bottom).





**Fig 2. Genome organisation in *T. brucei*. A. Structure and organisation of chromosomes in *T. brucei*.**

**Top:** The diagram depicts the megabase chromosomes. The green and blue boxes indicate Pol II transcribed PTUs. Arrays of VSG genes and pseudogenes are depicted by yellow and blue arrows with black circles depicting telomeric repeats. **Center:** The diagram depicts intermediate chromosomes. The VSG genes are depicted by yellow and blue arrows with black circles depicting the telomeric repeats. **Bottom:** The diagram depicts minichromosomes flanked by VSG genes on either end represented by blue and yellow arrows and a central large palindromic region of 177 bp depicted by black arrows. Arrows describe the direction of transcription.

**Fig. 2B. Structural organisation of BF and MF expression sites.**

**Top:** BF ESs in the 427 strain of *T. brucei* have an RNA Pol I promoter, up to 12 ESAGs, 70 bp repeat sequences (1-15 Kb), VSG gene and telomeric repeats. **Bottom:** The MF expression sites in the 427 strain of *T. brucei* have an RNA Pol I promoter, VSG gene and telomeric repeats.

ESs are located at telomeric regions found at the end of some of the 11 megabase and intermediate chromosomes in *T. brucei* and are about 30-60 Kb in length (52). There are about 15 BESs which can be expressed in BFs and five MESs that can be expressed in MFs of the parasite (Fig. 2B) (53). The polycistronic BES contains a Pol I promoter, followed by up to 12 expression site associated genes (ESAGs), 70 bp repeat sequences which are 1 to 15 kb long, the VSG gene followed by the telomeric repeats (52) (Fig. 2B, top). The MES which are active in MFs are consisting of a Pol I promoter and a telomeric VSG gene followed by the telomeric repeats VSG and in some strains (e.g. 427), they possess ESAG pseudogenes and residual 70 bp repeats (53, 54) (Fig. 2B, bottom). In the BES, the Pol I promoter is located 40-60 kb upstream of the VSG gene. Interestingly, the Pol I promoter of ESs are different from that of rRNA and procyclin genes (55). The Pol I promoter for rRNA genes and procyclin genes is about 250 bp in length and contain at the 5' end two regions

involved in the recruitment of Pol I transcription factors (55). The ES promoter is much shorter in length (70 bp) and also includes the two transcription factor recruitment regions.

Following the Pol I promoter, the BES also contains ESAGs (56). ESAG 4 encodes for a transmembrane protein with adenylate cyclase activity (56). ESAG 2, 3, and 11 encode for membrane associated or membrane targeted proteins (56). The products of ESAG 6 and 7 together form a transferrin binding protein complex (56). ESAG 9 and 10 encode for polypeptides that associate with the cell membrane (56). ESAG 5 and 8 encode for polypeptides that unlike the aforementioned ESAGs, are not associated with the cell membrane (56). ESAG 10 is homologous to *Leishmania* biopterin protein (56). Some of these ESAGs are yet to be functionally characterised.

### **8.2.2. Transcription of housekeeping and parasite specific protein coding genes**

All megabase chromosomes show similar structural organisation where the Pol II transcribed genes are organised in clusters of long, non-overlapping PTUs throughout the chromosome. These PTUs are arranged in a directional head-to-tail manner, adjacent to one another either on the same strand or on opposite strands. The PTUs are separated by characteristic regions known as strand switch regions (SSRs) which are also seen in other kinetoplastids (49, 50). The RNA Polymerase II (Pol II) mediated transcription of two neighboring PTUs can either be convergent or divergent. These parasites lack canonical Pol II promoters. The sites of transcription initiation and termination are associated with certain features of the chromatin. For instance, transcription start sites are characterised by the presence of GT rich sequences that act as promoter elements where transcription initiates (57). Notably, transcription initiates at SSRs found between divergent PTUs called divergent SSRs (dSSRs) and terminates at SSRs between convergent PTUs called convergent SSRs (cSSRs) (49). Notably, regions downstream of the convergent PTUs are enriched in tRNAs, another characteristic feature of the SSRs (58).

Transcriptome analysis by Kolev *et.al.* and Nilsson *et.al.*, using RNA sequencing led to mapping the sites of initiation of transcription (59) (60). These start sites are enriched in histone variants, namely: H410Kac, H2AZ and H2BV as well as proteins called bromodomain factors (BDFs) (61).

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) with antibodies for Bromodomain factor BDF3 revealed that it was enriched regions also enriched in H4K10ac (61). ChIP-seq data indicate that the Pol II transcription termination regions are also enriched in histone variants, namely: H3V and H4V (61), whereas ChIP-seq with antibodies specific to H3V mapped these variants to telomeric and subtelomeric regions (61).

Another well characterised feature of Pol II transcription is base J. Base J is a modification of thymine (T) which undergoes a two-step modification including an initial hydroxylation followed by glycosylation, ultimately replacing about 1% of Ts in *T. brucei* BFs (62, 63). Anti-J ( $\alpha$ -J) immunoblotting in combination with immunoprecipitation and P<sup>32</sup> labelling have identified base J in many trypanosome species (63). Moreover, these bases were found to be abundant in the telomeric repeat sequences (TTAGGG) of various analysed kinetoplastids (63). Interestingly, this modification can be found only in BFs but not in PFs (63, 64). In 2012, a study by van Leunen *et al.*, in *Leishmania* showed that base J played a role in the termination of Pol II transcription (65). Knockout of genes encoding base J binding proteins 1 and 2 by Schulz *et al.*, leading to the elimination of base J has brought to light the significance of base J in termination of transcription in *T. brucei* (66). Deletion of genes encoding base J binding proteins 1 and 2 and H3V showed that loss of base J and the histone variant H3V had a major impact on the termination of Pol II transcription. Precisely, the loss of base J and H3V resulted in a read through transcription from the transcribed PTU into the neighboring PTU (66). These results were further corroborated by the presence of polyadenylated antisense transcripts which are produced as a consequence of readthrough transcription from one PTU to the neighboring PTU (separated by a cSSR) upon depletion of base J and H3V (66).

Genes coding for tRNAs are transcribed by RNA Polymerase III (Pol III) in *T. brucei* (67). The tRNAs which are interspersed across PTUs which are transcribed by PTU are potential contributors to the termination of Pol II transcription due to their association with Pol III and the respective transcription factors (67).

### 8.2.3. Transcription by RNA polymerase I

RNA Polymerase I (Pol I) transcribes ribosomal RNA (rRNA) in most eukaryotes. Exceptionally in *T. brucei*, certain protein coding genes, namely VSG and procyclin genes, are also transcribed by Pol I (68). Experimental evidence in the 1980s and 1990s proved that these protein coding genes were transcribed by Pol I unlike other protein coding genes that are transcribed by Pol II. This evidence was initially based off nuclear run on assays in the presence of the Pol II inhibitor  $\alpha$ -amanitin. These studies showed that transcription of VSG genes and procyclins was by an  $\alpha$ -amanitin insensitive polymerase (69). In other words, Pol II transcription, but not Pol I transcription is inhibited by  $\alpha$ -amanitin (69). Interestingly, RNA interference of RNA Pol I in *T. brucei* led to a decrease in transcription of not just rRNA genes but also VSG and procyclin genes (70). Consistent with this, studies by Rudenko *et.al.*, on procyclin genes demonstrated that their transcription, like the VSGs and rRNA genes, also occurred in an alpha-amanitin resistant manner (71). Evidence from a study by Janz and Clayton was successful in demonstrating the interchangeability of rRNA and PARP promoters, both Pol I promoters, despite minimal sequence similarity between them (72). In 1998, in situ hybridisation studies by Chaves *et al.*, led to the discovery of an extranucleolar site for the transcription of a single expression site containing the VSG gene expressed by the parasite (73). In 2001, in situ hybridisation studies by Navarro and Gul revealed that Pol I colocalized with this extranucleolar locus now termed as the expression site body (ESB) (74).

Pol I transcription has various characteristics that make it a unique system. A major milestone in understanding the Pol I transcription machinery was the characterisation of Class I Transcription Factor A (CITFA) in *T. brucei* (75). The complex is absolutely necessary for Pol I transcription as was shown by in vitro transcription experiments (75). The transcription of Pol I promoter-equipped plasmids was ablated upon the addition of antibodies against the CIFTA-2 subunit of Pol I (75). Tandem affinity purification followed by mass spectrometry of CIFTA-2 resulted in the identification of the CITFA complex which contains six proteins (75). Of these six proteins, five were trypanosome-specific whereas DYNL1 (dynein light chain) was found to be closely related to human dynein DYNL1 and involved in the promoter binding activity of CITFA (76). More recently, *in vivo* as well as *in vitro* functional studies along with immunofluorescence studies have

led to the discovery of a novel CITFA subunit (CITFA-7) that is crucial for viability of the parasite as well as Pol I transcription (77). Notably, this complex is the first characterised protist factor to be associated with class I transcription (78).

#### **8.2.4. RNA processing, transport, and decay**

Transcription is immediately followed by processing primary transcripts by splicing, producing a mature messenger RNA (mRNA) (79). In trypanosomes, transcripts undergo trans-splicing, a mechanism discovered in trypanosomes while studying the transcription of VSGs (69). Transcripts of the VSG gene were found to have an identical 39 nucleotide long sequence at the 5' end which was absent in the VSG genomic locus (69). This sequence, referred to as the splice leader (SL) sequence, is derived from a SL-RNA. All trypanosome transcripts possess this sequence at the 5' end of their transcripts (69). During trans-splicing, the SL-RNA is involved in two processes (80). First, it acts as a substrate for splicing and thus participates in the processing of polycistronic mRNAs into monocistronic transcripts (80). Secondly, the capping of the mRNA with a 7-methyl guanosine (m7G) structure occurs through trans-splicing of SL-RNA (80). The m7G modification which occurs co-transcriptionally is crucial in regulating biological functions such as pre-mRNA processing, nuclear export and cap-dependent protein synthesis (81).

The export of nuclear mRNA into the cytoplasm is essential for gene expression in any organism as it bridges nuclear transcriptional with translation in the cytoplasm. RNAi based screening identified the export machinery in *T. brucei*, and some of the proteins are conserved with yeast orthologues: mRNA export factor 67 (Mex67), Yeast mRNA export factor 1 (Yra1), Suppressor of (Bad Response to Refrigeration protein 1) BRR1 protein 2 (Sub2), and Cap-specific mRNA (nucleoside-2'-O-)-methyltransferase 1 (Mtr1) (82). Tandem affinity purification and mass spectrometry identified the interacting partners Mtr1 and importin in *T. brucei*. Knockdown of Mtr1 or importin in *T. brucei* revealed that they were vital for the cells and led to the accumulation of polyadenylated mRNAs (83). Other proteins such as cap binding proteins (CBPs) and Poly-A binding proteins (PABPs) have also been reported to actively participate in mRNA export. In *T. brucei*, the CBP complex includes CPB20, also seen in higher eukaryotes. CPB20, along with three related proteins, is known to bind the SL CAP of the mRNA and participate in mRNA export (84).

Of the two PABPs in *T. brucei*, PABP 2 has been shown to bind CBPs indicating a potential role in mRNA export (85, 86).

*T. brucei* regulates gene expression primarily post transcriptionally. This involves the 3'-UTRs of short-lived mRNA (87). These mRNA harbor sequences rich in adenosine (A) and uridine (U)-AU rich elements (AREs) which are associated with RNA binding proteins (RBPs) (87). The association of such mRNA with RBPs through the AREs is known to target the mRNA towards a rapid degradation process (88). These RBPs act as cis-acting elements which along with their trans-acting factors regulate mRNA stability or degradation (89). These trans-acting factors include proteins such as phosphatases and kinases which act on the targeted mRNA (87). Deadenylation, followed by decapping and 5' - 3' exonuclease activities predominantly control degradation of trypanosome mRNA. Deadenylation complexes found in trypanosomes are chromatin assembly factor (CAF1) poly A nuclease 2/3 (PAN2/PAN3), and the poly-A ribonuclease (PARN) enzymes (90-92). Followed by deadenylation, the mRNA undergoes decapping, usually performed by MutT hydrolases in eukaryotes (93). However despite the presence of these hydrolases, trypanosomes utilize another protein: ApaH-like phosphatase 1 (ALPH1), a phosphatase that removes the cap and leaves behind a 5' diphosphate (93). The decapping prepares the mRNA for 5'-3' degradation mediated by the exonuclease activity of *T. brucei* exoribonuclease (XRNA) (93).

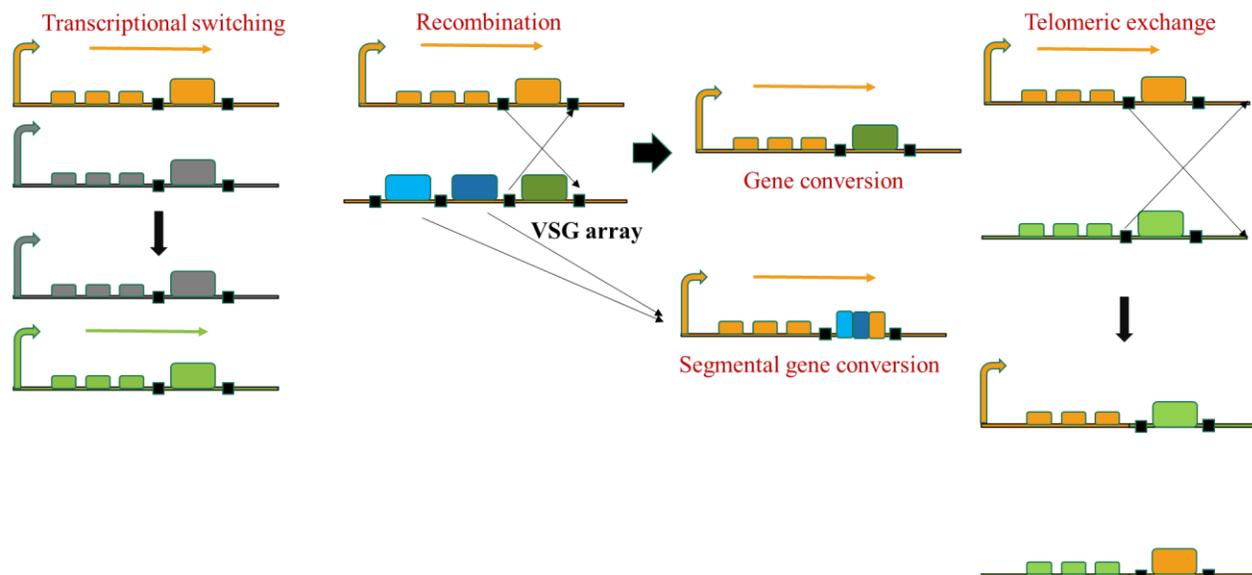
### 8.3. Antigenic variation in *T. brucei*

#### 8.3.1. What is antigenic variation?

*T. brucei* evades clearance by the host antibodies by periodically changing its VSG surface coat in a process known as antigenic variation (25). *T. brucei* expresses only one of its VSG genes from a repertoire of about 2,500 VSG genes and pseudogenes (25). VSG genes and pseudogenes can be found in sub-telomeric arrays of chromosomes and mini-chromosomes (101, 102). Chromatin conformation capture and high throughput sequencing (HI-C) based assembly of the genome of *T. brucei* Lister 427 resolved the organisation of ESs in the chromosomes which was previously unknown due the difficulties posed by the presence of large repeat sequences (103).

The VSGs are homodimers of 50-60 kDa, anchored to the membrane by glycosylphosphatidyl inositol (GPI) anchors and are known to form a surface coat of 12-15 nm thickness (104) The N-terminal domain of the VSGs is composed of 300-350 amino acids (105). The N-terminal domain is less conserved among VSGs with a sequence homology between 15-30% and currently available data indicate that they are exposed to the external environment (106) (107). They are classified into three classes (A to C), based on their cysteine residue patterns (105). The more conserved C-terminal domain is a 100 amino acid long domain (105). The conservation of the C-terminus does not affect the immunogenicity of the protein, as the scope of epitope recognition by host antibodies is limited to the N-terminal domains of the protein (105). Importantly, the C-terminus contains the GPI-addition site which mediates the attachment of VSGs to the GPI anchor, which is essential for the surface expression of VSGs (105). The GPI is a post-translational modification wherein a phosphoglyceride is added to the C-terminus of a protein (108). The GPI consists of a phosphatidyl inositol (PI) group with a carbohydrate linker such as mannose or glucosamine and this is linked to the C-terminus of a protein via an ethanolamine phosphate bridge (108). The fatty acid components of the PI group anchor the protein to the membrane (108).

### 8.3.2. Mechanisms of antigenic variation



**Fig. 3. Two mechanisms of antigenic variation in *T. brucei*.** Antigenic variation in *T. brucei* occurs by transcriptional switching or by VSG gene recombination. During transcriptional switching the actively transcribed ES is silenced with the simultaneous activation of a silent ES. In recombination switching, the actively transcribed VSG is replaced by a new VSG from other genomic regions resulting in the loss of the actively transcribed VSG. Recombination may also occur through segmental gene conversion where chimeric VSG sequences derived from multiple VSG genes or pseudogenes replace the actively transcribed VSG. Another form of recombination is telomeric exchange between the telomeric end of the transcribed ES and a silent ES resulting in the transcription of the newly recruited silent VSG gene.

VSG genes are transcribed from telomeric ESs where only a single VSG is transcribed at a time (Fig 2B). Antigenic variation in *T. brucei* is governed primarily by two mechanisms, namely transcriptional switching, and homologous recombination (Fig. 3). Transcriptional switching involves the silencing of an actively transcribed ES and the simultaneous activation of a silent ES (Fig. 3) (109). This leads to the expression of a new VSG gene without any change in DNA sequences (109). This mechanism of antigenic variation in trypanosomes limits the available pool of VSGs to those found on the 15 BESs. In order to access the larger repertoire of VSGs, homologous recombination is used (110). Recombination occurs by gene conversion, segmental gene conversion or telomeric exchange (Fig. 3) (110). Gene conversion, involves the replacement

of the VSG gene in the active ES by a VSG gene from other genomic regions and results in the loss of the former (Fig. 3) (111). Segmental gene conversion involves the formation of chimeric VSG gene which results from the recombination of multiple VSG sequences (Fig. 3) (112). This chimeric sequence replaces the actively expressed VSG genes through homologous recombination (110, 112). Interestingly, this mechanism seems to be more frequent during chronic infection (112). Another form of recombination entails telomeric exchange between the telomeric end of the transcribed ES and a silent ES resulting in the transcription of the recruited silent VSG gene (Fig. 3) (113). These mechanisms have potential to sustain the infection in the host, protecting the parasite from the host antibody response.

### **8.3.3. Structure of the ES chromatin and Pol I transcription**

A unique feature of the transcription of VSGs is that it occurs at a specialised extranucleolar location called the expression site body (ESB) (74). Immunofluorescence in *T. brucei* BFs revealed that the ESB was associated only with the actively transcribed ES in BFs whereas it is not found in PFs (74). In addition,  $\alpha$ -amanitin treatment to abrogate Pol II and III mediated transcription, suggested that the ES is transcribed by an  $\alpha$ -amanitin insensitive polymerase which could be Pol I (74). Immunofluorescence analysis reveal a single ESB during the S and G2 phases of the cell cycle (114). Just before cell division initiates, the sister chromatids containing the active ES segregate with Pol I (114). The division is completed with one ESB associated to one active ES in either daughter cells (114). Therefore, the ESB is associated with the single active ES that is inherited by the daughter cells in BFs. Efforts to express two ESs simultaneously under a drug selection pressure did not result in the stable transcription of both ESs (115). The two ESs simultaneously occupied the ESB and were transcribed only when they were in a close proximity of about 200 nm (115). Interestingly, this simultaneous occupation of the ESB was not a stable event (115). Instead, the 2 ESs alternate between the single ESB leading to the production of unstable transcripts i.e. transcripts with a much shorter half life (115). This is indicative that the resources such as RNA transcription and RNA processing machinery within the ESB are limited (115). For instance, transcripts produced simultaneously from 2 ESBs were not polyadenylated suggesting that the transcribed ESs do not have the required access to a robust RNA processing machinery (115). Single cell PCR and real time PCR (RT-PCR) analysis in *T. brucei* revealed that

transcription initiates for all ESs (116). However, analysis of the transcripts showed that transcription of all but one of the ESs ends before the VSG gene, allowing for the complete transcription of only one of the ESs and its associated VSG gene while others remain silent (116). This was indicative of a mechanism whereby transcription initiates at multiple ESs but only one of the ESs, the active one, is completely transcribed.

#### **8.3.4. Factors regulating ES transcription**

There are various factors that affect the transcription of the ES by Pol I. The Pol I mediated transcription of the ES is influenced by protein complexes as well as epigenetic modifications of the chromatin. Landeira *et.al.*, revealed that cohesins were crucial for accurate replication and segregation of the ES (114). In eukaryotes, cohesins have an important role in maintaining the structural integrity of sister chromatids (117). The RNAi-mediated knockdown of *T. brucei* sister chromatid cohesion 1 (SCC1), the orthologue of yeast cohesion subunit led to the abnormally fast separation of sister chromatids during cell division (114). Intriguingly, the depletion of SSC1 led to switching of the VSG expressed by a population of the cells suggesting a potential role of the protein in switching of the VSGs (114). were crucial for accurate replication and segregation of the ES (114). In eukaryotes, cohesins have an important role in maintaining the structural integrity of sister chromatids (117). The RNAi-mediated knockdown of *T. brucei* sister chromatid cohesion 1 (SCC1), the orthologue of yeast cohesion subunit led to the abnormally fast separation of sister chromatids during cell division (114). Intriguingly, the depletion of SSC1 led to switching of the VSG expressed by a population of the cells suggesting a potential role of the protein in switching of the VSGs (114).

Other factors contribute to the regulation of transcription at different levels through association with different regions of the ES chromatin. For instance, the ES promoter of the actively transcribed ES is enriched in the (CITFA), an indispensable factor for Pol I transcription (78, 118). Immunofluorescence analysis revealed that the CITFA complex was also localised in the nucleolus and the periphery of the nucleolus (77). ChIP assays using antibodies for the subunits 2 or 7 of the CITFA complex showed that it was enriched in the active ES and specifically at the promoter region (118). Moreover, immunofluorescence showed that the protein was localised at the ESB

(118). The knockdown of the subunit 7 proved lethal to the parasite and led to a reduction in the rRNA, VSGs, and procyclin transcript levels (75).

Transcription of the active ES is also subject to a post-translational modification called SUMOylation that potentially affects the interaction of RNA Pol I (119). SUMOylation is a reversible modification of proteins at the lysine residues and entails the addition of a 12 kDa SUMO proteins at these sites (120). SUMO proteins are involved in various cellular processes such as protein translocation, DNA repair, segregation of chromosomes in many higher organisms (120). Biochemical assays paired with *in silico* analysis led to the identification of key enzymes that catalyse this process: E1 activating enzyme, an E2 conjugating enzyme and an E3 ligase or SAP and Miz finger binding domain 1 - *TbSIZ1* (named after its yeast counterpart) (119). The E1 enzyme binds the SUMO protein and transfers the protein to enzyme E2 which conjugates the SUMO protein to a lysine residue in the target protein. Finally, the E3 ligase fuses the SUMO protein to the target producing a SUMOylated protein. In *T. brucei*, immunofluorescence assays on BFs using monoclonal antibodies developed against the *T. brucei* SUMO protein have identified that the protein localizes at the nucleus with a partial overlap at the ESB(121). In BFs, about 75% of the protein was found at one focus (overlapping with the ESB) with the rest of it distributed through the nucleus whereas similar assays in PFs (where all VSGs are silent) showed a dispersion of many SUMO foci throughout the nucleus with no major focus (121). The RNAi-mediated depletion of *TbSIZ1*, followed by ChIP analysis revealed a reduction in VSG ES chromatin SUMOylation (121). This led to a decreased recruitment of Pol I to the active ES with no differences in the silent ESs resulting in lower transcript reads as detected by RT-PCR analysis (121). In summary, the SUMO proteins play an important role by positively regulating the transcription of the VSG expression site as well as recruiting Pol I transcripts to the ES. Transcription of the active ES is also subject to a post-translational modification called SUMOylation that potentially affects the interaction of RNA Pol I (119). SUMOylation is a reversible modification of proteins at the lysine residues and entails the addition of a 12 kDa SUMO proteins at these sites (120). SUMO proteins are involved in various cellular processes such as protein translocation, DNA repair, segregation of chromosomes in many higher organisms (120). Biochemical assays paired with *in silico* analysis led to the identification of key enzymes that catalyse this process: E1 activating enzyme, an E2 conjugating enzyme and an E3 ligase or

SAP and Miz finger binding domain 1 - *TbSIZ1* (named after its yeast counterpart) (119). The E1 enzyme binds the SUMO protein and transfers the protein to enzyme E2 which conjugates the SUMO protein to a lysine residue in the target protein. Finally, the E3 ligase fuses the SUMO protein to the target producing a SUMOylated protein. In *T. brucei*, immunofluorescence assays on BFs using monoclonal antibodies developed against the *T. brucei* SUMO protein have identified that the protein localizes at the nucleus with a partial overlap at the ESB (121). In BFs, about 75% of the protein was found at one focus (overlapping with the ESB) with the rest of it distributed through the nucleus whereas similar assays in PFs (in which all VSGs are silent) showed a dispersion of many SUMO foci throughout the nucleus with no major focus (121). The RNAi-mediated depletion of *TbSIZ1*, followed by ChIP analysis revealed a reduction in VSG ES chromatin SUMOylation (121). This led to a decreased recruitment of Pol I to the active ES with no differences in the silent ESs resulting in lower transcript reads as detected by RT-PCR analysis (121). In summary, the SUMO proteins play an important role by positively regulating the transcription of the VSG expression site as well as recruiting Pol I transcripts to the ES.

