5-27-2015

# Lessons from Trypanosome TFIIH: Discovery of the Essential Roles of CRK9 in Gene Expression and of the Divergent Helicase Paralog XPB-R in Nucleotide Excision Repair

Nitika Badjatia
*University of Connecticut - Storrs*, nitikabadjatia@gmail.com

**Lessons from Trypanosome TFIIH:**

**Discovery of the Essential Roles of CRK9 in Gene Expression and of the Divergent Helicase Paralog XPB-R in Nucleotide Excision Repair**

**Nitika Badjatia**
**University of Connecticut, [2015]**

Eukaryotic TFIIH consists of a core of seven subunits, including the DNA helicase *Xeroderma Pigmentosum* B (XPB), and a cyclin-dependent kinase (CDK)-activating complex (CAK) that contains CDK7. XPB is crucial for DNA unwinding during transcription and nucleotide excision repair (NER) while the CAK complex phosphorylates RNA polymerase II (RNAPII) carboxy terminal domain (CTD), enabling the enzyme's promoter clearance.

*Trypanosoma brucei, Trypanosoma cruzi* and *Leishmania* spp. are lethal human parasites, belonging to the early-diverged order Kinetoplastida. They transcribe genes polycistronically and require spliced leader (SL) *trans* splicing for the maturation of mRNAs from polycistronic precursors. SL RNA gene (*SLRNA*) transcription by RNAPII depends on trypanosome TFIIH complex which contains orthologs of all core subunits but lacks the CAK complex. Despite this, the trypanosome CTD is phosphorylated and essential for *SLRNA* transcription. Gene silencing of CDC2-related kinase 9 (CRK9) revealed its importance for parasite survival and CTD phosphorylation. Interestingly, this loss of phosphorylation did not cause a specific RNAPII transcription defect. Instead, *CRK9* silencing blocked *trans* splicing and caused hypomethylation of SL RNA's extensively modified cap. Sedimentation analysis of tandem affinity purified CRK9 revealed a putative tripartite complex including a novel L-type cyclin (CYC12), and a CRK9-associated protein (CRK9AP). Silencing *CYC12* or *CRK9AP* recapitulated the defects observed upon CRK9 silencing, confirming that they functionally partner with CRK9 *in vivo*. CRK9AP depletion caused a rapid co-loss of CRK9 and CYC12, suggesting its role in complex assembly. This project identified CRK9 as crucial for parasite-

specific gene expression. As a CDK, CRK9 is a promising drug target against trypanosomes which was validated in a mouse model.

Another feature related to trypanosome TFIIH is the presence of two divergent paralogs of XPB in kinetoplastid genomes while only the larger XPB paralog consistently co-purified with TFIIH. Gene knockout of the second paralog, termed XPB-R showed that XPB-R is specialized in NER but dispensable for transcription. While XPB-R does not assemble into a TFIIH complex, reciprocal co-immunoprecipitations revealed an interaction with the p52 ortholog, a TFIIH component and known regulator of XPB in other systems, indicating that trypanosomes possess a TFIIH whose function is restricted to transcription and a XPB-R/p52 repair complex.

# Lessons from Trypanosome TFIIH:

# Discovery of the Essential Roles of CRK9 in Gene Expression and of the Divergent Helicase Paralog XPB-R in Nucleotide Excision Repair

**Nitika Badjatia**

**B.Sc., Devi Ahilya University, 2006**

**M.Sc., Devi Ahilya University, 2008**

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

[2015]

**APPROVAL PAGE**

Doctor of Philosophy Dissertation

**Lessons from trypanosome TFIIH:**

**Discovery of the Essential Roles of CRK9 in Gene Expression and**

**of the Divergent Helicase Paralog XPB-R in Nucleotide Excision Repair**

Presented by

Nitika Badjatia, B.Sc., M.Sc.

Major Advisor _____
Arthur Günzl, Ph.D.

Associate Advisor _____
Asis Das, Ph.D.

Associate Advisor _____
Blanka Rogina, Ph.D.

Associate Advisor _____
Bruce Mayer, Ph.D.

Associate Advisor _____
Gordon Carmichael

University of Connecticut
[2015]

# Table of Contents

# Publications and Contributions to the Thesis Projects

1. **Chapter II:** In this study, I functionally characterized the smaller *XPB* (*XPB-R*) paralog in *Trypanosoma brucei*. I performed all the experiments described in this study except the generation of phylogenetic trees with *XPB* sequences from 33 species. Dr. Ju Huck Lee generated the clonal cell line that exclusively expressed C-terminally PTP- tagged *XPB-R* and assisted me with *in vitro* transcription analysis of either mock or XPB-R depleted extracts. This work was published in *Badjatia N, Nguyen TN, Lee JH, Günzl A. Trypanosoma brucei harbours a divergent XPB helicase paralogue that is specialized in nucleotide excision repair and conserved among kinetoplastid organisms. Mol Microbiol* ***90 (2013)****, 1293-1308.*

2. **Chapter III:** To identify trypanosome cyclin dependent kinase that phosphorylates the carboxy terminal domain of trypanosome RPB1, Dr. Lee generated clonal RNAi cell lines for expression silencing of *cdc2-related kinase (CRK)1*, *CRK3*, *CRK7* and *CRK9* and performed initial RNA analyses after RNAi experiments with each of these cell lines. Thereafter, I confirmed the phenotypes observed upon silencing of CRK9 in independent experiments and performed qRT-PCR analysis. I also analyzed nascent RNA, methylation of cap4 and $m^7G$ cap in *CRK9* silenced cells. The loss of RPB1 phosphorylation upon *CRK9* silencing was studied by Dr. Daniela Ambrósio. To ensure that the phenotypes observed upon CRK9 silencing are not due to an off-target effect, I generated new RNAi cell lines in which we targeted 3′-UTR of the CRK9 mRNA and rescue cell lines that express *wild type* or T533A mutated *CRK9* which are resistant to the RNAi response. This work was published in *Badjatia N, Ambrósio DL, Lee JH, Günzl A. Trypanosome cdc2-*

*related kinase 9 controls spliced leader RNA cap4 methylation and phosphorylation of RNA polymerase II subunit RPB1.Mol Cell Biol* ***33 (2013),*** *1965-1975.*

3. **Chapter IV:** To biochemically characterize CRK9 enzyme complex, I generated a cell line that exclusively expressed C-terminally PTP tagged CRK9. Thereafter, Dr. Sung Hee Park and I performed PTP purifications of CRK9 in two independent experiments. The final eluates of CRK9 PTP purification were independently analyzed by the Keck Facility at Yale University. Dr. Daniela Ambrósio performed sedimentation analysis and *in vitro* kinase assays with PTP purified CRK9 complex. Dr. Park did mice infection studies with bloodstream form CRK9 UTR RNAi and CRK9-HA rescue cell lines. Justin Kirkham helped me in raising polyclonal immune serum against CRK9 and CRK9AP. I generated all the cell lines used in this project. I also performed RNA and protein analysis subsequent to expression silencing of *CYC12* and *CRK9AP* in procyclic and bloodstream form *T. brucei*. Lastly, I generated phylogenetic trees with the sequence alignment of cyclin sequences from model organisms and kinetoplastid CYC12 sequences. A manuscript is under preparation for this study.

4. **Chapter V:** In this study, Dr. Daniela Ambrósio characterized the trypanosome PRP19 complex. She performed all the experiments in this study with the exception of immunofluorescence assay and primer extension analysis of cap4 methylation in total RNAs from *CRK9-*, *ORC1-*, *SPF27-* and *LSm2-* silenced cells, which were done by me. This work was published in *Ambrósio DL, Badjatia N, Günzl A. The spliceosomal PRP19 complex of trypanosomes. Mol Microbiol* ***95 (2015),*** *885-901*.

# List of Tables

# List of Figures

# *Chapter I*

# Introduction

## 1. Trypanosomatids and human diseases

Trypanosomatids are eukaryotic unicellular flagellates that are obligate parasites of vertebrates, invertebrates and plants. They belong to the phylogenetic order Kinetoplastida and family Trypanosomatidae. Kinetoplastids are thought to have diverged early from the main eukaryotic lineage during evolution and they are characterized by the presence of the kinetoplast, a structure that refers to a network of catenated DNA in a single large mitochondrion located near the origin of flagellum. (Shlomai, 2004; Stechmann and Cavalier-Smith, 2002). The Trypanosomatidae include several species of *Trypanosoma* and *Leishmania* that cause severe vector-borne diseases in humans and livestock. *Trypanosoma brucei* and *Trypanosoma cruzi* cause Human African Trypanosomiasis (HAT), also known as (a.k.a.) Sleeping sickness, and Chagas disease, respectively, while several species of *Leishmania* cause a group of diseases called Leishmaniasis. These diseases, often lethal if untreated, collectively affect millions of people in tropical and sub-tropical regions of the world.

Infection with *Trypanosoma brucei* causes HAT in humans and Nagana in livestock. The parasite is transmitted to its vertebrate hosts by sanguivorous tsetse flies of the *Glossina* genus. Tsetse flies as well as the disease are prevalent in 36 countries in the sub-Saharan region of Africa. Currently just under 10,000 cases are reported each year although the exact number of infected individuals is not known as the disease is mainly prevalent in remote rural areas in Africa. *T. brucei* is classified in three subspecies: human-pathogenic, *T brucei gambiense* and

*T. brucei rhodesiense;* and human non-pathogenic*, T. brucei brucei. T. brucei gambiense* is found in west and central Africa, causes chronic infections and currently accounts for 97% of reported cases of sleeping sickness. On the other hand, *T. brucei rhodesiense* infection, prevalent in east and southern Africa, is usually acute, and accounts for fewer than 3% of reported cases. (Kennedy, 2013) *T. brucei brucei* typically doesn't infect humans but causes animal trypanosomiasis in cattle and wild animals along with other *Trypanosoma* species such as *T. vivax*. Wild animals can also host human pathogenic species and while they do not develop the disease, they serve as important reservoirs from which tsetse flies can gain an infection (Franco et al., 2014). The species of *Trypanosoma* including *T. brucei brucei* that are non-pathogenic in humans are highly susceptible to trypanosome lytic factors (TLFs) which are present in human serum. The two types of TLFs are TLF-1, which is high-density lipoprotein particle consisting of apolipoprotein L1 (APOL1) and haptoglobin-related protein (HPR), and TLF-2 which contains IgMs as a major component along with several other proteins including HPR and APOL1. APOL1 is considered as the trypanolytic component of both complexes. Human infective *T brucei gambiense* and *T. brucei rhodesiense* use distinct proteins to neutralize APOL1 trypanolytic action, which makes them resistant to human serum (Pays et al., 2014).

The disease progression in HAT can be divided into two major stages, the haemolymphatic stage and the meningoencephalitic or neurological stage. In the first stage, the parasites multiply in subcutaneous tissue, lymph and blood. The early stage symptoms that are typically observed 1-3 weeks after the insect bite are intermittent fever, headache, itching, malaise, fatigue etc. The recurrent intermittent fever in the first stage is associated with the waves of parasitemia caused due to antigenic variation which is a primary mechanism that enables parasites to evade the immune system in their hosts (Horn, 2014). As the disease develops, other clinical features such as lymphadenopathy; enlargement of the spleen and liver;

cardiac malfunction, ophthalmological features can become evident. The onset of the second stage of the disease correlates with the parasites crossing the blood-brain barrier to infect the central nervous system. The typical symptoms are mental disturbances, poor coordination, sensory and motor disturbances and abnormal reflexes. Sleep abnormalities that give the disease its name are also observed during late stage. This includes reversal of the sleep cycle and irrepressible episodes of sleep. Eventually the patient succumbs to coma and death if left untreated (Kennedy, 2013).

*Trypanosoma cruzi* infection causes Chagas disease in humans, also known as American trypanosomiasis. This disease is mainly endemic in Latin America but widespread emigration in last few decades of infected individuals from endemic areas to other nations, including the United States, Canada, and Western Europe, has caused the disease to spread elsewhere. According to the WHO, between 6-7 million people are estimated to be infected worldwide, mostly in Latin American nations. Transmission in humans is mainly through contact with the feces of an infected Triatomine bug also known as the kissing bug. Other ways of transmission are consumption of contaminated food, congenital transmission, blood transfusion, and organ transplants. Unlike *T. brucei*, *T. cruzi* is predominantly an intracellular parasite that can invade many different cell types and establish a chronic infection. Infection is manifested in two stages: acute and chronic stages. The acute Chagas phase, typically lasting for about two months after infection, can be mild or asymptomatic. However, 10-30% of infected individuals will display symptoms such as swelling or lesion at the site of infection, fever, headache, enlarged lymph glands, abdominal pain, anorexia etc. Rarely, myocarditis or meningoencephalitis during acute phase can result in death. The majority of infected people enter a prolonged asymptomatic phase with low or undetectable parasitemia in blood. About 30-40% of infected asymptomatic carriers develop chronic disease. Clinical manifestations of the chronic stage are dilated cardiomyopathy which can cause sudden death due to cardiac arrest

and/or digestive disorders typically due to megacolon and megaesophagus (Bonney, 2014; Nunes et al., 2013).

Leishmaniasis is caused by infection from over 20 different species of *Leishmania*. It is transmitted through the bite of females of certain species of sandfly. The disease affects over 80 countries worldwide but it is primarily endemic in tropical and sub-tropical countries. According to a recent World Health Organization (WHO) report, about 1.3 million new cases and 20,000-30,000 deaths are reported annually due to this disease. There are two major forms of this disease, namely cutaneous leishmaniasis, which is caused by most *Leishmania* species and visceral leishmaniasis, which is primarily caused by *Leishmania donovani* and *Leishmania infantum.* Cutaneous Leishmaniasis, the less severe and more common form of the disease, causes skin lesions such as ulcers which can produce lifelong scars. Muco-cutaneous leishmaniasis is a more serious form of the cutaneous disease, affecting mucous membranes of nose, mouth and throat. Visceral leishmaniasis, fatal if untreated, is characterized by prolonged oscillating fever, weight loss, anemia, and the enlargement of lymphatic organs (Chappuis et al., 2007; Murray et al., 2005).

Diseases caused by trypanosomatids are amongst the most important neglected tropical diseases, collectively affecting over 21 million people worldwide (McCall and McKerrow, 2014) and causing approximately 150,000 deaths annually (Nussbaum et al., 2010). Over 2 million new cases are reported each year (McCall and McKerrow, 2014). In the absence of vaccines, chemotherapeutic drugs are the only means of treatment. However, some of the current medications are unsatisfactory due to high toxicity, poor efficacy and increasing drug resistance.

Pentamidine and Suramin are the main drugs for first stage HAT treatment while Melarsoprol and Eflornithine are typically administered in the second stage. In 2009, a combination therapy of Eflornithine and Nifurtimox was also introduced for second stage of the

disease (Barrett and Croft, 2012). Most of these compounds cannot be administered orally, and have undesirable, sometimes fatal, side-affects. Moreover, poor efficacy and increasing resistance further limits the treatment with these drugs. Benznidazole and nifurtimox are the two main drugs for treatment of Chagas disease (Barrett and Croft, 2012). Although both are effective in curing the disease during the acute phase, their efficacy is reduced the longer a person has been infected. Additionally, both drugs are not suitable for individuals with kidney or liver malfunction and for pregnant women. This is significant since Chagas disease can be transmitted congenitally; an infected mother would invariably pass the disease to the infant unless treated. Pentavalent antimonials can be used for treatment of Leishmaniasis but resistance is on the rise in some areas (Barrett and Croft, 2012). Some other drugs are allopurinol and amphotericin B. Thus there is a pressing need for the development of new and more effective drugs for all three trypanosomatid diseases (Castillo et al., 2010; Nussbaum et al., 2010; Patterson and Wyllie, 2014). Since trypanosome parasites share several unique biological features that are distinct from their hosts, it is possible to envision a potential drug target common in all three parasites. One such shared feature is the mechanism of protein coding gene expression in trypanosomes and factors crucial for this process can be potentially suitable therapeutic targets for the treatment of trypanosomatid diseases.

## 2. Gene expression in trypanosomatids

Trypanosomatid genomes are organized into large gene clusters containing functionally unrelated, tandemly linked protein coding genes that are transcribed polycistronically by RNA polymerase (pol) II. This is highly unusual as transcription is typically monocistronic in eukaryotic organisms with specific initiation and termination sites. For each gene the binding of basal transcription factors to the core promoter recruits RNA pols to the transcription initiation site, and allows precise transcription initiation. On the other hand, polycistronic transcription is not initiated at a concrete transcription initiation site and, so far, no transcription factors or

promoter elements have been identified for transcription initiation at these gene arrays (Siegel et al., 2011). Instead, open chromatin formation in divergent strand switch regions appears to be sufficient to allow bidirectional RNA polymerase II transcription initiation to occur at multiple sites in these regions (Kolev et al., 2010; Siegel et al., 2009).

The polycistronic primary transcripts in trypanosomes are resolved into individual mRNAs by the coupled processes of spliced leader (SL) *trans* splicing and polyadenylation (LeBowitz et al., 1993; Matthews et al., 1994). The process of SL *trans* splicing was first discovered in *Trypanosoma brucei* (Boothroyd and Cross, 1982; Campbell et al., 1984; Kooter et al., 1984; Milhausen et al., 1984; Van der Ploeg et al., 1982) but later on it was also identified in other members of Euglenozoa as well as in some multicellular organisms such as Nematoda, Platyhelminthes, *Hydra* etc. Some instances of *trans* splicing have been observed in *Drosophila*, *C. elegans* and mosquitoes where exons from two different pre-mRNA molecules are joined together through *trans* splicing (Lasda and Blumenthal, 2011). Similar to *cis* splicing, SL *trans* splicing requires two transesterification reactions with a Y-structure intermediate which is analogous to the *cis* splicing specific lariat structure (Murphy et al., 1986; Sutton and Boothroyd, 1986).

During SL *trans* splicing, the 39-nucleotide long SL, also known as mini-exon, is fused to the 5' ends of all mRNAs in the cell. The SL is derived from a small nuclear RNA called SL RNA, also known as SL donor. Since maturation of each and every mRNA crucially depends on SL *trans* splicing, this RNA processing step is an absolutely essential process for the survival of trypanosomatid parasites. Whereas in most eukaryotes the mRNA capping enzymes associate with RNA pol II and co-transcriptionally cap pre-mRNA, in trypanosomes, the 5' cap is obtained as part of the SL post-transcriptionally during *trans* splicing (Figure 1.1). Thus, SL *trans* splicing uncouples capping from RNA pol II in trypanosomatids (Gunzl, 2010; Michaeli, 2011; Preußer et al., 2012). The SL harbors an extensively modified cap structure called cap4, consisting of a 7-

methylguanosine followed by four methylated nucleotides (Bangs et al., 1992). Cap4 appears to be essential for SL *trans* splicing (McNally and Agabian, 1992; Ullu and Tschudi, 1991) and it was also shown to be important for efficient translation (Zamudio et al., 2009a).

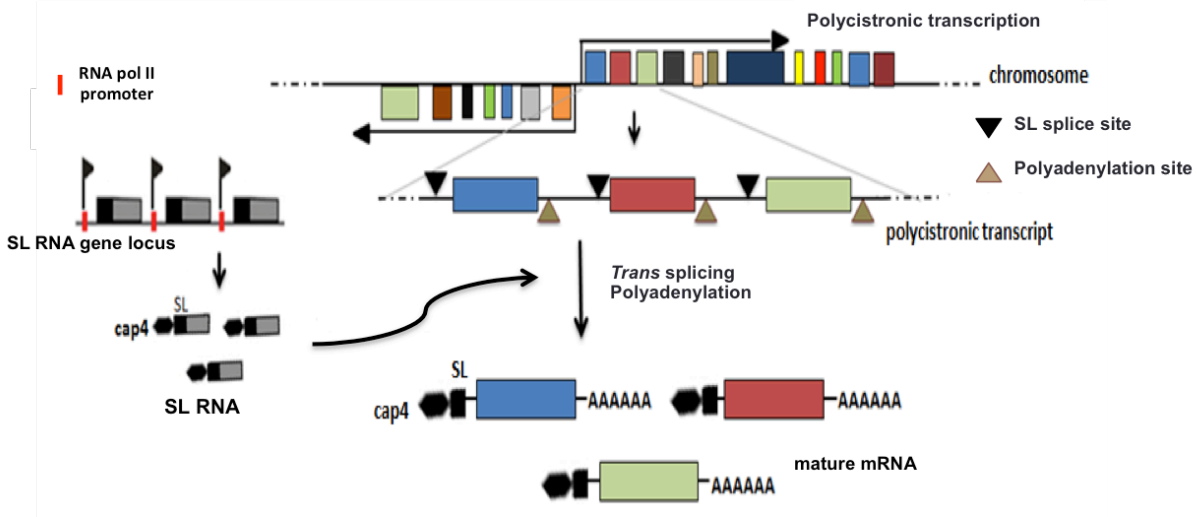



**Figure 1.1 Protein coding gene transcription and mRNA processing in trypanosomes.**

RNA pol II transcription of protein coding genes produces polycistronic transcripts that are processed spliced leader (SL) *trans* splicing and polyadenylation to generate mature individual mRNAs. During *trans* splicing, the capped SL, derived from the small nuclear SL RNA, is added to the 5' end of each mRNA. SL RNA genes (*SLRNA*) are transcribed monocistronically from a RNA pol II promoter. SL RNA acquires the complex cap4 structure $m^{7}G$ $pppm^{6,6}AmpAmpCmpm^{3}Um$ that is implicated in *trans* splicing.

In SL *trans* splicing, the conserved 5' splice site (SS) dinucleotide GU is located on the SL RNA while the 3' SS dinucleotide AG is present on the pre-mRNA. The region upstream of the 3' SS on the pre mRNA also contains the branchpoint (BP) adenosine and a large polypyrimidine tract ($Y_{\sim 20}$), typically just downstream of the BP. During the first step of *trans* splicing, the 2' hydroxyl (OH) group of the BP adenosine performs a nucleophilic attack at the 5' SS or donor SS which results in the release of the SL with a free 3' OH and the formation of a Y-structure intermediate. In the second step, the free 3' OH end of the SL does a nucleophilic attack on the 3' SS or acceptor SS which results in joining of the SL to the mRNA. Upon completion of the splicing process, the Y-structure intermediate is rapidly degraded (Figure 1.2).

Interestingly, while all mRNAs in trypanosomes are *trans* spliced, two trypanosomatid genes also harbor introns which are removed by *cis* splicing (Kolev et al., 2010; Mair et al., 2000a). Both *cis* and *trans* splicing are catalyzed by the spliceosome, consisting of five U-rich small nuclear RNAs (U snRNAs) and associated proteins. While trypanosomes possess all five U snRNAs, due to highly divergent amino acid sequences only a few splicing proteins were identified by standard genome annotation and most of the knowledge of splicing complexes comes from biochemical and functional characterizations.