Imitation switch (ISWI) family of chromatin modellers are ATP-dependent proteins conserved among eukaryotes. They also possess a characteristic HAND (because the four helical structures resemble an open hand)-SANT (Swi3 (switch 1 interacting protein 3)-Ada2 (Adenosine deaminase 2)-N-cor (Nuclear receptor co-repressor))-TFIIIB (Transcription factor IIIB)) -SLIDE (SANT-like ISW1 domain) domain that are involved in the DNA binding activity of the complex. Extensive work on the ISW1 complexes in yeast have revealed transcription-related functions. For example, the ISW1a complex in yeast binds to Pol II promoters and prevent transcription initiation (122). The *T. brucei* ISW1 protein was found to bind silent regions in the genome such as the 177 bp repeat arrays in the minichromosomes, 50 bp repeats upstream of the VSG gene ESs and also SSRs (123). RNAi mediated silencing of the ISWI in bloodstream or procyclic forms of the parasite led to expression of a green fluorescent protein (GFP) integrated in the otherwise silent 177 bp arrays (124). Also, RNAi mediated depletion of ISWI followed by flow cytometry analysis revealed that the *Dicosoma* species-red (DS-red) reporter gene integrated downstream of a promoter was derepressed (125). ChIP followed by qPCR (ChIP-qPCR) revealed ISW1 to be enriched in promoters of silent ES but was depleted in the ES that was actively transcribed (123). Interestingly, the depletion of the protein led to an increase in the expression of GFP coding gene integrated

immediately downstream of the promoter of a silent ES in both BFs and PFs (123). The protein was also found associated with Pol II transcribed PTUs, at the divergent SSRs which are involved in terminating Pol II transcription but its depletion did not affect the Pol II transcription (123). The *T. brucei* ISW1 complex was characterised in 2015 by Stanne *et.al* by immunoprecipitation experiments with ISW1 and identified the interacting partners in both BFs and PFs (126). These include: Nucleoplasmin like protein (NLP) (has a nucleoplasmin like domain), regulator of chromosome condensation 1-like protein (RCCP) (has 4 RCC1-regulator of cell cycle 1 domains) and phenylalanine/tyrosine-rich protein (FYRP) (has an N-terminal phenylalanine/tyrosine rich domain) (126). Moreover, RNAi mediated depletion of RCCP and FYRP proteins led to the expression of the GFP gene integrated into silent VSG ESs downstream of the promoter, indicating their importance in silencing the expression sites (126). This set of experimental data on the ISWI protein complex highlights its importance in the regulation of ES transcription in *T. brucei*.

Glover *et.al.*, identified that VSG exclusion 1 (VEX-1) localised in the active ES in BFs (127). Analysis of epitope-tagged VEX-1 by superresolution microscopy revealed that VEX-1 was concentrated in a single subnuclear compartment (127). Immunofluorescence analysis of Telomeric repeat binding factor-2 (TRF2), which localizes at the nucleus, revealed that one TRF-2 puncta was coincident with the VEX-1 puncta. Moreover, immunofluorescence analysis also revealed that VEX-1 was closely associated with the extranucleolar RNA Pol I focus which is the actively transcribed ES (127) (128). However, this interaction of VEX-1 was ablated when cells were treated with actinomycin D, a transcription inhibitor (127). Immunofluorescence analysis of VEX-1 in PFs showed a diffuse distribution of VEX-1 including the nucleus (127). Thus, VEX-1 associates with the active ES in BFs in a transcription dependent manner (127). Knockdown or overexpression of VEX-1 in the BFs led to the transcription of various silent VSGs and multiple VSG proteins were expressed by the cells (127). Further studies by the same group identified VEX-1 interacting partners: VEX-2, an orthologue of non-sense mediated mRNA decay ATP-dependent superfamily 1- type helicases which along with VEX-1 forms a VEX complex that is crucial to maintain the monoallelic transcription of the active VSG (128). The other enriched proteins were components of Chromatin Assembly Factor 1 (CAF-1), a conserved replication associated histone chaperone (128). Super-resolution microscopy revealed that VEX-2 occupied the same focus as VEX-1 in BFs, at the extranucleolar Pol I compartment (128). Knockdown of

VEX-2 resulted in the de-repression of various silent VSGs (128). Intriguingly, the simultaneous knockdown of the VEX-1 and VEX-2 proteins led to the formation of a very heterogeneous surface coat and led to cell death, highlighting the importance of these proteins in maintaining VSG monoallelic expression (128). In addition to this, there was also an increase in the number of VEX-1 puncta. The role of CAF-1 was also investigated since it was associated with the complex (128). Knockdown of the gene led to the increased expression of promoter proximal VSG genes and a relatively lower expression of promoter distal ones (128). Previous studies have reported a CAF-1 mediated de-repression of VSGs during S phase. Immunofluorescence analysis of the GFP-tagged CAF-1 subunit CAF-1b revealed that despite its overall nuclear localization, a major CAF-1b foci localises with a subnuclear compartment defined by both VEX-1 and VEX-2 (128). Moreover, this co-incident foci was further enriched during the S-phase suggesting an involvement of CAF-1 in the inheritance of VEX complex as well as the associated epigenetic state (128). The authors suggest a VSG exclusive expression model in which the association of VEX-1 with the Pol I compartment is VEX-2 dependent and its inheritance is CAF-1 dependant (128). However, this would require further investigation to establish a defined role of the proteins in this complex in regulating VSG gene expression (128).

### **8.3.5. Epigenetic modifications of the ES chromatin**

Apart from the above discussed factors that affect transcription of the ES, there are epigenetic modifications of the ES chromatin which influence the monoallelic nature of VSG expression. Navarro *et.al.*, discovered the role of epigenetics in the regulation of VSG gene expression (129). In order to study Pol I accessibility of an inactive ES, the authors inserted a T7 RNA polymerase (T7RNAP) promoter driven construct containing a bleomycin resistance (BLE) and LUC gene coding for luciferase upstream of the inactive ES (129). This allowed them to monitor the T7RNAP promoter activity (129). Based on luciferase assays in BFs they observed that despite the ES being silent, the regions around it (in this case the T7RNAP driven luciferase gene) were accessible to the polymerase (129). However, when they differentiated these BFs into PFs, the expression of luciferase was greatly diminished (129). Thus, the ES is efficiently repressed during differentiation from BFs to PFs and this may involve remodelling of the chromatin (129). In the regulation of VSG gene expression (129). In order to study Pol I accessibility of an inactive ES, the authors

inserted a T7 RNA polymerase (T7RNAP) promoter driven construct containing a bleomycin resistance (BLE) and Luciferase (LUC) gene upstream of the inactive ES (129). This allowed them to monitor the T7RNAP promoter activity (129). Based on luciferase assays in BFs they observed that despite the ES being silent, the regions around it (in this case the T7RNAP driven luciferase gene) were accessible to the polymerase (129). However, when they differentiated these BFs into PFs, the expression of luciferase was greatly diminished (129). Thus, the ES is efficiently repressed during differentiation from BFs to PFs and this may involve remodelling of the chromatin (129).

Other proteins that influence the epigenetics of the ES chromatin include histones, histone modifying proteins, BDFs, proteins involved in base J modifications, and histone chaperones. Active ES of trypanosomes are depleted of nucleosomes (124, 130). In eukaryotes, the chromatin is wound around a histone core consisting of histone proteins. This level of organisation allows for efficient packaging of the chromatin and also plays a crucial role in gene regulation. ChIP analysis revealed that the active ES was depleted of histones H3, H2 and H4A compared to the silent ESs (124). ChIP analysis revealed that the active ES was depleted of histones H3, H2 and H4A compared to the silent ESs (124). These results were further corroborated by micrococcal nuclease (MNase) digestion analysis of DNA (124). MNase preferentially digests chromatin at the linker regions present between nucleosomes and can be used to identify the spatial distribution of nucleosomes (124). In this study, permeabilized *T. brucei* cell lysates were subject to partial digestion with micrococcal nuclease (124), and the resultant material was then resolved on 5-30% sucrose gradients followed by qPCR analysis (124). The DNA from active ES was less enriched in di/ tri- nucleosome fractions compared to DNA from silent ESs (124). The findings in this study were consistent with the findings of a similar study conducted by Figueirido and Cross (130). This suggests that the silent ES adopt a more compact structure relative to active ES, brought about by the enrichment of histones. (MNase) digestion analysis of DNA (124). MNase preferentially digests chromatin at the linker regions present between nucleosomes and can be used to identify the spatial distribution of nucleosomes (124). In this study, permeabilized *T. brucei* cell lysates were subject to partial digestion with micrococcal nuclease (124), and the resultant material was then resolved on 5- 30% sucrose gradients followed by qPCR analysis (124). The DNA from active ES was less enriched in di/ tri- nucleosome fractions compared to DNA from silent ESs (124). The findings in this study were consistent with the findings of a similar study conducted by Figueirido

and Cross (130). This suggests that the silent ES adopt a more compact structure relative to active ES, brought about by the enrichment of histones.

A more recent study looks at how histone variants H3V and H4V act as architectural proteins that influence the genome of the parasite consequently affecting the expression and switching of VSGs (103). Single cell RNA sequencing from cells with mutant histones variants revealed that perturbing the function of these histone variants led to the transcription of multiple silent VSG genes at the same time with about 70% of cells switching to a different VSG (VSG 11), different from the initially active VSG (VSG 2) while the expression of other VSGs gradually declined (103). This indicated that the cells were switching VSGs in a manner that was not random and the expression of multiple VSGs was a state of transition allowing a new VSG to be expressed (103). Single Molecule Real Time (SMRT) sequencing revealed that there was a recombination process between the ESAG 8 gene pairs found on BES 1 and BES 15, the ESs housing VSG 2 and 11 resulting in the formation of a BES possessing 3 ESAG 8 genes: 2 from BES 15 and 1 from BES 1 (103). HI-C analysis revealed that there was an interaction between the end of chromosome 6 where VSG 2 is located and VSG 11 which led to the relocation of VSG 11 to chromosome 6 (103). Assays for transposase accessible chromatin (ATAC) with sequencing revealed that upon the mutation of H3V and H4V, there was an increase in the interactions between BESs that were repressed (103). This data further highlights the importance of histones in maintaining the monoallelic status of VSG expression and in mediating switching of VSGs. A more recent study looks at how histone variants H3V and H4V act as architectural proteins that influence the genome of the parasite consequently affecting the expression and switching of VSGs (103). Single cell RNA sequencing from cells with mutant histones variants revealed that perturbing the function of these histone variants led to the transcription of multiple silent VSG genes at the same time with about 70% of cells switching to a different VSG (VSG 11), different from the initially active VSG (VSG 2) while the expression of other VSGs gradually declined (103). This indicated that the cells were switching VSGs in a manner that was not random and the expression of multiple VSGs was a state of transition allowing a new VSG to be expressed (103). Single Molecule Real Time (SMRT) sequencing revealed that there was a recombination process between the ESAG 8 gene pairs found on BES 1 and BES 15, the ESs housing VSG 2 and 11 resulting in the formation of a BES possessing 3 ESAG 8 genes: 2 from BES 15 and 1 from BES 1 (103). HI-C analysis

revealed that there was an interaction between the end of chromosome 6 where VSG 2 is located and VSG 11 which led to the relocation of VSG 11 to chromosome 6 (103). Assays for transposase accessible chromatin (ATAC) with sequencing revealed that upon the mutation of H3V and H4V, there was an increase in the interactions between BESs that were repressed (103). This data further highlights the importance of histones in maintaining the monoallelic status of VSG expression and in mediating switching of VSGs.

Base J replaces 1% of the DNA base Ts in the telomeres, the expression sites and at sites involved in RNA Pol II transcriptional initiation and termination. (66). It influences the transcription of VSG- ESs (66). Base J is enriched in the telomeres and the silent ESs whereas it is depleted from the actively transcribed VSG-ESs (66). The depletion of base J by treatment with Dimethylloxalyglycine (DMOG), a compound that chemically reduces base J led to the de-repression of VSGs (66).

Facilitates Chromatin Transcription Complex (FACT), a complex enriched in silent ESs relative to active ES is also implicated in the regulation of VSG ES transcription (131, 132). The FACT complex in *T. brucei* is composed of Suppressor of transcription (Spt16) and Pol binding (Pob3) subunits (131, 132). RNAi-mediated knockdown studies on Spt16 led to a growth arrest in the G2/early M phase with the disruption of chromosomal segregation in both BFs and PFs (131). RNAi followed by FACS analysis using *T. brucei* cell lines with a DS-Red reporter gene revealed that along with the arrest in growth, the reporter gene in the silent VSG-ESs were also derepressed in both BFs and PFs (131). ChIP analysis revealed that the depletion of Spt16 led to a decrease in the levels of H3 and H2A histones at the silent ESs (132). Depletion of these histone variants at the silent ESs lead to a more open chromatin conformation at the silent ESs. These results indicate that the histone chaperoning activity of FACT affects the chromatin structure and consequently affects the expression of VSG- ESs. ChIP analysis revealed that the depletion of Spt16 led to a decrease in the levels of H3 and H2A histones at the silent ESs (132). Depletion of these histone variants at the silent ESs leads to a more open chromatin conformation at the silent ESs. These results indicate that the histone chaperoning activity of FACT affects the chromatin structure and consequently affects the expression of VSG- ESs.

### 8.3.6. Mechanisms involved in the switching of VSGs.

Various proteins have been implicated in regulating mechanisms of switching. For instance, histone methyltransferases such as DOT1B are known to influence the process of transcriptional switching (133). DOT1B was crucial for maintaining the kinetics of switching (133). The authors used wild type and mutant DOT1B cell lines that transcribe two VSGs under drug selection pressure. Upon removal of the drug selection pressure, cells expressing wild type DOT1B were able to immediately revert to a monoallelic state of VSG expression, expressing only either of the two VSG genes whereas cells expressing the mutant DOT1B took up to 30 days to revert to a monoallelic state of VSG expression (133). RAD-51, an integral member of the homologous recombination pathway is crucial for VSG switching, enabling recombination of VSG sequences from the large repertoire of available VSG gene sequences (134). Mutation of RAD-51 has been shown to decrease the rate of VSG switching to extents ranging from 2-fold up to 130-fold (134). Another protein integral for the recombination of VSGs is the telomere associated protein: telomeric repeat binding factor (TRF) which protects the telomeres and is known to impact the switching of VSGs (135). To investigate the role of TRF, RNAi was used to deplete the protein, and this was followed by assays to evaluate the effect on the expression of VSGs (135). These assays demonstrated that the depletion of TRF led to the switching of VSGs (135). About 65% of the cells switched their VSGs by expression site recombination (135), and 33% of the population switched via gene conversion (135). However, before the induction of RNAi, the majority of the cells (85%) underwent switching via VSG gene conversion events (135). This indicated that TRF influenced the mechanism of switching (135). Similarly, PIP5K, a phosphatidylinositol phosphate kinase was reported to play a role in the regulation of VSG gene switching (136). Knockdown of PIP5K in BFs led to the derepression of VSG genes and re-expression of PIP5K in these cells led to the expression of only a single VSG (136). However, most cells underwent a switch in VSG gene expression- 16% of the analysed clones transcribed VSG13, 2% transcribed VSG8, 2% transcribed VSG21, and 1% transcribed VSG19 while the remaining 43% continued to express the same VSG that they expressed before the knockdown of PIP5K (136). Among this population, 21% cells underwent recombination to express a VSG that was previously not found in any of the ES (136).

Models have been proposed explaining the mechanism of VSG switching have been proposed. One model emphasizes the role of the DNA double strand breaks (DSBs) (137). This model suggests that DSBs lead to the formation of DNA lesions in the ES that promote recombination events (137). Ligation mediated PCR has revealed that the 70 bp repeats of an active ES is more susceptible to such DSBs when compared to silent ESs (137). Such DSBs would promote homologous recombination as a repair mechanism which in turn leads to the replacement of the VSG gene through recombination. Further evidence by Glover *et. al.*, revealed that 70 bp repeats serve as a source for homologous recombination (138). R-loops (three-stranded RNA-DNA structures with an RNA-DNA hybrid and a displaced single-stranded DNA) have also been implicated in influencing the switching of VSGs (139). R-loops are stable structures that contribute to genomic instability interfering with the DNA replication fork progression, causing DSB formation and recombination (139). Various enzymes such as helicases and RNaseH are involved in resolving the structure or degrading the RNA in the loops (139). DSBs introduced due to R-loop formation at either the 70 bp repeat regions or telomeric repeat regions are potential inducers of homologous recombination and enhance the process of VSG switching (139). Evidently, artificially introducing an R-loop upstream of an active VSG gene led to a 250-fold increase in the VSG switching frequency (137, 139).

### **8.3.7. Telomeric position effect and silencing of ES transcription**

Telomeric positioning effect (TPE) is an inheritable and also reversible phenomenon where the transcription of genes lying in close proximity to the telomeres are suppressed (140). It was first reported in yeast by Gottschling *et.al.*, where Pol II transcribed genes lying within 5 kb from the telomeres were suppressed (140). David Horn and George Cross showed that a similar effect manifests in trypanosomes whereby VSG genes are suppressed (141). Three *T. brucei* cell lines were generated by transforming them with a construct possessing either an rRNA promoter, or a PARP promoter or an ES promoter. Each cell line was also transformed with a reporter for the endogenous ES promoter (downstream of the endogenous ES promoter) and another reporter for the integrated promoters (downstream of the respective promoters). None of the reporter genes were transcribed from the silent ES in BFs (141). A similar experiment was carried out in PFs (141). The reporter for the endogenous ES promoter mediated transcription remained suppressed

in PFs (141). The reporter for the integrated promoters was transcribed in the case of rRNA and PARP promoters but not in the case of the integrated ES promoter (141) This study revealed that *T. brucei*, like yeast exhibits a telomeric position effect (141).

#### **8.4. The role of Repressor activator protein 1 (RAP1) in VSG gene expression**

Yeast two hybrid screening using TRF as a bait led to the identification of RAP1 as a TRF-interacting protein (142). RAP1 is an essential telomeric protein found among many eukaryotes. It is a sequence specific DNA binding protein that has been implicated in various crucial cellular functions in eukaryotes (142). These include transcriptional repression/ activation, maintenance of telomeric structural integrity as well as protection of chromosome ends (142). RAP1 has been extensively studied in yeast (*S. cerevisiae*) where it directly binds duplex or double-stranded telomeric repeats (142). As described earlier, in yeast, it is known to play a major role in repressing gene expression by means of a telomeric position effect which silences genes that lie in close proximity to the telomeric region (142). Interestingly, in humans, RAP1 binds indirectly to the telomeric repeats by association with TRF2 (142).

*T. brucei* RAP1 consists of three domains that have been a recent subject of study. It consists of a N-terminal Breast cancer 1 (BRCA1) C-terminus (BRCT) domain, a central Myb (named after a functionally similar gene that causes avian myeloblastosis) and a C-terminal Myb-like (Mybl) domain (143). BRCT domains have been implicated in binding phosphopeptides and form an integral part of many proteins involved in cell cycle regulation as well as DNA repair (144). The Myb and Mybl domains are found in other DNA binding proteins and chromatin remodelling proteins (145). Previous work on RAP1 has shown that it is essential for the regulation of VSG expression in both BFs as well as PFs (142, 146). RNAi-mediated knockdown of RAP1 in BFs proved to be lethal to the parasite and it led to the expression of silent VSG genes (142). Ablation of RAP1 expression led to a 2 to 25-fold or 8 to 56-fold increase in the expression of silent VSGs 18 or 36 hrs post knockdown of RAP1 (142). IFA performed on these cells showed that multiple VSGs were expressed on the surface of the parasite (142). The most interesting finding in the study is that the silencing effect brought about by RAP1 is stronger at regions closer to the telomeres (142). Data from this study shows that RAP1 has the ability to silence the entire ES with a stronger

influence upon the regions closer to the telomeric region (142). Similarly, knockdown of RAP1 in PFs has shown that RAP1 is essential for the growth as well as the silencing of VSGs in PFs (146). qPCR data show a 10 to several 100-fold increase in the expression of silent VSGs when RAP1 was knocked down in PFs (146). Comparing BFs and PFs revealed a stronger silencing of VSGs in PFs, possibly due to a more compact chromatin structure in PFs relative to BFs (146). In summary, RAP1 plays a crucial role in silencing VSGs in both BFs and PFs.

Recent reports have shown that the RAP1 domains are crucial for certain protein-protein interactions such as its interactions with TRF or interactions between RAP1 molecules (147). Expression of a RAP1 allele without the Myb domain in *T. brucei* proved lethal to the parasite. RNA-seq data also show that the VSGs (about 2700) were upregulated upon the expression of this allele (147). Interestingly, co-immunoprecipitation with *T. brucei* TRF as well as immunofluorescence analysis in cells expressing RAP1 allele lacking the Myb domain revealed that the Myb domain is necessary for the interaction of RAP1 with TRF (147). The authors also produced cells expressing RAP1 alleles that lacked the Mybl domain (147). IFA in these cells revealed that the Mybl domain is important for the import of RAP1 into the nucleus by the importin protein (147). Sequencing analysis showed that the Mybl domain had a nuclear localization signal important for the import of RAP1 into the nucleus (147). Co-immunoprecipitation experiments also showed that RAP1 has a self-interacting ability which is dependent on the BRCT domain of the protein (147). This study sheds light on RAP1 and its domains in context to RAP1's interaction with TRF, its self-interacting ability and also the localization of RAP1 in the nucleus (147). These domains are all functionally important for RAP1's ability to perform its cellular functions, especially regulate VSG gene expression to sustain the monoallelic nature of its expression.

## **8.5. Phosphoinositides and the regulation of VSG expression**

Phosphatidylinositols (PIs) are molecules that possess a glycerol backbone, two non-polar fatty acid tails and a phosphate group with a polar inositol head group (148). When PIs undergo phosphorylation, they form PI phosphates (PIPs). An extensively studied aspect of PIP biology is their role as integral constituents of the cell membrane, protecting cellular components from harsh environmental changes such as high salinity, freezing temperatures, or even from dehydration

(148). PIs also play a major role as cell surface anchors in the form of GPIs in many eukaryotes ranging from single-celled parasites such as *T. brucei* and *Giardia lamblia* up to metazoans such as mammals (148). PIs also have various cellular functions including cell signalling and trafficking (148). For instance, phosphatidylinositol 4-phosphate (PI4P) and phosphatidylinositol 3-phosphate (PI3P) are essential for trafficking the precursors of complex glycosphingolipids through the Golgi complex and then to the plasma membrane (149). The loss of PI3P, for example, impedes the flux of these molecules out of the Golgi apparatus (149). Also, drastic changes in PIP concentrations are known to affect various cellular functions, consequently affecting cell survival (149).

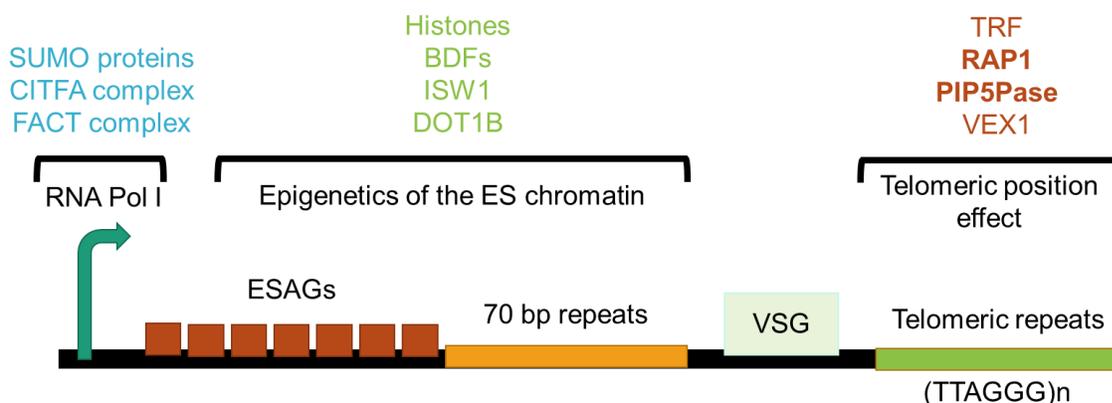
The genome of *T. brucei* has 26 genes that code for members of the pathway including inositol phosphates (IP) or PI kinases and phosphatases (23 in number), a phospholipase C, enzymes that are crucial for the synthesis of IP or PIs (4 in number), an IP symporter and an IP3 receptor (IP3R) (136, 150). The IP metabolites as well as the various enzymes of the IP pathway occupy various subcellular locations including the nucleus (PIP5Pase), plasma membrane (IPMK and PLC), endosomes (PIP5K), cytosol (IPMK) and acidocalcisomes (IP3RyR) (22, 136, 150-155). IPs and PIs are essential in *T. brucei* where members of the IP pathway have been studied in association with various cellular functions (150). Apart from functioning as structural molecules, their functions also include cellular signalling, regulators of cell development and energy metabolism and also contribute to the regulation of VSG gene expression (136, 151). Studies involving the knockdown or mutation of members in the pathway have shed light on their role in cellular viability, growth, and development. For instance, studies on inositol polyphosphate multikinase (IPMK) has been implicated in regulating life stage development from BFs to PFs (22). Specifically, the knockdown or the loss of catalytic activity of IPMK led to the upregulation of certain PF stage-specific genes, activation of the OXPHOS pathway, expression of the procyclins on the surface coat, and upregulation of genes coding for RBPs, phenotypes associated with the development of PFs from BFs (22). Another example is *T. brucei* Fab1, a PI(3,5)P2 kinase, a protein involved in endosomal and lysosomal trafficking in *T. brucei* (151). RNAi-mediated knockdown of PI(3,5)P2 kinase proved lethal to the parasite impeding the turnover of ubiquitinated membrane protein, highlighting the importance of the PI(3,5)P2 kinase in trypanosome endomembrane biology (151).

The genes encoding enzymes of the PIP pathway: phosphatidylinositol phosphate-5 kinase (PIP5K), phospholipase C (PLC) and phosphatidylinositol phosphate-5 phosphatase (PIP5Pase) are involved in the regulation of VSG expression (143) (136). Gene expression analysis show that the conditional knockdown of either PIP5Pase or PIP5K led to the expression of silent BF and MF ESs whereas the overexpression of PLC led to a similar phenotype (143) (136). The mRNA levels of VSGs in PIP5Pase or PIP5K knockdowns were about a 10,000-fold higher relative to the wild type cells that expressed a single VSG (136). Once the expression PIP5PK was restored, the monoallelic state of VSG expression was restored as well (136). Gene expression analysis of cells after restoration of PIP5K by qPCR revealed that 16% of the analysed clones transcribed VSG13, 2% transcribed VSG8, 2% transcribed VSG21, and 1% transcribed VSG19 while the remaining 43% continued to expressed the same VSG that they expressed before the knockdown of PIP5K (136). Immunofluorescence analysis of BFs showed that PLC, PIP5K and their substrates (PIPs) localised at the plasma membrane. However, PIP5Pase localizes in the nucleus and co-localises with telomeres (136). Interestingly, TRF and RAP1 also have a similar pattern of localisation and they partially localised with telomeres (136). Coimmunoprecipitation and colocalization experiments proved that RAP1 and PIP5Pase interact (143) (136). IP pathway and their association with telomeric proteins (RAP1 and TRF) was further studied using glycerol gradients, followed by western blot analysis of the glycerol gradient fractions (136). Sedimentation of RAP1 and TRF was initially determined using cell lysates from *T. brucei* BFs that express TAP tagged TbPIP5Pase (136). However, knock down of PIP5K affected the sedimentation both RAP1 and TRF (136). TRF sedimented in a higher sedimentation co-efficient whereas RAP1 shifted to a slightly lower sedimentation co-efficient (136). This suggests that the IP pathway may regulate the interaction between RAP1 and TRF (136). Immunofluorescence analysis of RAP1 after knocking down PIP5Pase revealed that the knockdown of PIP5Pase led to the formation of multiple RAP1 loci (136). Interestingly, FISH analysis by 3D-deconvolution microscopy using a telomeric repeat probe revealed that knockdown of PIP5Pase led to an increase in the number of telomeric foci from less than five up to ten telomeric foci (136). This data, along with data from VSG gene expression analysis suggests that the IP pathway regulates the association of these telomeric proteins with the telomeric ES and the disruption of their interaction may lead to the transcription of multiple ES (136).

PIP5Pase dephosphorylates PI(3,4,5)P3 yielding PI(3,4)P2 (136) (143). Biochemical assays using PI(3,4,5)P3 as a bait revealed that RAP1 also binds to PI(3,4,5)P3 (143). However, in the presence of PIP5Pase, PI(3,4,5)P3 is dephosphorylated, depriving RAP1 of PI(3,4,5)P3 (143). Interestingly, western blot analysis of *T. brucei* BF cell lysates run on a native gel revealed that PIP5Pase and RAP1 were part of a 0.9 MDa protein complex (143). Immunoprecipitation of PIP5Pase followed by MS analysis revealed that RAP1 was the top interacting candidate while immunoprecipitation of RAP1 followed by MS analysis revealed that PIP5Pase was the top interacting candidate (143). Other proteins identified through the MS analysis were: proteins involved in RNA processing, nuclear lamina proteins, regulatory proteins and proteins involved in RNA processing (143). In addition, ChIP analysis of RAP1 and PIP5Pase showed that PIP5Pase was enriched in 70 bp repeats whereas RAP1 was enriched in 70 bp and telomeric repeats (143). ChIP analysis of RAP1 in a cell line expressing a catalytic mutant of PIP5Pase (D360A/N362A) showed that the mutation of PIP5Pase increased the enrichment of RAP1 at the 70 bp repeats whereas it reduced its enrichment at the telomeric repeats (143). Amino acids D360 and N362 are crucial for the enzymological activity of *T. brucei* PIP5Pase and are conserved among yeast and human PI 5-pases (143). Enzymological analysis of the mutant PIP5Pase revealed that the mutations greatly reduced its dephosphorylation activity (143). In addition, silent VSGs were expressed in cells expressing the mutant PIP5Pase suggesting that the catalytic activity of PIP5Pase is crucial for the silencing of VSGs in BFs (143).

Based on these findings the authors suggest that RAP1 binds to PI(3,4,5)P3, a substrate of PIP5Pase (143). In the presence of PIP5Pase, RAP1 no longer binds PI(3,4,5)P3 (143). This may result in RAP1 binding the ES chromatin at the telomeric repeats and 70 bp regions that flank the VSG (143). This interaction of RAP1 may lead to the silencing of the transcription at all expression sites except one (143).

## 8.6. Concluding remarks



**Fig 4. ES associated proteins and their regulation of VSG gene expression.** The diagram depicts different proteins that regulate VSG gene expression by associating with RNA Pol I, the telomeric repeats or by regulating the epigenetics of the ES chromatin. RAP1 and PIP5Pase have been highlighted since the thesis focuses on their role in the regulation of VSG gene expression.

*T. brucei* is a protozoan parasite that infects mammals and thrives in the host bloodstream where it depends on antigenic variation to evade the host immune response and establish infection. Antigenic variation is tightly regulated wherein only one VSG gene is expressed from a repertoire of about 2500 VSGs. The transcription of VSGs is a multifactorial phenomenon depending on the regulation of transcription at various levels. Transcription is dependent on factors such as the regulation of Pol I transcription, structure of the chromatin as well as the telomeric position effect. All these facets of VSG transcription are further dependent on proteins and protein complexes. One such system implicated in the regulation of the VSG gene expression is the IP pathway, a complex pathway well known to participate in various cellular processes. Perturbing members of this pathway led to the de-repression of VSGs. PIP5Pase was implicated in regulating monoallelic VSG gene expression in BFs. It regulates VSG gene expression by regulating the interaction of RAP1 with the substrate PI(3,4,5)P<sub>3</sub>. The authors of this study suggest a model whereby the presence of PIP5Pase leads to the dephosphorylation of PI(3,4,5)P<sub>3</sub> to PI(4,5)P<sub>2</sub>, depriving RAP1 of its substrate. In turn, RAP1 binds to the ES chromatin whereby it mediates the silencing of VSG genes found at the associated ES. There are various questions regarding the regulation of VSG expression by PIP5Pase that are yet to be answered. This includes clarifying whether PIP5Pase plays a role in the developmental regulation of VSG expression. Moreover, the mechanism by

which PIP5Pase silences all VSGs except one. Further, there is also a need to understand how the catalytic mutation of PIP5Pase affects the association of RAP1 with the ES chromatin. This can be achieved by determining the importance of PIP5Pase in PFs, the insect stage of the life cycle to understand the role of VSGs in developmental regulation of VSG gene expression. Moreover, characterising the interactions involved is crucial. Specifically, this would involve the examining the interactions between the proteins: PIP5Pase and RAP1, determining the role of PIP5Pase in the developmental regulation of VSG transcription and also understanding the role of PIPs play in regulating the interactions of RAP1 with the ES chromatin DNA.

## 9. Project overview

### 9.1. Rationale and hypothesis

*T. brucei* BFs periodically switch the surface coat protein to evade the host antibody response. BFs express one among 2500 VSG genes while VSGs are completely silenced in PFs. Various factors have been implicated in the regulation the transcription of VSGs. Recently, the IP pathway has been implicated in regulating VSG gene expression. The IP pathway has various roles at the cellular level where it actively participates in trafficking, metabolism, and development of *T. brucei*. Interestingly PIP5Pase was identified as a part of a 0.9 MDa protein complex along with RAP1, a telomere associated protein implicated in regulating VSG expression. Both PIP5Pase and RAP1 bind to the ES chromatin wherein RAP1 is implicated in the silencing of the telomere proximal genes by TPE. Moreover, RAP1 binds to PI(3,4,5)P3, a substrate of PIP5Pase (143). ChIP data also shows that the ablation of PIP5Pase catalytic activity affected the binding of RAP1 to the ES DNA (143). In other words, the presence or the absence of PI(3,4,5)P3 affected the binding of RAP1 to the ES DNA (143).

There is a need to further understand how the interactions of PIP5Pase and RAP1 with PI(3,4,5)P3 and the ES DNA affect VSG gene expression. Moreover, it would be crucial to determine if PIP5Pase is essential in PFs and also how it affects the silent state of VSGs in the PFs. This would clarify the role of PIP5Pase in the developmental regulation of VSGs in *T. brucei*. Here, we hypothesize that PIP5Pase plays a role in the developmental regulation of VSG genes in *T. brucei*. Moreover, we hypothesize that the control of VSG silencing involves PIP5Pase regulation of RAP1 association with telomeric ES. RAP1 binds to PI(3,4,5)P3 which is PIP5Pase substrate, and this binding controls RAP1 association with the ES chromatin. This affects the process of VSG silencing in BFs and PFs.

### 9.2. Specific aims

The hypothesis will be tested through the three specific aims listed below.

1. Determine the role of PIP5Pase in VSG gene silencing in procyclic forms i.e. the insect stage of the parasite. Knockdown of PIP5Pase followed growth curve analysis and gene expression analysis

using qPCR will be used to determine the essentiality of PIP5Pase and its role in VSG gene silencing in procyclic forms.

2. Determine the domains of RAP1 that interact with PI(3,4,5)P3 and with the ES chromatin DNA. Recombinant RAP1 and its truncations will be purified. The purified recombinant proteins will be used in binding assays with PIP(3,4,5)P3 to determine the domain that binds with PI(3,4,5)P3. The recombinant proteins will also be used in gel shift assays and MST with synthetic ES chromatin DNA sequences to determine the domains that binds with ES chromatin DNA.
3. Determine the effect of PI(3,4,5)P3 on the interaction of RAP1 with the ES DNA. Recombinant proteins will be used in gel shift assays with synthetic ES chromatin DNA in the presence of PI(3,4,5)P3 to determine the effect of PI(3,4,5)P3 on the interaction of RAP1 with the ES DNA.

## 10. Methods

### 10.1. *T. brucei* cell culture and growth curve analysis.

*T. brucei* 29.13 (PFs) conditional nulls for PIP5Pase (*T. brucei* PIP5Pase CNs) were used for growth curve analysis. They were previously generated by Dr. Igor Cestari. These cells express a tetracycline regulatable V5 tagged PIP5Pase allele from the silent rDNA spacer. The endogenous alleles of PIP5Pase were sequentially replaced by a puromycin and blasticydine drug resistance cassette as described (136). The parasites were maintained in SDM-79 supplemented with 10% FBS, G418 (7.5  $\mu\text{g}/\text{mL}$ ), phleomycin (12.5  $\mu\text{g}/\text{mL}$ ), hygromycin (50  $\mu\text{g}/\text{mL}$ ) and tetracycline (0.5  $\mu\text{g}/\text{mL}$ ). They were seeded at the concentration of  $2 \times 10^6$  cells/mL and incubated at 25°C. Cells were diluted to  $2 \times 10^6$  cells/mL every 48 hours.

For growth curve analysis,  $2.5 \times 10^6$  cells were grown with or without tetracycline for a period of 12 days in SDM-79 supplemented with G418, phleomycin and hygromycin. Concentration of cells were determined daily using a cell counter (Beckman Coulter). The cells were diluted every two days to maintain the concentration between  $2 \times 10^6$  cells/mL and  $3 \times 10^7$  cells/mL. A cumulative curve was plotted with the time (hours) on the X-axis and cell concentration (cumulative cells/mL (Log10)) on the Y-axis using MS Excel. The experiments were performed using 3 biological replicates and the values represent 3 biological replicates +/- standard deviation of the mean (SDm).

### 10.2. Immunofluorescence analysis.

$2 \times 10^6$  mid log phase PIP5Pase CNs were collected for immunofluorescence analysis. They were fixed in 2% formaldehyde prepared in PBS supplemented with 6 mM glucose for 10 minutes. Cells were then loaded onto poly L- lysine (Thermo Fischer Scientific) treated 2 mm glass cover slips (Fischer Scientific). The cells were incubated for 10 mins, allowing them to adhere to the cover slip. The slips were washed three times with PBS-glucose. The cells were permeabilized with 0.2% NP-40 (Fischer Scientific) in PBS, for 10 mins. They were washed five times with PBS-glucose. They were then blocked for 1 hour with 10% milk, prepared in PBS-glucose. The cells were then

incubated in mouse mAb  $\alpha$ -V5 1:500 (Thermo Life Technologies) for 2 hours at RT. Cells were then washed five times with PBS-glucose. They were incubated in goat  $\alpha$ -mouse IgG (H+L) Alexa Fluor 488 (1:1000) (Invitrogen) for 2 hours at RT. Cells were then washed five times with PBS-glucose. The cells were then stained with 1  $\mu$ g/mL 4',6-diamidino-2-107 phenylindole (DAPI, Sigma-Aldirch) for 15 mins. The cells were washed four times in PBS-glucose and twice in water and then mounted onto microscopy glass slides using mounting medium (Southern Biotech). Epifluorescence microscopy and image acquisition was carried out on the Nikon E800 Upright Fluorescence microscope. The images were acquired using the 60X objective, with the excitation wavelength set at 499nm and emission wavelength set at 520nm.

### 10.3. Western blot analysis.

*T. brucei* PIP5Pase CNs were grown with or without 0.5  $\mu$ g/mL tetracycline. Cells were collected on 3 consecutive days, specifically: days one to three and spun down at  $4000 \times g$  for five mins in a benchtop centrifuge. Cells were washed twice in PBS-glucose and spun down at  $4000 \times g$ . The cell pellet was suspended in 1% Triton X-100 prepared in PBS-glucose supplemented with 1X EDTA free protease inhibitor cocktail tablet (Paclease extract, pronase, thermolysin, chymotrypsin and papain) (50X) (Roche) (one tablet per 50 mL). The suspensions were incubated on ice for 10 minutes. They were spun down at  $14000 \times g$  for 10 mins. The supernatant/lysate was collected, and the pellet was discarded. The lysates were then boiled in 1X Lamelli buffer [312 mM Tris-HCl (pH6.8), 10% (w/v) SDS, 45% (v/v) (Fischer scientific), Glycerol 1% (w/v) (Fischer Scientific), Bromophenol Blue 5% (v/v) (Fischer Scientific),  $\beta$ -mercaptoethanol 1% (v/v) (Fischer Scientific)] at  $95^{\circ}\text{C}$  for 5 mins. The denatured proteins in the lysate were resolved on 10% SDS-PAGE at 130 V in 1X SDS-PAGE running buffer [3.0 g of Tris base (Fischer Scientific), 14.4 g of glycine (Fischer Scientific), and 1.0 g of SDS (Fischer Scientific) in 1000 mL of water (pH 8.3)]. The resolved proteins were transferred onto nitrocellulose membrane for 1 hour at 85V in 1X transfer buffer [12.1g of Tris base (Fischer Scientific), 13.4g of glycine (Fischer Scientific) dissolved in 1000 mL of water (pH 7.6)] at  $4^{\circ}\text{C}$ . The membrane was blocked for 1hr at RT with 6% (w/v) non-fat dry milk prepared in PBS. Membranes were probed first with mouse  $\alpha$ -V5 1:2500 (Bioshop) in 6% milk prepared in PBS for 1hr at RT, followed by five washes of five minutes each with PBS supplemented with 0.05% Tween 20 (Pierce). Afterwards, it was probed with  $\alpha$ -mouse

IgG HRP 1:5000 (Biorad) for 1 hour at RT, followed by five washes of 5 minutes each with PBS supplemented with 0.05% Tween-20. Membranes were developed using Supersignal West Pico Chemiluminescent Substrate (Thermoscientific) and images acquired on a ChemiDoc MP imaging system (Biorad).

Loading control: After development and acquisition the membrane was stripped in low pH stripping solution (25mM glycine-HCl, pH 2). They were then blocked again before probing with mouse  $\alpha$ -HSP 70 1:1000 (gift from Ken Stuart laboratory, Seattle Children's Institute) for 1 hour at RT, followed by five washes of five minutes each - with PBS supplemented with 0.05% tween. Afterwards, it was probed with  $\alpha$ -mouse IgG HRP 1:5000 dissolved in 6% milk prepared in PBS and incubated at RT, followed by five washes of five minutes each with PBS supplemented with 0.05% tween. The membrane was developed as indicated above.

#### **10.4. RNA extraction and qPCR analysis.**

##### **10.4.1. RNA isolation**

$10^7$  *T. brucei* PIP5Pase CN cells grown in the presence or absence of tetracycline were collected on day 3. Day of sample collection was decided based on Western blot analysis that indicated complete loss of protein on day 3 post knockdown. Cells were spun down at  $4000 \times g$  for 5 minutes and the pellet was suspended in 1 mL TRIzol (Invitrogen). TRIzol suspensions were stored in  $-80^\circ\text{C}$  freezers until further use. The suspensions were treated with chloroform (200  $\mu\text{L}$  per mL of TRIzol used), mixed vigorously and spun down at  $14000 \times g$  for 10 mins at  $4^\circ\text{C}$ . The suspension separated into an upper aqueous phase and a lower organic phase. The aqueous phase was retained and treated with equal volumes of isopropanol (Fischer Scientific), gently mixed by tube inversion, incubated at room temperature for 10 minutes and spun down at  $14000 \times g$  for 15 minutes. The RNA pellet was retained and washed with 75% ethanol at  $7500 \times g$  for 5 minutes. The washed pellet was air dried for 5-10 minutes. The RNA was dissolved in 20  $\mu\text{L}$  nuclease free water. The extracted RNA was quantified on a nanodrop device (Thermo Fisher spectrophotometer ND-100) and visually analysed on a 1% agarose gel. It was stored at  $-80^\circ\text{C}$  freezers until further use.

#### 10.4.2. DNase treatment and cDNA synthesis

RNA extracted was treated with Ambion (AM2222) DNase treatment kit as per manufacturer's instructions. The RNA was again quantified using a nanodrop device (Thermo Fisher spectrophotometer ND-100). cDNA was then reverse transcribed from 1  $\mu$ G total RNA using Invitrogen Superscript III reverse transcriptase, using oligo dT primers as per manufacturer's instructions. The cDNA was stored in  $-80^{\circ}\text{C}$  until further use.

#### 10.4.3. Quantitative real-time PCR analysis

cDNA samples were diluted 20X in nuclease free water for qPCR analysis. qPCR was performed using Bright Green Express 2X qPCR Master Mix-ROX (Applied Biological Materials Inc.). Reactions were set up in 20  $\mu$ L reactions: 10  $\mu$ L master mix, 5  $\mu$ L cDNA and 5  $\mu$ L 10  $\mu$ M primers (forward + reverse primers) (Table 3). Primers for ES-VSGs and procyclins were obtained from Cestari and Stuart, 2015 (136). 18S, tubulin and GAPDH were also analysed as endogenous controls. Targets were analysed using a StepOnePlus RT PCR system (Applied Biosystems) using conditions:  $95^{\circ}\text{C}$  for 10 minutes (one cycle),  $95^{\circ}\text{C}$  for 15 seconds (45 cycles),  $60^{\circ}\text{C}$  for 60 seconds (45 cycles). Melting curve analysis was performed to ensure the amplification of single amplicons in each target sample. Relative quantification of RNA was performed using the  $\Delta\Delta\text{Ct}$  method (156). A bar graph representing the fold change in gene expression for each target was plotted with gene targets on the X-axis and the relative quantification on the Y-axis. The values represent 4 biological replicates +/- SDm.

Table 2. lists all the qPCR primers that were used to analyse BF VSGs, mVSGs, procyclins, 18s, tubulin and GAPDH. The second section contains DNA sequences of the biotinylated forward and non-biotinylated reverse complementary strands of the DNA used for the EMSA. The third section contains DNA sequences of the Cy5-labelled forward and unlabelled reverse complementary strands of the DNA used for the MST.

Table 2. List of qPCR primers.

qPCR Primer Sequences				
Description of Gene product or Oligonucleotide sequence	GENE ID (TritypDB.org)	Reference	Forward/Reverse	Primer sequence
VSG2	Tb427.BES40.22	Cestari and Stuart, 2015 (136)	Forward	GTCCTAGCCCAAGTTCTTC
			Reverse	GCTGTTGCAGTAGCTGTTAC
VSG3	Tb427.BES153.14/ Tb427.BES65.13	Cestari and Stuart, 2015 (136)	Forward	CAGTCTTGTGCGCACTAGCT
			Reverse	ATGCTGCTGCTGCTGTTACC
VSG6	Tb427.BES4.11/ Tb427.BES15.12	Cestari and Stuart, 2015 (136)	Forward	CAAGTTTGAGACGTGGGAGC
			Reverse	AGATCGGCTGCTATTGCTGC
VSG8	Tb427.BES29.9	Cestari and Stuart, 2015 (136)	Forward	CTACCAATTGCAGGGCTCAC
			Reverse	TTACGGCCAGCAGCTTGGAT
VSG9	Tb427.BES129.14	Cestari and Stuart, 2015 (136)	Forward	AGCTTATCTAGCAGACGCCG
			Reverse	TCCTCCCATTTCCTCCGATC
VSG11	Tb427.BES122.11/ Tb427.BES126.15	Cestari and Stuart, 2015 (136)	Forward	CATAGGAACTGGCGACAACG
			Reverse	GAGCCGCCGAATTGTGTTTC
VSG13	Tb427.BES51.12/ Tb427.BES59.12	Cestari and Stuart, 2015 (136)	Forward	ATAACGCATGGCCATCTTGAC
			Reverse	GTCGTTGCTGTGGATTGCTC
VSG14	Tb427.BES64.2	Cestari and Stuart, 2015 (136)	Forward	TGCTTATATCGAGGCGACGG
			Reverse	TTGTGGTTGGTGCATCTCG
VSG15	Tb427.BES134.6	Cestari and Stuart, 2015 (136)	Forward	AGCACTTTTTCCTGGACCTC
			Reverse	AGTCGGCGTAGGTTGCATTG
VSG16	Tb427.BES122.11	Cestari and Stuart, 2015 (136)	Forward	CGGATCAAAGATAGCAGGGC
			Reverse	GCCAGTGAACATACCTGTCTG
VSG17	Tb427.BES56.13	Cestari and Stuart, 2015 (136)	Forward	GGCCAACAAGCAAGGCTCAA/
			Reverse	TTTGTCAGGTCCCGAAGCAC
VSG18	Tb427.BES98.12	Cestari and Stuart, 2015 (136)	Forward	ACTGCTAAAGAAGGTCTGGG
			Reverse	TCGCCCTTTGAGATAGGTGG
VSG19	Tb427.BES10.10	Cestari and Stuart, 2015 (136)	Forward	ACAGGATGGCGGAAGTCTAC
			Reverse	GCGAGCAAAGCACATTTGGC
VSG21			Forward	ACATGCGTCACATAACGCGG

	Tb427.BES28.15/ Tb427.BES5.13	Cestari and Stuart, 2015 (136)	Reverse	TTGTTTGCTGGCCTGCTAGC
Metacyclic VSG 397	VSG 397	Kolev et al., 2012 (157)	Forward	TGAAGCTGTGAAAGGGACAG
			Reverse	GAGGGCGAATTGTTTGTTTAGG
Metacyclic VSG 531	VSG 531	Kolev et al., 2013	Forward	GACGAAAGCCTGGGTAACATAAA
			Reverse	CCGCAGCTCGTTGATAGTATTG
Metacyclic VSG 639	VSG 639	Kolev et al., 2014(158)	Forward	CCGACGATGAACACAGTTGA
			Reverse	TCTATGCCGTTTCGCCTTTAC
Metacyclic VSG 653	VSG 653	Ramey butler et al., 2015(159)	Forward	GGGCTGTTTCGCGACTAATA
			Reverse	CGTGGTGAAGTCTCCTGTTT
Metacyclic VSG 1954	VSG 1954	Savage et al., 2016(160)	Forward	GCAGAGGCCTTAGCACTAAAT
			Reverse	GGAGTTGACTTTCCTCCATCAG
Procyclin (EP1)	Tb427.10.10260	Urwyler, S. et al. 2005(161)	Forward	CTGTACTATATTGACTTCAATTACA CCAAAAAG
			Reverse	GGTCTCAGGCGATGGTTAATAAGG CATTAG
Procyclin (EP2)	Tb427.10.10250	Urwyler, S. et al. 2006(162)	Forward	CTGTACTATATTGACTTCAATTACA CCAAAAAG
			Reverse	CGTATATGCAAGTGCCTGTCGCC
Procyclin (EP3)	Tb427.06.520	Urwyler, S. et al. 2007(162)	Forward	CTGTACTATATTGACTTCAATTACA CCAAAAAG
			Reverse	TTATGCGTGTCAAGTGCCTGTCACA AAGGA
GPEET	Tb927.06.510	Urwyler, S. et al. 2008(163)	Forward	CTGTACTATATTGACTTCAATTACA CCAAAAAG
			Reverse	AGGTTTCGCGCTGAACAGAAG
$\beta$ -tubulin	Tb927.1.2330	Cestari and Stuart, 2015 (136)	Forward	TTCCGCACCCTGAAACTGA
			Reverse	TGACGCCGGACACAACAG
18S	Tb927.10.5330	Cestari and Stuart, 2015 (136)	Forward	CGGAATGGCACCACAAGAC
			Reverse	TGGTAAAGTTCCCCGTGTTGA
GAPDH	Tb427.06.4280	This was designed by Rishi Rajesh and Dr. Cestari	Forward	
			Reverse	