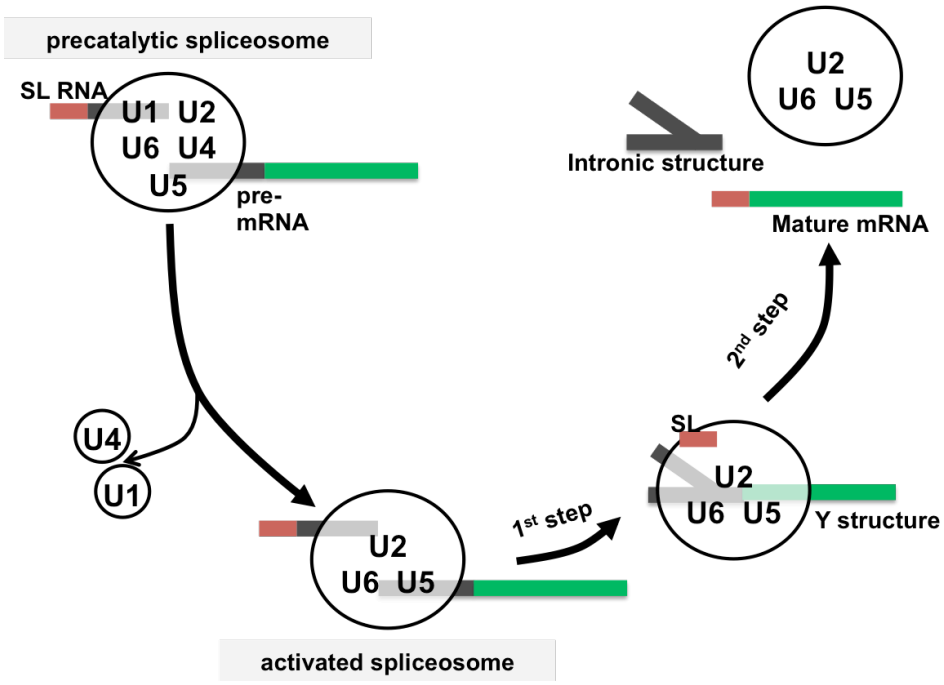



**Figure 1.2 Putative model for spliced leader (SL) *trans* splicing in trypanosomes.**

Analogous to *cis* splicing, SL *trans* splicing is achieved through two transesterification steps. Presumably, a pre-catalytic spliceosome is assembled on pre-mRNA that includes U1, U2, U5 andU4/U6 small nuclear ribo-nucleoproteins (snRNPs) as well as the SL snRNP. Subsequently, the spliceosome undergoes major structural and compositional changes to become activated during which U1 and U4 snRNAs leave the spliceosome. The activated spliceosome performs the first transesterification step which results in the release of the SL and the formation of a Y-structure intermediate. Thereafter, during the second catalytic step, mature spliced mRNA is generated that contains SL at its 5' end. The intronic structure is rapidly degraded whereas the snRNPs are released and recycled for a new round of splicing.

## 3. Splicing machinery in trypanosomes and other eukaryotes.

### 3.1 Eukaryotic Spliceosome

The process of pre-mRNA splicing is catalyzed by a macromolecular, dynamic complex called the spliceosome, which is composed of five small nuclear ribo-nucleoproteins (snRNPs) U1, U2, U4, U5, and U6, as well as various non-snRNP proteins. Uridine-rich (U-rich) snRNPs, composed of U-rich snRNAs and stably interacting proteins, are the most important component of the spliceosome. In humans, there are ~45 snRNA associated proteins amongst a total of ~170 proteins that are estimated to be associated with the spliceosome (Wahl et al., 2009). A large repertoire of spliceosome-associated proteins has regulatory roles while others serve to couple splicing to other processes of gene expression such as transcription and polyadenylation. Some are thought to provide flexibility and specificity to the spliceosome in order to rapidly respond to changes in cellular milieu. In accordance, several non-snRNP proteins are redundant and/or are loosely associated and most likely not needed for every splicing event (Hoskins and Moore, 2012; Matera and Wang, 2014; Wahl et al., 2009; Will and Luhrmann, 2001).

U snRNPs are the main building blocks of the spliceosome. Biogenesis of snRNPs occurs in a sequential manner in several distinct steps. In higher eukaryotes, all U snRNAs except U6 are transcribed by RNA pol II and acquire an $m^7G$ cap cotranscriptionally. These are then exported to cytoplasm, where they associate with seven common (Sm) proteins (B/B', D3, D2, D1, E, F and G) to form the core snRNP structure. These proteins form a heteromeric ring around the highly conserved Sm binding site in a process that is assisted by the survival of motor neurons (SMN) complex. Subsequently, the $m^7G$ cap is hypermethylated to 2,2,7-tri-methylated guanosine ($m^3G$) cap (Mouaikel et al., 2003) and the U snRNAs undergo 3′ end maturation (Seipelt et al., 1999). Thereafter, the core snRNPs are reimported into the nucleus where they bind to various snRNP-specific proteins in a special nuclear substructure called the

Cajal body. In contrast to other U snRNPs, the U6 snRNP lacks a cytoplasmic phase. It is transcribed by RNA pol III and acquires a γ-monomethyl cap. The U6 snRNA lacks the Sm binding site and, instead, binds to seven Sm-like proteins (LSm2-8) (Achsel et al., 1999; Vidal et al., 1999).

In addition to the seven common Sm or Lsm proteins, there are additional snRNP-specific proteins. Human U1 snRNP contains three U1 snRNP specific proteins namely, U1-70K, U1-C and U1-A (Will and Luhrmann, 2001). U2 snRNP in humans consists of at least 14 U2 snRNP-specific proteins including U2A' and U2B'' that directly bind to the U2 snRNA. U4 snRNA and U6 snRNA extensively base pair with each other resulting in formation of stable U4/U6 di-snRNP. Five proteins in addition to the Sm and LSm proteins are stably associated with the U4/U6 di-snRNP. These are: Snu13 (aka 15.5K), CypH, Prp4 , Prp31, and Prp3 (Liu et al., 2006; Will and Luhrmann, 2001). The U5 snRNP contains at least eight specific proteins of which three are especially important for the RNA/RNP rearrangements that are crucial for activating the splicing reaction (Sander et al., 2006; van der Feltz et al., 2012). These proteins are helicase Brr2, GTPase Snu114 and the highly conserved and crucially important Prp8.

In addition to snRNP-specific proteins, there are several non-snRNP proteins that are essential for splicing reaction such as SF1 that binds to the BP (Berglund et al., 1997; Kramer and Utans, 1991), U2 auxiliary factor (U2AF) subunits U2A65 and U2A35 that bind polypyrimidine tract and the 3' SS respectively (Berglund et al., 1998; Wu et al., 1999), and the heteromeric PRP19 complex (orthologous to the nineteen complex (NTC) in yeast) that is an integral part of the active spliceosome (Chan et al., 2003).

The spliceosome is assembled in a step-wise manner during RNA splicing. Firstly, the “early complex” or complex E is formed by the association of the U1 snRNP with the 5' SS via U1 snRNA base pairing. In addition, protein factor SF1 binds to the BP, and the U2AF

associates at both the BP and 3' SS. Next, in an ATP-dependent step, the U2 snRNA recognizes the BP adenosine and interacts with U1 snRNP to form the pre-spliceosome or complex A. SF1 gets displaced during this process with the help of SF3a and SF3b complexes that are associated with the U2 snRNP. Subsequently, the preassembled tri-snRNP, consisting of the U4/U6 di-snRNP and the U5 snRNP, is recruited with the help of the helicase Prp28, to form the precatalytic complex B. Thereafter, complex B undergoes extensive rearrangements, both in composition as well as conformation to form the active complex B*. This activation requires the U5 snRNP-specific proteins Brr2 and Snu114, and the protein factor Prp2. During this process, U4/U6 base pairing is disrupted, U1 and U4 are released and U2-U6 base pairing is established. Complex B* catalyzes the first transesterification step of splicing, which results in the formation of complex C containing the lariat intermediate. Complex C undergoes further ATP dependent remodeling with the help of Prp8 and Prp16 to perform the second step of splicing. The resulting post-spliceosomal complex contains spliced exons and the lariat intron. Both the release of the spliced product and the disassembly of the post-spliceosomal complex is ATP driven and require additional RNA helicases. U2, U5 and U6 snRNPs are discharged and recycled for new rounds of splicing (reviewed in (Matera and Wang, 2014; Wahl et al., 2009)).

Most of our current knowledge for RNA splicing comes from biochemical and structural characterizations of snRNPs and different spliceosomal complexes in yeast and human systems, and is related to *cis*-splicing, e.g. intron removal. In contrast to our detailed understanding of *cis* splicing, little is known about the particulars of the trypanosome *trans* splicing machinery. Previously, it was shown that U2, U4, U5 and U6 snRNAs are needed for *trans* splicing in *T. brucei* which indicated that the splicing machinery is similar for both forms of splicing (Tschudi and Ullu, 1990). On the other hand, it was shown in the nematode *Ascaris* that the U1 snRNP is dispensable for *trans* splicing *in vitro* (Hannon et al., 1991) and that the SL RNA, which is the SL donor, binds the Sm proteins and is assembled in a core snRNP itself

(Bruzik et al., 1988; Palfi et al., 1991). This suggests that there are some important differences between *cis* and *trans* spliceosomes. Thus, the detailed understanding of the *trans* splicing machinery may reveal *trans* splicing-specific factors that can be potential antiparasitic drug targets.

### 3.2 Splicing machinery of Trypanosomes

With the exception of two genes, trypanosomatid genes lack introns (Kolev et al., 2010). SL *trans* splicing is crucial for the maturation of each and every pre-mRNA in the cells. While trypanosomes harbor all five U-rich snRNAs, there are a few important differences between U snRNAs of humans and trypanosomatids. First, while in yeast and metazoans all U snRNAs with the exception of U6 snRNA, are synthesized by RNA pol II, in trypanosomatids all U snRNAs are transcribed by RNA pol III (Fantoni et al., 1994; Nakaar et al., 1994; Nakaar et al., 1997). Unlike human and yeast U5, trypanosomatid U5 snRNA lacks the $m^3G$ cap (Dungan et al., 1996; Xu et al., 1997). U1, U2 and U5 snRNAs in trypanosomatids are distinctively smaller than their counterparts in humans and yeasts, and lack the key stem-loop structures that are known to be important for RNA-protein interactions (Dungan et al., 1996; Hartshorne and Agabian, 1990; Tschudi et al., 1990; Xu et al., 1997). Lastly, the consensus Sm binding site that is typically well conserved in higher eukaryotes is relatively degenerate in trypanosomes (Preußer et al., 2012).

Since antibodies raised against the Sm core proteins that typically cross-react with Sm proteins in other organisms, did not recognize trypanosome Sm core, classical biochemistry and bioinformatics tools were used to identify the Sm and LSm core proteins in trypanosomes (Michaeli et al., 1990; Palfi and Bindereif, 1992; Palfi et al., 1991). Interestingly, while U1, U5 and U6 are assembled in a canonical core snRNP with trypanosomatid orthologs of the seven Sm and Lsm proteins, U2 and U4 snRNP exhibit core variations. In U2 snRNP, paralogs of SmB

and SmD3 called specific spliceosomal Sm2-1 (SSm2-1 a.k.a. Sm15k) and SSm2-2 (a.k.a. Sm16.5K) replace the canonical SmB and SmD3 respectively. Similarly paralogs SSm4 replaces SmD3 in trypanosomatid U4 snRNP (Tkacz et al., 2007; Wang et al., 2006). While orthologs of U2 snRNP specific U2A' (Cross et al., 1993) and U5 snRNP specific PRP8 (Lucke et al., 1997) were characterized almost two decades ago, most of the other snRNP specific proteins have been identified only in recent years. One method that has significantly contributed in increasing the repertoire of identified splicing proteins in trypanosomes is tandem affinity purification (TAP) using the modified TAP tag called the PTP (protein C epitope-TEV protease cleavage site-protein A domains) tag (Schimanski et al., 2005a). With the help of TAP and bioinformatic approaches, trypanosome orthologs of almost all known snRNP proteins have been identified (Gunzl, 2010; Preußer et al., 2012).

However a majority of non snRNP spliceosomal proteins have not been identified yet and remain under represented in the current repertoire of known spliceosomal proteins in trypanosomes. In humans, only ~45 proteins of the estimated 170 spliceosomal proteins are snRNP proteins. Since orthologs of almost all snRNP proteins are present in trypanosomes, it is likely that a comparable number of non-snRNP proteins exist in trypanosomes and are yet to be discovered (Gunzl, 2010).

### 3.3 The SL RNA and its assembly into the SL snRNP complex

One of the unique features of SL *trans* splicing is that the SL RNA, which is a splicing substrate itself, is assembled into a core snRNP and binds the common Sm proteins (Bruzik et al., 1988). Apart from the Sm proteins two other SL snRNP proteins, termed as SL175 and SL30 according to their molecular masses, were identified in the nematode *Ascaris* (Denker et al., 2002). These proteins are indispensable for *trans* splicing but are not required for *cis* splicing or non-SL *trans* splicing. SL RNA has the most extensively modified cap structure

known in any system. This complex cap structure that is composed of a $m^7G$ residue (cap0) followed by four 2'-O-methylated nucleotides (AACU) and additional base methylations on the first and the fourth position ($m^7G$ $pppm^{6,6}AmpAmpCmpm^3Um$) is termed cap4 (Bangs et al., 1992). The base methylations occurring on the first adenosine and the third uracil are unique to trypanosomes. The $m^7G$ cap is formed on nascent SL RNA cotranscriptionally by concerted action of two enzymes: the RNA triphosphatase TbCet1 (Ho and Shuman, 2001) and the bifunctional guanylytransferase/guanine N-7 methytransferase TbCgm1 (Ruan et al., 2007; Takagi et al., 2007). However it is not clear yet whether the remaining cap4 modifications occur cotranscriptionally. While it was thought initially that these modifications occur cotranscriptionally in a 5' to 3' direction and are independent of Sm binding (Hury et al., 2009; Mair et al., 2000b), it was shown subsequently that complete cap4 formation requires Sm core assembly (Zeiner et al., 2004a). In addition to cap4 methylations, *SL RNA* transcripts are also modified by pseudouridylation at nucleotide 28 (Liang et al., 2002) and 3' end trimming (Sturm et al., 1999). Pseudouridylation refers to the rotation of a uridine base relative to its ribose sugar ring. $\Psi_{28}$ formation on the SL RNA is directed by the SLA1 snoRNA (Hury et al., 2009; Liang et al., 2002; Watkins et al., 1994) and requires pseudouridine synthase TbCBF5 (Barth et al., 2005). $\Psi_{28}$ is not essential for parasite survival since mutation of the base U at position 28 of the SL RNA was not lethal. (Liang et al., 2002; Sturm et al., 1998), Surprisingly, silencing of *TbCBF5* resulted in elimination of $\Psi_{28}$ as well as incomplete cap4 formation (Barth et al., 2005). TbMTR1, the enzyme responsible for the cap1 ribose methylation and TbCBF5 are part of a single RNP complex called the SLA1 small nucleolar RNP (snoRNP) complex (Zamudio et al., 2009b). A knockout of TbMTR1 resulted in the specific loss of cap1 modification and accumulation of SL RNA intermediates with cap 0 and 3' end extension indicating a delay in downstream processing events (Zamudio et al., 2007). However, mRNAs from $TbMTR1^{-/-}$ cells had the remaining cap4 methylations, which clearly demonstrated that SL RNA lacking cap1 can be *trans* spliced and methylations at the remaining cap4 positions are independent of cap1

modification (Zamudio et al., 2009a). Two nuclear enzymes, TbMTR2 (a.k.a. TbMT48) and TbMTR3 (a.k.a. TbMT57) are responsible for ribose methylations at position 2 and 3 respectively (Arhin et al., 2006a; Arhin et al., 2006b; Hall and Ho, 2006; Zamudio et al., 2006). It has been speculated that TbMTR3 also catalyzes the methylation at position 4 since cap3 intermediates have not been detected yet. It appears that modification at positions 2, 3 and 4 requires binding of the Sm proteins to SL RNA. Knockout of either or both MTR2 and MTR3 did not affect culture growth and *trans* splicing (Zamudio et al., 2009a). However, neither MTR2 nor MTR3 could be deleted in TbMTR1$^{-/-}$ cells, indicating that at least a partly modified cap4 is essential for cell survival. Additionally cap4 also appears to be important for translation, but the mechanism for this remains to be determined (Zamudio et al., 2009a).

As mentioned above, assembly of the heptameric Sm proteins on SL RNA appears to occur after cap1 modification and prior to remaining cap modifications. It is not known if Sm assembly happens in the nucleus or cytoplasm and there are studies in support for both scenarios. Knockdown of Sm subunits or mutation of the Sm binding site results in an accumulation of immature SL RNA in the cytosol (Mandelboim et al., 2003; Zeiner et al., 2004a). These studies support the notion that the assembly of the SL RNP occurs in the cytoplasm in a manner that is similar to the association of Sm protein complex on U snRNAs in humans. On the contrary, nuclear colocalization of the SLA1 snoRNA with the SL RNA, SmE as well as the *SL RNA* transcription factor tSNAP42 suggests that SL RNA processing and Sm assembly occurs in a distinct site within the nucleus presumably at the site of *SL RNA* transcription (Hury et al., 2009; Tkacz et al., 2007). Nuclear localization of the Sm assembly factor SMN further supports this hypothesis (Jae et al., 2011; Palfi et al., 2009).

Finally, maturation of the SL RNA requires 3' end trimming of SL RNA, which involves removal of the last 4-6 nucleotides. Two nucleases are responsible for this, of which only one called SNIP is currently known (Zeiner et al., 2004b). The 3' end processing is dependent on the Sm-binding and is defined by the secondary structure of the SL RNA. SL RNA folds in a

secondary structure containing three stem loops, with the second and the third loop flanking a single stranded Sm-binding site (Bruzik et al., 1988).

## 4. SL RNA gene (*SLRNA*) transcription

Unlike other spilceosomal U snRNAs that are typically recycled, SL RNA is consumed in the *trans* splicing event. Consequently, it has a high turnover rate with a half life of ~4 minutes (Laird et al., 1987; Ullu and Tschudi, 1991). Most likely, to meet the constant high demand of SL RNA, there are ~100 *SLRNA* copies organized in a 1.4 kb long tandem array on chromosome 9 (Gilinger and Bellofatto, 2001). These genes are transcribed in a monocistronic fashion by RNA pol II, and SL RNA constitutes ~6% of total RNA synthesis in *T. brucei* (Campbell et al., 2000). Since a strong continuous supply of SL RNA is crucial for the survival of all trypanosomatid parasites, the processes of SL RNA synthesis and SL *trans* splicing are attractive targets for parasite control. Accordingly, proteins that are involved in SL RNA biosynthesis are typically encoded by essential genes.

*SLRNA* is the best-characterized transcription unit in trypanosomes. Unlike the protein coding genes, *SLRNA* units are transcribed in a monocistronic fashion from a definite transcription initiation site (TIS) (Hartree and Bellofatto, 1995) and transcription termination occurs at a poly T tract downstream of the *SLRNA* coding region (Sturm et al., 1999). Identification of the RNA pol that transcribes *SLRNA* has been controversial issue in the past. SL RNA has a $m^7G$ cap that is characteristic of RNA pol II transcripts (Laird et al., 1985) but termination of *SLRNA* transcription requires poly T tract (Sturm et al., 1999), which is a hallmark of RNA pol III transcription. While Tagetitoxin, a potent inhibitor of RNA pol III transcription, had no effect on SL RNA transcription (Grondal et al., 1989; Saito et al., 1994), α-amanitin that typically inhibits RNA pol II at low doses, did not affect *SLRNA* transcription in this manner (Crenshaw-Williams and Bellofatto, 1999; Gunzl et al., 1997; Saito et al., 1994). However, an *in vitro* transcription study unambiguously demonstrated that *SLRNA* is transcribed by RNA pol II

and not by RNA pol III (Gilinger and Bellofatto, 2001).

### 4.1 The *SLRNA* promoter

Mutational analyses in *Trypanosoma brucei*, *Leptomonas seymouri* and *Leishmania tarentolae* led to the identification of the tripartite *SLRNA* promoter elements that consist of two upstream sequence elements (USEs) and an initiator element (Inr) at the TIS (Gunzl et al., 1997; Hartree and Bellofatto, 1995; Yu et al., 1998). The relative spacing between these three promoter elements is important for efficient transcription. Also, mutations within the Inr abolished correct transcription initiation (Luo et al., 1999). Thus, the *SLRNA* promoter has three important sequence elements called USE1, USE2 and Inr that are essential for accurate and efficient *SLRNA* transcription.

### 4.2 *Trans*-acting factors binding to the *SLRNA* promoter

After characterization of the *SLRNA* promoter, the next obvious step was to identify a *trans*-acting factor that associates with it. Initial studies using conventional chromatography led to the purification of a tripartite complex that interacted specifically with USE1 of the *SLRNA* promoter (Luo et al., 1999). Of the two proteins that were identified in this study, one was a clear ortholog of the human snRNA-activating protein (SNAP)50 (Das and Bellofatto, 2003). SNAP50 is an essential subunit of the SNAP complex (SNAPc) which interacts with the proximal sequence element of the snRNA gene promoters to nucleate transcription pre-initiation complex formation in humans. Hence, this was the first evidence of a transcription factor in trypanosomes and suggested that a transcription pre-initiation complex (PIC), consisting of basal transcription factors (BTFs) and RNA pol II, is assembled at the *SLRNA* promoter. This was surprising since annotation of completed trypanosomatid genomes revealed only the trypanosome homolog of the TATA-binding protein (TBP), termed TBP-related factor 4 (TRF4) (Ruan et al., 2004), and the TFIIH helicases XPB and XPD as potential BTFs (Ivens et al.,

2005).

### 4.3 Basal factors for RNA pol II mediated transcription in humans

Promoter-dependent transcription initiation by RNA pol II in eukaryotes invariably depends upon the formation of the PIC at the core promoter. PIC assembly for protein coding genes is typically initiated by binding of the basal transcription factor TFIID to the TATA box. This interaction is mediated by TBP, which is a part of the TFIID complex. This is followed by the binding of the bipartite TFIIA to TBP that serves to counteract repressors that might associate with TBP and to strengthen the DNA-protein interaction of the complex. DNA is severely bent upon TBP binding, allowing the single subunit TFIIB to enter the PIC. TFIIB directly interacts with RNA pol II and aids the recruitment of TFIIF-bound RNA pol II to the promoter, positioning it at the correct TIS and, thus, directing accurate start site selection. The bipartite, multi-functional TFIIF complex tightly associates with RNA pol II and facilitates the recruitment of the enzyme to the TFIID-TFIIB complex (Flores et al., 1991). It also aids the polymerase in start site selection (Ghazy et al., 2004) and promoter escape (Yan et al., 1999), and enhances the enzyme's elongation efficiency (Price et al., 1989). TFIIF is also thought to recruit TFIIE to the PIC (Maxon et al., 1994). The two-subunit TFIIE complex brings TFIIH into the PIC and stimulates its helicase and kinase activity. TFIIH is a large complex consisting of a core of seven subunits, including the two helicases *Xeroderma pigmentosum* (XP)B and XPD, and a CDK-activating kinase (CAK) sub-complex of three subunits. The ATPase/helicase activity of XPB and XPD is needed for promoter opening and clearance and the kinase activity of CAK is required for CTD phosphorylation, facilitating promoter escape of RNA pol II [Reviewed in (Sikorski and Buratowski, 2009; Thomas and Chiang, 2006)].

Transcription initiation by RNA pol II at the snRNA genes is less well characterized (Hernandez, 2001; Jawdekar and Henry, 2008). In the human system, PIC formation is initiated

by SNAPc (Henry et al., 1995; Yoon et al., 1995). While human SNAPc consists of five subunits namely SNAP190, SNAP50, SNAP45, SNAP43 and SNAP19, only three of these factors: SNAP190, SNAP50 and SNAP43 are essential and conserved in other eukaryotes (Li et al., 2004a). An *in vitro* study showed that TBP, TFIIA, TFIIB and TFIIF are also essential for snRNA gene transcription (Kuhlman et al., 1999), but the role of TFIIH although suggested, has not been conclusively established (Jawdekar and Henry, 2008; Kuhlman et al., 1999). Nevertheless, these findings indicate that, except for the DNA-protein contact by SNAPc, an RNA pol II PIC is formed at snRNA promoters that is largely congruent to that studied in much more detail at protein coding genes.