## 10.5. Recombinant protein expression and purification.

Recombinant proteins (rRAP1 and its recombinant domains- rBRCT, rMyb, rMybl) were purified using plasmid constructs that were previously designed and generated by Dr. Cestari. The plasmid constructs consisted of a pET29-a expression vector into which the gene- RAP1 (Tritydb gene ID- Tb927.11.370) or its domains [BRCT (nucleotides 1-899), Myb (nucleotides 900-1680) and Mybl (nucleotides 1681-2569)] were cloned. These plasmids were transformed into NiCO(DE3) competent cells (New England Biolabs). Prior to induction, bacterial cultures were grown to a density of 0.6 OD in a shaker incubator (New Brunswick) at 37°C. The cultures were induced with 0.5 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) (Sigma Aldrich) overnight at 18°C for RAP1 or 37°C for 4 hours for the recombinant domains. Cultures were spun down at  $10,000 \times g$  for 15 minutes at 4°C. The supernatant was discarded, and the pellets were resuspended in lysis buffer (PBS, 5 mM DTT, 0.2 mg/mL lysozyme, 0.05% NP-40 and 10% glycerol) supplemented with 1mM PMSF or protease inhibitor cocktail (10 mL lysis buffer for every 1 L culture). For the purification of rBRCT, the lysis buffer was supplemented with 7 M urea (Fischer Scientific). The suspensions were incubated on ice for 10 minutes. The cells were lysed by sonication for 30 seconds (pulse on for 5 seconds, pulse off for 5 seconds) using a sonicator (Fischer Scientific). The lysed cells (20 mL) were spun down at 13,000 rpm for 5 minutes. The lysate was incubated with nickel beads (Biorad) (500  $\mu$ L for every 20mL lysate volume) overnight at 4°C, on a rotating tube holder. Beads were washed with wash buffer: two washes with 10 mL protein purification wash buffer [50 mM sodium phosphate dibasic (Fischer Scientific), 300 mM NaCl (Fischer Scientific), 20 mM Imidazole (Fischer Scientific), pH 8] and five washes with 1 mL protein purification wash buffer. Beads were spun at  $1000 \times g$  after each wash step and the flow through was discarded. The protein was eluted from the beads using protein elution buffer (50 mM sodium phosphate dibasic, 300 mM NaCl, 300 mM Imidazole, pH 8). The beads were incubated in 500  $\mu$ L elution buffer for five minutes and spun down at  $1000 \times g$  for 1 minute. The eluate was collected into a new eppendorf tube. Eight such elutions were performed for each purification. To visualize the eluted protein, 20  $\mu$ L from each fraction was run on a 10% SDS-PAGE at 130 V in SDS-PAGE running buffer, followed by Coomassie staining. The fractions containing proteins were pooled. The purified protein was dialysed using dialysis buffer (25 mM HEPES (Invitrogen), 150 mM NaCl, 10% glycerol, pH 7-7.5) overnight at 4°C. The dialysed proteins were once again resolved

on a 10% SDS-PAGE and visualised by Coomassie staining. The dialysed protein was mixed with 10% glycerol and stored at  $-80^{\circ}\text{C}$  until further use. The proteins were then quantified using Pierce BCA Protein Assay Kit (Sigma Aldrich).

### 10.6. Phosphatidylinositol binding assays.

1  $\mu\text{g}$  of HIS-tagged recombinant protein (rRAP1, rBRCT, rMyb, or rMybl) was diluted in 400  $\mu\text{L}$  of binding buffer (25 mM HEPES pH 7.5, 150mM NaCl, 0.2% Nonidet P-40 and 0.1% Tween 20) for binding assays as described (165). 5% of volume was collected for Western blot analysis (input). 1  $\mu\text{M}$  biotin conjugated Dic8-PI(3,4,5)P<sub>3</sub>, Dic8-PI(4,5)P<sub>2</sub>, InsP(1,4,5)P<sub>3</sub> or InsP(1,3,4,5)P<sub>4</sub> (all from Echelon biosciences) was incubated with the diluted protein in a rotating eppendorf tube holder at RT for 1 hour. After 1 hour, magnetic streptavidin beads (blocked overnight with 3% bovine serum albumin (BSA) (Biobasic) at  $4^{\circ}\text{C}$ ) were added and incubated at  $4^{\circ}\text{C}$  for 1 hour in a rotating eppendorf tube holder. The tubes were placed in a magnetic stand. The flow through was separated into an eppendorf tube for Western blot analysis. The beads were washed five times with wash buffer (25 mM HEPES pH 7.5, 300 mM NaCl, 0.2% Nonidet P-40 and 0.1% Tween 20). The beads were separated from the suspension using a magnetic stand after each wash and the flow through was discarded. The beads were resuspended in 50  $\mu\text{L}$  elution buffer (2x Laemmli buffer supplemented with 710 mM 2-mercaptoethanol) and boiled at  $95^{\circ}\text{C}$  for 5 minutes. The beads were separated using a magnetic stand and the eluate was collected into an eppendorf tube. For competition assays, recombinant proteins, biotin-labelled PI(3,4,5)P<sub>3</sub> were incubated with or without 5-50  $\mu\text{M}$  unlabelled Dic8-PI(3,4,5)P<sub>3</sub> (Echelon biosciences) as a competitor and the same protocol was followed. The input, flow through and eluate were resolved on 10% SDS-PAGE at 130 V in SDS-PAGE running buffer. The resolved proteins were transferred onto a nitrocellulose membrane at 85 V for 1 hour at  $4^{\circ}\text{C}$  in transfer buffer. The membrane was blocked for 1 hour at RT with 6% milk prepared in PBS. The membrane was probed first with mouse  $\alpha$ -HIS 1:5000 (Genescript) diluted in 6% milk prepared in PBS for 1 hour at RT, followed by five 5-minute washes with PBS supplemented with 0.05% tween. Afterwards, it was probed with  $\alpha$ -mouse IgG HRP 1:5000 dissolved in 6% milk prepared in PBS for 1 hour at RT, followed by five washes of 5 minutes each in PBS supplemented with 0.05% tween. The membrane was

developed using Supersignal West Pico Chemiluminescent Substrate (Thermoscientific) and the image was acquired on a ChemiDoc MP imaging system (Biorad).

### 10.7. Electrophoretic mobility shift (EMSA) or gel shift assays.

Biotinylated DNA sequences (Table 3.) were denatured in a thermocycler using the program: 95°C for 10 minutes (one cycle), 94°C for 50 seconds (-1°C per cycle for 72 cycles) and 22°C forever. Reactions were set up with 100 nM DNA, 2 mg/mL yeast tRNA, 1 µg recombinant protein in DNA binding buffer (20 mM HEPES, 40 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM CaCl<sub>2</sub> and 0.2% Nonidet P-40). The reactions were incubated on a thermomixer (Eppendorf) at 37°C for 1 hour. The binding was crosslinked using the “optimal crosslink” setting on a benchtop UV-crosslinker (Spectrolinker XL-1 500). 10 µL of the total reaction was resolved on a 10% native gel at 100 V in 0.5 X TAE. The reactions were transferred onto a nylon membrane at 100 V for 30 minutes in 0.5 X TAE. The membrane was blocked for 1 hour at RT with nucleic acid detection blocking buffer (Life Technologies). It was probed with streptavidin 1:5000 (Genescript) diluted in nucleic acid detection blocking buffer (Life Technologies) for 1 hour at RT, followed by five minute washes with PBS supplemented with 0.05% tween. The membrane was developed using Supersignal West Pico Chemiluminescent Substrate (Thermoscientific) and the image was acquired on a ChemiDoc MP imaging system (Biorad).

**Table 3. List of Oligonucleotide sequences for EMSA.**

<b>Oligonucleotide Sequences for EMSA</b>				
Telomeric repeats	NA	This was designed by Rishi Rajesh and Dr. Cestari	Forward strand (G-rich strand)	<b>BIOTIN-</b> TTAGGGTTAGGGTTAGGGTTAGGGT TAGGGTTAGGGTTAGGGTTAGGTTA GGGTTAGGG
			Reverse complementary strand	CCCTAACCCCTAACCCCTAACCCCTAAC CCTAACCCCTAACCCCTAACCCCTACCC TAACCCTAA
70 bp repeats	NA	Miner et al. 2016 (164)	Forward strand	<b>BIOTIN-</b> ATGATAATGATAATAATAATAGGA GAGTGTGTGAGAGTGTATATACG AATATTATAATAAGAGCAGTAATA ATA

			Reverse complementary strand	TATTATTACTGCTCTTATTATAATAT TCGTATATACACTCTCACAACTC TCCTATTATTATTATCATTATCAT
Scrambled sequence	NA	This was derived from sequence of telomeric repeats	Forward strand	<b>BIOTIN-</b> AGGTGGGTTGTGGTGTGAGTTGTG TATAGATGGGGGAGGTGGGAGGTT AATGGGATTGT
			Reverse complementary strand	ACAATCCCATTAACCTCCCACCTCC CCCATCTATACAACTCAACACCA CAACCCACT

## 10.8. Microscale thermophoresis (MST).

### 10.8.1. DNA acquisition and preparation.

Nucleic acid substrates (Table 4.) were designed with a Cy5 fluorophore conjugated to the 5' hydroxyl group and were prepared by Integrated DNA Technologies (IDT) for analysis with Monolith NT.115 Red (NanoTemper®). Non-fluorescent antisense DNA was prepared in order to prepare double stranded DNA substrates for the analysis. The 2 strands were annealed by mixing equimolar concentrations in a buffer containing 10 mM Tris-HCl pH 7.4, 2 mM MgCl<sub>2</sub>, 50 mM NaCl. followed by incubation in the thermocycler using the following program: using the program: 95°C for 10 minutes (one cycle), 94°C for 50 seconds (-1°C per cycle for 72 cycles) and 22°C forever. All DNA samples were stored at -20°C. Before each assay the DNA was re-folded using the program: 95°C for 10 minutes (one cycle), 94°C for 50 seconds (-1°C per cycle for 72 cycles) and 22°C forever.

**Table 4. List of Oligonucleotide sequences for MST.**

Oligonucleotide Sequences for MST				
Telomeric repeats	NA	This was designed by Rishi Rajesh and Dr. Cestari	Forward strand (G-rich strand)	<b>Cy5-</b> TTAGGGTTAGGGTTAGGGTTAGGGT TAGGGTTAGGGTTAGGGTTAGGT GGGTTAGGG
			Reverse complementary strand	CCCTAACCCCTAACCCCTAACCCCTAAC CCTAACCCCTAACCCCTAACCCCTACCC TAACCCTAA

70 bp repeats	NA	Miner et al. 2016 (164)	Forward strand	<b>Cy5-</b> ATGATAATGATAATAATAATAGGA GAGTGTGTTGTGAGAGTGTATATACG AATATTATAATAAGAGCAGTAATA ATA
			Reverse complementary strand	TATTATTACTGCTCTTATTATAATAT TCGTATATACACTCTCACAACTC TCCTATTATTATTATCATTATCAT
Scrambled sequence	NA	This was derived from sequence of telomeric repeats	Forward strand	<b>Cy5-</b> AGGTGGGTTGTGGTGTGAGTTGTG TATAGATGGGGGAGGTGGGAGGTT AATGGGATTGT
			Reverse complementary strand	ACAATCCCATTAACCTCCACCTCC CCCATCTATACAACTCAACACCA CAACCCACCT

### 10.8.2. Binding checks to confirm binding.

Following annealing, binding checks were performed. 5 nM DNA was incubated with 1  $\mu$ M rRAP1 in a buffer containing 50 mM HEPES pH 7.4, 5 mM  $MgCl_2$ , 100mM NaCl and 0.25% (V/V) NP-40 in a total reaction volume of 50  $\mu$ L. The reaction was incubated at 37°C for 1 hour. After incubation, the samples (5-10  $\mu$ L) were loaded, by capillary action on to capillary tubes (4 capillaries containing the DNA with protein and 4 capillaries containing control reactions that were set up with no protein) and analysed using the Monolith NT. 115 Red (NanoTemper®). Reactions were also set up with the scrambled sequence derived from the telomeric repeats. The fluorescence change between two time points was measured for each sample to confirm the binding.

### 10.8.3. Equilibrium dissociation constant ( $k_d$ ) measurements.

16 reactions containing 5 nm annealed DNA was incubated with 16, 1:2 serial dilutions of rRAP1 (3.125  $\mu$ M was the highest concentration of rRAP1 used). The total reaction volume was 50  $\mu$ L. The reaction was incubated at 37°C for 1 hour. After incubation, the samples (approximately 5-10  $\mu$ L) were loaded, by capillary action on to capillary tubes. Reactions were incubated at 37°C for 1 hour followed by which they were loaded into capillary tubes and analysed using the Monolith NT. 115 Red (NanoTemper®). The fluorescent change between two time points was measured for each reaction and a binding affinity curve was generated by plotting the fraction bound (Y-axis)

against each dilution point (X-axis). This was performed using the Graphpad Prism software version 9.

### **10.9. Statistical analysis.**

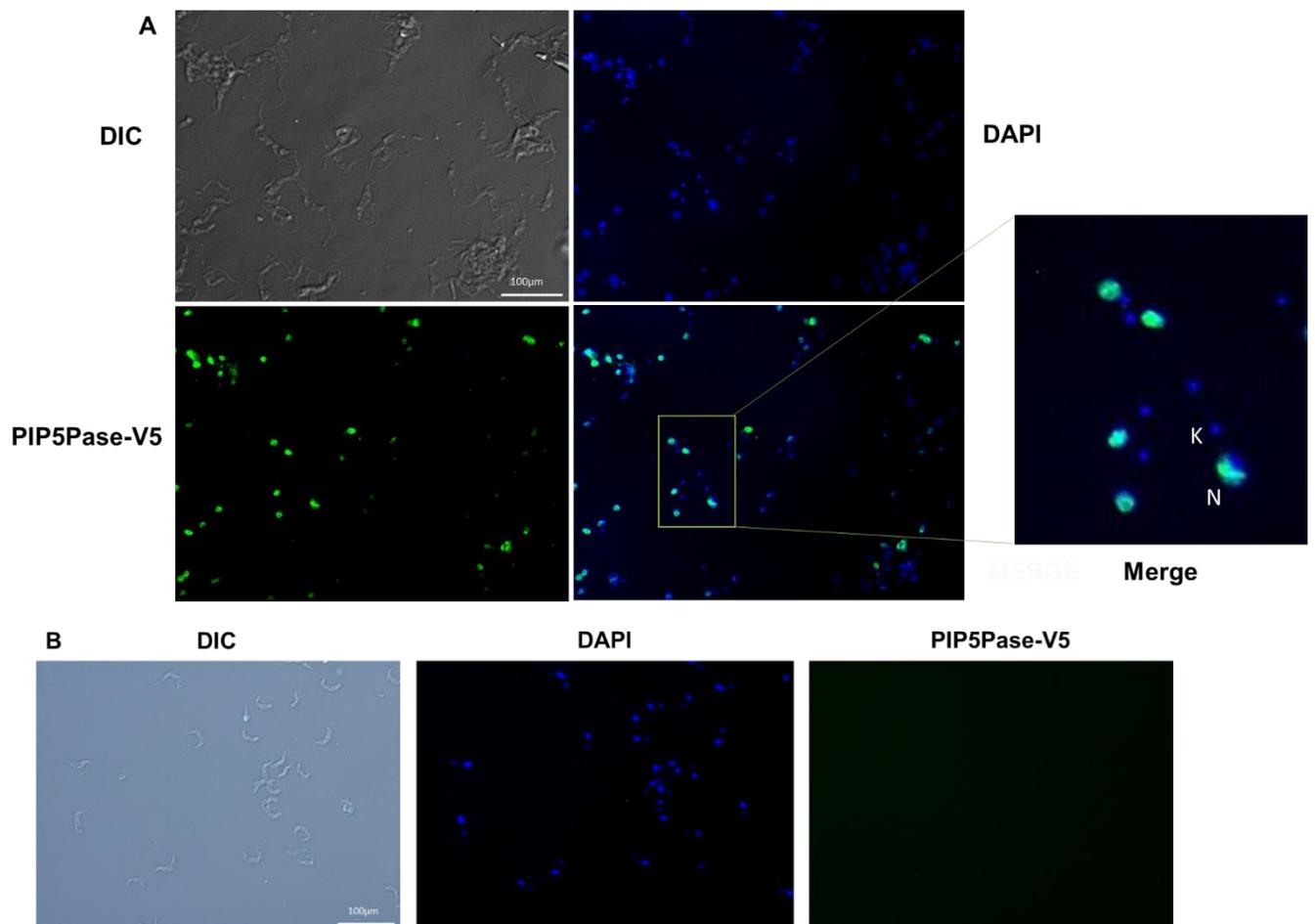
All data are representatives of multiple replicates. Data from growth curve analysis and gene expression analysis by qPCR are representatives of at least 3 biological replicates +/- SDm. Data from Western blots, phosphatidylinositol binding assays and EMSA are representatives of at least 2 replicates.

## 11. Results:

### 11.1. PIP5Pase is important for the developmental regulation of VSGs

#### 11.1.1. PIP5Pase localizes in the nucleus in PFs

To determine the localization of PIP5Pase in PFs, immunofluorescence using PIP5Pase conditional null cell line was performed. This cell line expresses a V5-tagged PIP5Pase from the silent rDNA spacer in presence of tetracycline. Immunofluorescence assays in BFs showed that PIP5Pase localizes in the nucleus (136).



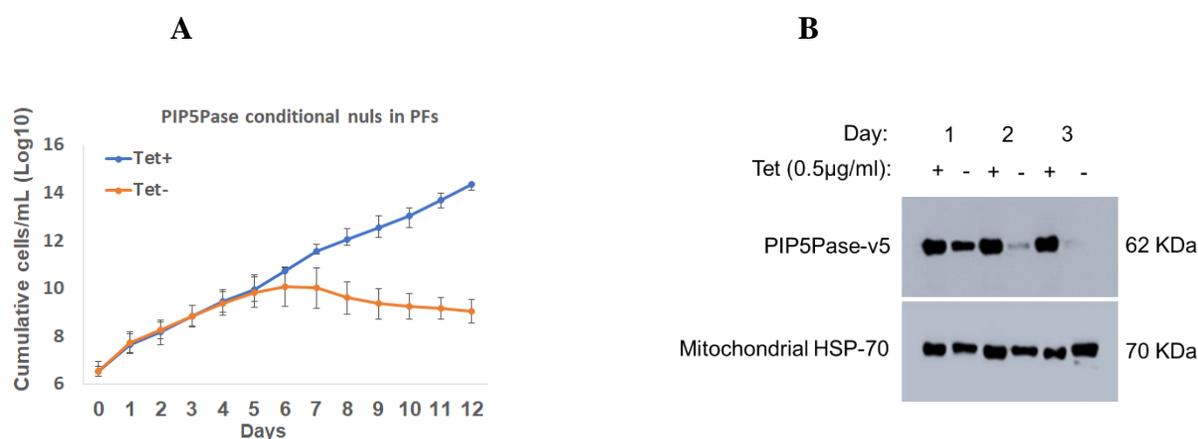
**Fig. 5. PIP5Pase localises in the nucleus in PFs. A. Immunofluorescence analysis of V5- tagged PIP5Pase in *T. brucei* PFs.** Differential interference contrast shows the cell population analysed by immunofluorescence microscopy. DAPI was used to stain the nucleus. Mouse Anti V5 antibodies followed by anti-mouse IgG alexa 488 antibodies were

used to detect V5-tagged PIP5Pase. Merge shows the localization of PIP5Pase at the nucleus. K denotes the kinetoplastid and N denotes the nucleus. The red square indicates the magnified image. Bar 100  $\mu\text{m}$  **B. Control: Immunofluorescence analysis of V5- tagged PIP5Pase in the absence of tetracycline in *T. brucei* PFs.** DIC shows the cell population analysed by immunofluorescence microscopy. DAPI was used to stain the nucleus. Mouse Anti V5 antibodies followed by mouse IgG alexa 488 antibodies were used to detect V5-tagged PIP5Pase (undetected here). Bar 100  $\mu\text{m}$ .

The localization data in PFs show that PIP5Pase-V5 coincides with DAPI staining of the nuclear DNA (Fig. 5A), and thus indicates that PIP5Pase in PFs also localizes in the nucleus. No significant signal for PIP5Pase-V5 was detected outside of the nucleus or kDNA (Fig. 5A). Immunofluorescence analysis performed in control showed no signal for V5 (Fig. 5B). Hence, PIP5Pase localizes in the nucleus of PFs as seen in BF.

### 11.1.2. PIP5Pase is essential for the growth of PFs

To determine if PIP5Pase is essential for PFs growth, PIP5Pase was knocked down in PFs and growth curve analysis was performed in which cells were counted daily. A growth defect was detected starting at day five after which cells with the PIP5Pase-V5 KD started to die whereas the cells expressing PIP5Pase continued to grow normally (Fig. 6A).



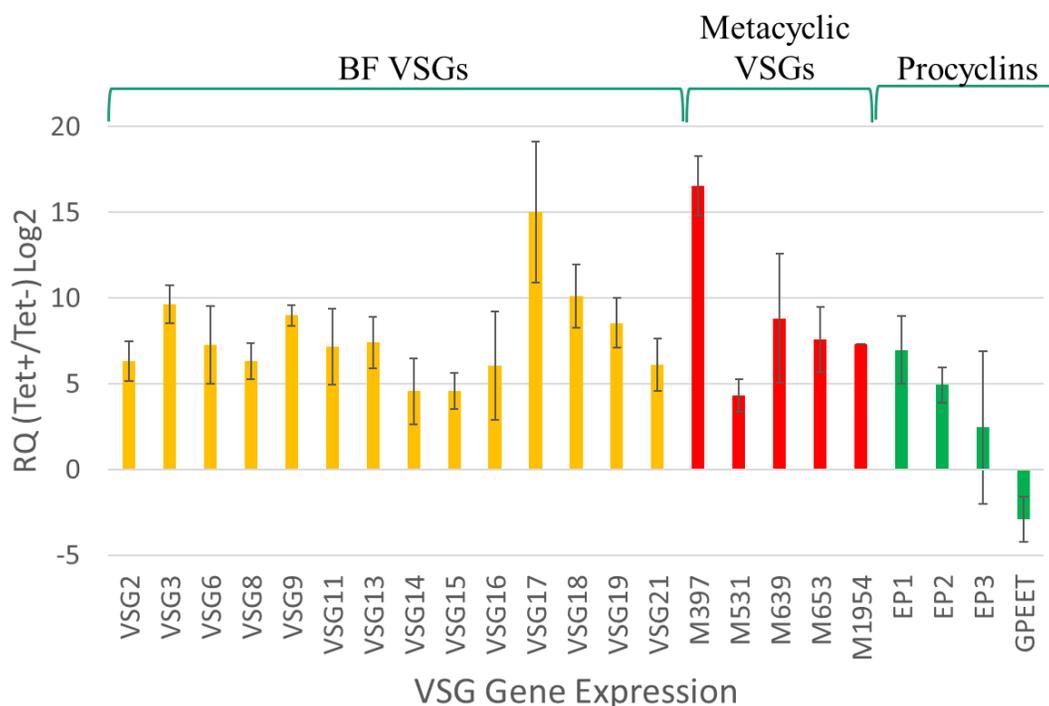
**Fig. 6. PIP5Pase is essential for the growth of PFs. A. Growth curve analysis after PIP5Pase knock down in *T. brucei* PFs.** Growth curve of cells expressing PIP5Pase-V5. Cells were grown in the absence or presence of 0.5  $\mu\text{g}/\text{mL}$  of tetracycline, and counted daily with a Beckman particle counter. Data shows the mean of three biological experiments, +/- SDm. **B. Western blot analysis of V5-tagged PIP5Pase in *T. brucei* PFs.** Cell lysates were extracted

from cells grown with or without tetracycline. Proteins were resolved in 10% SDS/PAGE and transferred to nitrocellulose membrane. PIP5Pase-V5 was detected using monoclonal mouse anti-V5 antibodies and secondary antibodies: goat anti-mouse IgG-HRP by chemiluminescence. The membrane was stripped and re-probed with mouse anti-mitochondrial HSP-70 antibodies. It was detected using secondary goat anti mouse IgG-HRP antibodies by chemiluminescence.

To determine the time required for the complete loss of protein, cells were collected every 24 hrs and cell lysates prepared for Western blot analysis from 24 to 120 hrs (Fig. 6B). There was a gradual decrease in PIP5Pase-V5 expression, however, the protein was almost completely depleted by day three (Fig. 6B). Interestingly, the cells remained viable at this time point. The data indicate that PIP5Pase is essential for PFs growth.

### 11.1.3. Knockdown of PIP5Pase results in the expression of silent VSGs in PFs

Previously published RNA-seq data from BFs show that the KD of PIP5Pase results in transcription of VSGs present in silent ESs as well as those found in sub-telomeric arrays (143). In PFs, all VSGs are silent, and the PFs express procyclins on their surface which protect the parasites against the acidic nature of the insect gut (26). To determine the effect of PIP5Pase KD on the expression of VSGs in PFs, qPCR was performed (Fig. 7).



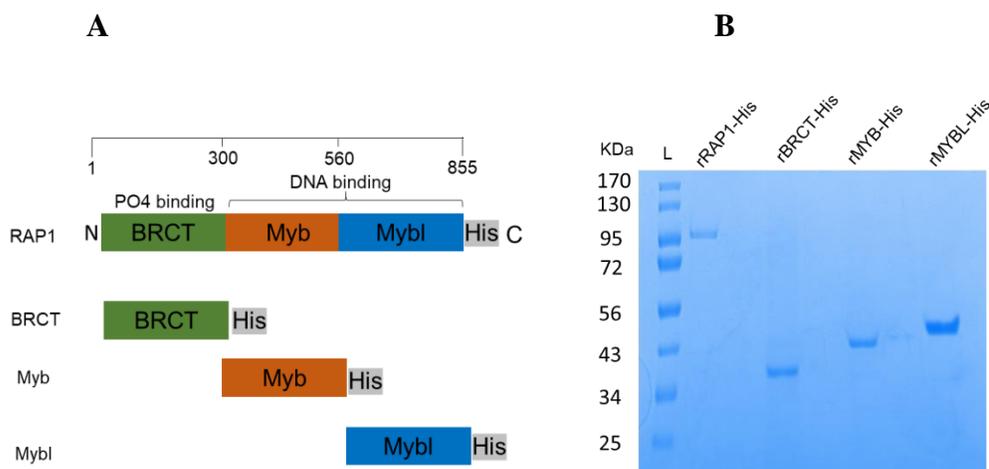
**Fig7. VSG gene expression analysis by qPCR after knockdown of PIP5Pase in *T. brucei* PFs.** The yellow bars represent the bloodstream form VSGs, the red bars represent the metacyclic form VSGs and the green bars represent the procyclins.

The analysis was done on cells collected on day three, a time point in which the protein was completely knocked down and the cells were still viable. qPCR was performed with primers specific to the VSGs found on the BF and MF ESs and procyclins. 18s rDNA and tubulin were used as endogenous controls. KD of PIP5Pase resulted in the expression of all silent VSGs found in the both BF ESs and MF ESs (Fig. 7). Upon KD, the change in expression was between 5-20-fold higher for BF and MF ES VSGs compared to cells that express PIP5Pase (Fig. 7). Interestingly, the expression of procyclins were also affected upon KD of the gene. Normally, all procyclins are transcribed in PFs and they are regulated post-transcriptionally. The cell line under study predominantly express EP2. First, EP1, EP2 and EP3 are overexpressed relative to the levels in tet+ cells with about 5-fold change in their expression (Fig. 7). Second, GPEET is downregulated relative to tet- cells with almost about 5-fold change in expression (Fig. 7). This data indicates that PIP5Pase is essential for silencing of VSGs in the PFs. PIP5Pase KD also affects the expression of procyclins.

### 11.2.1. Purification of RAP1 and its domains

Studies showed that RAP1 and PIP5Pase interact and migrate in a 0.9 MDa protein complex in native gel (143), and that RAP1 interacts with PI(3,4,5)P3, a substrate of PIP5Pase (25, 143). Interestingly, RAP1-PI(3,4,5)P3 interaction was ablated in the presence of PIP5Pase (143). ChIP, followed by qPCR also showed that RAP1 interacts with the ES chromatin at the 70 bp repeats and the telomeric repeats whereas PIP5Pase interacts with the 70 bp as well (25, 143). To validate the interactions *in vitro* between RAP1 and ES DNA sequences and to identify the RAP1 domains interacting with PI(3,4,5)P3, recombinant RAP1 (rRAP1) or RAP1 domains were purified to homogeneity (Fig 8B). *T. brucei* RAP1 is a 91 KDa protein with an N-terminal breast cancer 1 carboxy-terminal domain (BRCT), a central Myb domain and a C-terminal Myb-like (Mybl) domain (Fig. 8A). BRCT domains can be found among proteins involved in the cell cycle,

recombination as well as those involved in DNA-damage response (166). A study by Williams *et al.*, showed that BRCT binds to phosphorylated proteins via the recognition of phosphoserine residues (166). On the other hand, Myb and Mybl domains are found among proteins involved in transcriptional regulation and are known to bind DNA (167).



**Fig. 8. Purification of recombinant RAP1 and its truncations.** A. Diagram of recombinantly purified *T. brucei* His-tagged RAP1 (rRAP1-His) and its domains- rBRCT-His, rMyb-His and rMybl-His. B. 10% SDS/PAGE shows rRAP1-His and its domains purified to homogeneity. Recombinant proteins were resolved in a 10% SDS-PAGE, followed by Coomassie blue staining.

A number of conditions were tested to optimize the purification of each protein, summarized in table 3. The table lists the bacterial strains, induction temperatures and induction times tested and the conditions used for the purification if each protein.

**Table 5. Conditions tested to optimise the purification of proteins- rRAP1, rBRCT, rMyb, rMybl.**

Protein	Bacterial strain*		Induction temperature (°C)		Induction time (hours)	
	Tested	Used	Tested	Used	Tested	Used
rRAP1	1,2	1	18, 30, 37	18	2, 4, ON	ON

rBRCT	1,2,3	1	18, 30, 37	2, 4, ON	4
rMyb	1,2	1	18, 30, 37	2, 4	4
rMybl	1,2	1	18, 30, 37	2, 4	4

\*-1: NiCO(DE3), 2: Rosetta(DE3), 3: BL21(DE3)

ON indicates overnight.

Notable, NiCO(DE3) showed better yield than other cell lines for the expression of each protein and hence was used in the purification protocol. rRAP1 was induced at 18°C overnight (ON) while rMyb and rMybl were induced at 37°C for 4 hours. They were also soluble and were purified using Nickel beads. rBRCT was best expressed when induced at 37°C for 4 hours. However, unlike rRAP1 or the other domains, rBRCT was insoluble and the yield post purification was low. In order to increase the solubility and thereby the purification yield, a range of urea concentrations: 1M, 3M or 7M was used in addition to the lysis buffer. rBRCT was purified in denaturing conditions using 7 M urea. The proteins were dialysed against a pH 7.4 HEPES buffer to remove urea, allowing the protein to refold. This makes the protein suitable for use in experiments. The yields have been listed in table 4. The table lists the yield in  $\mu\text{g}/\mu\text{L}$  obtained for each protein after affinity purification. The purified proteins were then quantified using Pierce BCA Protein Assay Kit (Sigma Aldrich).

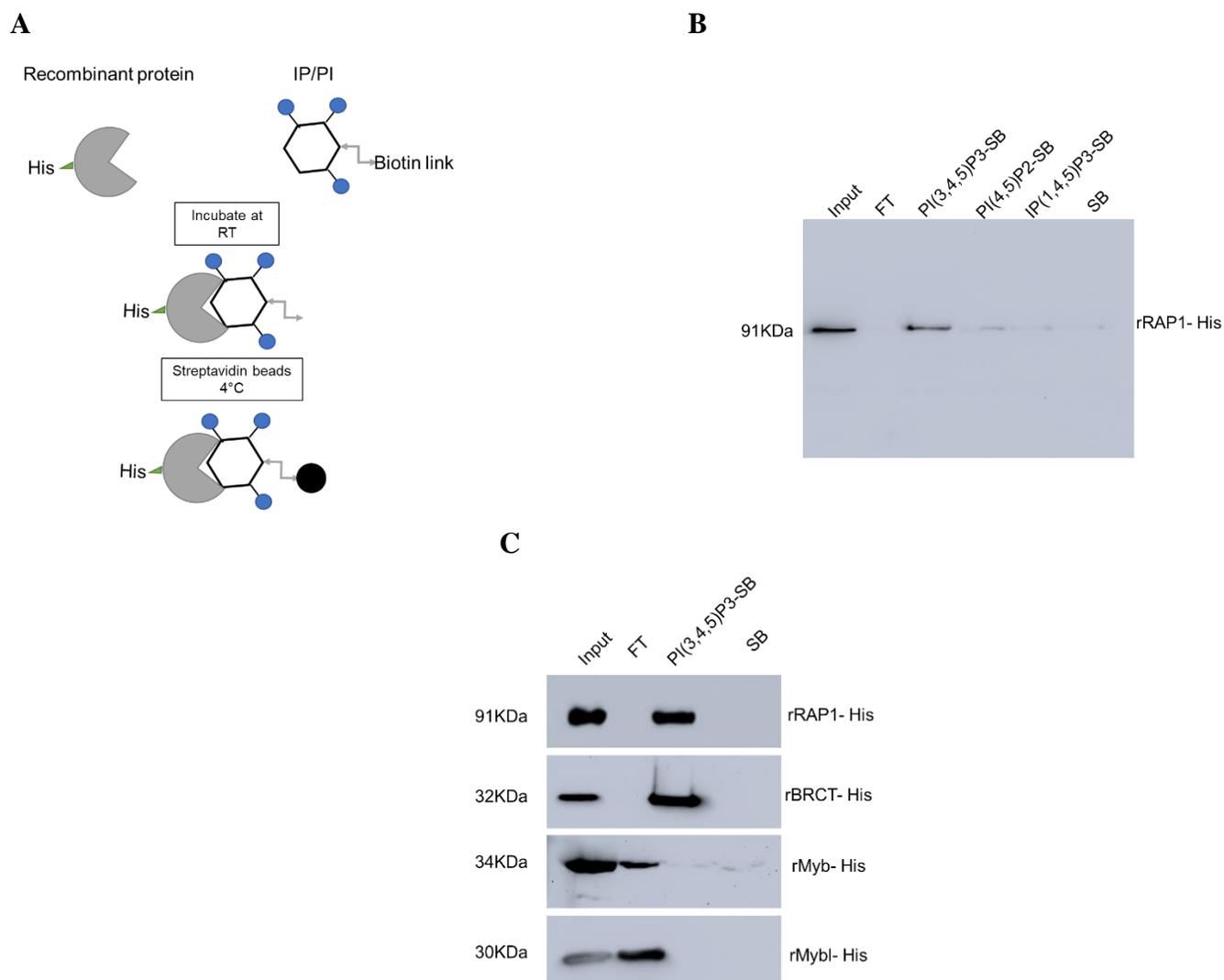
**Table 6. Yield of proteins (rRAP1, rBRCT, rMyb, rMybl) after affinity purification.**

Protein	Yield ( $\mu\text{g}/\mu\text{L}$ )
rRAP1	1.3
rBRCT	0.7
rMYB	1.1
rMYBL	1.02

### 11.2.2. BRCT domain of RAP1 binds to PI(3,4,5)P3

Pulldown assays with PI(3,4,5)P3 and other PI/IP metabolites revealed that RAP1 bound preferentially to PI(3,4,5)P3 (143). Purified recombinant proteins (rRAP1 and its domains) were

utilised in pulldown assays with biotinylated PIPs/IPs. After incubation with proteins, these metabolites were pulled down with streptavidin beads (Fig. 9A). The bound proteins were eluted from the beads and used in Western analysis. The flow through from each of these reactions were also collected for Western blot analysis using  $\alpha$ -His antibodies.



**Fig. 9. Phosphoinositide binding assays revealed that the rBRCT domain of RAP1 binds with PI(3,4,5)P3. A. Diagram of phosphoinositide binding assays used to study the binding of rRAP1-His, rBRCT-His, rMyb-His, rMybl-His to PIPs/IPs.** The recombinant protein and the biotin-labelled PIPs/IPs are incubated at room temperature for 1 hour, followed by which the PIPs/IPs are pulled down using streptavidin beads. Any bound protein is eluted and resolved on a 10% SDS-PAGE, followed by transfer to a nylon membrane. The protein is then detected using monoclonal mouse anti-His antibodies and secondary antibodies: goat anti-mouse IgG-HRP by chemiluminescence. **B. Phosphoinositide binding assay of rRAP1-His and PIPs/IPs.** The Western blot shows the binding assay performed

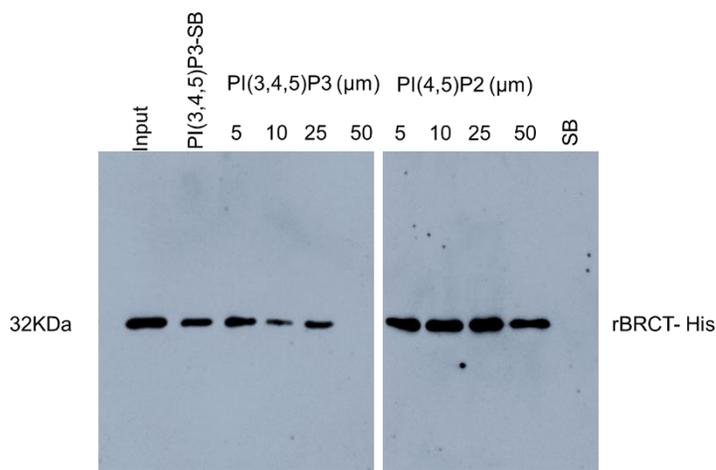
with His-tagged rRAP1 and biotinylated PIPs/IPs. **C. PI(3,4,5)P3 binding assays with rRAP1-His and its domains.** Western analysis of eluates from phosphoinositide binding assays with rRAP1, rBRCT, rMyb and rMybl. The binding assay with rRAP1 was used as the positive control and the binding assays with each of the proteins with streptavidin beads was used as the negative control.

First, binding assays showed that rRAP1 binds to PI(3,4,5)P3, but not PI(4,5)P2 or IP(1,4,5)P3 (Fig. 9B). There was no nonspecific binding of the proteins to the beads as confirmed by a negative control using only rRAP1 and streptavidin beads (Fig. 9B). To determine which domain of RAP1 bound to PI(3,4,5)P3, assays were performed with rBRCT, rMyb and rMybl recombinant proteins (Fig. 9C). Reactions with rRAP1 were used as positive controls (Fig. 9C). Reactions incubated with beads were used to rule out nonspecific binding of the proteins to the beads (Fig. 9C). The data showed that the BRCT domain of RAP1 bound to PI(3,4,5)P3; however, neither Myb nor Mybl domains bound to the PI(3,4,5)P3 and they could be detected in the flow through (Fig. 9C).

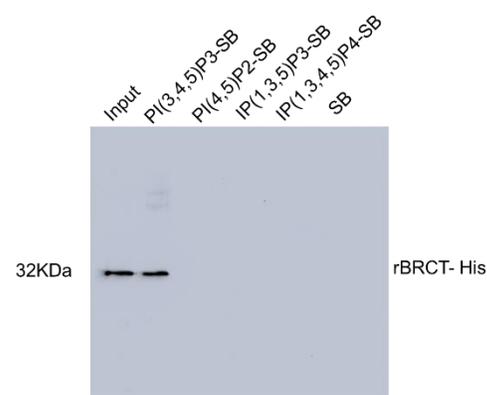
### 11.2.3. BRCT domain of RAP1 binds specifically to PI(3,4,5)P3

To determine if rBRCT binds specifically to PI(3,4,5)P3-biotin, competition assays were performed by addition of increasing molar excess concentrations of non-biotinylated PI(3,4,5)P3 or PI(4,5)P2 to the binding reactions (Fig. 10A).

**A**



**B**



**Fig. 10. BRCT domain of RAP1 binds specifically to PI(3,4,5)P3. A. PI(3,4,5)P3 competition assay.**

Phosphoinositide binding assay of His-tagged rBRCT with PI(3,4,5)P3 from binding assays with PI(3,4,5)P3 in the presence of free PI(3,4,5)P3 and free PI(4,5)P2. Proteins were resolved in a 10% SDS/PAGE and transferred to a nitrocellulose membrane. rBRCT was detected using monoclonal mouse anti-His antibodies and secondary goat anti-mouse IgG-HRP antibodies by chemiluminescence. **B. Binding assay with rBRCT-His and PIPs/IPs.** Phosphoinositide binding assay of His-tagged rBRCT with PI(3,4,5)P3, PI(4,5)P2, IP(1,3,5)P3, and IP(1,3,4,5)P4 to identify PIPs/IPs that bind to rBRCT. rBRCT was detected using monoclonal mouse anti-His antibodies and secondary goat anti-mouse IgG-HRP by chemiluminescence.

Interestingly, the binding of rBRCT to biotinylated PI(3,4,5)P3 was competed out by molar excess of PI(3,4,5)P3 (Fig. 10A). The binding was partially competed out at 10 and 25 $\mu$ M of non-biotinylated PI(3,4,5)P3 and was completely outcompeted at a concentration of 50 $\mu$ M of non-biotinylated PI(3,4,5)P3 (Fig. 10A). The binding remained unaffected by the addition of molar excess of PI(4,5)P2 (Fig. 10A). The data indicated that the binding between PI(3,4,5)P3 and rBRCT is specific.

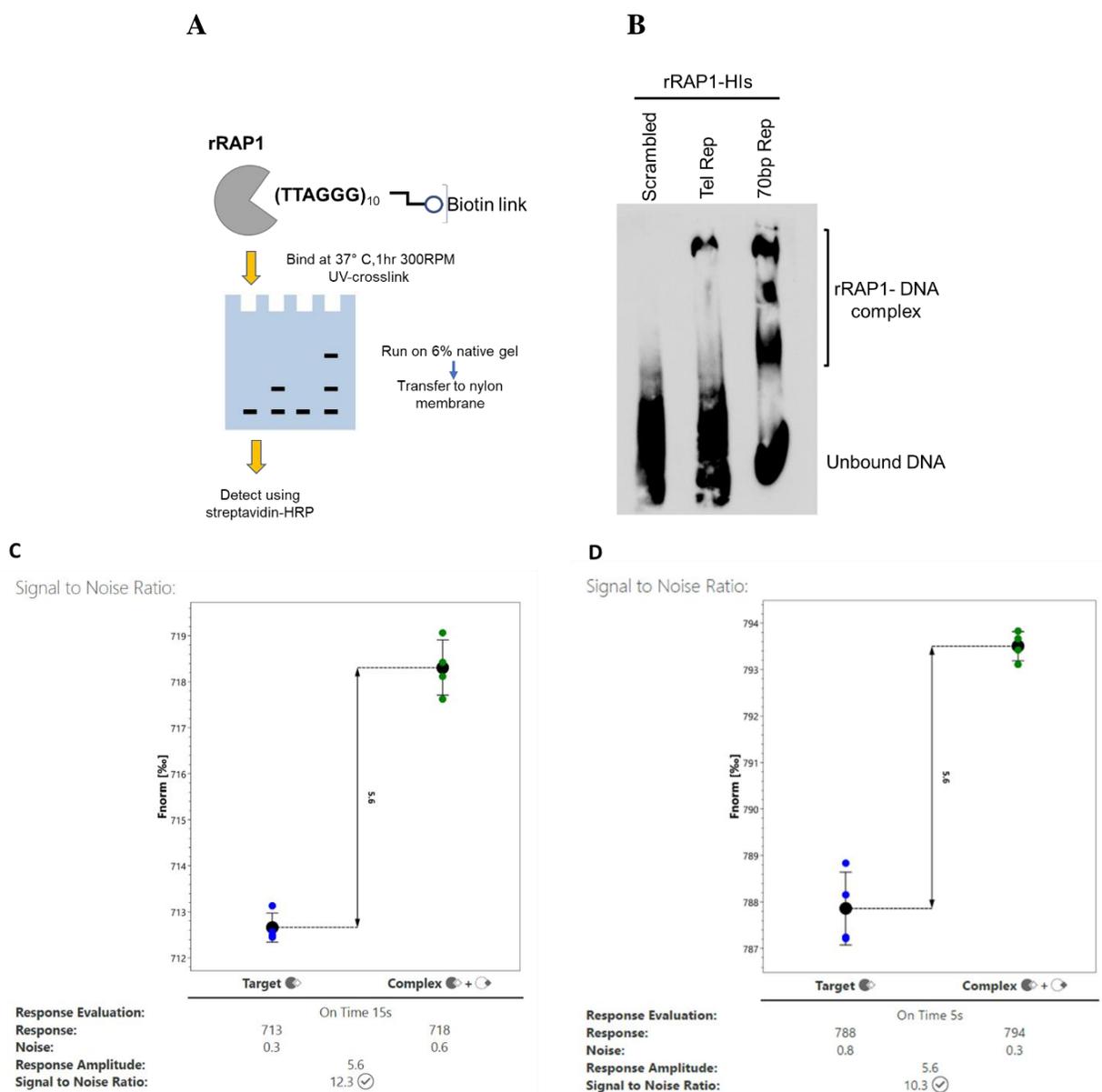
Data from previous work show that RAP1 binds specifically to PI(3,4,5)P3 and in trace amounts with PI(4,5)P2 while it does not bind other PIP/IP metabolites (143). To determine if rBRCT bound other PI/IP metabolites, similar binding assays were used. rBRCT bound specifically to PI(3,4,5)P3 but not other PIP/IP metabolites (Fig 10B). Unlike RAP1, not even trace amounts of rBRCT bound to PI(4,5)P2 (Fig 10B). Hence, RAP1 binds specifically to PI(3,4,5)P3 via the N-terminal BRCT domain.

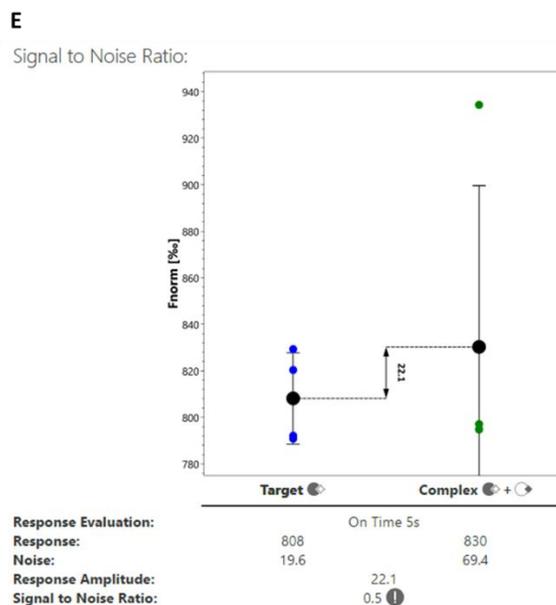
### **11.3. Myb and Mybl domains of RAP1 bind ES chromatin DNA.**

#### **11.3.1. RAP1 binds directly to 70 bp repeats and telomeric repeats *in vitro*.**

ChIP followed by qPCR revealed that RAP1 binds to ES chromatin at the 70 bp and telomeric repeats (143). RAP1 is also associated with PIP5Pase which also interact with 70 bp and telomeric repeat chromatin (143). Moreover, these proteins are a part of a 0.9 MDa protein complex (143). To determine if RAP1 bound directly to ES DNA repeats, electrophoretic mobility shift assays (EMSA) were performed. Reactions were performed with purified rRAP1 and biotinylated DNA

sequences i.e., synthetic telomeric repeats (ten repeats of TTAGGG) and 70 bp repeats (one sequence of 70 bp length). In this assay, the DNA bound to the protein will migrate slower than unbound DNA, appearing as a shift in the migration. To control for nonspecific DNA-protein binding a scrambled sequences derived from telomeric repeats was used. These reactions were then run on native gels, transferred to nylon membranes and detected using the streptavidin based-chemiluminescence. First, the assay was performed with telomeric repeats, 70 bp repeats and the scrambled sequence (Fig. 11B).