### 4.4 Transcription factors essential for *SLRNA* transcription in trypanosomes

As mentioned above, genomic annotation of the completed genome of the *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania major* (also referred as Tritryp) did not identify orthologs of BTFs for RNA pol II other than the multifunctional TBP and TFIIH helicases XPB and XPD (Ivens et al., 2005). This was surprising since BTFs are typically highly conserved, and the apparent lack of the genes encoding for BTFs in the genome suggested that RNA pol II recruitment to DNA and transcription initiation does not require PIC formation in trypanosomatids. However, since these organisms diverged very early from the main eukaryotic lineage, it was possible that, due to sequence divergence, *in silico* identification of the corresponding BTF genes failed. Accordingly, BTF identification in trypanosomes relied on biochemical analysis. Since previous studies had identified SNAP50 (Das and Bellofatto, 2003) and TRF4 (Ruan et al., 2004) as transcription factors that bind to the *SLRNA* promoter and since both were needed for *SLRNA* transcription, these factors were tagged and tandem affinity-purified to further characterize the transcription machinery. In both purifications, a complex consisting of tripartite SNAPc, TRF4, a clear ortholog of the smaller TFIIA subunit, and a sixth protein that is most likely a highly divergent ortholog of the larger TFIIA subunit, were identified

(Das et al., 2005; Schimanski et al., 2005b). This complex bound to the *SLRNA* USE and its depletion from extract completely abolished *SLRNA* transcription activity while control transcription by RNA pol I remained unaffected (Schimanski et al., 2005b). Next, a highly divergent putative TFIIB ortholog was identified by a more thorough *in silico* search. Functional analysis showed that this protein was recruited to the PIC and interacted with both RNA pol II and TRF4. Expression silencing of *TFIIB* was lethal for the cells and clearly affected *SLRNA* transcription (Schimanski et al., 2006). Subsequent biochemical characterization of the TFIIH complex by tagging and TAP of the XPD helicase revealed a full core complex of seven subunits (Lee et al., 2007) and two additional *bona fide* subunits termed as trypanosomatid-specific protein 1 (TSP1) and TSP2 that most likely represent divergent trypanosome orthologs of the two TFIIE subunits (Lee et al., 2009). Gene silencing of both TSP1 and 2 affected trypanosome culture growth and SL RNA abundance in the cells. Moreover, depletion of TSP2 from extract completely abolished transcription from the *SLRNA* promoter (Lee et al., 2009). Interestingly, trypanosome TFIIH lacked the CAK complex, which consists of CDK7, cyclin H and MAT1. Potential homologs of these proteins did not co-purify with TFIIH subunits and the electron microscopic structure of TFIIH in *T. brucei* lacked the characteristic knob-like structure that corresponds to the CAK complex in the human TFIIH. This suggested that, unlike in the model organisms, trypanosomatid TFIIH does not have kinase activity to phosphorylate the polymerase's CTD (Lee et al., 2009). Nonetheless, the TFIIH complex was indispensable for *SLRNA* transcription, and localized to the putative sites of *SLRNA* synthesis (Lee et al., 2009; Lee et al., 2007). Finally, the trypanosome TFIIH complex associated with another protein complex of at least nine subunits termed Med-T that is essential for cell-survival, *SLRNA* transcription and PIC assembly. Molecular structure analysis by electron microscopy revealed that Med-T is similar to the head module of the mediator complex of higher eukaryotes. This was the first evidence that the mediator complex plays a role in the transcription of an snRNA gene (Lee et al., 2010).

Taken together, the current repertoire of BTFs for *SLRNA* transcription in trypanosomes includes 25 unique proteins with orthologs of all known BTFs except TFIIF. Thus, the transcription machinery in trypanosomes is not as simplified as it was initially speculated. Since gene sequences in trypanosomes are highly divergent from those of the model organisms, standard genome annotations failed to identify these factors. However, functional characterizations have clearly demonstrated that a PIC is assembled at the *SLRNA* promoter that, albeit extremely divergent in amino acid sequence, closely resembles the PIC in other eukaryotes.

While a PIC is formed for *SLRNA* transcription in trypanosomes, there are some important distinctions between BTFs of trypanosomes and humans. For example, high-resolution crystal structure of the C-terminal domain of trypanosome TFIIB revealed trypanosome specific structural attributes that are absent in human TFIIB (Ibrahim et al., 2009). Additionally, trypanosome TFIIH has unique features that distinguish it from human TFIIH. Firstly, while two divergent paralogs of the TFIIH helicase *Xeroderma pigmentosum* B (XPB) are present in the genome, only the larger of the two was identified as a TFIIH subunit. Secondly, trypanosome TFIIH lacks the CAK complex (Lee et al., 2009; Lee et al., 2007), which has essential function in RNA pol II CTD phosphorylation and promoter escape (Compe and Egly, 2012). Despite the apparent lack of CAK complex, the CTD of trypanosome RPB1 is phosphorylated and important for *SLRNA* transcription (Chapman and Agabian, 1994; Das and Bellofatto, 2009).

## 5. The TFIIH complex is essential for transcription and DNA repair

### 5.1 The multifunctional TFIIH complex of eukaryotes

The multisubunit TFIIH is an essential component of the RNA pol II PIC amongst all eukaryotes. Apart from its essential role in transcription initiation by RNA pol II, it has also been

shown to function as an elongation factor in RNA pol I transcription (Assfalg et al., 2012) and as a DNA repair factor for nucleotide excision repair of helix-distorting lesions that are generated after exposure to UV light or certain chemical agents (Compe and Egly, 2012). Conserved from yeast to mammals, TFIIH is composed of ten subunits that are present in two subcomplexes. The core complex consists of the two DNA helicases XPB & XPD, p62, p52, p44, p34 and the repair specific subunit TTDA. XPD connects the core to the tripartite CAK subcomplex of CDK7, cyclin H and MAT1.

Mutations in TFIIH subunits XPB, XPD and TTD-A are responsible for autosomal recessive disorders of *Xeroderma pigmentosum* (*XP*), Trichothiodystrophy (TTD) and Cockayne Syndrome. Patients with *XP* have a 1000-fold increase risk of developing skin cancer in addition to neurological and sexual developmental defects. Cockayne syndrome is associated with dwarfism, mental retardation, microcephaly etc. Lastly, patients with TTD experience neurological problems including retardation, tremors and ataxia, have short stature and develop sterility. For a long time clinical symptoms of these diseases were attributed to impairment of nucleotide excision repair (NER) of damaged DNA in patients, but it is now becoming clear that at least some of the symptoms are due to transcriptional defects of specific genes (Compe and Egly, 2012).

TFIIH was initially identified as a factor that is indispensable for basal transcription by RNA pol II (Flores et al., 1992). The enzymatic activities of the ATP dependent DNA helicase XPB and the cyclin dependent kinase CDK7 are essential for basal RNA pol II transcription. The ATPase activity of XPB facilitates melting of promoter DNA that allows RNA pol II to read the coding strand and initiate transcription. Subsequently, CDK7 of the CAK complex phosphorylates the CTD of RPB1, the largest subunit of RNA pol II (Compe and Egly, 2012). In most eukaryotes, the CTD is composed of 5-52 repeats of the heptad sequence motif ‘YSPTSPS’. Post-translational modifications of the CTD regulate different stages of RNA pol II

transcription (Buratowski, 2009). Specific post-translational modifications also facilitate association of the CTD with various mRNA processing factors. RNA pol II that is recruited to a promoter is typically hypo-phosphorylated at the CTD. Phosphorylation of the CTD at Serine 5 of the heptad repeat by CDK7 enables promoter clearance of RNA pol II and allows the enzyme to dissociate from most BTFs (Egloff et al., 2010; Phatnani and Greenleaf, 2006). Additionally, CDK7 is implicated in the phosphorylation of the serine at position 7 of the CTD heptad repeat. The exact role of serine 7 phosphorylation for transcription of protein coding genes is not clear, although it is suggested that the bivalent marks might play a role in recruiting specific RNA processing factors (Akhtar et al., 2009; Glover-Cutter et al., 2009). On the other hand, serine 7 phosphorylation by CDK7 at U snRNA genes is important for recruitment of the Integrator complex, which plays a role in 3′ processing of snRNAs (Egloff et al., 2007; Kim et al., 2009). In summary, TFIIH is a BTF that plays an essential role in transcription initiation, promoter clearance and recruitment of various processing factors during RNA pol II transcription.

Even before XPB and XPD helicases were identified as subunits of TFIIH, they were known to be important for NER of damaged DNA. Both of these helicases belong to the SF2 family of helicases and contain seven conserved helicase motifs. But while XPB is a $3^{/}$ - $5^{/}$ helicase, XPD is a $5^{/}$ - $3^{/}$ one. Interestingly, while ATPase activity of XPB is essential for both transcription and repair, its helicase activity is not required for DNA repair (Coin et al., 2007; Lin et al., 2005). Conversely, the helicase activity of XPD is essential for NER (Coin et al., 2007; Tirode et al., 1999). It is proposed that XPB is involved in anchoring of TFIIH complex to the sites of DNA damage (Oksenych et al., 2009) while XPD opens the damaged DNA that allows subsequent NER factors to repair the damage (Coin et al., 1998a; Coin et al., 1998b). Another subunit of TFIIH that is especially important for DNA repair is TTD-A (a.k.a. Tfb5). While recombinant TFIIH complex containing all nine subunits except for TTD-A, was competent in transcription assays, its repair activity was much reduced. Addition of TTD-A restored the *in-vitro* repair activity of the recombinant TFIIH, clearly demonstrating the importance of TTD-A in

NER (Coin et al., 2006). In addition to the XPB, XPD and TTD-A subunits that appear to have a direct role in DNA repair, other subunits of TFIIH perform regulatory functions. p52 and TTD-A stimulate the ATPase activity of XPB while p44 and p34 are known to stimulate the helicase activity of XPD.

The main function of TFIIH in NER is to unwind the DNA at the site of damage, allowing subsequent repair factors to access the damaged DNA and repair the defect (Compe and Egly, 2012). NER is triggered when XPC-RAD23B recognizes the DNA lesions that are generated due to exposure to UV light or certain genotoxic chemicals such as cisplatin (Mocquet et al., 2007; Sugasawa et al., 2001). These lesions include covalently linked bulky DNA adducts that distort the double helical structure. UV-induced lesions are also recognized by UV-damaged DNA binding (UV-DDB) proteins, which in turn, recruit XPC to the damage site (Fei et al., 2011; Fitch et al., 2003; Wang et al., 2004). In the transcription-coupled NER (TC-NER) pathway, which involves repair of DNA lesions on genes that are being transcribed, damage detection is mediated by stalled RNA pol II along with the Cockayne syndrome type B (CSB) protein (Hanawalt and Spivak, 2008). Subsequent to damage recognition, TFIIH enters the site of lesion, and unwinds the DNA with the help of XPB and XPD helicases (Compe and Egly, 2012; Oksenych and Coin, 2010). XPA and Replication Protein A (RPA) are then recruited to stabilize the TFIIH-XPC complex and protect the single-stranded DNA respectively (Aboussekhra et al., 1995; Krasikova et al., 2010). Next, the NER endonucleases XPF and XPG incise the damaged DNA at 5′ and 3′ sides, respectively (Evans et al., 1997). This results in removal of an oligonucleotide that is about 27 nucleotides long, encompassing the DNA lesion. After XPF generates the 5′ incision, XPC, TFIIH and XPA are released from the reaction and recycled (Coin et al., 2008). Prior to the 3′ incision, RPC and Proliferating Cell Nuclear Antigen (PCNA) and DNA polymerases are recruited to the site of damage (Staresincic et al., 2009). Subsequently, DNA polymerases fill the gap and DNA ligases seal the newly synthesized DNA

(Lehmann, 2011). It appears that in non-dividing cells, DNA polymerase δ (Pol δ) and Pol κ synthesize DNA (Ogi et al., 2010), and DNA ligase III–X-ray repair cross-complementing protein 1 (XRCC1) performs ligation while in diving cells synthesis also needs Pol ε (Ogi et al., 2010) and ligation is mediated by both DNA ligase III–XRCC1 and DNA ligase I–flap endonuclease 1 (FEN1) (Moser et al., 2007). In a final step, chromatin assembly factor 1 (CAF1) repositions the nucleosomes on the repaired DNA (Gaillard et al., 1996) (Reviewed in (Compe and Egly, 2012)).

**5.2 Unique features of trypanosome TFIIH**

As mentioned earlier, trypanosome TFIIH is composed of the orthologs of XPD, XPB, p52, p62, p34, p44 and TTD-A, and two trypanosomatid specific subunits which are indispensable for *SLRNA* transcription (Lee et al., 2009; Lee et al., 2007). Surprisingly, annotation of completely sequenced genomes showed that trypanosomatids have two divergent *XPB* genes. The sequences of the *XPB* paralogs from *T. brucei* exhibit only 24.2% identity and 37.1% similarity, indicating that this XPB dualism is not the result of a recent gene duplication event. Moreover, two divergent *XPB* genes are present in all kinetoplastids whose genome has been completed, suggesting that the XPB dualism is not restricted to trypanosomatid organisms. Interestingly, only the larger of the two XPB proteins co-purified with *T. brucei* TFIIH leaving the functional role of the smaller *XPB* paralog undetermined.

**The study described in Chapter II is the functional characterization of the smaller *XPB* paralog termed *XPB-R* (R for repair). This study identified that while XPB-R was dispensable for *SLRNA* and protein coding gene transcription, it was essential for NER in *T. brucei.* Our results also suggest that although XPB-R is not assembled in a TFIIH complex, it interacts with the p52 subunit of TFIIH. In other eukaryotes, it has been shown that p52 regulates XPB activity and interacts directly with XPB. Consistent with**

**this, a knockdown of *p52* also impaired NER in *T. brucei.* To our knowledge, this is the first study that demonstrates a TFIIH independent function of XPB for NER in any eukaryotic organism.**

Another interesting feature of trypanosome TFIIH is the apparent lack of the CAK sub-complex that is typically associated with TFIIH in all eukaryotes (Lee et al., 2009). As introduced above, the CAK enzyme CDK7 plays a crucial role in RNA pol II transcription by phosphorylating the CTD of RPB1. Interestingly, a comparative genomics study indicated that organisms that lack the heptad repeat motif in the CTD do not possess orthologs of CDK7 (Guo and Stiller, 2004). Accordingly, the finding that trypanosome TFIIH is not associated with a CAK complex was consistent with the fact that there are no heptad repeats in trypanosomatid RPB1s. Despite the absence of heptad repeats, the trypanosome CTD of RPB1 is phosphorylated (Chapman and Agabian, 1994; Das and Bellofatto, 2009). Moreover, it was shown that deletion of a major part of *T. brucei* CTD is lethal for cells and abolishes RNA pol II transcription (Das and Bellofatto, 2009). However, the kinase that phosphorylates the CTD and the functional significance of CTD phosphorylations in trypanosomes is not yet known although a recent study suggests that chromatin-associated RNA pol II is phosphorylated at the CTD (Rocha et al., 2014). In summary, these results indicated that trypanosomes possess a so-called transcriptional CDK that controls RNA pol II phosphorylation and function. This prospect was exciting since CDKs, in contrast to almost all other transcription factors, are promising targets of chemotherapy.

### 5.3 CDKs are essential regulators of the cell cycle and of gene expression

CDKs form a special class of serine-threonine kinases that require association with their regulatory partners called cyclins for kinase activity. CDK/cyclin complexes were first identified in yeast as key regulators of cell cycle progression (Beach et al., 1982; Nasmyth and Reed, 1980). Although now it is well established that several CDK/cyclin complexes have crucial

functions in transcription and RNA splicing (Lim and Kaldis, 2013; Loyer et al., 2005). CDKs share a conserved catalytic domain that consists of the ATP binding pocket, PSTAIRE-like helix domain responsible for cyclin binding and an activating T-loop motif. Association of cyclin to the PSTAIRE domain confers spatial reorientation of the T loop that enables substrate binding (Cao et al., 2014; Malumbres et al., 2009). While *Saccharomyces cerevisiae* encodes for six members of the CDK family and about 23 cyclins, the human CDK family is composed of 13 established members that associate with 29 cyclins or cyclin-related proteins (Malumbres and Barbacid, 2009).

Pioneering studies in yeast showed that a single CDK − Cdc28p in *Saccharomyces cerevisiae* and cdc2 in *S. pombe* − controls the progression through the cell cycle by binding to specific cyclins at different stages (Beach et al., 1982; Evans et al., 1983; Nurse and Thuriaux, 1980; Reed et al., 1982). In contrast to yeasts, many more CDKs and cyclins are involved in cell cycle regulation in mammals. These include CDK2, CDK4 and CDK6 also referred as interphase CDKs, CDK1 (a.k.a. cell division control protein 2 (CDC2)), and ten cyclins belonging to the classes of A-, B-, D- and E-type cyclins (Malumbres and Barbacid, 2009). According to the classical model, specific CDK-cyclin pairs control distinct events during the mammalian cell cycle. In response to mitogenic stimulus, D-type cyclins are expressed that preferentially bind to CDK4 and CDK6 during G1 phase. Activation of CDK4 and CDK6 stimulates the expression of E-type cyclins which bind and activate CDK2. The expression and availability of E-type cyclin is tightly restricted to the early stage of DNA replication. During the late stages, CDK2 interacts with cyclin A2 to drive the transition from S to G2 phase of cell cycle. Next, CDK1 activation by A-type cyclins drives initiation of mitosis. Subsequent to the breakdown of nuclear envelope, A-type cyclins are rapidly degraded and CDK1 interacts with cyclin B to drive mitosis (reviewed in (Malumbres and Barbacid, 2009)). However, in recent years the classical model has been challenged by the studies that show that CDK2, CDK4 and CDK6 are not essential for the cell cycle (Berthet et al., 2003; Malumbres et al., 2004; Martin et al., 2003; Mettus and Rane, 2003;

Ortega et al., 2003; Rane et al., 1999). Instead, they regulate specific functions in specialized cell types. For instance CDK4 is important for development of pancreatic β cells (Rane et al., 1999) while CDK2 is essential for meiotic division in germ cells (Berthet et al., 2003; Ortega et al., 2003).

While CDKs that are important for the cell cycle are regulated by their sequential binding to different cyclins, each of which are transcribed and degraded at definite stages during the cell cycle, the CDKs and cyclins that are involved in gene expression appear to be constitutively expressed throughout the cell cycle. CDK7 is a unique CDK that has roles in both cell cycle and transcription regulation (Fisher, 2005). It associates stoichiometrically with Cyclin H and a RING finger domain containing protein, MAT1 forming a trimeric complex of CDK7/Cyclin H/MAT1 called the CDK activating kinase (CAK) complex because of its ability to phosphorylate and activate key cell cycle CDKs at the T loop (Fisher, 2005; Roy et al., 1994; Serizawa et al., 1995; Shiekhattar et al., 1995). MAT1 functions as an assembly factor (Fisher et al., 1995; Tassan et al., 1995) and interacts directly with both Cyclin H and CDK7 (Busso et al., 2000).

Another CDK/cyclin complex that appears to function both in gene expression and cell cycle is the CDK11/cyclin L complex. As a result of alternative splicing and exon skipping, different isoforms of CDK11 are produced which are grouped as $CDK11^{p110}$ and $CDK11^{p58}$ proteins based on their size. While $CDK11^{p58}$ isoforms are only expressed during G2/M phase and are implicated in mitosis, $CDK11^{p110}$ forms are constitutively expressed and have a role in transcription and pre-mRNA splicing (Hu et al., 2003; Loyer et al., 1998; Trembley et al., 2004). $CDK11^{p110}$ associates with RNA pol II as well as the splicing factors RNPS1 and 9G8, and it was shown to phosphorylate the SR protein 9G8 (Hu et al., 2003). Phosphorylation of 9G8 is critical for its role in pre-mRNA splicing and export of mRNA from nucleus to cytoplasm. Thus, $CDK11^{p110}$ might regulate splicing by controlling the phosphorylation status of 9G8 (Dickinson et al., 2002; Hu et al., 2003; Loyer et al., 2008; Loyer et al., 1998).

Two other CDK/cyclin complexes known to function in transcription are CDK8/cyclin C and CDK9/cyclin T. CDK8/cyclin C associates with the Mediator complex and functions as a transcriptional repressor by phosphorylating cyclin H, which deceases the activity of TFIIH on the CTD (Akoulitchev et al., 2000). Conversely, CDK9/cyclin T, a.k.a. positive transcription elongation factor b (P-TEFb), promotes transcription elongation by RNA pol II by phosphorylating the CTD at Serine 2 of the heptad repeats (Ramanathan et al., 2001; Zhou et al., 2000).

In summary, CDKs are crucial regulators of the cell cycle and gene expression. CDKs as well as other kinases that control the cell cycle are often dysregulated during cancer leading to unrestricted cell proliferation (Malumbres and Barbacid, 2001). Since, kinases can be effectively inactivated by small molecule inhibitors, they are considered as relevant drug targets (Shapiro, 2006). Pan-CDK inhibitor flavopiridol was the first CDK inhibitor to enter clinical trials. It showed anti-tumor effects in leukemia cell and non-small cell lung cancer cell lines (Konig et al., 1997; Shapiro et al., 1999) as well as in patients with renal, prostate and colon cancer, metastatic gastric cancer and non-Hodgkin's lymphoma during phase I clinical trials (Senderowicz et al., 1998; Tan et al., 2002; Thomas et al., 2002). However it exhibited several adverse effects in patients and did not show significant advantages in phase II and III clinical trials (Aklilu et al., 2003; Burdette-Radoux et al., 2004; Schwartz et al., 2001; Shapiro et al., 2001). As of 2013, there were at least 40 ongoing clinical studies with several first and second-generation CDK inhibitors (Pitts et al., 2014). Currently, more attention is focused on the development of selective ATP-noncompetitive CDK inhibitors that do not exhibit toxicity that pan CDK showed in previous clinical trials (Abate et al., 2013).

### 5.4 *T. brucei* CRK9 is a transcriptional CDK

*T. brucei* contains eleven CDKs, termed cdc2-related kinases (CRK)1-4 and CRK6-12 and 10 cyclins (CYC2-11) (Hammarton, 2007; Naula et al., 2005). Due to high sequence

divergence of trypanosome CDKs, it has been impossible to identify their putative orthologs in other eukaryotes without functional validation. Due to their potential druggability, trypanosomatid kinases have constantly been a subject of interest. A number of studies have been performed to identify essential CDKs in *T. brucei* (Jones et al., 2014; Mackey et al., 2011; Merritt and Stuart, 2013). RNAi mediated gene knockdowns identified CRK1-4, CRK6, CRK9, and CRK12 as crucial for cell survival in *T. brucei* (Jones et al., 2014). CRK3 appears to be the functional homolog of CDK1 in trypanosomes. It partners with CYC6 (a.k.a. cyclin B2) and regulates mitosis (Hammarton et al., 2003; Li and Wang, 2003; Tu and Wang, 2004). CRK1 is important for G1/S progression and interacts with CYC2 in a yeast two-hybrid screen although interaction was not observed *in vivo* (Gourguechon et al., 2007; Tu and Wang, 2004; Van Hellemond et al., 2000). In addition, CRK2, CRK4 and CRK6 appear to have accessory functions in the cell cycle although their cyclin partners are unknown (Tu and Wang, 2005). Recently CRK12 was shown to associate with CYC9 during TAP. Both CYC9 and CRK12 were essential in the human-infective bloodstream stage of *T. brucei*, but gene silencing resulted in distinct phenotypes suggesting that while CYC9 is important for cytokinesis, CRK12 has a role in endocytosis (Monnerat et al., 2013).