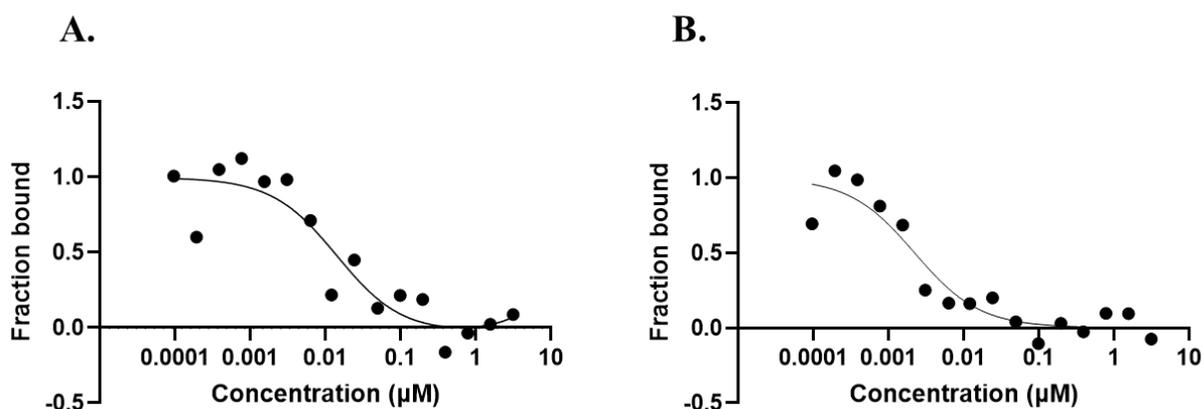




**Fig. 11. RAP1 binds directly to 70 bp repeats and telomeric repeats in vitro. A. Diagram of the EMSA used to study the binding of rRAP1-His and its domains to 70 bp repeats and the telomeric repeats B. EMSA of rRAP1-His with biotin-labelled 70 bp or telomeric repeat DNA.** rRAP1 and the biotin labelled DNA sequences are incubated for 1 hour at 37°C in the appropriate buffer. The reactions are then UV-crosslinked and resolved on a 6% native gel, followed by transfer to a nylon membrane. Afterwards, the DNA is detected using streptavidin-HRP with the chemiluminescent nucleic acid detection module (Thermo Fischer Scientific). **C. Binding check by MST to confirm the binding of rRAP1-His to telomeric repeats. D. Binding check by MST to confirm the binding of rRAP1-His to 70 bp repeats. E. Binding check by MST to rule out non-specific binding of rRAP1-His using the scrambled sequence.** The curve shows the binding check performed by MST using the Monolith NT.115 to confirm the interaction between rRAP1 and the Cy5-labelled DNA sequences. The Y-axis represents Fnorm (the normalised fluorescence i.e., the difference in fluorescence before and after activation of the IR laser). The X-axis represents the samples: without (left) or with (right) protein. Binding is detected when there is a change in DNA fluorescence (Fnorm) of the Cy5-labelled telomeric repeats when it is bound to a protein (rRAP1). Further, the signal obtained from binding of telomeric repeats and 70 bp repeats to RAP1 was larger than the signal obtained from the binding of the scrambled sequence to RAP1, confirming the binding of rRAP1 to telomeric repeats and 70 bp repeats.

Two major observations were made from this assay. First, rRAP1 bound to both 70 bp repeats and telomeric repeats whereas it did not bind the scrambled control (Fig. 11B). Secondly, rRAP1 interactions with the telomeric repeats were different from its interaction with the 70 bp repeats (Fig. 11B). rRAP1 migrates as a single complex with the telomeric repeats (Fig. 11B). With the 70 bp repeats, RAP1 forms multiple DNA-protein complexes (Fig. 11B).

MST analysis using rRAP1 and the three DNA sequences further corroborate these findings (Fig. 11C, 11D, 11E). Binding analysis performed with rRAP1 and the telomeric repeats confirm the interaction of rRAP1 and the telomeric repeats *in vitro* (Fig.11C). Similarly, binding analysis performed with rRAP1 and the 70 bp repeats confirm the interaction between rRAP1 and 70 bp repeats *in vitro* (Fig.11D). Finally, control assays performed with rRAP1 and the scrambled sequence concluded that rRAP1 binds specifically with the telomeric repeats and the 70 bp repeats *in vitro*. The data indicates that RAP1 can bind directly to the telomeric repeats or 70 bp repeats *in vitro*.



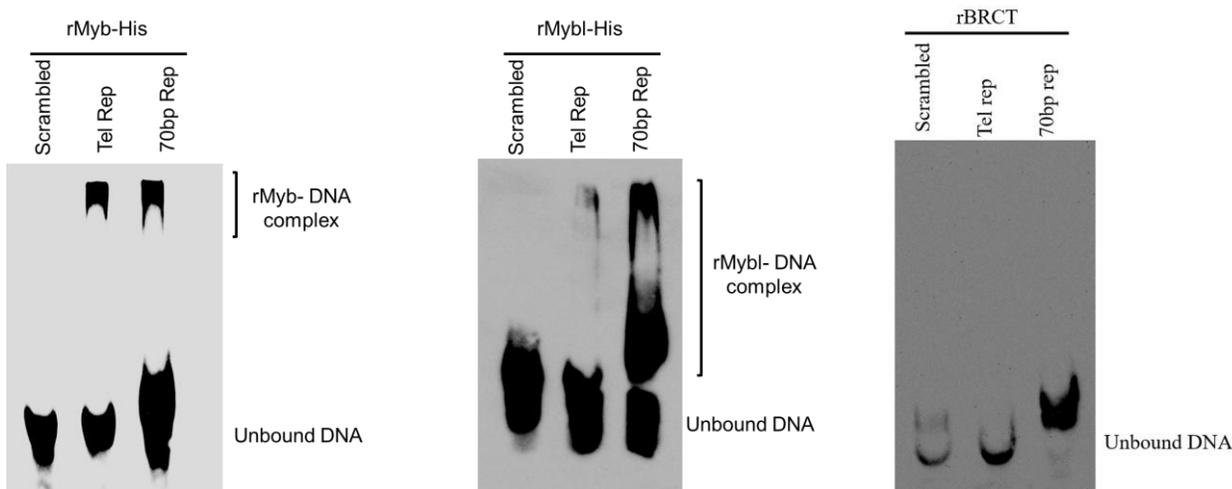
**Fig 12. RAP1 binds with different affinities to the tel repeats and 70 bp repeats *in vitro*. A. Equilibrium dissociation constant ( $k_d$ ) calculated for the binding of rRAP1-telomeric repeats from dose response curve obtained from MST.** MST analysis was performed with 5 nM telomeric repeats and 1:2 serial dilutions of rRAP1 (3.125 µM was the highest concentration used). The curve represents the dose response obtained from MST for rRAP1 and telomeric repeats. The Y-axis represents the fraction bound. The X-axis represents the concentration of rRAP1. **B. Equilibrium dissociation constant ( $k_d$ ) calculated for the rRAP1-70 bp repeats from dose response curve obtained from MST.** The curve represents a replicate of the dose response curve obtained from MST for rRAP1 and 70 bp repeats. The Y-axis represents the fraction bound. The X-axis represents the concentration of rRAP1 in each reaction.

The equilibrium dissociation constants for rRAP1-telomeric repeats as well as rRAP1-70 bp repeats were measured using MST (Fig. 12A and 12B). A dose response curve was generated in each case and the  $k_d$  was measured (Fig. 12A and 12B) and the average of the  $k_d$  was calculated. For the binding of rRAP1 with telomeric repeats, the  $k_d$  was determined to be 0.03127 µM. For

the binding of rRAP1 with 70 bp repeats, the  $k_d$  was determined to be  $0.3940 \mu\text{M}$ . This data indicates a difference in the affinities of each of the DNA sequences to rRAP1 *in vitro*.

### 11.3.2. Myb and Mybl domains of RAP1 bind to the ES chromatin DNA *in vitro*.

RAP1 has a central Myb and a C-terminal Mybl domain. Myb domains are DNA binding domains found in proteins that are involved in transcriptional regulation (167). To determine the domains of RAP1 that bind DNA, EMSA was performed with purified rMyb, rMybl or rBRCT domains (Fig. 13).



**Fig. 13. EMSA with rMyb-His, rMybl-His and rBRCT-His with telomeric repeats and 70 bp repeats.** Gel shift assays were performed with rMyb, rMybl and rBRCT domains of RAP1 and DNA sequences. Reactions were resolved on a 6% polyacrylamide native gel and transferred onto a nylon membrane. The biotin-labelled DNA is detected by streptavidin-HRP using chemiluminescence.

Both rMyb and rMybl domains of RAP1 bound to both the telomeric repeats as well as the 70 bp repeats whereas it did not bind to the scrambled sequence whereas rBRCT did not bind any DNA sequence (Fig. 13). Interestingly, there was an apparent difference on the domains binding the DNA sequences. rMyb bound to both sequences forming single protein-DNA complexes whereas rMybl bound to the telomeric repeats forming a single DNA-protein complex but it formed multiple protein-DNA complexes with the 70 bp repeats (Fig. 13). Together, this data indicate that

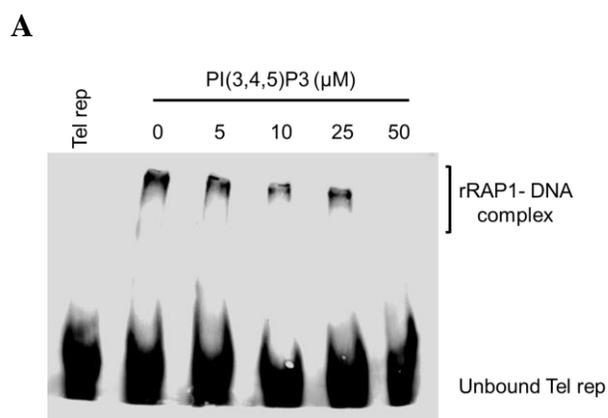
the Myb and Mybl domains of RAP1 bind the telomeric repeats as well as the 70 bp repeats. Interestingly, these domains bind apparently differently to the telomeric repeats and the 70 bp repeats.

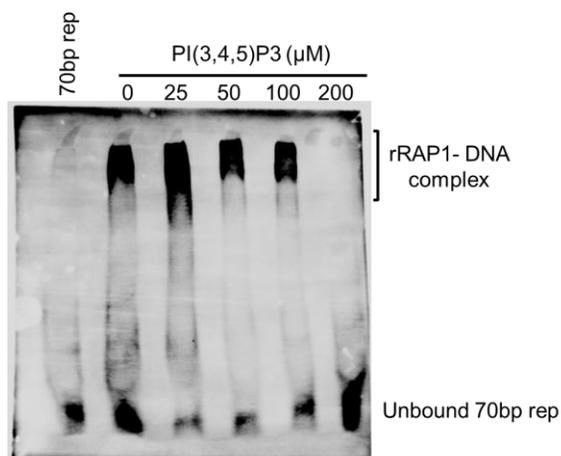
### 11.3.3. PI(3,4,5)P3 disrupts the binding of RAP1 to telomeric repeats and 70 bp repeats.

Previous data showed that PIP5Pase KD or catalytic mutation affects transcription of ESs (Cestari *et. al.*, 2019 and Fig. 7) and affects association of RAP1 with ES DNA, and thus ES transcription. Hence, EMSA was performed to test whether PI(3,4,5)P3 affects RAP1 binding to ES repeats.

First, EMSA was performed with telomeric repeats and rRAP1 with the addition of increasing molar concentrations of PI(3,4,5)P3.

Increase in PI(3,4,5)P3 concentration disrupted the binding of rRAP1 with the telomeric repeats in a dose-dependent fashion (Fig. 14A). At 5 $\mu$ M, there was a slight change in rRAP1-DNA binding whereas the increase in PI(3,4,5)P3 concentrations to 10 and 25  $\mu$ M resulted in the loss of DNA-protein complex at the top of the gel (Fig. 14A). Moreover 50  $\mu$ M of PI(3,4,5)P3 completely inhibited the binding of rRAP1 to telomeric repeats (Fig. 14A). The data indicates that PI(3,4,5)P3 affects the binding of RAP1 to telomeric repeats.



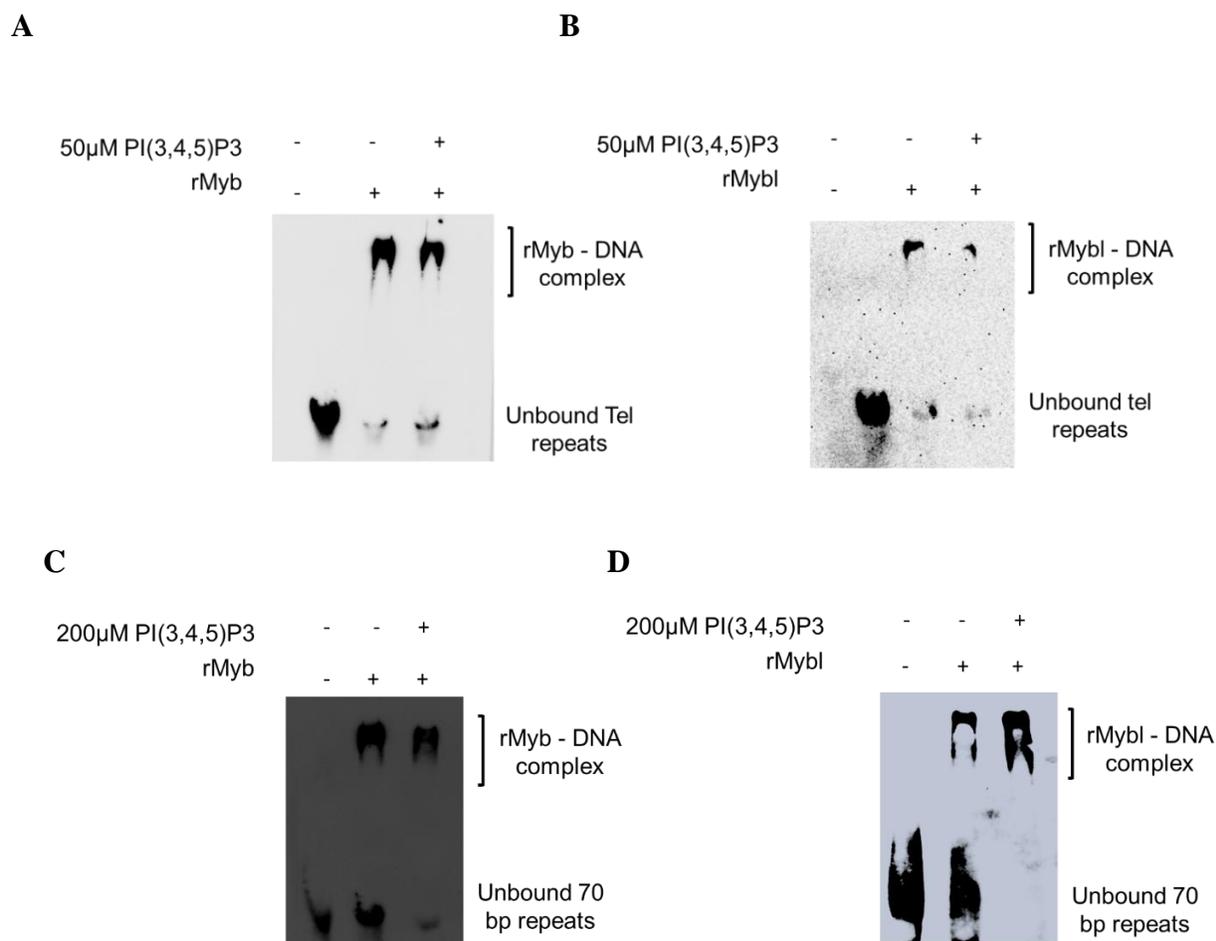


**Fig. 14. Binding of RAP1 to ES chromatin DNA *in vitro* is affected by PI(3,4,5)P3.** EMSA was performed with RAP1 and telomeric repeats or 70 bp repeats in the presence of increasing concentrations PI(3,4,5)P3. The reactions were resolved on a 6% polyacrylamide native gel and transferred to a nylon membrane. The biotin-labelled DNA was detected using streptavidin-HRP by chemiluminescence. **A. Binding of rRAP1-His with telomeric repeats in the presence of PI(3,4,5)P3.** **B. Binding of rRAP1-His with 70 bp repeats in the presence of PI(3,4,5)P3.**

EMSA was also performed with rRAP1 and 70 bp repeats with the addition of molar excess PI(3,4,5)P3 (Fig. 14B). Interestingly, 50  $\mu\text{M}$  of PI(3,4,5)P3 did not affect the binding of rRAP1 to 70 bp repeats (Fig. 14B). To determine if a further increase in the PI(3,4,5)P3 concentration would affect the binding, EMSA was performed with PI(3,4,5)P3 concentrations ranging from 25-200  $\mu\text{M}$  (Fig. 14B). The binding of rRAP1 to 70 bp repeats was only disrupted at 200  $\mu\text{M}$  (Fig. 14B). Together, the data indicate that the *in vitro* binding of RAP1 to either DNA sequences was affected by the addition of PI(3,4,5)P3. However, the binding of RAP1 to telomeric repeats was at least four times more sensitive to PI(3,4,5)P3 than 70 bp repeats. This suggests that the binding of RAP1 to the ES chromatin DNA is affected by the presence of PI(3,4,5)P3.

To test if the binding of the RAP1 domains Myb and Mybl is affected by the presence of PI(3,4,5)P3, assays were performed with the recombinant domains and the DNA sequences (Fig. 15). Specifically, assays with 70 bp repeats were performed in the presence of 200  $\mu\text{M}$  PI(3,4,5)P3, the concentration required to disrupt the binding of rRAP1 to the 70 bp repeats whereas assays with the telomeric repeats were performed in the presence of 50  $\mu\text{M}$  PI(3,4,5)P3 (Fig. 15).

None of the protein DNA interactions were affected by the presence of PI(3,4,5)P3. The data suggests that PI(3,4,5)P3 affects the binding of RAP1 to the DNA sequences through its interaction with the BRCT domain of RAP1 and not via Myb or Mybl domains.



**Fig. 15.** PI(3,4,5)P3 affects the binding of RAP1 to the DNA sequences through its interaction with the BRCT domain of RAP1 and not via Myb or Mybl domains. EMSA was performed with rMyb or rMybl and telomeric repeats or 70 bp repeats in the presence of increasing concentrations of PI(3,4,5)P3. The reactions were resolved on a 6% polyacrylamide native gel and transferred to a nylon membrane. The biotin-labelled DNA was detected using streptavidin-HRP by chemiluminescence. **A.** Binding of rMyb-His with telomeric repeats in the presence of PI(3,4,5)P3 **B.** Binding of rMybl-His with telomeric repeats in the presence of PI(3,4,5)P3. **C.** Binding of rMyb-His with 70 bp repeats in the presence of PI(3,4,5)P3. **D.** Binding of rMybl-His with 70 bp repeats in the presence of PI(3,4,5)P3.

## 12. Discussion

### 12.1. PIP5Pase plays a key role in regulating VSG expression in different developmental stages.

PIP5Pase is a nuclear protein and is one among the key members of IP pathway (136). The conditional knockdown of PIP5Pase was lethal to *T. brucei* PFs. *T. brucei* PFs reside in the insect gut and unlike BFs, do not express VSGs on their surface (168). Instead, they express a different coat known as procyclins which protect the parasite from the protease content of the insect gut (168). Interestingly, the knockdown of PIP5Pase led to the expression of silent VSG genes in procyclic forms and also affected the transcription of procyclins- both of which are transcribed by RNA polymerase I. Previous data show that PIP5Pase is essential in BFs of the parasite, the knockdown of which led to the derepression of the silent VSGs (136). The knockdown of PIP5Pase in PFs occurred within 3 days after depriving the media of tetracycline. This differed from what was observed in the BFs where the protein is depleted within a few hours (136). The differences can be attributed to the differences in the division time between the BFs and PFs. BFs have a doubling time of about 6 hours whereas PFs have a longer doubling time of about 11 hours. Secondly, the growth defect appears to start 5 days after the knockdown of the protein in PFs whereas data on BFs show that the growth defect occurs 24 hours after the knockdown of PIP5Pase (136). The longer time required for procyclic forms to divide could account for the longer interval that exists between the time of tetracycline removal and establishment of the growth defect. Specifically, the differences in the rate of cell division could affect the dilution of protein over the course of multiple cell divisions. Moreover, the turnover of proteins from the two life stages maybe different. This could possibly account for the longer time required for the depletion of the protein in PFs than in BFs. This phenomenon has been observed in the case of some RBPs. For instance, protein turnover rate for KREPB4 is about 2-4 days in BFs whereas it's about 4-8 days in the case of PFs (169). In BFs, with divisions occurring at every few hours, any remaining PIP5Pase is subject to dilution at a much faster rate. On the other hand, protein dilution can occur at a much smaller rate in PFs as the rate of division is relatively lower.

VSGs are expressed in a monoallelic fashion in BFs whereas in PFs they are silent. This status of VSG expression is stringently maintained by various proteins and protein complexes. Perturbation of some of these factors affects the status of VSG transcription (25). Previous studies on the inositol phosphate pathway provide evidence that argues that the integrity of this pathway is key in maintaining VSG monoallelic expression in BFs (136). Further studies on PIP5Pase in BFs and the current study on the PFs also suggests that PIP5Pase, a member of the pathway is crucial for the appropriate expression and silencing of VSGs in either stage (143). Interestingly, the magnitude of derepression of VSGs varies greatly between the BFs and PFs. RNA analysis from BFs after conditional knockdown of PIP5Pase showed that the MF and BF ES VSGs were up to 10,000-fold abundant relative to cells which actively expressed PIP5Pase (143). RNA analysis from PFs after conditional knockdown of PIP5Pase shows a much lower abundance of about 5 to 20-fold relative to the cells which actively express PIP5Pase. This derepression of VSGs upon knockdown of PIP5Pase may be a consequence of the chromatin organisation in either of the developmental stages. Studies have shown that the chromatin structure plays a major role in regulating the transcription and silencing of the ESs. Studies by Cross and Rudenko laboratories on the chromatin of active and silent ESs highlight the structural differences between the chromatin of actively transcribed ESs and silent ESs (124, 130). For instance, the chromatin of an active ES is depleted of nucleosomes and thus it has a more open structure compared to a silent ES that has a more compact structure (124, 130). Furthermore, recent studies on BFs have revealed the association of various chromatin remodelling factors that regulate ES transcription by maintaining the structure of the active and silent ES. This includes the ISW1 complex, FACT complex, BDFs, DOT1b, histones, which actively participate in maintaining the compact chromatin structure as seen in silent ESs in BFs (126) (132) (61) (170). Perturbation of any of these proteins or the associated complexes either led to the transcription of silent VSGs or switching of VSGs in these silent ESs. Similarly, studies on PFs have also provided evidence indicating the importance of chromatin integrity in maintaining the silent state of VSGs in PFs. Work by Navarro et al., has shown that T7RNAP mediated transcription from a T7 promoter integrated in the chromosome is completely suppressed in BES in PFs but not in the PFs suggesting a tighter chromatin conformation in PFs (171). Formaldehyde assisted isolation of regulatory elements or FAIRE analysis of ES chromatin in PFs also suggests the presence of a more compact ES chromatin structure in PFs (146). In the same study, the depletion of the telomere associated RAP1 led to a

strong derepression of VSGs from the ESs (146). PIP5Pase, like RAP1, is associated with the ES chromatin and interestingly, are both are part of a 0.9 MDa protein complex. This suggests a possibility whereby the depletion of PIP5Pase affects the chromatin structure of the associated ES, resulting in the derepression of VSGs. In addition to this, a possible explanation for the differences observed in the magnitudes of derepression between the two life stages could be the differences in chromatin structures between the BFs and PFs as discussed earlier (146).

PIP5Pase associates with the ES chromatin as suggested by previously published data (143). The possible influence over the chromatin conformation is further corroborated by evidence collected from these recent studies. ChIP data on PIP5Pase shows that PIP5Pase associates to the ES chromatin, highly enriched at the 70 bp repeats. In addition to this, PIP5Pase is a part of a major protein complex. All current data on PIP5Pase seems to suggest a model whereby PIP5Pase catalytic activity regulates VSG expression by influencing the chromatin structure. However, the current data does not clarify the role of the other proteins found in association with PIP5Pase. Other proteins in this complex involve regulatory proteins, proteins of the nuclear lamina, RNA processing proteins and DNA binding proteins. Some of these proteins have been associated with the regulation of gene expression in other organisms. For example, homologs of CRK9 and PP1 (components of the 0.9 MDa multiprotein complex) in *Saccharomyces pombe* have been implicated in regulating RNA Pol II transcription (172). RAP1, another associated member of the complex, also binds telomeric repeats in *S. cerevisiae* mediating a telomeric silencing effect (173). NOP54 may also play a role in the developmental regulation of VSG expression since it plays an important role in chromosomal segregation during cell cycle (174). Studies focusing on these interacting proteins in either life stages of *T. brucei* will be crucial to gain clarity on the role of the 0.9 MDa multiprotein complex in the regulation of VSG gene expression.

## **12.2. Characterization of RAP1 and its domains**

Data from previous studies show that RAP1 interacts with PIP5Pase as a part of a multiprotein complex (143). RAP1 also binds the ES chromatin, being highly enriched at the 70 bp repeats as well as the telomeric repeats (143). In addition RAP1 also binds PI(3,4,5)P<sub>3</sub>, a substrate of

PIP5Pase (143). *In vitro* assays from this study aimed at characterizing the domains of RAP1 that were responsible for these interactions. This provides a better understanding of the mechanism by which these interactions influence the ES transcription. Recombinant proteins (RAP1 and its domains) were purified for these *in vitro* assays. Data from the previous study on RAP1-PI(3,4,5)P3 interaction was reproduced using biotinylated PI(3,4,5)P3 as a bait in pull down assays (143). Interestingly, pull down assays with each of the domains revealed that only the N-terminal BRCT domain of RAP1 bound to PI(3,4,5)P3 while the Myb domain or the Myb1 domain did not. This is the first report that demonstrates the PIP binding activity of the BRCT domain. The BRCT domain has been implicated in recognizing and binding phosphopeptides (144) but never has it been shown to bind phosphoinositides. Notably, a more recent study on *T. brucei* RAP1 has also revealed that the BRCT domain is crucial for the protein-protein “self-interaction” between RAP1 molecules which could account for the enrichment of the protein at the ES chromatin (147). Put together, these studies along with current data suggests a role for the BRCT domain of RAP1 in the regulation of VSG transcription which involves its interaction with PI(3,4,5)P3. Assays with few other PIPs and IPs showed that it bound specifically to PI(3,4,5)P3 and none of the others including PI(4,5)P2, IP(1,3,5)P3, and IP(1,3,4,5)P4. Further, data showing that molar excess of PI(3,4,5)P3 and not PI(4,5)P2 out competes the BRCT-PI(3,4,5)P3 interaction corroborates the specificity of this interaction. However, more detailed analysis of this interaction is needed to determine the key amino acid residues involved in this interaction. Structural characterization of the BRCT domain of BRCA, a protein implicated in tumor suppression, revealed amino acids responsible for its phosphopeptides binding ability including methionine (M1775), valine (V1696 and V1809), and arginine (R1699) residues, possibly forming a phosphopeptide binding motif (144). Mutation of any of these amino acids (R1699Q, R1699W, V1696L, M1775R and V1809F) ablated the protein’s interaction with phosphopeptides (144). Multiple sequence analysis of *T. brucei* RAP1 BRCT domain and other RAP1 BRCT domains revealed certain conserved amino acid residues- P81, C118, Y144. Further characterization of the domain by mutation of these conserved amino acids (P81, C118, Y144) could shed light on the important amino acids needed for this interaction. It would also be essential to determine if this interaction is specific to *T. brucei* RAP1 or if it is a universal property of RAP1 conserved among other eukaryotes.