**In an attempt to identify the trypanosome CDK that is involved in CTD phosphorylation, we performed RNAi-mediated gene knockdowns of CRK1, CRK3, CRK7 and CRK9 since CRK1, CRK3 and CRK9 were shown to be most important for trypanosome survival** (Gourguechon and Wang, 2009; Tu and Wang, 2004) **and CRK7 was speculated to be the ortholog of CDK7 based on sequence similarity (Parsons et al., 2005). In a study described in chapter III, we identified that CRK9 is required for RPB1 phosphorylation in *T. brucei*. Interestingly, *CRK9* silencing did not lead to a specific defect in RNA pol II transcription; instead it caused SL RNA cap4 hypomethylation and a block at the first step of SL *trans* splicing. This study is the first identification of a**

**trypanosome CDK that is essential for phosphorylation of RNA pol II and parasite specific process of *trans* splicing.**

**In a next step, I characterized the enzyme complex of this crucial CDK by TAP of CRK9. This led to identification of a putative tripartite complex comprised of CRK9, a previously unannotated cyclin termed CYC12, and a third protein termed CRK9AP with no obvious conserved sequence motifs. Gene silencing analysis identified CYC12 and CRK9AP as functional partners of CRK9 *in vivo*. A phylogenetic analysis suggested that CYC12 is an L-type cyclin, which, in other eukaryotes, associates with CDK11, a CDK that is involved in transcription and RNA splicing. However, the presence of CRK9AP in the enzyme complex distinguishes CRK9 from CDK11. CRK9AP appears to be important for enzyme complex assembly or stability, since its ablation led to rapid co-depletion of the other two proteins. Moreover, we validated in the mouse host that this unusual key enzyme is a suitable target for chemotherapy. This work is discussed in Chapter IV.**

**Lastly, due to CRK9's putative role in the *trans* splicing process we renewed our interest in the trypanosome spliceosome, focusing on the non-snRNP PRP19 complex, which is important for the activation of the spliceosome. Most interestingly, we discovered that ablation of the PRP19 complex, in contrast to that of an essential U6 snRNP protein, led to similar defects in *trans* splicing, cap4 formation and RPB1 phosphorylation as we observed upon CRK9 depletion, indicating that CRK9 acts upstream of spliceosome activation. The functional and biochemical characterization of the trypanosome PRP19 complex was a collaborative study with Dr. Daniela Ambrósio and is described in Chapter V.**

# *Chapter II*

# *Trypanosoma brucei* harbors a divergent XPB helicase paralog that is specialized in nucleotide excision repair and conserved among kinetoplastid organisms

## 1. Abstract

Conserved from yeast to humans, TFIIH is essential for RNA polymerase II transcription and nucleotide excision repair (NER). TFIIH consists of a core that includes the DNA helicase *Xeroderma pigmentosum* B (XPB) and a kinase subcomplex. *Trypanosoma brucei* TFIIH harbors all core complex components and is indispensable for RNA polymerase II transcription of spliced leader RNA genes (*SLRNAs*). Kinetoplastid organisms, however, possess two highly divergent XPB paralogs with only the larger being identified as a TFIIH subunit in *T. brucei*. Here we show that a knockout of the gene for the smaller paralog, termed *XPB-R* (R for repair) resulted in viable cultured trypanosomes that grew slower than normal. XPB-R depletion did not affect transcription *in vivo* or *in vitro* and XPB-R was not found to occupy the *SLRNA* promoter which assembles a RNA polymerase II transcription pre-initiation complex including TFIIH. However, *XPB-R*$^{-/-}$ cells were much less tolerant than wild-type cells to UV light- and cisplatin-induced DNA damage, which require NER. Since *XPB-R*$^{-/-}$ cells were not impaired in DNA base excision repair, XPB-R appears to function specifically in NER. Interestingly, several other protists possess highly divergent XPB paralogs suggesting that XPBs specialized in transcription or NER exist beyond the Kinetoplastida.

## 2. Introduction

The multi-subunit transcription factor TFIIH is of central importance to transcription initiation by RNA polymerase II and DNA nucleotide excision repair (NER) in yeast and mammalian systems (Compe and Egly, 2012; Feaver et al., 1993; Schaeffer et al., 1993). It consists of two sub-complexes: the core and the cyclin activating kinase (CAK) complex. The core complex includes the two ATP-dependent DNA helicases *Xeroderma pigmentosum* B (XPB, also known as ERCC3, RAD25, SSL2), and XPD (also known as ERCC2, RAD3), their respective regulators p52 and p44 along with p62, p34 and p8 (also known as TTDA, TFB5). XPD connects the core complex to the CAK complex which consists of cyclin-dependent kinase 7 (CDK7), cyclin H and MAT1. Both XPB and XPD belong to the SF2 superfamily of helicases and contain seven characteristic helicase motifs, namely Walker motifs I, Ia, II, III, IV, V, and VI. While XPB is a $3^{/}$ - $5^{/}$ helicase, XPD exhibits $5^{/}$ - $3^{/}$ activity. Recent studies in human cells have demonstrated that these helicases have distinct roles in NER and transcription. The ATPase activity of XPB is essential for DNA opening in both repair and transcription while its helicase activity is specifically important for transcription: by unwinding DNA it facilitates RNA pol II escape from the promoter (Coin et al., 2007; Lin et al., 2005). The TFIIH subunit p52 directly interacts with an N-terminal region of XPB, termed XPB domain, and regulates its ATPase activity (Coin et al., 2007; Jawhari et al., 2002). In contrast, XPD helicase activity is essential for NER whereas it appears to be dispensable for transcription initiation (Coin et al., 2007; Tirode et al., 1999). Given the importance of these helicases in NER, it is not surprising that mutations in XPB and XPD cause deficiency in NER leading to the human diseases *Xeroderma pigmentosum*, Cockayne Syndrome and Trichothiodystrophy (Compe and Egly, 2012).

XPB homologs have been identified in both bacteria and archaea (Balasingham et al., 2012; Biswas et al., 2009; Richards et al., 2008). Recent data suggest that archaeal XPB functions in NER. Crystal structure data of XPB from the archaeon *Archaeoglobus fulgidus*

revealed a DNA damage recognition domain (Fan et al., 2006), and a functional analysis of the association between *Sulfolobus solfataricus* XPB and BAX1 endonuclease showed that this complex can recognize, unwind and cleave model NER substrates *in vitro* (Rouillon and White, 2010). Moreover, homologs of other eukaryotic NER proteins have been identified in archaeal organisms as well (Rouillon and White, 2011). Conversely, the bacterial NER pathway is controlled by so-called Uvr proteins whose amino acid sequences are different from those of eukaryotic NER proteins (Truglio et al., 2006). Interestingly, bacterial and archaeal XPB appears to function independently of a TFIIH complex since, except for XPD, TFIIH subunit homologs have remained elusive in these organisms (Rouillon and White, 2011).

In eukaryotes, XPB exerts its dual function in transcription and DNA repair exclusively within a TFIIH complex (Compe and Egly, 2012). Notably, the completed genomes of members of the protistan family Trypanosomatidae, which include the human parasites *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania major,* harbor two distinct *XPB* genes that encode two divergent XPB paralogs (Ivens et al., 2005). The *T. brucei* TFIIH complex has been functionally, biochemically and structurally characterized, and has been shown to have a basal function in RNA polymerase II transcription as in other eukaryotes (Lecordier et al., 2007; Lee et al., 2010). This function, however, may be restricted to a single gene, namely the spliced leader (SL) RNA gene (*SLRNA*). In trypanosomatids, protein coding genes are arranged in long tandem arrays that are transcribed polycistronically. Individual mRNAs are then processed from precursor RNA by polyadenylation and SL *trans* splicing (Gunzl, 2010). Transcription of the gene arrays initiates predominantly in divergent strand-switch regions (dSSRs) in which the arrays are arranged head-to-head (Martinez-Calvillo et al., 2010; Martinez-Calvillo et al., 2003). While dSSRs typically harbor two peaks of open chromatin marks, it remains unclear whether basal transcription factors such as TFIIB or TFIIH are required for RNA pol II transcription initiation in these regions (see reviews by (Alsford et al., 2012) and Günzl, 2012) In contrast,

*SLRNA genes*, which are tandemly linked on chromosome 9 and encode the SL donor in the *trans* splicing process, are transcribed monocistronically by RNA pol II from a concrete transcription initiation site. Accordingly, the *SLRNA* promoter assembles a conventional, albeit highly divergent, transcription pre-initiation complex which includes TFIIH (Preußer et al., 2012). The *T. brucei* TFIIH was characterized biochemically by tandem affinity purification revealing a full core of seven subunits and two additional subunits, termed TSP1 and TSP2, which most likely represent divergent subunit orthologs of the factor TFIIE (Lecordier et al., 2007; Lee et al., 2009; Lee et al., 2007). Importantly, only the larger of the two XPB paralogs was identified in the TFIIH complex regardless of whether the complex was purified through tagging XPD or TSP2 (Lecordier et al., 2007; Lee et al., 2009).

Here, we functionally characterized the smaller *T. brucei* XPB paralog which has previously been named XPBz (Lecordier et al., 2007). Our results showed that this helicase was neither important for transcription nor essential for the viability of the insect stage, procyclic form of the parasite in culture although the gene knockout did retard trypanosome proliferation. Instead, we found that the helicase specifically functioned in NER and, correspondingly, was localized throughout the nucleus. We therefore renamed this helicase XPB-R (R for repair). Furthermore, our results strongly indicate that XPB-R does not assemble into a TFIIH complex yet forms a repair complex with the trypanosome p52 ortholog. Interestingly, comparative genomics showed that the two divergent XPB orthologs are present in all kinetoplastids including the bodonid *Bodo saltans* and a phylogenetic analysis revealed that highly divergent XPB paralogs are present in distantly related protistan organisms suggesting that a bifunctional TFIIH complex is not the norm among deep branching eukaryotes.

## 3. Materials and Methods

### 3.1 DNAs

For *XPB-R* and *p52* silencing, the coding region from position 8 to position 487 and from position 1 to 500, respectively, were integrated in a stem-loop arrangement (sense-stuffer-antisense) into the pT7-stl construct (Brandenburg et al., 2007). For C-terminal PTP tagging of XPB-R, 499 bp of the 3′ terminal XPB-R coding sequence including an internal BsgI linearization site were cloned into the pC-PTP-NEO vector using the vector's ApaI and NotI restriction sites (Schimanski et al., 2005a). DNA oligonucleotides used for semi-quantitative RT-PCR or PCR are specified in Table S1 of the supplemental material. Plasmids used in dot blots were described previously and contained the complete coding regions of GPEET procyclin, α tubulin, actin, HSP70 and 18S rRNA or the threonine tRNA/U6 snRNA gene association (Schimanski et al., 2006). The template plasmids for *in vitro* transcription, GPEET-trm and SLins19, have been described previously (Gunzl et al., 1997; Laufer et al., 1999).

### 3.2 Cells

*T. brucei brucei* 427 cell culture, stable transfection by electroporation, and the generation of clonal cell lines by selection and limiting dilution was carried out as described (Lee et al., 2007). In RNAi experiments, dsRNA synthesis was induced with 2 µg/ml of doxycycline. For measurement of procyclic and bloodstream form culture growth, cells were counted and diluted daily to $2 \times 10^6$ cells/ml and $2 \times 10^5$ cells/ml, respectively. The procyclic clonal XPB-$R^{+/-}$ cell line was generated by knocking out one of the *XPB-R* alleles with a PCR product in which 101 bp of *XPB-R* 5′ and 3′ gene flanks were fused to the hygromycin phosphotransferase coding region. The clonal cell line that exclusively expressed XPB-R-PTP was obtained by targeted integration of the linearized plasmid XPB-R-PTP-NEO into the remaining *XPB-R* allele of the XPB-$R^{+/-}$ cell line. For HA tagging of p52, the construct p52-HA-BLA (Lee et al., 2007) was

linearized with StuI and transfected into a clonal cell line that exclusively expressed XPB-R-PTP. Immunoblotting was performed to confirm correct tagging of the protein in each clonal cell line. The XPB-$R^{-/-}$ KO-1 and KO-2 knockout cell lines were created by removing the remaining *XPB-R* allele in XPB-$R^{+/-}$ cells with a PCR product of the blasticidin-S deaminase coding region surrounded by 101 bp of *XPB-R* 5′ and 3′ gene flanks, and by limiting dilution of the cells immediately after transfection. Correct DNA integrations were ensured in clonal cell lines by PCR amplification of genomic DNA using oligonucleotides that hybridize outside of the transfected DNA constructs.

### 3.3 RNA analysis

The relative abundance of RNAs in cells was determined by preparing total RNA from 8 x $10^7$ cells using Trizol reagent (Invitrogen), reverse transcription and semi-quantitative PCR. Reverse transcription was carried out with SuperScript II reverse transcriptase (Invitrogen) according to the manufacturer's protocol using oligo-dT as primer. For semi-quantitative PCR, the number of cycles for the linear amplification range was determined empirically for each oligonucleotide pair. Labeling and analysis of nascent RNA using a permeabilized cell system was carried out as described previously (Schimanski et al., 2006). Quantification was performed by densitometry using ImageJ for SL RNA signals and by scintillation counting for dot blot signals.

### 3.4 Protein Analysis

Immunoprecipitations (IPs) were performed as described previously (Lee *et al.*, 2007). Briefly, for reciprocal co-IPs, 100 µl of crude extract was incubated with either 40 µl settled volume of IgG beads (Amersham, Piscataway, NJ) which interact with the PTP tag, or 25 µl settled volume of anti-HA antibody (Bedez et al., 2013) immobilized on protein G-Sepharose (Amersham). Once bound to protein G, the anti-HA antibody is unable to interact with the

protein A domains of the PTP tag, ensuring that these beads precipitate HA-tagged protein only. After washing the beads six times with 700 µl of TET100 buffer (100 mM NaCl, 20 mM Tris-HCl, pH 8.0, 3 mM $MgCl_2$, 0.05% Tween 20), PTP-tagged protein was released by AcTEV protease (Invitrogen) digestion, whereas HA-tagged protein was directly eluted into SDS loading buffer at 100°C for 5 min. Precipitated complexes were analyzed by immunoblotting; PTP- and HA-tagged proteins were detected on blots with the anti-protein C antibody HPC4 (Bedez et al.) or a monoclonal rat anti-HA antibody (Bedez et al.), and TSP2 with a polyclonal rat immune serum (Lee et al., 2009) in combination with the BM chemiluminescence blotting substrate (Bedez et al.) according to the manufacturer's specifications.

XPB-R-PTP was tandem affinity purified according to the standard PTP purification protocol (Schimanski et al., 2005a). Purified proteins were separated on a 10 to 20% SDS-polyacrylamide gradient gel and stained with SYPRO Ruby (Invitrogen) according to the manufacturer's protocol. Proteins were trypsin digested, eluted from the gel and subjected to liquid chromatography-tandem mass spectrometry.

For sedimentation analysis, 200 µl of crude extract prepared from a clonal cell line expressing XPB-R-PTP and p52-HA was subjected to ultracentrifugation in 4 ml 10-40% linear sucrose gradients at 41,000 rpm for 16 h at 4°C as described previously (Lee *et al.*, 2010). Twenty fractions were collected from top to bottom and protein in each fraction was collected with hydrophobic StrataClean resin (Stratagene, La Jolla, CA) as previously described (Schimanski et al., 2005a). Proteins were separated by denaturing PAGE and detected by immunoblotting.

### 3.4 *In vitro* transcription

Preparation of extract active in transcription and the carrying out of *in vitro* transcription assays were performed as described previously (Laufer and Gunzl, 2001; Laufer et al., 1999).

Newly synthesized RNAs of the GPEET-trm and SLins19 templates were detected by primer extension of $^{32}$P-end-labeled oligonucleotides Tag-PE and SLtag, which hybridize to unrelated oligonucleotide tags of the GPEET-trm and SLins19 RNAs. Primer extension products were resolved on 6% polyacrylamide-50% urea gels and visualized by autoradiography.

### 3.5 Chromatin Immunoprecipitation

ChIP assays were performed as described previously (Lee et al., 2010). XPD-PTP- and XPB-R-PTP-bound chromatin fragments were immunoprecipitated overnight at 4°C with a polyclonal rabbit antibody directed against the protein A domains of the PTP tag (Kartalou and Essigmann). Immunoprecipitated DNA was analyzed by semiquantitative and quantitative real-time PCR of the *SLRNA* promoter region (positions −90 to +63 relative to the transcription initiation site), the *SLRNA* intergenic region (+407 to +540), and the α-tubulin-coding region (+737 to +860 relative to the translation initiation codon) using the primer pairs listed in Table S1. For quantification of immunoprecipitated DNA, qPCR was performed using the SsoFast EvaGreen Supermix (BioRad) on a CFX96 cycler (BioRad) according to the manufacturer's recommendations. Two independent ChIP experiments were performed for both XPD-PTP and XPB-R-PTP cell lines. For each amplification and ChIP experiment, triplicate qPCR samples were analyzed using the Bio-Rad CFX Manager software package. The specificity of each amplification product was ensured by agarose gel electrophoresis and melting curve analysis. Standard curves for oligonucleotide pairs were obtained from serial dilution of input DNA. Their coefficient of determination ($R^2$) value was above 0.98. Samples were standardized according to the starting quantities of tubulin DNA in each experiment and the fold enrichment values were calculated as the ratio between starting quantities of positive and control precipitations.

### 3.6 Survival assays

Sensitivity of cells towards DNA damaging agents were tested at a cell density of 2 × $10^6$ cells/ml. For assessing UV irradiation effects, the wildtype (427) and knockout cells were exposed to doses of 300, 500 or 1000 J/$m^2$ using the *Stratalinker® UV Crosslinker* (Stratagene). For Cisplatin (Kartalou and Essigmann) and MMS (Kartalou and Essigmann) sensitivity assays, cells were cultured with Cisplatin ranging from 0.5 to 10 µM and with MMS from 0.0001 to 0.0004% for two days. In each case, treated and untreated control cell cultures were diluted daily to a concentration of 2 × $10^6$ cells/ml. Total cell count was determined after two days and the survival was calculated as a percentage of untreated controls. In order to assess the sensitivity of RNAi cell lines towards UV irradiation, procyclic cell culture at the density of 2 × $10^6$ cells/ml was induced for dsRNA synthesis with 2 µg/ml of doxycycline. After 24 hours, induced and non-induced cells were diluted to a concentration 2 × $10^6$ cells/ml and exposed to 500 and 1000 J/$m^2$ of UV light. The cells were counted 24 hours after exposure and the survival was determined as a percentage of untreated controls.

### 3.7 Light Microscopy

Immunolocalizations were carried out as described previously (Luz Ambrósio et al., 2009). XPB-R-PTP and XPD-PTP were detected with a rabbit polyclonal anti-protein A immune serum (Kartalou and Essigmann) followed by an Alexa 594-conjugated anti-rabbit secondary antibody (Invitrogen) whereas Tbp52-HA was visualized with a mouse monoclonal anti-HA antibody conjugated with fluorescein isothiocyanate (FITC; Sigma). Images were acquired on a Zeiss Axiovert 200 microscope equipped with DAPI (4′,6′-diamidino-2-phenylindole), EGFP, and Texas-Red filters using a 100× (1.3-numerical-aperture) oil immersion objective.

## 4. Results

### 4.1 XPB-specific domains are conserved in both trypanosome XPB and XPB-R.

Genome annotation of *T. brucei* identified two genes, *Tb927.3.5100* and *Tb927.11.16270* (accession numbers are from www.genedb.org; (Logan-Klumpler et al., 2012)), that encode XPB proteins with calculated masses of 105 kDa and 89 kDa, respectively (Berriman et al., 2005). The sequences of these two XPB paralogs are divergent, exhibiting only 24.2% identity and 37.1% similarity between them according to a pair-wise alignment using the EMBOSS Needle server (http://www.ebi.ac.uk/Tools/psa/emboss_needle/). Despite this divergence both proteins are clearly XPB helicases: they harbor all seven Walker motifs of the SF2 superfamily of helicases and they possess an N-terminal XPB domain that interacts with the regulatory TFIIH subunit p52 (Coin et al., 2007; Jawhari et al., 2002), an XPB-specific R-E-D residue loop implicated in recognition of damaged DNA sites (Fan et al., 2006; Oksenych et al., 2009), and an XPB signature sequence motif that was recently identified by a comparative genomics analysis (Figure. 2.1A and 2.1B) (Bedez et al., 2013). Both XPB paralogs are present in all sequenced kinetoplastid genomes, including that of the bodonid *B. saltans*, and a multiple sequence alignment showed that XPB domains are equally well conserved among the paralogs sequences (Attachment 1, Fig.S1). Moreover, the archaeal XPB structure revealed the presence of a Thumb-like domain (ThM) between Walker motifs III and IV that is structurally conserved in human XPB and thought to cooperate with the R-E-D loop in anchoring XPB to sites of DNA damage (Fan et al., 2006; Oksenych et al., 2009). While sequence conservation between human and archaeal ThM domains is very limited, there is clear sequence conservation between the human ThM sequence and the aligned kinetoplastid XPB sequences further supporting the notion that kinetoplastid organisms harbor two functional XPB paralogs (Attachment 1, Fig.S1).

To further verify the nature of XPB-R, we tested whether this protein interacts with its functional partner p52. We created an insect-stage, procyclic *T. brucei* cell line in which XPB-R and p52 (accession number Tb927.10.5210) were C-terminally tagged with the composite PTP tag and the HA tag, respectively. Previously, we have shown that p52-HA interacted with XPD as part of the TFIIH complex and correctly localized to the nucleus, indicating that the tag does not interfere with p52 function (Lee et al., 2007). Similarly, cell proliferation assays (see below) indicated that XPB-R-PTP was functional as well. In whole cell lysates XPB-R-PTP was expressed as a single protein of correct size in two independently derived cell lines. However, in the extract sample, the appearance of a second band, which was 15 kDa smaller than the full length protein, suggested that some of XPB-R-PTP was cleaved at the N-terminus during extract preparation (Attachment 1, Fig.S4). Both bands were precipitated with IgG beads which bind to the protein A domains of the PTP tag (Figure 2.1C). p52-HA was co-precipitated but TSP2, a trypanosome TFIIH subunit partner of p52, was not. In the reciprocal anti-HA immunoprecipitation experiment, full length XPB-R-PTP was efficiently co-precipitated confirming the XPB-R/p52 interaction. The smaller, N-terminally degraded XPB-R-PTP form, however, did not interact with p52-HA which indicated that trypanosome XPB-R does possess a functional N-terminal p52 interaction domain. This negative co-precipitation result confirmed our previous finding (Nguyen et al., 2006) that protein G-bound anti-HA antibody does not co-precipitate PTP-tagged protein non-specifically. As expected, TSP2 did co-precipitate with p52, verifying that both proteins are part of the same TFIIH complex. Together, these results showed that kinetoplastid organisms harbor two highly divergent XPB paralogs with conserved functional domains. Although *T. brucei* XPB-R was not detected in TFIIH complexes that were purified via tagged XPD or TSP2 (Lee et al., 2009), its interaction with p52 suggested that it is a functional enzyme in this parasite.

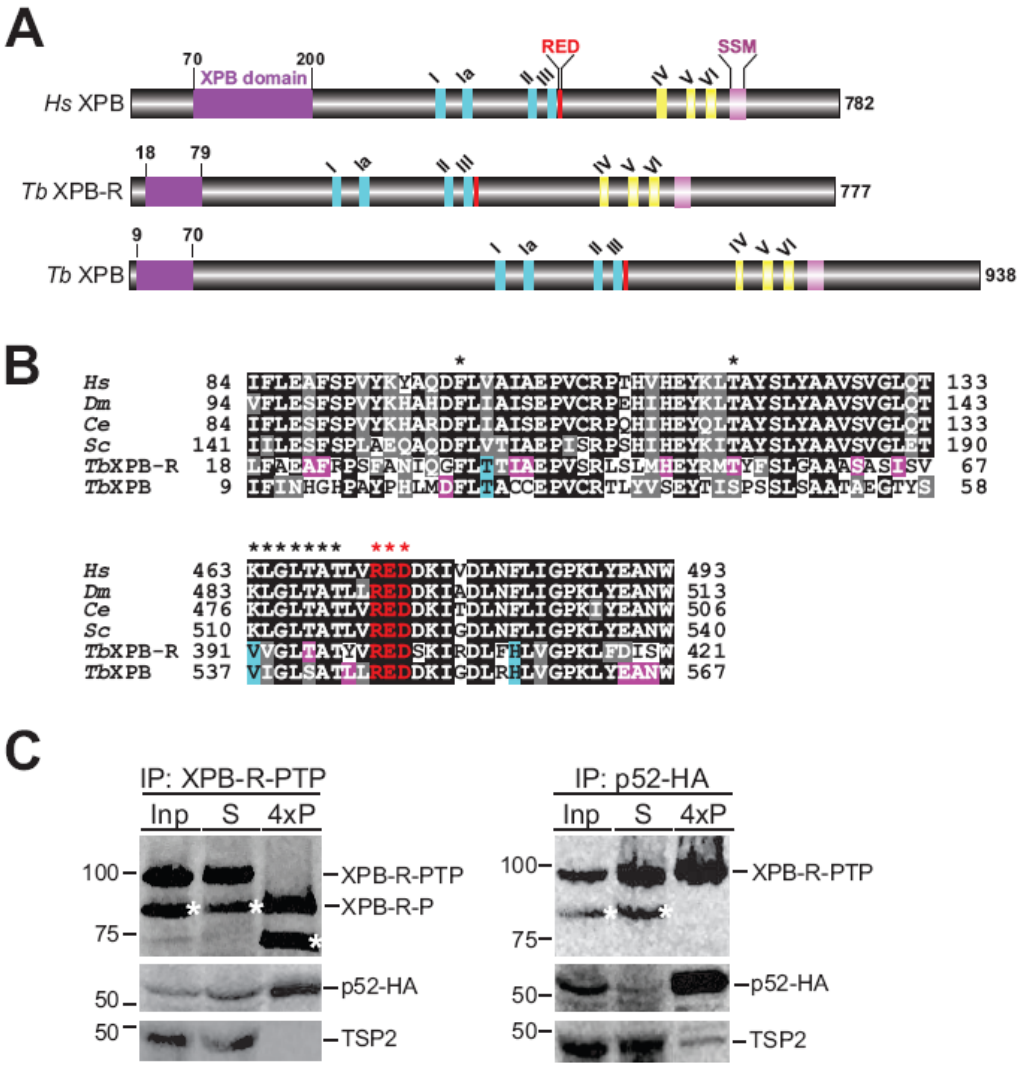



**Figure 2.1 Trypanosomatids possess two different XPB helicases.**

**A**. Schematic (to scale) of human XPB (Hs XPB) and of *T. brucei* (Egloff et al.) XPB-R and XPB showing the XPB and RED domains, Walker motifs I, Ia, II-VI, and an XBP-specific sequence signature motif (Hegele et al.).

**B.** Alignment of XPB and RED domain sequences from humans (Ivens et al.)*, Drosophila melanogaster* (Ivens et al.), *Caenorhabditis elegans* (Ce), *Saccharomyces cerevisiae* (Sc), and *T. brucei*. Positions that are identical and similar across the model organisms are shaded in black and gray, respectively. Specific conservations in either TbXPB-R or TbXPB are shaded in pink and positions that are identical in trypanosome XPBs and different from those of the model organisms are shaded in cyan. Stars in the upper panel identify the phenylalanine and threonine residues found mutated in many *Xeroderma pigmentosum* patients. Stars in the lower panel identify Walker motif III and the R-E-D residue loop.

**C.** Immunoprecipitation (IP) of either XPB-R-PTP (left panels) or p52-HA (right panels).Equivalent amounts of the input extract (Inp) and the supernatant (S) were analyzed whereas four times of the precipitate (4xP) were loaded. XPB-R-PTP was detected with the monoclonal anti-protein C epitope HPC4 antibody, p52-HA with a monoclonal anti-HA antibody, and the TFIIH subunit TSP2 with a polyclonal immune serum. Precipitated XPB-R-PTP was released from beads by TEV protease digest which removed the protein A domains, resulting in the smaller XPB-R-P protein. Asterisks identify the N-terminally shortened XPB-R species seen in extract.

### 4.2 Knockout of *XPB-R* retards trypanosome proliferation in culture.

Previously, *XPB-R* silencing in procyclic trypanosomes did not affect cell proliferation (Lecordier et al., 2007). We confirmed this result (data not shown) and, in addition, analyzed *XPB-R* silencing in bloodstream form trypanosomes by inducing synthesis of *XPB-R* dsRNA from a tetracycline operator-controlled stem-loop construct that had been integrated in tetracycline repressor-expressing single marker cells (Wirtz et al., 1999). Although semi-quantitative reverse transcription (RT)-PCR showed that *XPB-R* mRNA was reduced after one and two days of induction, *XPB-R* silencing did not affect trypanosome proliferation suggesting that *XPB-R* is not an essential gene in either procyclic or bloodstream form trypanosomes (Attachment 1, Fig. S2). Accordingly, we succeeded in generating a procyclic *XPB-R*$^{-/-}$ knockout cell line in which the two *XPB-R* alleles were replaced by the hygromycin (*HYG*$^{R}$) and blasticidin (*BLA*$^{R}$) resistance markers (Figure 2.2A). After the second transfection of a single allele knockout XPB-R$^{+/-}$ cell line, we obtained two independently derived, double resistant *XPB-R*$^{-/-}$ cell lines which we termed KO-1 and KO-2. The gene knockouts were confirmed by a PCR of genomic DNA prepared from wild-type, XPB-R$^{+/-}$ and KO cell lines using primers that annealed to the 5$^{/}$ and 3$^{/}$ *XPB-R* gene flanks outside of the transfected, linear DNA constructs. While the KO cells gained *HYG*$^{R}$- and *BLA*$^{R}$-specific amplification products, they completely lost the wild-type product (Figure 2.2B). Correspondingly, an RT-PCR analysis with oligonucleotides specific to the *XPB-R* coding region showed that the KO cells did not express *XPB-R* mRNA any longer (Figure 2.2C). The successful *XPB-R* knockout demonstrated that this gene was dispensable for cultured procyclics. However, the KO cell lines clearly grew slower than wild-type cells. While our wild-type strain had a doubling time of 8.5 hours, KO-1 and KO-2 cells doubled every 13.8 and 15.6 hours, respectively (Figure 2.2D). This slow growth phenotype remained constant over a period of three months and was not affected by the presence or absence of antibiotics (data not shown). To determine whether the knockout delayed the cell cycle in a particular stage, we

determined the number of kinetoplasts and nuclei in DAPI-stained wild-type and *XPB-R*$^{-/-}$ cells (n=100 for each cell line). Compared to the wild-type cell line, we found a strong accumulation of 2K2N cells (25% versus 6%) and a concomitant decrease of 1K1N trypanosomes (44% versus 68%) indicating a cell cycle-relevant function of XPB-R in late mitosis/cytokinesis (Attachment 1, Fig. S3)

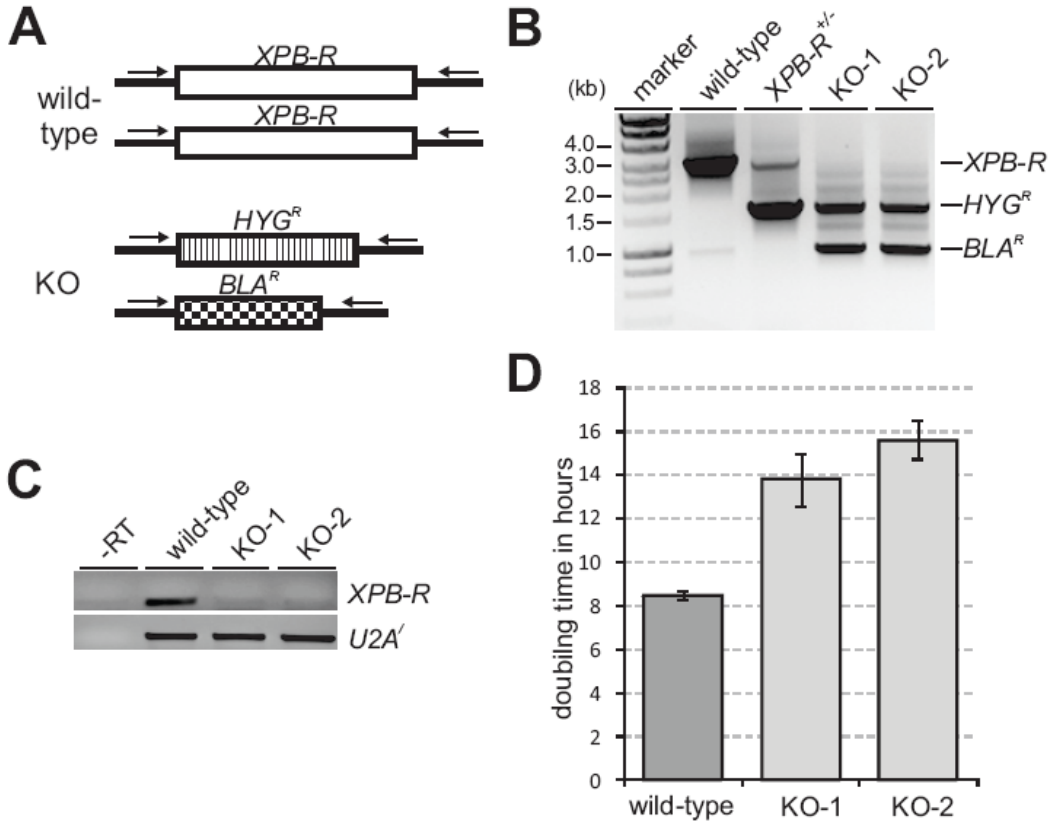



**Figure 2.2 A knockout of *XPB-R* increases the doubling time of trypanosomes in culture.**

**A.** Schematic (not to scale) of the *XPB-R* locus in wild-type and *XPB-R*$^{-/-}$ (KO) cells. In the latter, the *XPB-R* coding regions of the two alleles were replaced with those of hygromycin phosphotransferase (*HYG*$^{R}$) and blasticidin-S deaminase (*BLA*$^{R}$). Oligonucleotides used for competitive PCR analysis of *XPB-R*, *HYG*$^{R}$ and *BLA*$^{R}$ alleles are indicated by arrows.

**B.** Corresponding competitive PCR analysis of total DNA prepared from wild-type cells, cells in which one *XPB-R* allele was knocked out (*XPB-R*$^{+/-}$), and two independently obtained *XPB-R*$^{-/-}$ knockout cell lines (KO-1, KO-2)

**C.** RT-PCR of *XPB-R* and, as a control, of *U2A*$^{/}$ mRNA. A control reaction with wild-type RNA in the absence of reverse transcriptase (-RT) demonstrated effective removal of genomic DNA from the RNA preparation.

**D.** KO-1 and KO-2 cells have a prolonged doubling time versus wild-type cells.

### 4.3 XPB-R does not function in transcription

To our knowledge eukaryotic XPB helicases function exclusively in transcription and NER. It was therefore possible that XPB-R had an important, albeit non-essential, function in transcription, possibly in a regulatory role. Since the *SLRNA* promoter is the only trypanosome promoter known to assemble an RNA pol II transcription pre-initiation complex including TFIIH (Lee et al., 2009), we first tested whether XPB-R occupies the *SLRNA* promoter by chromatin immunoprecipitation (ChIP). We previously showed that chromatin bound by PTP-tagged proteins can be highly enriched by a ChIP-grade, anti-protein A antibody (Lee et al., 2010). Accordingly, when we used a cell line that exclusively expressed functional XPD-PTP (Lee et al., 2007), we observed a 21 fold enrichment of the *SLRNA* promoter over a precipitation with a comparable non-specific immune serum (Figure 2.3A). This occupancy was strongly reduced in the *SLRNA* intergenic spacer approximately 500 bp downstream of the promoter, revealing the specificity of the ChIP assay. To analyze XPB-R in an equivalent way, we generated a cell line that exclusively expressed XPB-R-PTP. This was achieved by targeted integration of plasmid XPB-R-PTP-NEO into the remaining *XPB-R* allele of the single knockout XPB-R$^{+/-}$ cell line (Attachment 1, Fig. S4). To assess whether the PTP tag interferes with XPB-R function, we determined the doubling time of the XPB-R-PTP-expressing cells which, with 11.1 hours, was intermediate between wild-type and XPB-R$^{-/-}$ cells. Interestingly, single knockout XPB-R$^{+/-}$ cells exhibited a similarly prolonged doubling time (12 h) suggesting that maximal trypanosome proliferation requires a sufficiently high level of XPB-R and that the slower proliferation rate of XPB-R-PTP-expressing cells is due to haplo-insufficiency rather than a functionally deleterious PTP tag. Nevertheless, XPB-R-PTP did not occupy the *SLRNA* promoter, indicating that this helicase has no function in pre-initiation complex formation and RNA pol II transcription initiation (Figure 2.3A).

To further validate this notion, we prepared transcriptionally active extract from XPB-R-PTP-expressing cells and depleted the helicase by IgG affinity chromatography. Although the extract was almost fully depleted of XPB-R, it supported transcription of the *SLRNA* promoter template SLins19 equally well as mock-depleted extract did (Figure 2.3B). Accordingly, adding back tandem affinity-purified XPB-R (see below) to the extract had no impact on transcription efficiency.

Finally, we employed permeabilization of wild-type and XPB-R$^{-/-}$ cells to label nascent RNA which, in this system, increases linearly for up to 20 minutes (Ullu and Tschudi, 1990). Due to a very strong synthesis rate, newly synthesized SL RNA can be directly visualized after separating labeled RNA by denaturing PAGE (Figure 2.3C). To assess protein coding gene transcription in these experiments, we hybridized the labeled RNA to dot blots of plasmids that contained whole coding regions (Figure 2.3D). Quantification of SL RNA bands by densitometry and of dots by scintillation counting in three independent experiments did not reveal any nascent RNA synthesis defect in XPB-R$^{-/-}$ cells in either KO cell line (Figure 2.3E). This was true for the RNA pol I transcripts 18S rRNA and *GPEET* procyclin mRNA, for the RNA pol II-synthesized heat shock protein 70, actin and α tubulin mRNAs, and for the RNA pol III-derived threonine tRNA and U6 snRNA. Therefore, we concluded that XPB-R does not function in transcription.

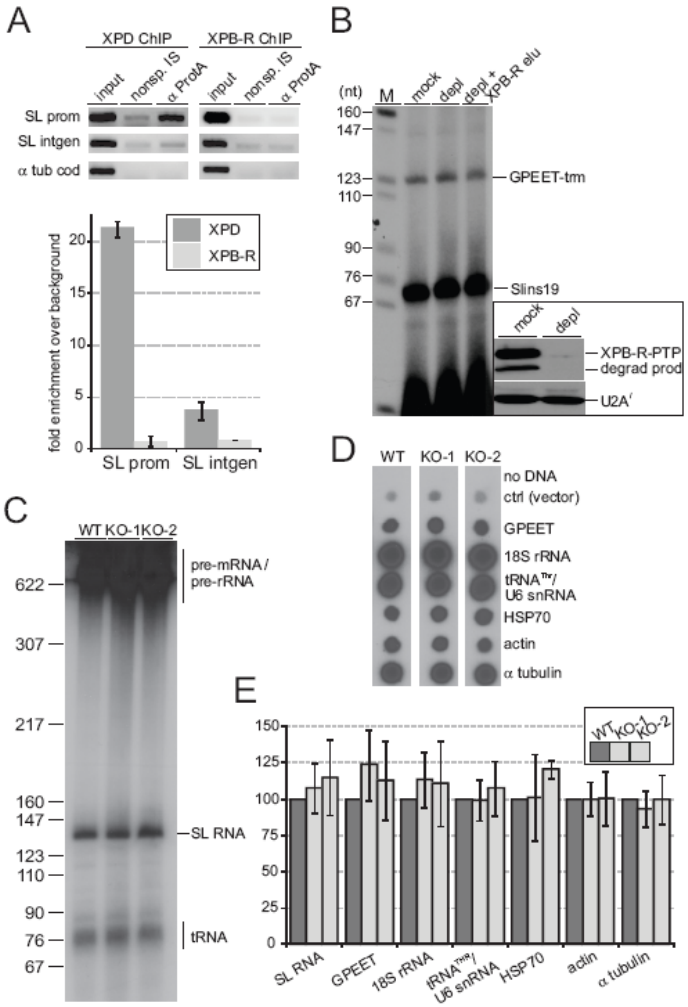



**Figure 2.3 *XPB-R* does not function in transcription.**

**A.** Anti-XPD-PTP and anti-XPB-R-PTP ChIP with a ChIP-grade, polyclonal anti-protein A antibody (α ProtA) and a comparable non-specific immune serum (nonsp. IS). Top panels, semiquantitative PCR of the *SLRNA* promoter (SL prom), the *SLRNA* intergenic region (SL intgen), and the α tubulin coding region (α tub cod). Lower panel, corresponding qPCR analysis showing the corrected fold difference of precipitated DNA over the nonspecific immunoprecipitation control. **B.** Co-transcription of the RNA pol I-recruiting GPEET-trm template DNA and the *SLRNA* promoter template SLins19 in extract prepared from cells that exclusively express XPB-R-PTP. The extract was either mock-treated or depleted of XPB-R-PTP by IgG affinity chromatography (depl). In a third reaction, tandem affinity-purified XPB-R was added to the XPB-R-PTP-depleted extract (depl + XPB-R elu). Main panel, transcription signals were obtained by primer extension of newly synthesized RNA with radiolabeled oligonucleotides specific for GPEET-trm or for SLins19 RNA. Left panel, detection of XPB-R-PTP including its partially degraded product (degrad prod) and, as a loading control, of U2A′ in mock-treated and XPB-R-PTP-depleted extract by immunoblotting. **C.** Nascent RNA was radiolabeled in permeabilized wild-type (Logan-Klumpler et al.), KO-1 and KO-2 cells, separated by denaturing PAGE, and visualized by autoradiography. Signals that correspond to tRNAs, the SL RNA, and pre-mRNA/pre-rRNA are indicated on the right. **D.** Labeled nascent RNA was hybridized to dot blots of plasmids containing the complete coding regions of GPEET procyclin, 18S rRNA, heat shock protein 70 (HSP70), actin and a tubulin. In addition, a plasmid harboring the coupled genes of threonine tRNA and U6 snRNA ($tRNA^{THR}$/U6 snRNA) and a control plasmid without insert (ctrl) were analyzed. **E.** Quantification of RNA signals from three independent experiments. In each case the wild-type signal was set to 100. Error bars represent standard deviations.

### 4.4 XPB-R is important for Nucleotide Excision Repair

NER is essential for the removal of covalently cross-linked bulky DNA adducts that cause a severe distortion of the DNA double helix. Formation of such DNA adducts can be induced by exposure to UV light or by treatment with cisplatin, strategies that have been employed in various systems (Kamileri et al., 2012) including trypanosomes (Sheader et al., 2004). Ultraviolet (Van Hellemond et al.) radiation causes two major types of DNA lesions, namely *cis-syn* cyclobutane-pyrimidine dimers and 6,4 pyrimidine-pyrimidine photoproducts, both of which distort the helical structure of DNA (Thoma, 1999). To determine the role of XPB-R in the repair of UV-induced DNA lesions, wild-type and XPB-$R^{-/-}$ cells were treated with increasing doses of UV radiation. (Figure 2.4A) Two days after the exposure, cell densities were determined and compared to those of untreated control cells. Even at the lowermost dose of 300 J/$m^2$, XPB-$R^{-/-}$ cells showed only 17.4% survival as compared to untreated XPB-$R^{-/-}$ cells while 75.3% of wild-type cells that had received the same dose survived relative to the untreated wild-type control. This indicated that the repair of DNA lesions that require NER is strongly impaired in XPB-$R^{-/-}$ cells. To validate this result, wild-type and XPB-$R^{-/-}$ cells were exposed to increasing concentrations of cisplatin, a chemical that induces the formation of bulky DNA lesions. The majority of cisplatin-induced lesions are 1,2-intrastrand crosslinks between adjacent guanines or between guanine and adenine both of which are efficiently removed by NER (Kartalou and Essigmann, 2001). As shown in figure 2.4B, XPB-$R^{-/-}$ cells were more sensitive towards cisplatin treatment than wild-type cells. For example, while treatment with 1 µM of cisplatin reduced wild-type cell survival to 65.5%, only 33.5% of XPB-$R^{-/-}$ cells survived the same treatment. Taken together, the increased sensitivity of XPB-$R^{-/-}$ cells towards UV radiation and cisplatin strongly indicated direct involvement of XPB-R in NER.

To assess whether XPB-R's role in DNA repair is specific to NER, cells were treated with methylmethane sulfonate (MMS), an alkylating agent that methylates DNA on $N^7$-deoxyguanine

and $N^3$-deoxyadenine. MMS causes formation of DNA base lesions that are repaired by the base excision repair (BER) pathway. If left unrepaired, these MMS-induced lesions can cause stalling of replication forks and lead to single or double strand DNA breaks (Brem et al., 2008). Despite using a range of different MMS concentrations, we did not detect a significant difference between XPB-$R^{-/-}$ cells and wild-type cells in their sensitivity towards MMS (Figure 2.4C). We therefore concluded that XPB-R is specifically important for NER and not required for BER.

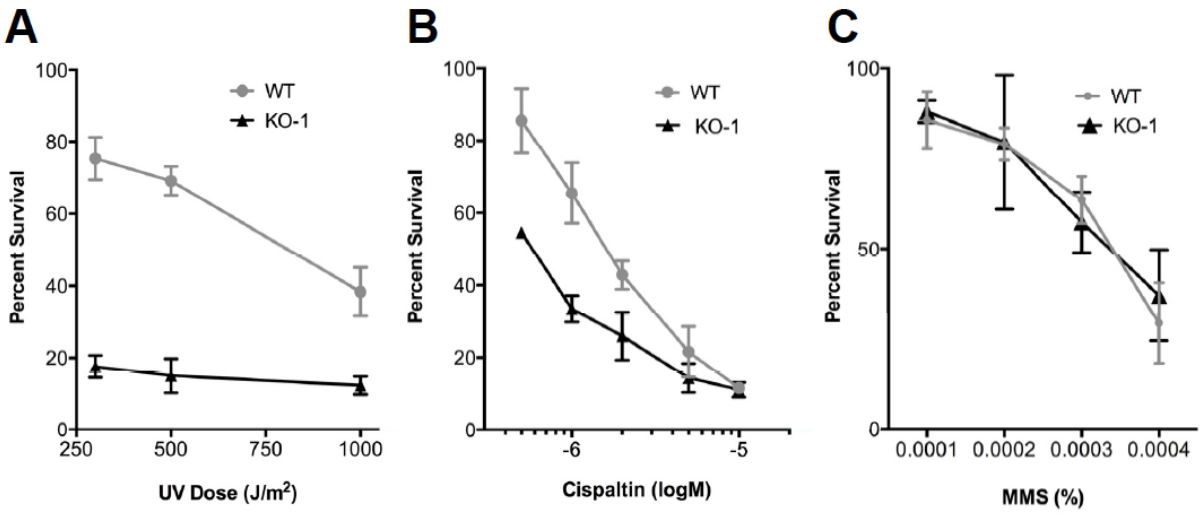



**Figure 2.4 *XPB-R*$^{-/-}$ cells are specifically sensitive to DNA lesions requiring NER.**

Wild-type (Logan-Klumpler et al.) and KO-1 cell survival after treatment with **(A)** UV light, **(B)** cisplatin or **(C)** methyl methanesulfonate (MMS) was determined relative to untreated wild-type and KO-1 cells, respectively. The results were obtained from three independent experiments. Error bars represent standard deviations.