EMSA used to determine the domains involved in RAP1 protein-DNA interaction have revealed interesting details about the protein-DNA interaction. Firstly, RAP1 directly interacts with both the telomeric as well as the 70 bp repeats. RAP1 forms a large protein complex with many other proteins, some of which (for example, TRF) are known to interact directly with DNA (22). Hence, it is important to establish if the interaction of RAP1 with the telomeric ES is direct or not (22). Thus, the EMSA establishes that the interaction of RAP1 with the telomeric repeats and the 70 bp repeats can be direct. Secondly, EMSA with the RAP1 domains revealed that the domains responsible for the interaction of RAP1 with the DNA sequences are the central Myb and the N-terminal Mybl domains, both of which are conserved among RAP1 in other eukaryotes which are known to interact with DNA (142). The EMSA show that the Myb domain interact with the telomeric repeats as well as the 70 bp repeats. Moreover, it appears to interact with both the sequences, similarly, forming a single protein-DNA complex. Interestingly, Mybl, which also binds both sequences, forms multiple protein-DNA complexes with the 70 bp repeats whereas it appears to have a much weaker interaction with the telomeric repeats. The multiple complexes formed with the 70 bp repeats suggest the possibility that stoichiometric differences exist between these sequences. The formation of multiple complexes with 70 bp repeats may also reflect the differences in affinity between the sequences.

To gain a better understanding of the affinities of RAP1 to the telomeric repeats and the 70 bp repeats, MST was used. Interestingly, MST revealed a clear difference between the binding of RAP1 to the telomeric repeats and the 70 bp repeats. RAP1 binds with telomeric repeats with a higher affinity ( $0.03127 \mu\text{M}$ ) compared to the 70 bp repeats ( $0.3940 \mu\text{M}$ ). A much higher concentration ( $250 \mu\text{M}$ ) of PI(3,4,5)P3 was required to disrupt the binding of RAP1 to the 70 bp repeats when compared to telomeric repeats which required only  $50 \mu\text{M}$  of PI(3,4,5)P3. The differences may be significant to understand the mechanisms of RAP1 and PIP5Pase regulation of ES silencing.

### 12.3. PI(3,4,5)P3 regulates RAP1 binding to the ES chromatin.

Interestingly, the presence of PI(3,4,5)P3 appeared to affect the binding of RAP1 to the telomeric repeat sequences at a concentration much lower (50  $\mu$ M) than that required to disrupt the binding of the 70 bp repeat sequences (250  $\mu$ M). This *in vitro* data correlates to the published ChIP data which showed that the loss of PIP5Pase catalytic activity affected the binding of RAP1 to the 70 bp and telomeric repeats *in vivo* (143). More specifically, the loss of PIP5Pase catalytic activity led to a three-fold increase in the binding of RAP1 to the 70 bp repeats accompanied by a 650-fold decrease in its binding to the telomeric repeats (143). A model proposed by the authors suggests that PI(3,4,5)P3, a substrate of RAP1, is dephosphorylated by PIP5Pase catalytic activity to PI(3,4)P2, depriving RAP1 of its substrate (143). In turn, RAP1 binds to the ES chromatin leading to a change in the conformation of the chromatin, silencing the associated VSG gene (143). In addition to characterizing the domains of RAP1, this thesis also sheds light on the mechanistic aspect of this proposed model. In absence of PI(3,4,5)P3, RAP1 binds the chromatin via the Myb and Mybl DNA binding domains. In presence of PI(3,4,5)P3, the BRCT domain of RAP1 binds to PI(3,4,5)P3 which potentially brings about a conformational change in RAP1 which disrupts its the binding to the ES chromatin. The presence of PI(3,4,5)P3 may also influence the oligomerization of RAP1 molecules which has been identified to occur by the interaction between the BRCT domains of RAP1 molecules (175). Oligomeric RAP1 might play an important role in the binding of RAP1 to the telomeric repeats or 70 bp repeats. The possible change to the RAP1 oligomeric state by the binding of the PI(3,4,5)P3 to the BRCT domain may affect RAP1-DNA interaction.

### 13. Conclusions

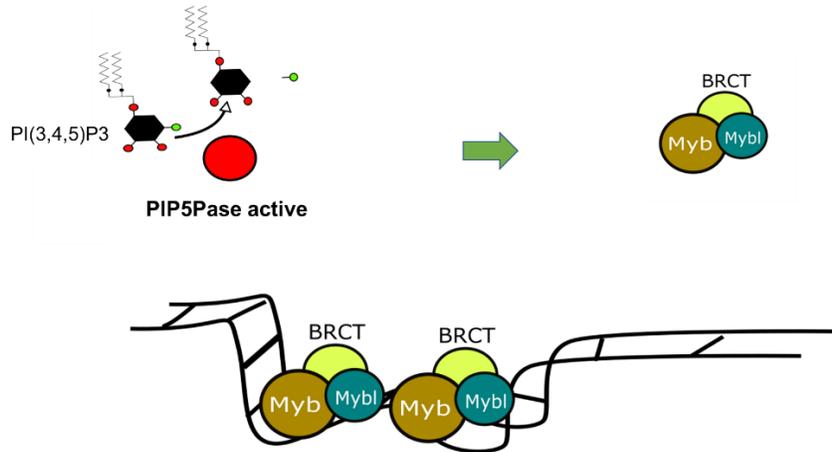
The process of antigenic variation in *T. brucei* has been a subject of great interest for many decades. Precisely, the regulation of VSG gene expression has been extensively studied leading to the discovery and implication of multiple proteins and protein complexes. Recently, the IP pathway has been implicated in the regulation of VSG gene expression. The IP pathway plays a major role in various cellular pathways such as cell growth, differentiation, endocytosis, and apoptosis. However, until recently the role of the IP pathway in regulating VSG gene expression was unknown. Specifically, PIP5Pase was implicated in the regulation of monoallelic expression of VSG genes in BFs. It associates with RAP1, as a part of a larger multiprotein complex at the telomeric ES. Previously, a mechanism of regulation was proposed wherein the regulation of VSG gene expression was regulated by interactions of PIP5Pase and RAP1.

The aim of the current project was to gain further resolution on the developmental regulation of VSG gene expression as well as to understand better the mechanism involved. To carry out this study, a hypothesis was proposed where RAP1 binds to PI(3,4,5)P3 which is a PIP5Pase substrate, and this binding controls RAP1 association with the ES chromatin. To test this, the role of PIP5Pase in the silencing of VSGs in PFs was examined. This involved knocking down of PIP5Pase in PFs, followed by analysis of VSG gene expression. Further, the interaction of RAP1 with PI(3,4,5)P3 and the ES chromatin DNA was studied.

The key findings from this study were:

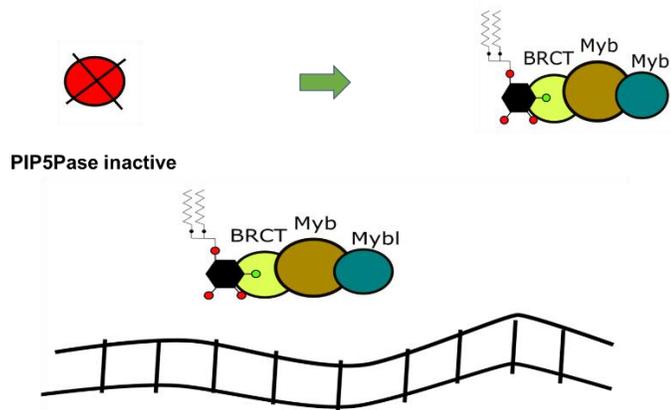
1. PIP5Pase is essential for PFs and is necessary for the silencing of VSG genes in PFs.
2. The N-terminal BRCT domain of *T. brucei* RAP1 binds PI(3,4,5)P3 *in vitro*.
3. RAP1 binds directly to telomeric repeats and 70 bp repeats *in vitro*.
4. Myb and Mybl domains of RAP1 are the DNA binding domains of RAP1.
5. The presence of PI(3,4,5)P3 affects the binding of RAP1 to the ES chromatin DNA whereas the presence of PI(3,4,5)P3 does not affect the binding of either Myb or Mybl domains to the ES chromatin DNA.

Data from previous studies put together with data obtained from the current study has provided a better understanding of the process by which RAP1 and PIP5Pase interact and thereby regulate VSG gene expression. The diagram (Fig. 16 and 17) suggests how the interactions of each RAP1 domain with either the substrate PI(3,4,5)P3 or the ES chromatin DNA affects the transcription of VSGs



**Fig. 16. A silent ES.**

In the case of a silent ES (Fig. 16), the presence of PIP5Pase dephosphorylates PI(3,4,5)P3 to PI(3,4)P2, depriving RAP1 of its substrate. In turn, RAP1 binds to the ES chromatin DNA via the Myb and the Mybl domains. This interaction of RAP1 with the ES disrupts the transcription of the ES by Pol I, rendering the ES and the associated VSG gene inactive.



**Fig. 17. An active ES**

However, in the case of PIP5Pase knockdown or inactivation (Fig. 17), RAP1 binds PI(3,4,5)P3 via the BRCT domain, leading to a possible conformational change that disrupts its interaction with the ES chromatin DNA. This could potentially lead to some change in the ES chromatin DNA, allowing the active transcription of the entire ES including the VSG gene.

Based on the model (Fig. 16 and 17) and data from previous studies, it can be said that the regulation of PIP5Pase may play a crucial role in the regulating the monoallelic expression of VSGs (136, 143). Moreover, PIP5Pase could also be involved in the switching of VSGs (136). PIP5Pase is a part of a 0.9 MDa protein complex and the members of this complex may regulate the activity of PIP5Pase (143). Apart from members of the 0.9 MDa complex, other PIP proteins (found in the plasma membrane) such as PIP5K and PLC or other signalling pathways may also be involved in the regulation of PIP5Pase activity. It would be crucial to understand how these proteins and pathways may play a role in regulating PIP5Pase activity.

## 14. References

1. Jackson Andrew P, Otto Thomas D, Aslett M, Armstrong Stuart D, Bringaud F, Schlacht A, et al. Kinetoplastid Phylogenomics Reveals the Evolutionary Innovations Associated with the Origins of Parasitism. *Current Biology*. 2016;26(2):161-72.
2. Kaufer A, Ellis J, Stark D, Barratt J. The evolution of trypanosomatid taxonomy. *Parasit Vectors*. 2017;10(1):287.
3. Franco JR, Simarro PP, Diarra A, Jannin JG. Epidemiology of human African trypanosomiasis. *Clin Epidemiol*. 2014;6:257-75.
4. Altamura F, Rajesh R, Catta-Preta CMC, Moretti NS, Cestari I. The current drug discovery landscape for trypanosomiasis and leishmaniasis: Challenges and strategies to identify drug targets. *Drug Dev Res*. 2020.
5. Dagnachew S, Bezie M, Terefe G, Abebe G, Barry JD, Goddeeris BM. Comparative clinico-haematological analysis in young Zebu cattle experimentally infected with *Trypanosoma vivax* isolates from tsetse infested and non-tsetse infested areas of Northwest Ethiopia. *Acta Vet Scand*. 2015;57:24.
6. Aksoy S, Caccone A, Galvani AP, Okedi LM. *Glossina fuscipes* populations provide insights for human African trypanosomiasis transmission in Uganda. *Trends Parasitol*. 2013;29(8):394-406.
7. Lehane M, Alfaroukh I, Bucheton B, Camara M, Harris A, Kaba D, et al. Tsetse Control and the Elimination of Gambian Sleeping Sickness. *PLoS neglected tropical diseases*. 2016;10(4):e0004437-e.
8. Rocha G, Martins A, Gama G, Brandao F, Atouguia J. Possible cases of sexual and congenital transmission of sleeping sickness. *Lancet*. 2004;363(9404):247.
9. Martin-Davila P, Fortun J, Lopez-Velez R, Norman F, Montes de Oca M, Zamarron P, et al. Transmission of tropical and geographically restricted infections during solid-organ transplantation. *Clin Microbiol Rev*. 2008;21(1):60-96.
10. Herwaldt BL. Laboratory-acquired parasitic infections from accidental exposures. *Clin Microbiol Rev*. 2001;14(4):659-88, table of contents.
11. Gardiner PR. Recent studies of the biology of *Trypanosoma vivax*. *Adv Parasitol*. 1989;28:229-317.
12. Osorio AL, Madruga CR, Desquesnes M, Soares CO, Ribeiro LR, Costa SC. *Trypanosoma* (*Duttonella*) *vivax*: its biology, epidemiology, pathogenesis, and introduction in the New World--a review. *Mem Inst Oswaldo Cruz*. 2008;103(1):1-13.
13. MacGregor P, Szoor B, Savill NJ, Matthews KR. Trypanosomal immune evasion, chronicity and transmission: an elegant balancing act. *Nat Rev Microbiol*. 2012;10(6):431-8.
14. Vassella E, Reuner B, Yutzy B, Boshart M. Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. *J Cell Sci*. 1997;110 ( Pt 21):2661-71.
15. Rojas F, Silvester E, Young J, Milne R, Tettey M, Houston DR, et al. Oligopeptide Signaling through TbGPR89 Drives Trypanosome Quorum Sensing. *Cell*. 2019;176(1-2):306-17.e16.
16. Sbicego S, Vassella E, Kurath U, Blum B, Roditi I. The use of transgenic *Trypanosoma brucei* to identify compounds inducing the differentiation of bloodstream forms to procyclic forms. *Molecular and biochemical parasitology*. 1999;104(2):311-22.
17. Engstler M, Boshart M. Cold shock and regulation of surface protein trafficking convey sensitization to inducers of stage differentiation in *Trypanosoma brucei*. *Genes Dev*. 2004;18(22):2798-811.
18. Overath P, Czichos J, Haas C. The effect of citrate/cis-aconitate on oxidative metabolism during transformation of *Trypanosoma brucei*. *Eur J Biochem*. 1986;160(1):175-82.

19. Szöör B, Wilson J, McElhinney H, Taberner L, Matthews KR. Protein tyrosine phosphatase TbPTP1: A molecular switch controlling life cycle differentiation in trypanosomes. *J Cell Biol.* 2006;175(2):293-303.
20. Dean S, Marchetti R, Kirk K, Matthews KR. A surface transporter family conveys the trypanosome differentiation signal. *Nature.* 2009;459(7244):213-7.
21. Szöör B, Wilson J, McElhinney H, Taberner L, Matthews KR. Protein tyrosine phosphatase TbPTP1: A molecular switch controlling life cycle differentiation in trypanosomes. *J Cell Biol.* 2006;175(2):293-303.
22. Cestari I, Anupama A, Stuart K. Inositol polyphosphate multikinase regulation of *Trypanosoma brucei* life stage development. *Mol Biol Cell.* 2018;29(9):1137-52.
23. Sharma R, Peacock L, Gluenz E, Gull K, Gibson W, Carrington M. Asymmetric Cell Division as a Route to Reduction in Cell Length and Change in Cell Morphology in Trypanosomes. *Protist.* 2008;159(1):137-51.
24. Roditi I, Lehane MJ. Interactions between trypanosomes and tsetse flies. *Curr Opin Microbiol.* 2008;11(4):345-51.
25. Cestari I, Stuart K. Transcriptional Regulation of Telomeric Expression Sites and Antigenic Variation in Trypanosomes. *Curr Genomics.* 2018;19(2):119-32.
26. Roditi I, Clayton C. An unambiguous nomenclature for the major surface glycoproteins of the procyclic form of *Trypanosoma brucei*. *Molecular and biochemical parasitology.* 1999;103(1):99-100.
27. Creek DJ, Mazet M, Achcar F, Anderson J, Kim DH, Kamour R, et al. Probing the metabolic network in bloodstream-form *Trypanosoma brucei* using untargeted metabolomics with stable isotope labelled glucose. *PLoS Pathog.* 2015;11(3):e1004689.
28. Smith TK, Bringaud F, Nolan DP, Figueiredo LM. Metabolic reprogramming during the *Trypanosoma brucei* life cycle. *F1000Res.* 2017;6:F1000 Faculty Rev-683.
29. Rico E, Rojas F, Mony BM, Szöör B, Macgregor P, Matthews KR. Bloodstream form pre-adaptation to the tsetse fly in *Trypanosoma brucei*. *Front Cell Infect Microbiol.* 2013;3:78.
30. Rico E, Rojas F, Mony BM, Szöör B, Macgregor P, Matthews KR. Bloodstream form pre-adaptation to the tsetse fly in *Trypanosoma brucei*. *Front Cell Infect Microbiol.* 2013;3:78-.
31. Mantilla BS, Marchese L, Casas-Sanchez A, Dyer NA, Ejeh N, Biran M, et al. Proline Metabolism is Essential for *Trypanosoma brucei brucei* Survival in the Tsetse Vector. *PLoS Pathog.* 2017;13(1):e1006158.
32. Nes CR, Singha UK, Liu J, Ganapathy K, Villalta F, Waterman MR, et al. Novel sterol metabolic network of *Trypanosoma brucei* procyclic and bloodstream forms. *Biochem J.* 2012;443(1):267-77.
33. Millerioux Y, Ebikeme C, Biran M, Morand P, Bouyssou G, Vincent IM, et al. The threonine degradation pathway of the *Trypanosoma brucei* procyclic form: the main carbon source for lipid biosynthesis is under metabolic control. *Mol Microbiol.* 2013;90(1):114-29.
34. Capewell P, Cren-Travaillé C, Marchesi F, Johnston P, Clucas C, Benson RA, et al. The skin is a significant but overlooked anatomical reservoir for vector-borne African trypanosomes. *eLife.* 2016;5:e17716.
35. Stich A, Abel PM, Krishna S. Human African trypanosomiasis. *BMJ (Clinical research ed).* 2002;325(7357):203-6.
36. Checchi F, Filipe JAN, Barrett MP, Chandramohan D. The natural progression of Gambiense sleeping sickness: what is the evidence? *PLoS neglected tropical diseases.* 2008;2(12):e303-e.
37. Kennedy PGE, Rodgers J. Clinical and Neuropathogenetic Aspects of Human African Trypanosomiasis. *Front Immunol.* 2019;10:39.
38. Checchi F, Chappuis F, Karunakara U, Priotto G, Chandramohan D. Accuracy of five algorithms to diagnose gambiense human African trypanosomiasis. *PLoS Negl Trop Dis.* 2011;5(7):e1233.

39. Bonnet J, Boudot C, Courtioux B. Overview of the Diagnostic Methods Used in the Field for Human African Trypanosomiasis: What Could Change in the Next Years? *BioMed research international*. 2015;2015:583262-.
40. Buscher P, Mumba Ngoyi D, Kabore J, Lejon V, Robays J, Jamonneau V, et al. Improved Models of Mini Anion Exchange Centrifugation Technique (mAECT) and Modified Single Centrifugation (MSC) for sleeping sickness diagnosis and staging. *PLoS Negl Trop Dis*. 2009;3(11):e471.
41. Camara O, Camara M, Lejon V, Ilboudo H, Sakande H, Leno M, et al. Immune trypanolysis test with blood spotted on filter paper for epidemiological surveillance of sleeping sickness. *Tropical Medicine & International Health*. 2014;19(7):828-31.
42. Lejon V, Büscher P, Magnus E, Moons A, Wouters I, Van Meirvenne N. A semi-quantitative ELISA for detection of *Trypanosoma brucei gambiense* specific antibodies in serum and cerebrospinal fluid of sleeping sickness patients. *Acta Tropica*. 1998;69(2):151-64.
43. Lejon V, Legros D, Richer M, Ruiz JA, Jamonneau V, Truc P, et al. IgM quantification in the cerebrospinal fluid of sleeping sickness patients by a latex card agglutination test. *Tropical Medicine & International Health*. 2002;7(8):685-92.
44. Lejon V, Lardon J, Kenis G, Pinoges L, Legros D, Bisser S, et al. Interleukin (IL)-6, IL-8 and IL-10 in serum and CSF of *Trypanosoma brucei gambiense* sleeping sickness patients before and after treatment. *Trans R Soc Trop Med Hyg*. 2002;96(3):329-33.
45. Tiberti N, Hainard A, Lejon V, Courtioux B, Matovu E, Enyaru JC, et al. Cerebrospinal Fluid Neopterin as Marker of the Meningo-Encephalitic Stage of *Trypanosoma brucei gambiense* Sleeping Sickness. *PLOS ONE*. 2012;7(7):e40909.
46. Deeks ED. Fexinidazole: First Global Approval. *Drugs*. 2019;79(2):215-20.
47. Jacobs RT, Nare B, Wring SA, Orr MD, Chen D, Sligar JM, et al. SCYX-7158, an orally-active benzoxaborole for the treatment of stage 2 human African trypanosomiasis. *PLoS Negl Trop Dis*. 2011;5(6):e1151.
48. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, et al. The genome of the African trypanosome *Trypanosoma brucei*. *Science*. 2005;309(5733):416-22.
49. Martinez-Calvillo S, Yan S, Nguyen D, Fox M, Stuart K, Myler PJ. Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. *Mol Cell*. 2003;11(5):1291-9.
50. El-Sayed NMA, Ghedin E, Song J, MacLeod A, Bringaud F, Larkin C, et al. The sequence and analysis of *Trypanosoma brucei* chromosome II. *Nucleic acids research*. 2003;31(16):4856-63.
51. Wickstead B, Ersfeld K, Gull K. The small chromosomes of *Trypanosoma brucei* involved in antigenic variation are constructed around repetitive palindromes. *Genome Res*. 2004;14(6):1014-24.
52. Berriman M, Hall N, Sheader K, Bringaud F, Tiwari B, Isobe T, et al. The architecture of variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. *Molecular and biochemical parasitology*. 2002;122(2):131-40.
53. Pedram M, Donelson JE. The anatomy and transcription of a monocistronic expression site for a metacyclic variant surface glycoprotein gene in *Trypanosoma brucei*. *J Biol Chem*. 1999;274(24):16876-83.
54. Hertz-Fowler C, Figueiredo LM, Quail MA, Becker M, Jackson A, Bason N, et al. Telomeric expression sites are highly conserved in *Trypanosoma brucei*. *PLoS one*. 2008;3(10):e3527-e.
55. Zomerdijk JC, Kieft R, Shiels PG, Borst P. Alpha-amanitin-resistant transcription units in trypanosomes: a comparison of promoter sequences for a VSG gene expression site and for the ribosomal RNA genes. *Nucleic acids research*. 1991;19(19):5153-8.
56. Redpath MB, Windle H, Nolan D, Pays E, Voorheis HP, Carrington M. ESAG11, a new VSG expression site-associated gene from *Trypanosoma brucei*. *Molecular and biochemical parasitology*. 2000;111(1):223-8.