### 4.5 An XPB-R/p52 complex functions independently of a TFIIH complex

Previously, the *T. brucei* TFIIH complex was isolated and mass spectrometrically analyzed in four independent experiments: it was purified using XPD as bait by two different research groups (Lee et al., 2009; Lee et al., 2007) or through tagging TSP2 (Lee et al., 2009), and it was co-purified as an associated complex of the multi-subunit mediator complex (Lee et al., 2010). In each case, XPB was identified as a TFIIH subunit whereas XPB-R was never found. Although unlikely, it is possible that XPB-R is part of a distinct TFIIH complex of minor abundance that was not detected in these biochemical protein complex characterizations. To facilitate XPB-R detection in isolated TFIIH complexes, we tried unsuccessfully to raise a specific polyclonal immune serum in rats against two partial, recombinant XPB-R proteins that were fused to the GST tag and purified from *Escherichia coli* (data not shown). Next, we carried out a tandem affinity purification of XPB-R-PTP(Attachment 1, Fig. S4). Although the purification was efficient and the final eluate revealed several distinct protein bands, mass spectrometry identified only proteins that typically co-purify as contaminants such as α/β tubulin and HSP70 but no other proteins (Attachment 1, Figs. S4C to S4E). Since p52, which did interact with XPB-R-PTP in reciprocal co-IPs (Figure 2.1C), was not detected in this analysis, it was possible that tagging XPB-R at the C-terminus had a deleterious effect on the XPB-R/p52 interaction. It therefore appeared that either XPB-R tagging partially interfered with the XPB-R/p52 interaction, or that this interaction is weak and does not withstand purification. To obviate the latter concern, we sedimented crude extract, freshly prepared from cells expressing XPB-R-PTP and p52-HA, through a linear sucrose gradient by ultracentrifugation, a process which is much less stringent than tandem affinity purification. An immunoblot analysis of proteins collected from gradient fractions that were taken from top to bottom showed that p52-HA sedimentation peaked in fractions 6/7 and in fractions 14-16, between the 17S apoferritin and the 19S thyroglobulin markers (Figure 2.5A). The latter sedimentation peak corresponded to the previously

determined ~17.5 S sedimentation coefficient of the TFIIH complex (Lee et al., 2009) and, accordingly, the TFIIH subunit TSP2 did co-peak with p52-HA in these fractions. The additional peak of TSP2 in fraction 11 likely corresponds to a XPB/TSP1/TSP2 subcomplex that we previously found missing from a partial TFIIH complex (Lee et al., 2009). Importantly though, neither band of XPB-R-PTP showed a peak in fractions 14-16 and they were, in fact, hardly detectable in these fractions, further supporting the notion that XPB-R does not function within a TFIIH complex. Its sedimentation in fractions 5 and 6 suggested that it peaks there by itself and/or in a complex with p52-HA. Together, these data strongly indicated that XPB-R is not part of a TFIIH complex comprising XPD and the other trypanosome TFIIH orthologs of p62, p44, and p34.

To corroborate these findings at the functional level, we generated procyclic cell lines for conditional *p52* silencing by RNAi analogously to the *XPB-R* RNAi cell lines described above. Examination of three independently derived cell lines demonstrated that cultures stopped growing 3 days after induction of *p52* dsRNA synthesis (Figure 2.5B). These growth defects closely resembled those observed previously upon silencing of *XPD* (Lee et al., 2007), indicating that p52 is an essential component of TFIIH. To assess whether RNAi-mediated depletion of either XPD, XPB-R or p52 in cells led to increased UV susceptibility, we silenced the expression of the respective genes in procyclic cell lines, exposed induced and non-induced cells to 500 or 1000 J/m2 UV light, and recorded cell survival one day after the treatment. While *XPD* silencing was as efficient as previously published (data not shown), UV exposure did not affect the survival rate of *XPD*-silenced cells (Figure 2.5C, top panel), strongly indicating that trypanosome TFIIH is specialized in transcription and, unlike its bifunctional counterparts in yeast and mammals, not involved in DNA repair. Conversely and as expected, *XPB-R* silencing significantly increased the UV susceptibility of trypanosomes (Figure 2.5C, middle panel). Since UV treatment was similarly lethal to *p52*-silenced cells (Figure 2.5C, bottom panel), these

results identify p52 as a functional partner of XPB-R in DNA repair and as the only trypanosome protein so far that is likely involved in both RNA pol II transcription and NER. They also confirm our co-IP results (Figure 2.1C).

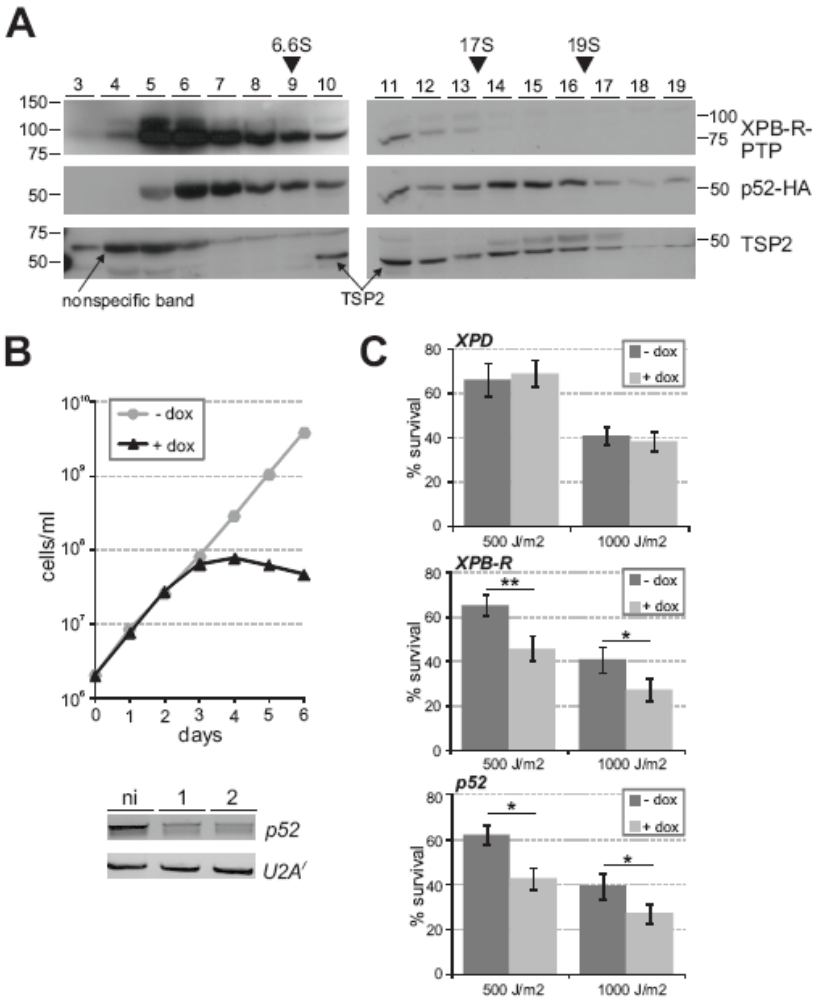



**Figure 2.5 XPB-R and p52 but not XPD function together in NER.**

**A.** Crude extract from XPB-R-PTP-expressing trypanosomes was sedimented through a linear 10-40% sucrose gradient by ultracentrifugation. Fractions were taken from top to bottom and analyzed by immunoblotting as described above. As sedimentation markers IgG (6.6S), apoferritin (17S) and thyroglobulin (19S) were co-analyzed.

**B.** Upper panel, growth curve of a representative, procyclic cell line (one out of three) in which doxycycline induces *p52* silencing. Lower panel, semi-quantitative RT-PCR analysis of p52 mRNA and, as a control, of U2A′ mRNA in total RNA prepared from non-induced cells and from cells that were induced for one or two days.

**C.** Diagrams showing the survival of cells treated with UV relative to untreated cells. Three cell lines were investigated in which *XPD* (upper panel), *XPB-R* (middle panel) or *p52* (lower panel) were silenced (+ dox) or not (- dox). Each experiment was conducted three times, and the results were statistically analyzed by an unpaired, two-tailed student's t test assuming unequal variances. One and two asterisks indicate p values that are smaller than 0.05 and 0.01, respectively.

### *4.6* XPB-R-PTP is localized in the nucleus and partially co-localizes with p52-HA

NER factors including various TFIIH subunits typically exhibit a diffuse nuclear staining pattern (Hoogstraten et al., 2008; Vermeulen et al., 2000). We have previously shown that in *T. brucei* p52-HA localizes to the nucleus with the strongest accumulation in one or two perinucleolar foci. These foci most likely represent the ~100 tandemly linked *SLRNA* copies on chromosomes 9 that are transcribed monocistronically, thereby requiring the independent formation of transcription pre-initiation complexes on each and every *SLRNA* promoter (Lee et al., 2010). Accordingly, when we used a cell line in which the TFIIH helicase XPD and p52 were PTP- and HA-tagged, respectively (Lee et al., 2007), we observed that XPD-PTP predominantly localized to one or two perinucleolar foci in the nucleus with additional faint staining of other nuclear regions (Figure 2.6 and Attachment1 Fig.S5, left panels). p52-HA invariably co-localized with XPD-PTP in these foci presumably as part of the TFIIH complex, yet it also appeared to be more widely distributed than XPD within the nucleus. When we co-localized XPB-R-PTP with p52-HA in the cell line described above (Figure 2.6 and Attachment 1, Fig. S5, right panels), we detected a partial overlap of the two proteins which supports our finding that XPB-R and p52 do interact. On the other hand, the clearest difference from the XPD analysis was that XPB-R-PTP most often did not co-localize with the brightest p52-HA spots, supporting our finding that XPB-R does not occupy the *SLRNA* promoter and is not required for RNA pol II transcription initiation at *SLRNA* genes. These results support a model in which trypanosomes possess a TFIIH complex with a basal function in RNA pol II transcription initiation, and an XPB-R/p52 complex specifically functioning in NER.

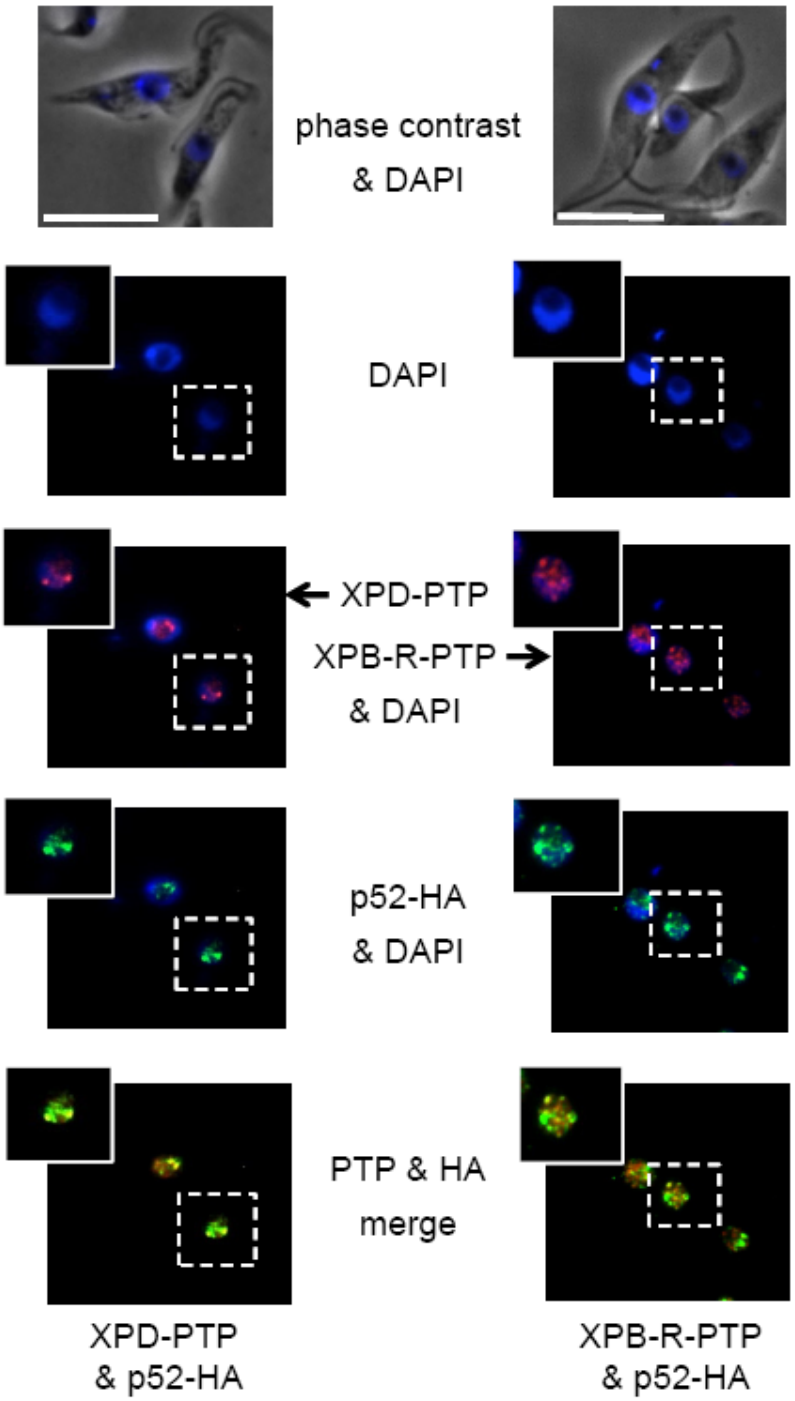



**Figure 2.6 XPB-R-PTP co-localizes with p52-HA outside putative *SLRNA* expression foci in the nucleus.** p52-HA (green) was co-localized with XPD-PTP (red; left panels) or with XPB-R-PTP (red, right panels). White bars represent 10 μm.

### *4.7* Highly divergent *XPB* paralogs are present in other protistan lineages

We have shown that *T. brucei* possesses two highly divergent XPB paralogs. One participates in the transcriptionally active TFIIH complex while the other is important for DNA repair and, likely, functions independently of TFIIH. Since TFIIH analyses have mainly been restricted to yeast and higher eukaryotes, we wanted to know whether divergent XPB paralogs could be found in other eukaryotes. We carried out a phylogenetic analysis of *XPB* sequences of 33 species whose genomes were completely sequenced (Attachment 1, Fig. S6). Each XPB sequence that was retrieved gave an E value smaller than $e^{-95}$ during a protein-protein BLAST analysis against the human XPB sequence and they all harbored the R-E-D motif, except for one of the two sequences found in the ciliate *Paramecium tetraurelia* that contained an R-Q-D motif instead (Attachment 1, Fig. S1). Despite the small E-values, multiple sequence alignment revealed several blocks of ambiguous alignment due to poorly conserved regions. To restrict our analysis to appropriately aligned sequences we applied the Gblock algorithm (Talavera and Castresana, 2007) in combination with maximum likelihood calculations. As expected, metazoa, yeasts and fungi have only a single XPB sequence which correlates with the dual function of TFIIH described in these organisms. The two plant species investigated, *Oryza sativa* and *Arabidopsis thaliana*, do possess two *XPBs* but their encoded sequences are nearly identical and more closely conserved within each species. This strongly suggests that these *XPB* genes stem from recent gene duplications that have not functionally diverged. In contrast, all kinetoplastid organisms possess two highly divergent XPBs that clearly segregate into two separate branches, one with *T. brucei* XPB and one with *T. brucei* XPB-R. Even the only bodonid species investigated, *B. saltans*, encoded unambiguously identifiable orthologs of these two helicases. The root of the two clusters is very deep, raising the possibility that the presence of two functionally distinct XPB helicases is an ancient eukaryotic trait. This hypothesis is supported by the presence of similarly divergent XPB paralogs in the ciliates *Tetrahymena*

*thermophila* and *P. tetraurelia* although the ciliate XPB sequences do not form two separate clusters. Ciliates are grouped together with Apicomplexa in the supergroup Chromalveolata. However, all apicomplexan species investigated have only a single *XPB* gene. On the other hand, *Entamoeba histolytica and Entamoeba dispar* again have two highly divergent *XPB* genes that form two deeply rooted branches, while other members of the supergroup Amoebozoa, *Dictyostelium discoideum* and *Acanthamoeba castellanii,* have only a single *XPB* gene. Finally, among early diverged protists, the *Giardia lamblia* genome revealed only a single *XPB* gene whereas *Trichomonas vaginalis* and *Naegleria gruberi* encode two divergent XPB paralogs, although their roots are not as deep as those of ciliates and *Entamoeba*.

If distinct XPBs for transcription and DNA repair were a plesiomorphic trait in eukaryotes, we would expect a phylogenetic tree with separate XPB-R and XPB branches. This was never observed despite applying several different phylogenetic approaches (data not shown). On the other hand, the bootstrap values at the base of the tree in Attachment 1, fig. S6 are not high and it is possible that specific XPB-R and XPB branches do exist. Accordingly, when we carried out the same phylogenetic analysis with an archaeal sequence as an outgroup, the basal branching of the tree changed while the XPB and XPB-R dichotomies described here remained (Attachment 1, Fig. S7). Therefore, it will require more whole genome sequences from protistan organisms and a more comprehensive approach to solve the basal eukaryotic branching of the XPB tree unambiguously. Until then our data suggest that the presence of two functionally distinct XPBs is a phenomenon that exists beyond the Kinetoplastida.

## 5. Discussion

This is the first report of a eukaryotic XPB paralog that is specialized in DNA repair while having no role in transcription. We showed that transcription of genes by all three nuclear RNA pols was not impaired in XPB-R$^{-/-}$ cells. Moreover and in contrast to TFIIH subunits, XPB-R did

not occupy the *SLRNA* promoter, the only trypanosome promoter known to assemble a RNA pol II transcription pre-initiation complex, and depletion of XPB-R from extract did not affect *SLRNA* transcription *in vitro*. On the other hand, removal of the *XPB-R* gene rendered trypanosomes highly sensitive to UV irradiation and cisplatin, but not to methylmethane sulfonate, strongly indicating that XPB-R, like its XPB homologs from yeast to humans, specifically functions in NER and not in other DNA repair pathways. Interestingly, although not absolutely required for cell viability, the expression level of XPB-R directly influenced the trypanosome doubling time. Knockout of a single *XPB-R* allele increased the 8.5 h doubling time of wild-type to 12 h and a complete knockout to ~14.7 h. Given that a specific down-regulation of XPB-R would result in slow growth and impaired DNA repair, we speculate that regulation of its expression may play a role in pathways of programmed cell death that have been described in trypanosomes (Michaeli, 2012; Welburn et al., 2006).

An interesting question emanating from our study is why XPB-R$^{-/-}$ cells take longer to double than wild-type cells. One possibility is that XPB-R$^{-/-}$ cells accumulate mutations that lead to increased cell death thus slowing culture growth. If this was the case, one would expect that XPB-R$^{-/-}$ cell lines would grow slower over time because all cells would accumulate mutations simultaneously. However, the doubling time of the knockout cells did not change over a period of ~150 generations. In addition, we did not detect a substantial increase of aberrant or dead cells in XPB-R$^{-/-}$ cell cultures when compared to wild-type cultures suggesting that accumulation of mutations is not the major cause of the growth defect (Attachment 1, Fig. S3 and data not shown). Instead, our results indicated that XPB-R$^{-/-}$ cells take substantially longer to complete mitosis and/or initiate cytokinesis because a quarter of XPB-R$^{-/-}$ cells, compared to 6% of wild-type cells, was found to be in the precytokinesis stage with two kinetoplasts and two nuclei. Given that XPB-R is a DNA-dependent helicase, this finding suggests that XPB-R has a role in late mitosis. While this could be a trypanosome-specific function, a recent study did find human

XPB but not other TFIIH subunits to be associated with centrosomes and adjacent parts of the mitotic spindle during mitosis (Weber et al., 2010).

Thus far, eukaryotic XPB function has exclusively been described in the context of the TFIIH complex (Compe and Egly, 2012; Naegeli and Sugasawa, 2011). The two TFIIH subunits that interact with XPB and modify its enzymatic activities are p52 and p8 (Coin et al., 2007; Coin et al., 2006; Fregoso et al., 2007). Accordingly, our analysis showed that trypanosome XPB-R does interact with p52, and while trypanosomes do possess a p8 ortholog that is part of the TFIIH complex (Lee et al., 2009), it remains to be determined whether it interacts with XPB-R. However, our study strongly indicates, in accordance with previous comprehensive TFIIH characterizations (Lecordier et al., 2007; Lee et al., 2009; Lee et al., 2007), that XPB-R does not assemble into a TFIIH complex. In the sedimentation analysis, tagged XPB-R was found almost exclusively near the top of the sucrose gradient and did not exhibit faster co-sedimentation with the TFIIH subunits p52-HA and TSP2. Accordingly, TSP2 did not co-immunoprecipitate with XPB-R from extract. Furthermore, isolation of XPB-R by tandem affinity purification did not co-purify any TFIIH subunit suggesting that tagged XPB-R is not assembled into a stable protein complex. Finally, while the TFIIH subunits XPD and TSP2 were predominantly localized in one or two perinucleolar spots, XPB-R exhibited a more diffuse nuclear staining and rarely co-localized with p52-HA in its perinucleolar spots, further indicating that XPD and XPB-R are not part of the same complex.

How does XPB-R function in NER in the absence of other TFIIH subunits? There are two distinct pathways of NER, namely transcription-coupled NER and global genome NER, which use different proteins to detect DNA lesions and recruit TFIIH to the sites of damage (Naegeli and Sugasawa, 2011). In the Tritryp genome projects, most of the factors involved in NER factors have been identified suggesting that both pathways are functional in trypanosomes (Berriman et al., 2005; Ivens et al., 2005); recently reviewed in (Passos-Silva et al., 2010). A key

factor in global genome NER is XPC (accession number for the *T. brucei* ortholog is Tb927.9.11930) which, by directly interacting with XPB, recruits TFIIH to sites of DNA damage. Moreover, it was shown in the human system that the R-E-D residue loop, the ThM domain and the ATPase activity of XPB are important for recruitment of TFIIH to damaged sites (Oksenych et al., 2009). All of these domains are conserved in XPB-R suggesting that it can be recruited to DNA lesions by XPC in the absence of a TFIIH complex. Transcription-coupled NER removes DNA lesions that stall RNA pol II at actively transcribed genes (Hanawalt and Spivak, 2008; Laine and Egly, 2006). A key factor in this pathway is the so-called Cockayne syndrome type B or CSB protein (Tb927.7.4080), which is important for the recruitment of TFIIH and other NER factors to damaged DNA. Although it is not entirely clear how TFIIH is recruited in this pathway, early evidence revealed a direct link between CSB and XPB (Tantin, 1998), which would allow direct recruitment of XPB-R by trypanosome CSB for transcription-coupled NER.

After recruitment to DNA lesions, the function of TFIIH in global genome NER is to open up the site of damage by XPB ATPase and XPD helicase activity. The ATPase activity of XPB apparently drives the protein like a wedge into the damaged bubble enabling XPD, whose helicase activity alone is important for NER, to unwind the DNA (Coin et al., 2007). Since trypanosome XPB-R apparently does not function together with XPD, the XPB-R ATPase-driven opening of the damaged DNA site may be sufficient in trypanosomes for nucleotide repair. Alternatively, XPB-R may be a more processive helicase than its mammalian counterpart whose activity is important for NER. Mammalian XPD has also been implicated in DNA damage recognition and verification (Mathieu et al., 2010). Since there is no evidence that trypanosome XPB-R and XPD form a common complex, this function is either absent in trypanosomes or achieved independently of XPB-R. In the absence of a stable XPB-R complex, it appears that XPB-R functions like its archaeal ortholog. *In vitro* assays have shown that XPB of the archaeon *Sulfolobus solfataricus* can open up DNA lesions by 6-8 bp which is sufficient for the

endonuclease Bax1 to cleave off the damaged DNA (Rouillon and White, 2010). Bax1 is thought to be the functional homolog of the eukaryotic NER endonuclease XPG (Rouillon and White, 2010) which is encoded in the *T. brucei* genome (Tb927.9.11760; E value of $8e^{-20}$ derived from a protein-protein blast with the human XPG amino acid sequence) and likely functions together with XPB-R. In sum, it appears that trypanosomes have a simplified NER machinery that is separate from the transcriptionally essential TFIIH complex.

We have been able to combine gene silencing with UV exposure to assess the function of proteins in NER. This approach clearly demonstrated that *XPD* silencing did not alter cell survival after UV exposure whereas *XPB-R* and *p52* silencing clearly increased the UV susceptibility of cells (Figure 2.5C). These data strongly support a model in which trypanosome TFIIH has an essential role in RNA pol II transcription while NER is carried out through a simplified, XPB-R-based machinery. This dichotomy in trypanosomes provides a unique opportunity to decipher specific DNA repair and transcription functions of XPB. For example, amino acids specifically required for DNA repair should be conserved in kinetoplastid XPB-R sequences as well as in the bifunctional XPBs of model organisms but not in kinetoplastid XPBs. Accordingly, several such positions were identified in the XPB multiple sequence alignment (Attachment 1, Fig. S1). Moreover, the bifunctional nature of human and yeast XPB has hampered XPB analysis *in vivo* since all mutations affecting the transcriptional function of XPB are lethal, an obstacle absent in future XPB-R analyses.

Finally, specific NER machinery may not be restricted to Kinetoplastida. Our survey of completely sequenced genomes revealed the presence of two highly divergent XPBs in several protistan taxa that belong to different supergroups (Dacks et al., 2008). In addition to the Excavata, the supergroup of kinetoplastids, highly divergent XPBs were found in the Ciliophora of the supergroup Chromalveolata and in the genus *Entamoeba* which is part of the Amoebozoa supergroup. Although our initial XPB phylogenetic analysis did not provide unambiguous

branching at the base of the tree, which may be the reason why we did not observe an overall dichotomy of XPB and putative XPB-R sequences, our results raise the possibility that a DNA repair-dedicated XPB helicase is an ancestral eukaryotic trait.

# Chapter III

# Trypanosome cdc2-related kinase 9 controls spliced leader RNA cap4 methylation and phosphorylation of the RNA polymerase II subunit RPB1

## 1. Abstract

Conserved from yeast to mammals, phosphorylation of the heptad repeat sequence $Tyr^1$-$Ser^2$-$Pro^3$-$Thr^4$-$Ser^5$-$Pro^6$-$Ser^7$ in the carboxy-terminal domain (CTD) of the largest RNA polymerase II subunit RPB1 mediates the enzyme's promoter escape and binding of RNA processing factors such as the $m^7G$ capping enzymes. The first critical step, $Ser^5$ phosphorylation, is carried out by cyclin-dependent kinase 7 (CDK7), a subunit of the basal transcription factor TFIIH. Many early-diverged protists such as the lethal human parasite *Trypanosoma brucei*, however, lack the heptad repeats and, apparently, a CDK7 ortholog. Accordingly, characterization of trypanosome TFIIH did not identify a kinase component. The *T. brucei* CTD, however, is phosphorylated and essential for transcription. Here we show that silencing the expression of *T. brucei* cdc2-related kinase 9 (CRK9) led to a loss of RPB1 phosphorylation. Surprisingly, this event did not impair RNA pol II transcription or co-transcriptional $m^7G$ capping. Instead, we observed that *CRK9* silencing led to a block of spliced leader (SL) *trans* splicing, an essential step in trypanosome mRNA maturation, that was caused by hypomethylation of the SL RNA's unique cap4.

## 2. Introduction

In eukaryotes, RNA polymerase II transcription of protein coding and small nuclear RNA genes is a highly conserved process involving a plethora of transcription and co-transcriptionally active RNA processing factors. A key regulatory structure in this process is the carboxy-terminal domain (CTD) of the largest RNA pol II subunit RPB1. Conserved from some protists to humans, the CTD consists of 5 to 52 repeats of the heptad sequence $Tyr^1$-$Ser^2$-$Pro^3$-$Thr^4$-$Ser^5$-$Pro^6$-$Ser^7$ whose regulated serine phosphorylation cycle is essential for RNA pol II function (Buratowski, 2009; Chapman et al., 2008; Hsin and Manley, 2012). RNA pol II that is recruited to the core promoter has hypophosphorylated heptad repeats. Transcription initiation and RNA pol II escape from the promoter then essentially depends on phosphorylation of $Ser^5$ by cyclin-dependent kinase 7 (CDK7; the yeast ortholog is termed Kin28) which is a component of the basal transcription factor TFIIH. The same phosphorylation event leads to binding of the $m^7G$ capping enzyme to the CTD and co-transcriptional capping of pre-mRNA (Fabrega et al., 2003). Further downstream of the transcription initiation site, phosphorylation of $Ser^2$ by CDK9 (Bur1) and CDK12/13 (Ctk1) transform RNA pol II into an actively elongating enzyme (Karagiannis and Balasubramanian, 2007; Liu et al., 2009; Qiu et al., 2009). Finally, $Ser^7$ phosphorylation appears to be particularly important for transcription of snRNA genes since it mediates the interaction of the integrator 3′ end RNA processing complex to RNA pol II. $Ser^7$ is primarily phosphorylated by CDK7 (Akhtar et al., 2009; Glover-Cutter et al., 2009; Kim et al., 2009) although CDK9 (Kim et al., 2009; Tietjen et al., 2010) and DNA-activated protein kinase (Egloff et al., 2010) may play a role as well.

The repetitive CTD structure, however, is not a universal feature of eukaryotes (Guo and Stiller, 2004). Notably, protistan organisms such as amoebozoa, kinetoplastids, trichomonads and diplomonads, which appear to have diverged very early from other eukaryotic lineages, lack repetitive CTD motifs. Moreover, comparative genomics suggested that these organisms also

lack a CDK7 ortholog (Guo and Stiller, 2004). It therefore remains to be determined whether phosphorylation of non-repetitive CTDs is important for RNA pol II-mediated gene expression or whether the essential coordinating function of the canonical CTD has evolved later in evolution.

The kinetoplastid *Trypanosoma brucei* is a lethal, vector borne parasite of humans and livestock in sub-Saharan Africa. It has emerged as a model organism for deviating gene expression mechanisms found in early-diverged eukaryotes. Its protein coding genes are arranged in long directional tandem arrays which are transcribed polycistronically. The resulting precursor RNAs are processed into individual mRNAs by spliced leader (SL) *trans* splicing and polyadenylation. The SL is derived from the small nuclear SL RNA and *trans* spliced onto each and every mRNA (Gunzl, 2010). The SL is 39 nt-long, pseudouridylated at position 28, and carries the most extensively modified cap found in any eukaryote. The cap structure is $m^7G(5')ppp(5')m^{6,6}AmpAmpCmpm^3Ump$ and was termed cap4 (Bangs et al., 1992). Trypanosome RNA pol II transcribes the protein coding gene arrays and the ~100 tandem copies of the SL RNA gene (*SLRNA*) on chromosome 9 whereas other small nuclear RNAs in trypanosomes are synthesized by RNA pol III. *SLRNA* transcription is monocistronic, initiates at a distinct initiation site, and requires the assembly of a conventional, albeit extremely divergent pre-initiation complex (Das et al., 2008; Lee et al., 2010; Lee et al., 2009). The *T. brucei* genome encodes a full set of RNA pol II subunits, termed RPB1 to RPB12 (Kelly et al., 2005), and biochemical characterization of the enzyme complex identified all subunits except RPB10 (Das et al., 2006; Devaux et al., 2006; Martinez-Calvillo et al., 2007). In *T. brucei* there are two slightly different *RPB1* genes (relevant *T. brucei* gene accession numbers are listed in supplementary Table S1) which appear to be functionally redundant since deletion of three *RPB1* alleles did not impair trypanosome proliferation in culture (Ruan et al., 2004). All RPB1 homologs possess eight highly conserved regions, termed A-H, and the sequence downstream of region H is generally considered to be the CTD (Chapman et al., 2008). The conserved

regions can be clearly recognized in trypanosome RPB1 defining the CTD from residue 1481 to residue 1765. Despite the lack of the heptad repeat motif, this region is serine-rich, phosphorylated and essential for transcription (Chapman and Agabian, 1994; Das and Bellofatto, 2009).

Characterization of the trypanosome phosphoproteome revealed 14 phosphorylated RPB1 residues which all reside in the CTD (Nett et al., 2009), yet the functional significance of CTD phosphorylation in trypanosomes remains unclear. In accordance with the notion that organisms with a non-canonical CTD lack a CDK7 ortholog, biochemical and structural characterization of trypanosome TFIIH, which is a basal factor for *SLRNA* transcription, revealed a full core complex of seven subunits but no kinase component (Lee et al., 2009; Lee et al., 2007). These findings suggested that RNA pol II transcription initiation in trypanosomes does not depend on TFIIH-based CTD phosphorylation as it does in eukaryotes with canonical CTDs. Moreover, CTD phosphorylation seems not to be important for co-transcriptional RNA processing either. Since SL *trans* splicing uncouples mRNA capping from RNA pol II transcription, trypanosomes can utilize RNA pols other than RNA pol II to produce functional mRNA. *T. brucei* has evolved a multifunctional RNA pol I which transcribes the large ribosomal gene units as in other eukaryotes and, in addition, gene units that encode their major cell surface proteins procyclin and variant surface glycoprotein (Gunzl et al., 2003). In addition, genetically modified trypanosome cell lines that express T3 or T7 RNA polymerase are able to produce functional mRNA from the corresponding phage promoters (Wirtz et al., 1994). Thus, the use of promoters for different RNA pols showed that processing of a reporter pre-mRNA was not coupled to RNA pol II (Stewart et al., 2010). On the other hand, there is strong evidence that $m^7G$ capping and cap4 formation of the SL RNA occurs co-transcriptionally (Hury et al., 2009; Mair et al., 2000b).

Canonical CTDs are primarily phosphorylated by CDKs. *T. brucei* harbors 11 enzymes of this class termed cdc2-related kinase (CRK)1-4 and CRK6-12. Sequence divergence

prevents an unambiguous assignment of these enzymes to putative counterparts in other eukaryotes although CRK3 appears to be the functional homolog of CDK1 (Hammarton, 2007; Parsons et al., 2005). Here, we show that silencing *CRK9* expression leads to both a loss of RPB1 phosphorylation and a *trans* splicing defect due to hypomethylation of the first four SL RNA nucleotides. Most interestingly and consistent with a CDK7-less TFIIH, loss of RPB1 phosphorylation did not impair RNA pol II transcription or $m^7G$ capping of SL RNA.

## 3. Materials and methods

### 3.1 DNAs

For *CRK9, CRK1, CRK3* and *CRK7* silencing, the coding regions from position 9 to position 508, from position 5 to position 505, from position 431 to position 930, and from position 6 to 506, respectively, were integrated in a stem-loop arrangement (sense-stuffer-antisense) into the pT7-stl construct (Brandenburg et al., 2007). The RNAi vector targeting the *CRK9* 3′ UTR was generated by inserting 501 bp of the *CRK9* 3′ UTR from positions 2326 to 2827 relative to the translation initiation codon into pT7-stl. For HA tagging of RPB1, 1077 bp of the 3′-terminal RPB1 coding sequence was inserted into the pC-HA-BLA vector (Lee et al., 2007) utilizing the vector's ApaI and NotI restriction sites. The resulting vector was named pRPB1-HA-BLA. pCRK9-PTP-NEO was generated by inserting 1,532 bp of the *CRK9* coding region (position 791 to position 2322) into the ApaI and NotI sites of pC-PTP-NEO (Schimanski et al., 2005a). pCRK9-HA-BLA and pCRK9-$HA^{T533A}$-BLA were generated by inserting 1,532 bp of the *CRK9* coding region (position 791 to position 2322) into the ApaI and NotI sites of pC-HA-BLA. An A1597G transition was generated through overlapping PCR which changed the threonine codon 533 to an alanine codon. For transfection, pRPB1-HA-BLA, pCRK9-PTP-NEO, pCRK9-HA-BLA, and pCRK9-$HA^{T533A}$-BLA were linearized with restriction enzymes NarI, XhoI, SphI, and SgrA1, respectively. DNA oligonucleotides that were used in semi-quantitative and quantitative reverse transcription (RT)-PCR and in primer extension assays are specified in

supplementary Table S2. Plasmids used in dot blots were described previously and contained the complete coding regions of GPEET procyclin, α tubulin, HSP70, and 18S rRNA or the $tRNA^{Thr}$/U6 snRNA gene association (Schimanski et al., 2006).

### 3.2 Cells

*T. brucei* cell culture, transfection, and the generation of stable cell lines by selection and limiting dilution was carried out as described (Lee et al., 2007; Park et al., 2011). In RNAi experiments, dsRNA synthesis was induced with 2 μg/ml of doxycycline. Cells were counted and diluted daily to $2 \times 10^6$ cells/ml for procyclic cell culture and to $2 \times 10^5$ cells/ml for bloodstream form culture. A procyclic clonal cell line that exclusively expressed CRK9 with a C-terminally fused PTP tag, was generated by targeted integration of linearized pCRK9-PTP-NEO in one *CRK9* allele and by knocking out the remaining allele with a PCR product in which 101 bp of *CRK9* 5′ and 3′ gene flanks were fused to the hygromycin phosphotransferase coding region. Clonal *CRK9* RNAi cell lines that expressed RPB1 with a C-terminal HA tag were generated by transfection of linearized pRPB1-HA-BLA.

### 3.3 RNA analysis

To determine the relative abundance of RNAs in RNAi cells, total RNA prepared from 8 x $10^7$ cells using TRIzol reagent (Invitrogen) was analyzed by semi-quantitative and quantitative RT-PCR. Reverse transcriptions were carried out with SuperScript II reverse transcriptase (Invitrogen) according to the manufacturer's protocol and with oligo-dT and random hexanucleotides (Roche) for the analysis of mature and pre- mRNAs, respectively. For each semi-quantitative PCR, the number of cycles for the linear amplification range was determined empirically. For RNA quantifications, cDNA preparations from two independent RNAi experiments were analyzed by qPCR assays using the SsoFast EvaGreen Supermix (BioRad) on a CFX96 cycler (BioRad) according to the manufacturer's recommendations. For each

amplification and RNAi experiment, triplicate qPCR samples were analyzed using the CFX Manager software package (BioRad). Each amplification product was evaluated for specificity by both agarose gel electrophoresis and melting curve analysis. Standard curves for oligonucleotide pairs were obtained from serial dilutions of non-induced cDNA samples and ranged in their coefficient of determination ($R^2$) value from 0.98 to 1.0. Samples were standardized with the 18S rDNA result from random hexamer-derived cDNA.

Labeling and analysis of nascent RNA using a permeabilized cell system was carried out as described previously (Schimanski et al., 2006). For steady state analysis of SL RNA and U2 snRNA abundances 10 µg of total RNA prepared from induced or non-induced trypanosomes was assayed by primer extension using the 5′ $^{32}$P-endlabeled oligonucleotides SLf and U2f (Gunzl et al., 1992). The methylation status of SL RNA cap4 was analyzed by a modified primer extension assay in which 4 µg of total RNA was reverse transcribed by unmodified Moloney murine leukemia virus reverse transcriptase (Invitrogen) at a concentration of individual dNTPs that was reduced from 0.5 mM to 0.1 mM. DNA sequence ladders were generated by using the Thermo Sequenase Cycle sequencing kit (USB) and radiolabeled SLf according to the manufacturer's specifications. Sequencing ladder and cap4-specific primer extension products were separated on high resolution 50% urea-6% polyacrylamide gels and visualized by autoradiography.

The $m^7$G capping status of SL RNA was analyzed by immunoprecipitation with a mouse monoclonal anti-2,2,7-trimethylguanosine antibody coupled to agarose beads that also recognizes the $m^7$G cap nucleotide but not an unmethylated guanosine cap (Calbiochem). In immunoprecipitations, 5 µg of total RNA were diluted in 220 µl of buffer PA-150 (20 mM Tris-HCl, pH 7.7; 150 mM NaCl; 3 mM $MgCl_2$, 0.5 mM dithiothreitol; 0.1% Tween 20), mixed with a 30 µl of a 1:1 slurry of beads, and incubated for 1 h at 4 °C. Immunoprecipitates were washed six times with 0.8 ml of PA-150 buffer. Bound RNA was extracted from the beads using the

TRIzol reagent and resuspended in 20 µl of RNase-free water. 10 µl of bound and unbound fractions (equal amounts) were used for primer extension analysis using Superscript II reverse transcriptase as described above.

### 3.4 Protein analysis

For the generation of a polyclonal anti-RPB1 immune serum, the CTD portion of RPB1, comprising the C-terminal 228 amino acids, was expressed with an N-terminal glutathione *S*-transferase tag in *Escherichia coli*, purified by glutathione affinity chromatography, and used as a GST fusion protein to immunize rats according to a previously published protocol (Schimanski et al., 2006).

RPB1 was detected on immunoblots by 1:2000 dilution of the anti-RPB1 immune serum followed by a 1:5000 dilution of a monoclonal, peroxidase-labeled anti-rat IgG secondary antibody (Vector Laboratories). HA-tagged protein was detected with a commercial monoclonal rat anti-HA antibody (Roche) and RPA1 with a polyclonal rabbit immune serum as described previously (Schimanski et al., 2003). Blots were developed with BM chemiluminescence blotting substrate (Roche) according to the manufacturer's protocol.

To demonstrate that the upper RPB1 band detected in immunoblots with anti-HA or anti-RPB1 antibodies corresponded to phosphorylated RPB1, 8 x $10^7$ parasites were resuspended in 25 µl of lysis buffer (2% SDS; 100 mM Tris-HCl, pH 6.8) and boiled for 5 min. To enable alkaline phosphatase treatment the lysate was diluted twentyfold with tryp wash solution (100 mM NaCl; 3 mM $MgCl_2$, 20 mM Tris–HCl, pH 7.5) to reduce the SDS concentration to 0.05% and subsequently reduced to a volume of 100 µl using an Amicon ultra 0.5 ml filter device (Millipore). The phosphatase reaction was carried out in a volume of 30 µl containing 20 µl of sample, 1x NEB3 buffer, and 10 units of calf intestinal phosphatase (New England Biolabs), and by

incubation at 37 °C for one hour. To block the phosphatase, 0.5 µl of phosphatase inhibitor cocktail 2 (Sigma Co.) were added to the reaction.

## 4. Results

### 4.1 *CRK9* silencing blocked SL *trans* splicing.

The research group of Dr. C.C. Wang (University of California San Francisco) had previously studied all eleven *T. brucei* CRKs by RNAi-mediated gene knockdown. Their data indicated that CRK1, CRK3 and CRK9 were the most important CRKs for trypanosome viability (Gourguechon and Wang, 2009; Tu and Wang, 2004). Moreover, based on sequence similarity it was previously proposed that CRK3 (Ivens et al., 2005) or CRK7 (Parsons et al., 2005) may represent CDK7 orthologs. To explore the possibility that these CRKs regulate gene expression and/or are involved in CTD phosphorylation, we generated clonal cell lines of the insect-stage, procyclic form of *T. brucei,* in which each of these four *CRK* genes can be conditionally silenced by doxycycline-induced dsRNA expression (Wirtz et al., 1999). The knockdowns of *CRK9* (Figure 3.1A) and of *CRK1* and *CRK3* (Attachment 2, Fig. S1) were lethal, halting culture growth after the first or second day of induction. While these growth defects were consistent with previous results (Gourguechon and Wang, 2009; Tu and Wang, 2004), the much stronger growth inhibition in our experiments was likely the consequence of our use of the well-regulated pT7-stl RNAi construct (Brandenburg et al., 2007) which caused a rapid drop of the target mRNA levels within 24 hours of induction (boxed panels in Figures 3.1B and Attachment 2, Fig. S1). Analysis of RNAs synthesized by all three RNA pols showed that the *CRK1* and *CRK3* knockdowns had no strong effects on RNA abundances (Attachment 2, Fig. S1) whereas *CRK9* silencing consistently resulted in a clear reduction of α tubulin and *GPEET* procyclin mRNA abundances and a concomitant increase of the corresponding pre-mRNAs. Since procyclin and α tubulin genes are transcribed by RNA pols I and II, respectively, it was unlikely that these effects were connected to the transcription of these genes. In addition to these changes in pre-

and mature mRNA levels, we found that SL RNA abundance sharply increased 48 hours after induction of *CRK9* silencing whereas there was no consistent effect on the abundances of rRNA and U2 snRNA which, in trypanosomes, are synthesized by RNA pols I and III, respectively (Figure 3.1B). According to reverse transcription-quantitative real-time (RT-q)PCR experiments these were strong effects: after three days of induction the pools of mature α tubulin and *GPEET* mRNA were reduced to approximately one quarter while the levels of SL RNA and pre-mRNA were increased up to ~nine-fold (Figure 3.1C). Since SL RNA is consumed in the first transesterification step of the splicing process, these findings strongly indicated that *CRK9* silencing led to inhibition of SL *trans* splicing before the first step. In accordance with this notion, a primer extension assay showed that the SL intron, the product of the first splicing step, disappeared specifically upon *CRK9* silencing (Figure 3.1B, bottom panel; compare with the corresponding panels in Attachment 2, Fig. S1) CRK9 appears to have the same important functional role in gene expression in the mammalian-infective bloodstream trypanosome because a *CRK9* knockdown in this life cycle stage was as lethal as in the procyclic form and also led to a decrease of α tubulin mRNA and a concomitant increase of the corresponding pre-mRNA (Attachment 2, Fig. S2). Importantly, this result is in direct contrast to a previous report which found CRK9 to be essential only in the procyclic form (Gourguechon and Wang, 2009). Again, this discrepancy may be explained by our use of the more suitable pT7-stl RNAi vector (Brandenburg et al., 2007).

In accordance with previously published results (Gourguechon and Wang, 2009), silencing of *CRK7* did not impair trypanosome proliferation in culture, and abundances of rRNA, *GPEET* mRNA and *α* tubulin mRNA remained unaffected by the knockdown (Attachment 2, Fig. S3). Since these results strongly suggested that CRK7 is not a functional homolog of eukaryotic CDK7 we did not further analyze this kinase. It should be noted though that these results are in

contrast to a genome-wide RNAi screen which found *CRK7* to be important for trypanosome viability across different life cycle stages (Alsford et al., 2011).