57. Wedel C, Förstner KU, Derr R, Siegel TN. GT-rich promoters can drive RNA pol II transcription and deposition of H2A.Z in African trypanosomes. *The EMBO Journal*. 2017;36(17):2581-94.
58. Clayton C. Regulation of gene expression in trypanosomatids: living with polycistronic transcription. 2019;9(6):190072.
59. Kolev NG, Franklin JB, Carmi S, Shi H, Michaeli S, Tschudi C. The Transcriptome of the Human Pathogen *Trypanosoma brucei* at Single-Nucleotide Resolution. *PLOS Pathogens*. 2010;6(9):e1001090.
60. Nilsson D, Gunasekera K, Mani J, Osteras M, Farinelli L, Baerlocher L, et al. Spliced Leader Trapping Reveals Widespread Alternative Splicing Patterns in the Highly Dynamic Transcriptome of *Trypanosoma brucei*. *PLOS Pathogens*. 2010;6(8):e1001037.
61. Siegel TN, Hekstra DR, Kemp LE, Figueiredo LM, Lowell JE, Fenyo D, et al. Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev*. 2009;23(9):1063-76.
62. Cross M, Kieft R, Sabatini R, Dirks-Mulder A, Chaves I, Borst P. J-binding protein increases the level and retention of the unusual base J in trypanosome DNA. *Mol Microbiol*. 2002;46(1):37-47.
63. van Leeuwen F, Taylor MC, Mondragon A, Moreau H, Gibson W, Kieft R, et al. beta-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. *Proc Natl Acad Sci U S A*. 1998;95(5):2366-71.
64. Gommers-Ampt J, Lutgerink J, Borst P. A novel DNA nucleotide in *Trypanosoma brucei* only present in the mammalian phase of the life-cycle. *Nucleic Acids Res*. 1991;19(8):1745-51.
65. van Luenen Henri GAM, Farris C, Jan S, Genest P-A, Tripathi P, Velds A, et al. Glucosylated Hydroxymethyluracil, DNA Base J, Prevents Transcriptional Readthrough in *Leishmania*. *Cell*. 2012;150(5):909-21.
66. Schulz D, Zaringhalam M, Papavasiliou FN, Kim HS. Base J and H3.V Regulate Transcriptional Termination in *Trypanosoma brucei*. *PLoS Genet*. 2016;12(1):e1005762.
67. Maree JP, Patterton HG. The epigenome of *Trypanosoma brucei*: A regulatory interface to an unconventional transcriptional machine. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. 2014;1839(9):743-50.
68. Günzl A, Bruderer T, Laufer G, Schimanski B, Tu LC, Chung HM, et al. RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. *Eukaryot Cell*. 2003;2(3):542-51.
69. Kooter JM, Borst P. Alpha-amanitin-insensitive transcription of variant surface glycoprotein genes provides further evidence for discontinuous transcription in trypanosomes. *Nucleic Acids Res*. 1984;12(24):9457-72.
70. Günzl A, Bruderer T, Laufer G, Schimanski B, Tu L-C, Chung H-M, et al. RNA Polymerase I Transcribes Procyclin Genes and Variant Surface Glycoprotein Gene Expression Sites in *Trypanosoma brucei*. *Eukaryotic Cell*. 2003;2(3):542.
71. Rudenko G, Bishop D, Gottesdiener K, Van der Ploeg LH. Alpha-amanitin resistant transcription of protein coding genes in insect and bloodstream form *Trypanosoma brucei*. *Embo j*. 1989;8(13):4259-63.
72. Janz L, Clayton C. The PARP and rRNA promoters of *Trypanosoma brucei* are composed of dissimilar sequence elements that are functionally interchangeable. *Mol Cell Biol*. 1994;14(9):5804-11.
73. Chaves I, Zomerdijk J, Dirks-Mulder A, Dirks RW, Raap AK, Borst P. Subnuclear localization of the active variant surface glycoprotein gene expression site in *Trypanosoma brucei*. *Proceedings of the National Academy of Sciences*. 1998;95(21):12328.
74. Navarro M, Gull K. A pol I transcriptional body associated with VSG mono-allelic expression in *Trypanosoma brucei*. *Nature*. 2001;414(6865):759-63.

75. Nguyen TN, Nguyen BN, Lee JH, Panigrahi AK, Gunzl A. Characterization of a novel class I transcription factor A (CITFA) subunit that is indispensable for transcription by the multifunctional RNA polymerase I of *Trypanosoma brucei*. *Eukaryot Cell*. 2012;11(12):1573-81.
76. Kirkham JK, Park SH, Nguyen TN, Lee JH, Gunzl A. Dynein Light Chain LC8 Is Required for RNA Polymerase I-Mediated Transcription in *Trypanosoma brucei*, Facilitating Assembly and Promoter Binding of Class I Transcription Factor A. *Mol Cell Biol*. 2016;36(1):95-107.
77. Nguyen TN, Nguyen BN, Lee JH, Panigrahi AK, Günzl A. Characterization of a novel class I transcription factor A (CITFA) subunit that is indispensable for transcription by the multifunctional RNA polymerase I of *Trypanosoma brucei*. *Eukaryotic cell*. 2012;11(12):1573-81.
78. Brandenburg J, Schimanski B, Nogoceke E, Nguyen TN, Padovan JC, Chait BT, et al. Multifunctional class I transcription in *Trypanosoma brucei* depends on a novel protein complex. *Embo j*. 2007;26(23):4856-66.
79. Lei Q, Li C, Zuo Z, Huang C, Cheng H, Zhou R. Evolutionary Insights into RNA trans-Splicing in Vertebrates. *Genome biology and evolution*. 2016;8(3):562-77.
80. Preusser C, Jae N, Bindereif A. mRNA splicing in trypanosomes. *Int J Med Microbiol*. 2012;302(4-5):221-4.
81. Mair G, Ullu E, Tschudi C. Cotranscriptional cap 4 formation on the *Trypanosoma brucei* spliced leader RNA. *J Biol Chem*. 2000;275(37):28994-9.
82. Serpeloni M, Vidal NM, Goldenberg S, Avila AR, Hoffmann FG. Comparative genomics of proteins involved in RNA nucleocytoplasmic export. *BMC evolutionary biology*. 2011;11:7-.
83. Dostalova A, Kaser S, Cristodero M, Schimanski B. The nuclear mRNA export receptor Mex67-Mtr2 of *Trypanosoma brucei* contains a unique and essential zinc finger motif. *Mol Microbiol*. 2013;88(4):728-39.
84. Li H, Tschudi C. Novel and essential subunits in the 300-kilodalton nuclear cap binding complex of *Trypanosoma brucei*. *Mol Cell Biol*. 2005;25(6):2216-26.
85. Kramer S, Bannerman-Chukualim B, Ellis L, Boulden EA, Kelly S, Field MC, et al. Differential localization of the two *T. brucei* poly(A) binding proteins to the nucleus and RNP granules suggests binding to distinct mRNA pools. *PLoS One*. 2013;8(1):e54004.
86. Zoltner M, Krienitz N, Field MC, Kramer S. Comparative proteomics of the two *T. brucei* PABPs suggests that PABP2 controls bulk mRNA. *PLoS Negl Trop Dis*. 2018;12(7):e0006679.
87. Otsuka H, Fukao A, Funakami Y, Duncan KE, Fujiwara T. Emerging Evidence of Translational Control by AU-Rich Element-Binding Proteins. 2019;10(332).
88. Barreau C, Paillard L, Osborne HB. AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Research*. 2005;33(22):7138-50.
89. Rao SJ, Chatterjee S, Pal JK. Untranslated regions of mRNA and their role in regulation of gene expression in protozoan parasites. *Journal of Biosciences*. 2017;42(1):189-207.
90. Erben E, Chakraborty C, Clayton C. The CAF1-NOT complex of trypanosomes. 2014;4(299).
91. Jonas S, Christie M, Peter D, Bhandari D, Loh B, Huntzinger E, et al. An asymmetric PAN3 dimer recruits a single PAN2 exonuclease to mediate mRNA deadenylation and decay. *Nat Struct Mol Biol*. 2014;21(7):599-608.
92. Utter CJ, Garcia SA, Milone J, Bellofatto V. PolyA-specific ribonuclease (PARN-1) function in stage-specific mRNA turnover in *Trypanosoma brucei*. *Eukaryotic cell*. 2011;10(9):1230-40.
93. Kramer S. The ApaH-like phosphatase TbALPH1 is the major mRNA decapping enzyme of trypanosomes. *PLoS Pathog*. 2017;13(6):e1006456.
94. Simpson L. The Mitochondrial Genome of Kinetoplastid Protozoa: Genomic Organization, Transcription, Replication, and Evolution. *Annual Review of Microbiology*. 1987;41(1):363-80.
95. Aphasizhev R, Aphasizheva I. Uridine insertion/deletion editing in trypanosomes: a playground for RNA-guided information transfer. *Wiley Interdiscip Rev RNA*. 2011;2(5):669-85.

96. Aphasizheva I, Alfonzo J, Carnes J, Cestari I, Cruz-Reyes J, Göringer HU, et al. Lexis and Grammar of Mitochondrial RNA Processing in Trypanosomes. *Trends in Parasitology*. 2020;36(4):337-55.
97. Benne R, Van den Burg J, Brakenhoff JP, Sloof P, Van Boom JH, Tromp MC. Major transcript of the frameshifted *coxII* gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell*. 1986;46(6):819-26.
98. Panigrahi AK, Gygi SP, Ernst NL, Igo RP, Jr., Palazzo SS, Schnauffer A, et al. Association of two novel proteins, TbMP52 and TbMP48, with the *Trypanosoma brucei* RNA editing complex. *Mol Cell Biol*. 2001;21(2):380-9.
99. Seiwert SD, Heidmann S, Stuart K. Direct Visualization of Uridylate Deletion In Vitro Suggests a Mechanism for Kinetoplastid RNA Editing. *Cell*. 1996;84(6):831-41.
100. Seiwert SD, Stuart K. RNA editing: transfer of genetic information from gRNA to precursor mRNA in vitro. *Science*. 1994;266(5182):114.
101. Cross GAM, Kim H-S, Wickstead B. Capturing the variant surface glycoprotein repertoire (the VSGnome) of *Trypanosoma brucei* Lister 427. *Molecular and biochemical parasitology*. 2014;195(1):59-73.
102. McCulloch R, Field MC. Quantitative sequencing confirms VSG diversity as central to immune evasion by *Trypanosoma brucei*. *Trends Parasitol*. 2015;31(8):346-9.
103. Müller LSM, Cosentino RO, Förstner KU, Guizetti J, Wedel C, Kaplan N, et al. Genome organization and DNA accessibility control antigenic variation in trypanosomes. *Nature*. 2018;563(7729):121-5.
104. Schwede A, Macleod OJS, MacGregor P, Carrington M. How Does the VSG Coat of Bloodstream Form African Trypanosomes Interact with External Proteins? *PLOS Pathogens*. 2016;11(12):e1005259.
105. Hutchinson OC, Smith W, Jones NG, Chattopadhyay A, Welburn SC, Carrington M. VSG structure: similar N-terminal domains can form functional VSGs with different types of C-terminal domain. *Molecular and biochemical parasitology*. 2003;130(2):127-31.
106. Hutchinson OC, Smith W, Jones NG, Chattopadhyay A, Welburn SC, Carrington M. VSG structure: similar N-terminal domains can form functional VSGs with different types of C-terminal domain. *Molecular and biochemical parasitology*. 2003;130(2):127-31.
107. Schwede A, Jones N, Engstler M, Carrington M. The VSG C-terminal domain is inaccessible to antibodies on live trypanosomes. *Molecular and biochemical parasitology*. 2011;175(2):201-4.
108. Paulick MG, Bertozzi CR. The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins. *Biochemistry*. 2008;47(27):6991-7000.
109. Myler PJ, Allison J, Agabian N, Stuart K. Antigenic variation in African trypanosomes by gene replacement or activation of alternate telomeres. *Cell*. 1984;39(1):203-11.
110. Vink C, Rudenko G, Seifert HS. Microbial antigenic variation mediated by homologous DNA recombination. *FEMS Microbiol Rev*. 2012;36(5):917-48.
111. Pays E, Van Assel S, Laurent M, Darville M, Vervoort T, Van Meirvenne N, et al. Gene conversion as a mechanism for antigenic variation in trypanosomes. *Cell*. 1983;34(2):371-81.
112. Roth C, Bringaud F, Layden RE, Baltz T, Eisen H. Active late-appearing variable surface antigen genes in *Trypanosoma equiperdum* are constructed entirely from pseudogenes. *Proc Natl Acad Sci U S A*. 1989;86(23):9375-9.
113. Rudenko G, McCulloch R, Dirks-Mulder A, Borst P. Telomere exchange can be an important mechanism of variant surface glycoprotein gene switching in *Trypanosoma brucei*. *Molecular and biochemical parasitology*. 1996;80(1):65-75.
114. Landeira D, Bart J-M, Van Tyne D, Navarro M. Cohesin regulates VSG monoallelic expression in trypanosomes. *Journal of Cell Biology*. 2009;186(2):243-54.
115. Budzak J, Kerry LE, Aristodemou A, Hall BS, Witmer K, Kushwaha M, et al. Dynamic colocalization of 2 simultaneously active &em&gt;VSG&lt;/em&gt; expression sites within a single expression-site

- body in *Trypanosoma brucei*. Proceedings of the National Academy of Sciences. 2019;116(33):16561.
116. Kassem A, Pays E, Vanhamme L. Transcription is initiated on silent variant surface glycoprotein expression sites despite monoallelic expression in *Trypanosoma brucei*. Proceedings of the National Academy of Sciences. 2014;111(24):8943.
  117. Mehta GD, Rizvi SMA, Ghosh SK. Cohesin: A guardian of genome integrity. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. 2012;1823(8):1324-42.
  118. Nguyen TN, Muller LS, Park SH, Siegel TN, Gunzl A. Promoter occupancy of the basal class I transcription factor A differs strongly between active and silent VSG expression sites in *Trypanosoma brucei*. *Nucleic Acids Res*. 2014;42(5):3164-76.
  119. Ye K, Zhang X, Ni J, Liao S, Tu X. Identification of enzymes involved in SUMOylation in *Trypanosoma brucei*. *Scientific Reports*. 2015;5(1):10097.
  120. Yang Y, He Y, Wang X, liang Z, He G, Zhang P, et al. Protein SUMOylation modification and its associations with disease. 2017;7(10):170167.
  121. López-Farfán D, Bart J-M, Rojas-Barros DI, Navarro M. SUMOylation by the E3 Ligase TbSIZ1/PIAS1 Positively Regulates VSG Expression in *Trypanosoma brucei*. *PLOS Pathogens*. 2014;10(12):e1004545.
  122. Vary JC, Jr., Gangaraju VK, Qin J, Landel CC, Kooperberg C, Bartholomew B, et al. Yeast Isw1p forms two separable complexes in vivo. *Molecular and cellular biology*. 2003;23(1):80-91.
  123. Stanne TM, Kushwaha M, Wand M, Taylor JE, Rudenko G. TbISWI regulates multiple polymerase I (Pol I)-transcribed loci and is present at Pol II transcription boundaries in *Trypanosoma brucei*. *Eukaryot Cell*. 2011;10(7):964-76.
  124. Stanne TM, Rudenko G. Active VSG expression sites in *Trypanosoma brucei* are depleted of nucleosomes. *Eukaryot Cell*. 2010;9(1):136-47.
  125. Hughes K, Wand M, Foulston L, Young R, Harley K, Terry S, et al. A novel ISWI is involved in VSG expression site downregulation in African trypanosomes. *The EMBO journal*. 2007;26(9):2400-10.
  126. Stanne TM, Narayanan MS, Ridewood S, Ling A, Witmer K, Kushwaha M, et al. Identification of the ISWI Chromatin Remodeling Complex of the Early Branching Eukaryote *Trypanosoma brucei*. *J Biol Chem*. 2015;290(45):26954-67.
  127. Glover L, Hutchinson S, Alsford S, Horn D. VEX1 controls the allelic exclusion required for antigenic variation in trypanosomes. *Proc Natl Acad Sci U S A*. 2016;113(26):7225-30.
  128. Faria J, Glover L, Hutchinson S, Boehm C, Field MC, Horn D. Monoallelic expression and epigenetic inheritance sustained by a *Trypanosoma brucei* variant surface glycoprotein exclusion complex. *Nat Commun*. 2019;10(1):3023.
  129. Navarro M, Cross GA, Wirtz E. *Trypanosoma brucei* variant surface glycoprotein regulation involves coupled activation/inactivation and chromatin remodeling of expression sites. *Embo j*. 1999;18(8):2265-72.
  130. Figueiredo LM, Cross GA. Nucleosomes are depleted at the VSG expression site transcribed by RNA polymerase I in African trypanosomes. *Eukaryot Cell*. 2010;9(1):148-54.
  131. Denninger V, Fullbrook A, Bessat M, Ersfeld K, Rudenko G. The FACT subunit TbSpt16 is involved in cell cycle specific control of VSG expression sites in *Trypanosoma brucei*. *Molecular microbiology*. 2010;78(2):459-74.
  132. Denninger V, Rudenko G. FACT plays a major role in histone dynamics affecting VSG expression site control in *Trypanosoma brucei*. *Mol Microbiol*. 2014;94(4):945-62.
  133. Figueiredo LM, Janzen CJ, Cross GAM. A Histone Methyltransferase Modulates Antigenic Variation in African Trypanosomes. *PLOS Biology*. 2008;6(7):e161.
  134. McCulloch R, Barry JD. A role for RAD51 and homologous recombination in *Trypanosoma brucei* antigenic variation. *Genes Dev*. 1999;13(21):2875-88.

135. Jehi SE, Li X, Sandhu R, Ye F, Benmerzouga I, Zhang M, et al. Suppression of subtelomeric VSG switching by *Trypanosoma brucei* TRF requires its TTAGGG repeat-binding activity. *Nucleic acids research*. 2014;42(20):12899-911.
136. Cestari I, Stuart K. Inositol phosphate pathway controls transcription of telomeric expression sites in trypanosomes. *Proceedings of the National Academy of Sciences*. 2015;112(21):E2803.
137. Boothroyd CE, Dreesen O, Leonova T, Ly KI, Figueiredo LM, Cross GAM, et al. A yeast-endonuclease-generated DNA break induces antigenic switching in *Trypanosoma brucei*. *Nature*. 2009;459(7244):278-81.
138. Glover L, Alsford S, Horn D. DNA break site at fragile subtelomeres determines probability and mechanism of antigenic variation in African trypanosomes. *PLoS Pathog*. 2013;9(3):e1003260.
139. Saha A, Nanavaty VP, Li B. Telomere and Subtelomere R-loops and Antigenic Variation in Trypanosomes. *Journal of Molecular Biology*. 2020;432(15):4167-85.
140. Gottschling DE, Aparicio OM, Billington BL, Zakian VA. Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell*. 1990;63(4):751-62.
141. Horn D, Cross GA. A developmentally regulated position effect at a telomeric locus in *Trypanosoma brucei*. *Cell*. 1995;83(4):555-61.
142. Yang X, Figueiredo LM, Espinal A, Okubo E, Li B. RAP1 is essential for silencing telomeric variant surface glycoprotein genes in *Trypanosoma brucei*. *Cell*. 2009;137(1):99-109.
143. Cestari I, McLeland-Wieser H, Stuart K. Nuclear Phosphatidylinositol 5-Phosphatase Is Essential for Allelic Exclusion of Variant Surface Glycoprotein Genes in Trypanosomes. *Molecular and Cellular Biology*. 2019;39(3):e00395-18.
144. Williams RS, Lee MS, Hau DD, Glover JNM. Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1. *Nature Structural & Molecular Biology*. 2004;11(6):519-25.
145. Klempnauer KH, Sippel AE. The highly conserved amino-terminal region of the protein encoded by the *v-myb* oncogene functions as a DNA-binding domain. *The EMBO journal*. 1987;6(9):2719-25.
146. Pandya UM, Sandhu R, Li B. Silencing subtelomeric VSGs by *Trypanosoma brucei* RAP1 at the insect stage involves chromatin structure changes. *Nucleic acids research*. 2013;41(16):7673-82.
147. Afrin M, Kishmiri H, Sandhu R, Rabbani MAG, Li B. *Trypanosoma brucei* RAP1 Has Essential Functional Domains That Are Required for Different Protein Interactions. *mSphere*. 2020;5(1):e00027-20.
148. Michell RH. Inositol derivatives: evolution and functions. *Nature Reviews Molecular Cell Biology*. 2008;9(2):151-61.
149. Simonsen A, Wurmser AE, Emr SD, Stenmark H. The role of phosphoinositides in membrane transport. *Current Opinion in Cell Biology*. 2001;13(4):485-92.
150. Cestari I. Phosphoinositide signaling and regulation in *Trypanosoma brucei*: Specialized functions in a protozoan pathogen. *PLOS Pathogens*. 2020;16(1):e1008167.
151. Gilden JK, Umaer K, Kruzel EK, Hecht O, Correa RO, Mansfield JM, et al. The role of the PI(3,5)P(2) kinase TbFab1 in endo/lysosomal trafficking in *Trypanosoma brucei*. *Molecular and biochemical parasitology*. 2017;214:52-61.
152. Cestari I, Haas P, Moretti NS, Schenkman S, Stuart K. Chemogenetic Characterization of Inositol Phosphate Metabolic Pathway Reveals Druggable Enzymes for Targeting Kinetoplastid Parasites. *Cell Chem Biol*. 2016;23(5):608-17.
153. Hall BS, Gabernet-Castello C, Voak AA, Goulding DA, Natesan SKA, Field MCJJoBC. TbVps34, the Trypanosome Orthologue of Vps34, Is Required for Golgi Complex Segregation\*. 2006;281:27600 - 12.
154. Rodgers MJ, Albanesi JP, Phillips MA. Phosphatidylinositol 4-kinase III-beta is required for Golgi maintenance and cytokinesis in *Trypanosoma brucei*. *Eukaryotic cell*. 2007;6(7):1108-18.

155. Huang G, Bartlett PJ, Thomas AP, Moreno SNJ, Docampo R. Acidocalcisomes of *Trypanosoma brucei* have an inositol 1,4,5-trisphosphate receptor that is required for growth and infectivity. *Proc Natl Acad Sci U S A*. 2013;110(5):1887-92.
156. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the  $2^{-\Delta\Delta CT}$  Method. *Methods*. 2001;25(4):402-8.
157. Kolev NG, Ramey-Butler K, Cross GAM, Ullu E, Tschudi C. Developmental progression to infectivity in *Trypanosoma brucei* triggered by an RNA-binding protein. *Science*. 2012;338(6112):1352-3.
158. Kolev NG, Ullu E, Tschudi C. The emerging role of RNA-binding proteins in the life cycle of *Trypanosoma brucei*. *Cell Microbiol*. 2014;16(4):482-9.
159. Ramey-Butler K, Ullu E, Kolev NG, Tschudi C. Synchronous expression of individual metacyclic variant surface glycoprotein genes in *Trypanosoma brucei*. *Molecular and biochemical parasitology*. 2015;200(1-2):1-4.
160. Savage AF, Kolev NG, Franklin JB, Vigneron A, Aksoy S, Tschudi C. Transcriptome Profiling of *Trypanosoma brucei* Development in the Tsetse Fly Vector *Glossina morsitans*. *PLOS ONE*. 2016;11(12):e0168877.
161. Urwyler S, Vassella E, Van Den Abbeele J, Renggli CK, Blundell P, Barry JD, et al. Expression of procyclin mRNAs during cyclical transmission of *Trypanosoma brucei*. *PLoS pathogens*. 2005;1(3):e22-e.
162. Urwyler S, Studer E, Renggli CK, Roditi I. A family of stage-specific alanine-rich proteins on the surface of epimastigote forms of *Trypanosoma brucei*. 2007;63(1):218-28.
163. Güther MLS, Beattie K, Lamont DJ, James J, Prescott AR, Ferguson MAJ. Fate of Glycosylphosphatidylinositol (GPI)-Less Procyclin and Characterization of Sialylated Non-GPI-Anchored Surface Coat Molecules of Procyclic-Form *Trypanosoma brucei*. 2009;8(9):1407-17.
164. Hovel-Miner G, Mugnier MR, Goldwater B, Cross GAM, Papavasiliou FN. A Conserved DNA Repeat Promotes Selection of a Diverse Repertoire of *Trypanosoma brucei* Surface Antigens from the Genomic Archive. *PLoS Genet*. 2016;12(5):e1005994-e.
165. Au - Cestari I. Identification of Inositol Phosphate or Phosphoinositide Interacting Proteins by Affinity Chromatography Coupled to Western Blot or Mass Spectrometry. *JoVE*. 2019(149):e59865.
166. Williams RS, Lee MS, Hau DD, Glover JN. Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1. *Nat Struct Mol Biol*. 2004;11(6):519-25.
167. Klempnauer KH, Sippel AE. The highly conserved amino-terminal region of the protein encoded by the v-myb oncogene functions as a DNA-binding domain. *The EMBO journal*. 1987;6(9):2719-25.
168. Rodrigues JA, Acosta-Serrano A, Aebi M, Ferguson MAJ, Routier FH, Schiller I, et al. Parasite Glycobiology: A Bittersweet Symphony. *PLOS Pathogens*. 2015;11(11):e1005169.
169. McDermott SM, Stuart K. The essential functions of KREPB4 are developmentally distinct and required for endonuclease association with editosomes. *RNA*. 2017;23(11):1672-84.
170. Figueiredo LM, Janzen CJ, Cross GAM. A histone methyltransferase modulates antigenic variation in African trypanosomes. *PLoS biology*. 2008;6(7):e161-e.
171. Navarro M, Cross GA, Wirtz E. *Trypanosoma brucei* variant surface glycoprotein regulation involves coupled activation/inactivation and chromatin remodeling of expression sites. *The EMBO journal*. 1999;18(8):2265-72.
172. Parua PK, Booth GT, Sansó M, Benjamin B, Tanny JC, Lis JT, et al. A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II. *Nature*. 2018;558(7710):460-4.
173. Kyrion G, Liu K, Liu C, Lustig AJ. RAP1 and telomere structure regulate telomere position effects in *Saccharomyces cerevisiae*. *Genes Dev*. 1993;7(7a):1146-59.
174. Zhou Q, Lee KJ, Kurasawa Y, Hu H, An T, Li Z. Faithful chromosome segregation in *Trypanosoma brucei* requires a cohort of divergent spindle-associated proteins with distinct functions. *Nucleic acids research*. 2018;46(16):8216-31.

175. Afrin M, Kishmiri H, Sandhu R, Rabbani MAG, Li B. [Trypanosoma brucei](#); RAP1 Has Essential Functional Domains That Are Required for Different Protein Interactions. *mSphere*. 2020;5(1):e00027-20.