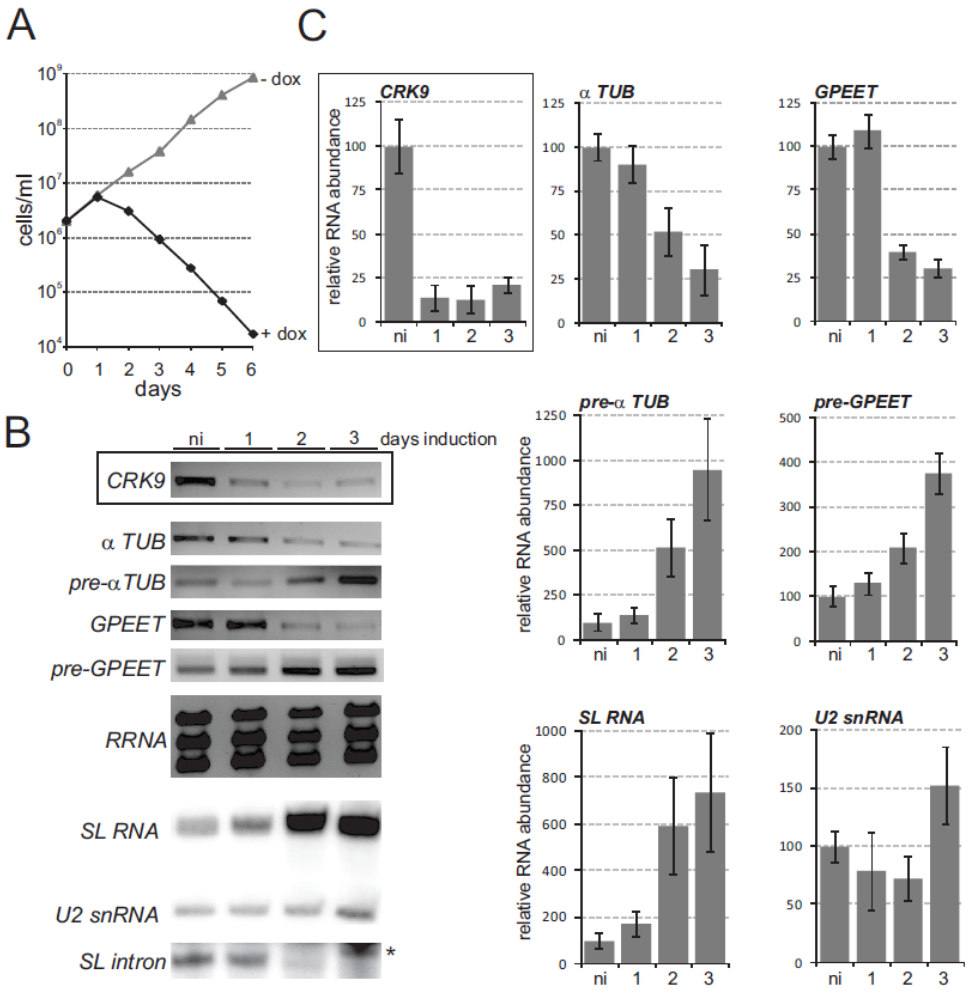



**Figure 3.1 *CRK9* expression silencing caused a decrease of mature mRNA levels and concomitant accumulations of SL RNA and pre-mRNAs.**

(**A**) A representative growth curve from one of three independently derived clonal cell lines was obtained in the absence and presence of the dsRNA synthesis-inducing reagent doxycycline (-/+ dox).

(**B**) Relative *CRK9* mRNA levels were determined by semi-quantitative RT-PCR of total RNA prepared from non-induced (ni) cells and in cells that were silenced for 1, 2 or 3 days (boxed panel). In the same RNA preparations, the relative abundances of α tubulin (α TUB) and of *GPEET* procyclin mature and pre-mRNAs were determined by semi-quantitative RT-PCR, those of the large ribosomal RNAs by ethidium bromide staining, and those of the SL RNA, the U2 snRNA, and the SL intron by primer extension assays. Note that the SL intron signals were obtained through overexposure of the same gels; the asterisk indicates overexposed U2 signal just above the SL intron signal.

(**C**) Corresponding RT-qPCR analyses. The relative level of the analyzed RNAs in non-induced cells was arbitrarily set to 100. The quantitative data were calculated from two independent experiments each of which was analyzed three times by qPCR. The boxed panel shows the silenced *CRK9* mRNA levels.

### 4.2 *CRK9* silencing led to a loss of RPB1 phosphorylation.

Since we had speculated that RPB1 CTD phosphorylation is mediated by a CDK, we wanted first to know whether *CRK9* silencing did affect the phosphorylation status of the CTD. We expressed a GST-CTD fusion protein in *Escherichia coli* and raised a rat polyclonal anti-CTD immune serum against the purified recombinant protein. On immunoblots, the antibody specifically detected two specific protein bands in the range of 200 kDa in whole cell lysates. However, the larger band was consistently absent in cell-free extract suggesting that in extract RPB1 is rapidly dephosphorylated (Figure 3.2A). This finding was in accordance with previous studies in which detection of RPB1 phosphorylation in extract proved to be difficult (Chapman and Agabian, 1994; Das and Bellofatto, 2009). Although we tested several phosphatase inhibitor cocktails, the loss of the upper band in extract could not be prevented (data not shown). To prove that the upper band detected with our anti-CTD antibody represented phosphorylated RPB1, we lysed trypanosomes in an SDS buffer and then reduced the detergent concentration to enable alkaline phosphatase treatment. As shown in Figure 3.2B, the upper band was retained after mock incubation without enzyme while addition of alkaline phosphatase converted the upper band into the lower band, and this conversion was completely blocked with phosphatase inhibitor. Therefore, we concluded that the upper and lower bands in our immunodetections represented phosphorylated (pRPB1) and dephosphorylated RPB1, respectively. Immunoblotting of whole lysates from non-induced and induced cells then clearly showed that the pRPB1 signal became diminished upon *CRK9* silencing whereas dephosphorylated RPB1 increased (Figure 3.2C). Quantification of immunoblot signals revealed that the pRPB1 signal decreased to 31% and 21% of non-induced levels on days 2 and 3, respectively, whereas the signal of dephosphorylated RPB1 increased to ~2.5 fold during the same period (Figure 3.2D). To confirm this result, we generated a clonal procyclic cell line that expressed RPB1 with a C-terminal HA tag. Anti-HA immunoblotting detected the same two

RPB1 bands as our polyclonal antibody and *CRK9* silencing again led to a signal decrease of the upper band and a concomitant signal increase of the lower band (Figure 3.2D). To make sure that loss of RPB1 phosphorylation is a specific CRK9-dependent effect and not a general phenotype of dying trypanosomes, we analyzed RPB1 in *CRK3*-silenced cells which, like *CRK9*-silenced cells, ceased to proliferate after the first day of induction (Attachment 2, Fig. S1). Although we noted a general loss of RPB1 on day 3 of induction, the relative amounts of upper and lower bands remained at a similar level during the course of this experiment (Figure 3.2E) and allowed us to conclude that CRK9 is essential for the majority of, if not all, RPB1 phosphorylations.

Although our data showed that RPB1 phosphorylation depended on CRK9 they did not show whether CRK9 directly phosphorylates the CTD. To detect a physical link between CRK9 and RNA pol II we generated a cell line which exclusively expressed CRK9 C-terminally tagged with the composite PTP tag (Schimanski et al., 2005a). Since CRK9 is essential for trypanosome viability we concluded that the tag did not interfere with the function of the kinase. A pull-down of CRK9 under low stringency conditions, however, did not co-precipitate detectable amounts of RPB1 (data not shown). Since this result may be due to a transient or low affinity interaction between the kinase and RPB1, we speculated that if RPB1 was phosphorylated at *SLRNAs* then the kinase should be cross-linkable to *SLRNA* chromatin. However, chromatin immunoprecipitation (ChIP) of PTP-tagged CRK9 or equivalently tagged RPB9, a specific subunit of RNA pol II, showed high occupancy of the *SLRNA* promoter by RPB9 whereas a CRK9 association with this promoter could not be detected (Attachment 2, Fig. S4). Although these experiments do not rule out that CRK9 directly phosphorylates RPB1 they raise the possibility that a different, CRK9-dependent kinase carries out these phosphorylations.

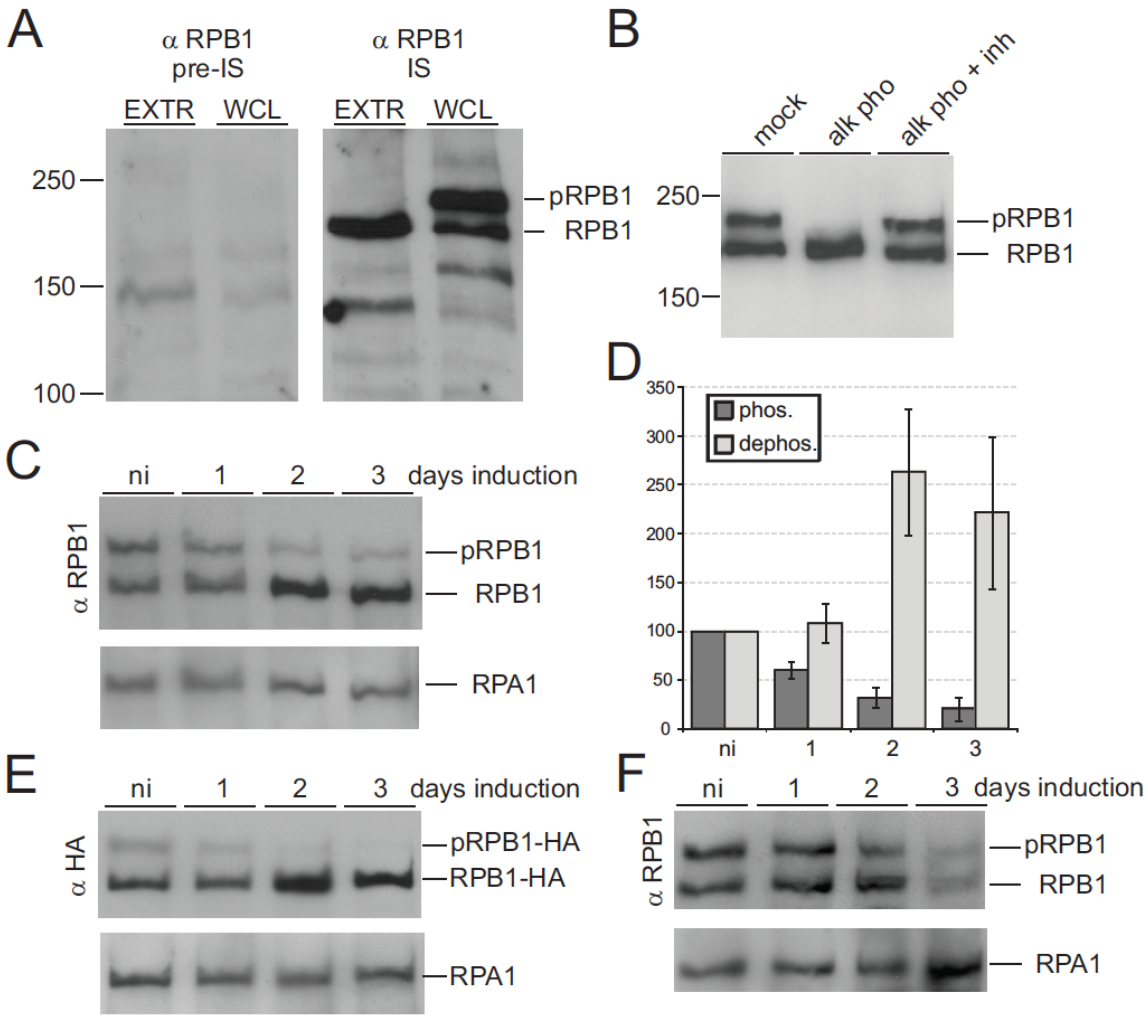



**Figure 3.2 *CRK9* silencing caused a loss of RPB1 phosphorylation.**

(**A**) Cell-free extract (EXTR) or whole cell lysate (WCL) was separated by SDS-PAGE, blotted and probed with anti-RPB1 pre-immune (pre-IS) or immune serum (IS). The immune serum specifically detected phosphorylated (pRPB1) and dephosphorylated RPB1.

(**B**) Diluted whole cell lysate was mock-treated or incubated with alkaline phosphatase (alk pho) in the absence or presence (+inh) of a phosphatase inhibitor cocktail.

(**C**) Anti-RPB1 immunoblot analysis of whole cell lysates prepared from *CRK9*-silenced cells. Detection of the similar-sized RNA pol I subunit RPA1 served as a loading control.

(**D**) Densitometric quantification of the phosphorylated (phos.) and dephosphorylated (dephos.) RPB1 bands upon *CRK9* silencing from three independent experiments. The signal of non-induced cells was arbitrarily set to 100.

(**E**) Anti-HA immunoblot of a corresponding experiment with *CRK9* RNAi cells that expressed RPB1-HA.

(**F**) Anti-RPB1 immunoblot analysis of whole cell lysates prepared from *CRK3*-silenced cells.

### 4.3 *CRK9* silencing does not impair RNA pol II transcription.

Since CTD phosphorylation is essential for transcription initiation in yeast and mammals, we explored whether CRK9-dependent RPB1 phosphorylation is required for RNA pol II transcription in trypanosomes. We employed a permeabilized cell system which supports labeling of nascent RNA in a linearly increasing fashion for a period of 15-20 min (Ullu and Tschudi, 1990). Upon separation of labeled RNA by denaturing PAGE, the 142 nt-long SL RNA can be readily detected due to its strong overall synthesis rate (Figure 3.3A). The amount of SL RNA was roughly the same in non-induced cells and in cells in which the *CRK9* knockdown was induced for one day. After two days of induction, the SL RNA was reduced by approximately one half but the signals for tRNA and larger RNAs were similarly reduced, most likely due to an increasing number of dying cells. Accordingly, when we used densitometry to adjust the tRNA signal, labeled SL RNA remained at the same high level two days after induction (Figure 3.3A, right panel). Signal quantification from independent experiments confirmed that nascent SL RNA synthesis is not perturbed upon CRK9 silencing (Figure 3.3C). Due to a large number of dead or dying cells, the experiment could not be performed meaningfully with cells that were *CRK9*-silenced for three days. However, since trypanosome growth ceased already after one day of *CRK9* silencing and since the levels of pre-mRNA, mature mRNA and SL RNA were already clearly perturbed after two days of induction (Figure 3.1), we concluded that the loss of CRK9-dependent RPB1 phosphorylation did not impair SL RNA synthesis. Importantly, the *SLRNA* promoter is the only characterized RNA pol II promoter in trypanosomes that assembles a transcription pre-initiation complex including TFIIH, the factor responsible for initial CTD phosphorylation in other eukaryotes. Therefore, this finding strongly indicated that, in trypanosomes, TFIIH-mediated CTD phosphorylation is not required for RNA pol II escape from the *SLRNA* promoter.

To explore the possibility that RPB1 phosphorylation is important for transcription of protein coding genes, we hybridized labeled nascent RNA to dot blots of plasmids carrying trypanosome gene inserts (Figure 3.3B). The experiment was repeated three times and revealed that labeling efficiency was reduced throughout the array two days after induction. For comparison, we normalized all signals with the signal of the RNA pol III-transcribed threonine tRNA/U6 snRNA gene association (Nakaar et al., 1997) which on day 2 was reduced on average by ~35%. Quantification of hybridization signals did not detect a specific defect in RNA pol II-mediated transcription of protein coding genes. Two days after induction, the α tubulin and *HSP70* signals were reduced by 14% and 3% respectively, RNA pol I-synthesized *GPEET* procyclin RNA was reduced by 21%. While none of these reductions proved to be significant (unpaired Student's T test assuming equal variance and two-tailed distribution), the strongest signal reduction of 33% was observed in *RRNA* transcription. Since it is well-known from other systems that inhibition of rRNA synthesis in stationary or stressed cells is often a first measure to down-regulate gene expression, this signal reduction could reflect the poor health of *CRK9*-silencend cells.

In summary, these experiments failed to detect a specific defect in RNA pol II transcription and thus strongly indicate that CRK9-dependent RPB1 phosphorylation is neither required for transcription initiation at the *SLRNA* promoter nor an essential modification for transcription elongation within the protein coding gene arrays.

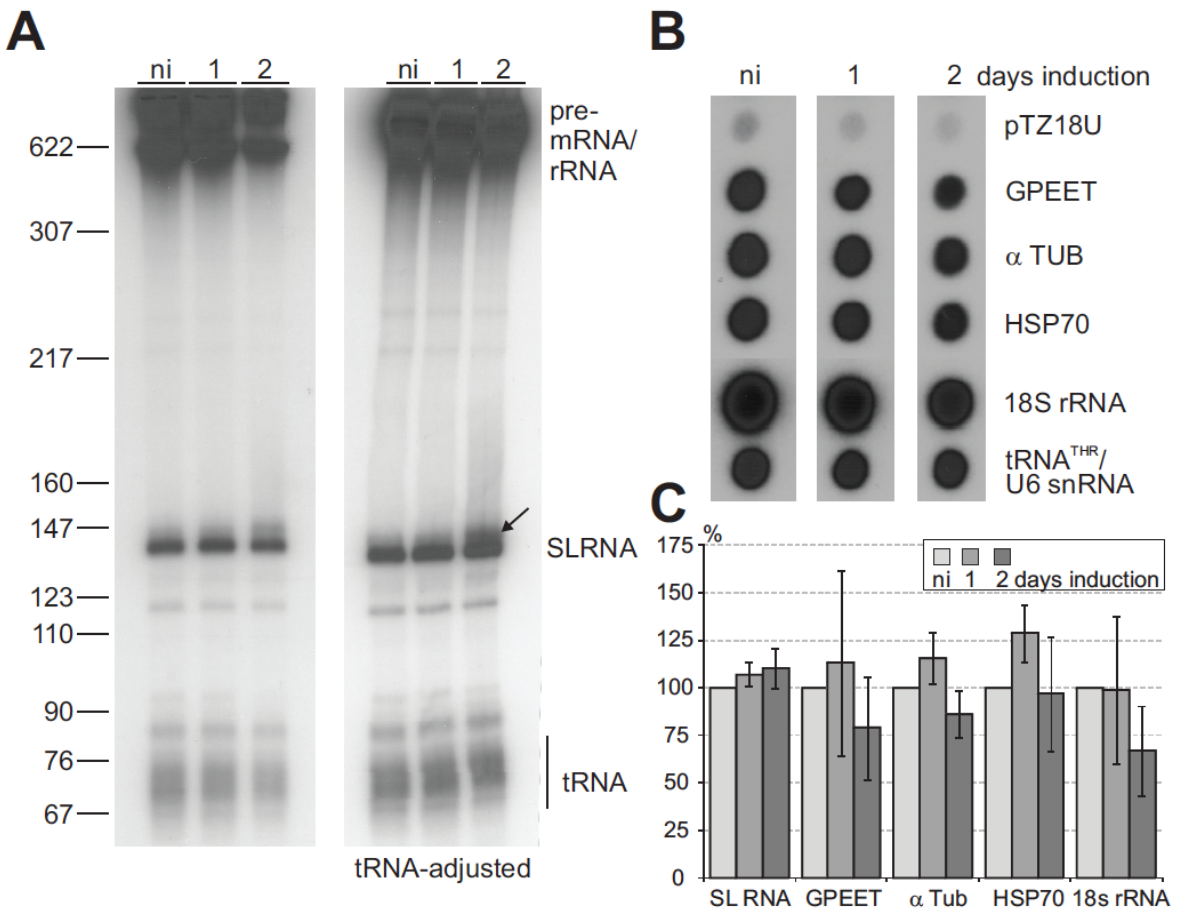



**Figure 3.3 *CRK9* silencing does not affect RNA pol II transcription.**

(**A**) Nascent RNA of cells that were non-induced (ni) or *CRK9*-silenced for 1 or 2 days, was radio-labeled in a permeabilized cell system, extracted, separated on a 50% urea-6% polyacrylamide gel, and visualized by autoradiography. In the left panel, the loading of RNA was standardized to the number of cells and on the right panel, the loading was standardized according to the tRNA signal as indicated. The arrow points to an RNA band that most likely represents a previously described SL RNA that is extended at the 3′ end.

(**B**) The same labeled RNA preparations were used to probe dot blots of plasmids containing the coding regions of *T. brucei* GPEET procyclin, α tubulin (α TUB), heat shock protein 70 (HSP70), and 18S rRNA as well as the associated threonine tRNA/U6 snRNA genes ($tRNA^{THR}$/U6 snRNA). The plasmid without insert (pTZ18U) served as a negative control. For each lane, upper and lower panels are from the same image.

(**C**) The radioactive signals from four independent experiments were quantified by scintillation counting and normalized to the RNA pol III-mediated $tRNA^{THR}$/U6 snRNA signal because CRK9 silencing did not affect the abundance of the RNA pol III-synthesized U2 transcript. The SL RNA signal was quantified by densitometry and normalized to the tRNA signal as shown in (A). The error bars represent standard deviations.

### 4.4 *CRK9* silencing impairs cap4 formation of the SL RNA.

As detailed above, reporter gene expression by different RNA pols demonstrated that pre-mRNA processing in trypanosomes is uncoupled from RNA pol II transcription (Stewart et al., 2010). In accordance with these results, *CRK9* silencing affected RNA pol I-synthesized GPEET procyclin RNA in the same way as it did RNA pol II-synthesized α tubulin mRNA (Figure 3.1). However, these results did not exclude the possibility that SL RNA modification depended on RNA pol II because SL *trans* splicing is an essential maturation step for all trypanosome mRNAs independent of the transcribing RNA pol. Since *trans* splicing depends on the methylation status of the SL RNA cap4 structure (McNally and Agabian, 1992; Ullu and Tschudi, 1991) and since there is evidence that SL RNA cap4 formation occurs co-transcriptionally (Mair et al., 2000b), it was possible that the *trans* splicing block observed upon *CRK9* silencing was due to inappropriate cap4 formation. To monitor cap4 methylations on the SL RNA we employed a modified primer extension assay with a reduced concentration of deoxyribonucleotides and with unmodified MMLV reverse transcriptase that terminates polymerization before the methylated nucleotide under these conditions (Mandelboim et al., 2002; Zamudio et al., 2006). In non-induced cells, the SL RNA-specific extension products spanned six positions: cap0 - cap4 in which, as previously noted (Zamudio et al., 2007), a cap3-specific extension product is typically not observed because MTR3 methylates positions 3 and 4, and cap4 can generate two extension products (Figure 3.4A). The cap4 signals were the strongest in these cells and the pattern did not change after one day of *CRK9* silencing. Conversely, on days two and three, the cap4 signals faded whereas the main extension product of the accumulating SL RNA was the cap0 product indicating a complete loss of cap4 methylations (Figure 3.4A). Furthermore, loss of cap1 modification was previously shown to correlate with ineffective trimming of the SL RNA 3′ end by the nuclease TbSNIP leading to slightly larger SL RNAs (Zamudio et al., 2007). In accordance with this finding, we detected an

accumulating signal above the correct SL RNA band in labeled nascent RNA (Figure 3.4A, arrow). In summary, we concluded that *CRK9* silencing caused a degree of cap4 hypomethylation that impaired SL *trans* splicing leading to SL RNA and pre-mRNA accumulations and decreasing levels of mature mRNAs.

Next, we addressed the question whether cap4 hypomethylation is linked to the loss of RPB1 dephosphorylation. We first analyzed the U1 snRNA cap. While the *T. brucei* U1 gene appears to be transcribed by RNA pol III (Djikeng et al., 2001), the U1 snRNA has a cap1 structure consisting of a trimethylguanosine (TMG) cap and a methylated adenosine at position 1 that leads to a premature primer extension stop at position 2 (Djikeng et al., 2001; Zamudio et al., 2007). Importantly, cap1 methylation of both SL RNA and U1 snRNA is carried out by the same methyltransferase termed MTR1 (Zamudio et al., 2007). Hence, if *CRK9* silencing led to a direct deactivation of MTR1 we would expect to see a loss of cap1 in U1 snRNA. To test this, we carried out the modified primer extension assay with total RNA prepared from *CRK9*- and *CRK3*-silenced cells using a U1-complementary oligonucleotide (Figure 3.4B). We reproducibly detected three U1-specific products in our assay: the two lower bands were specific for cap1 and unmodified cap0 whereas the fainter product with an extra nucleotide (position -1) was possibly generated due to the TMG cap (as opposed to the SL RNA $m^7G$ cap). As expected, treating the cells for six hours with sinefungin resulted in a signal shift from cap1 to cap0/pos-1 products. This shift was not observed in *CRK9*- or *CRK3*-silenced cells even after RNAi induction for three days indicating that the MTR1 enzyme activity remained active. Consequently, the specific loss of SL RNA cap1 methylation upon *CRK9* silencing may be due to a failure in MTR1 recruitment to the SL RNA 5′ end.

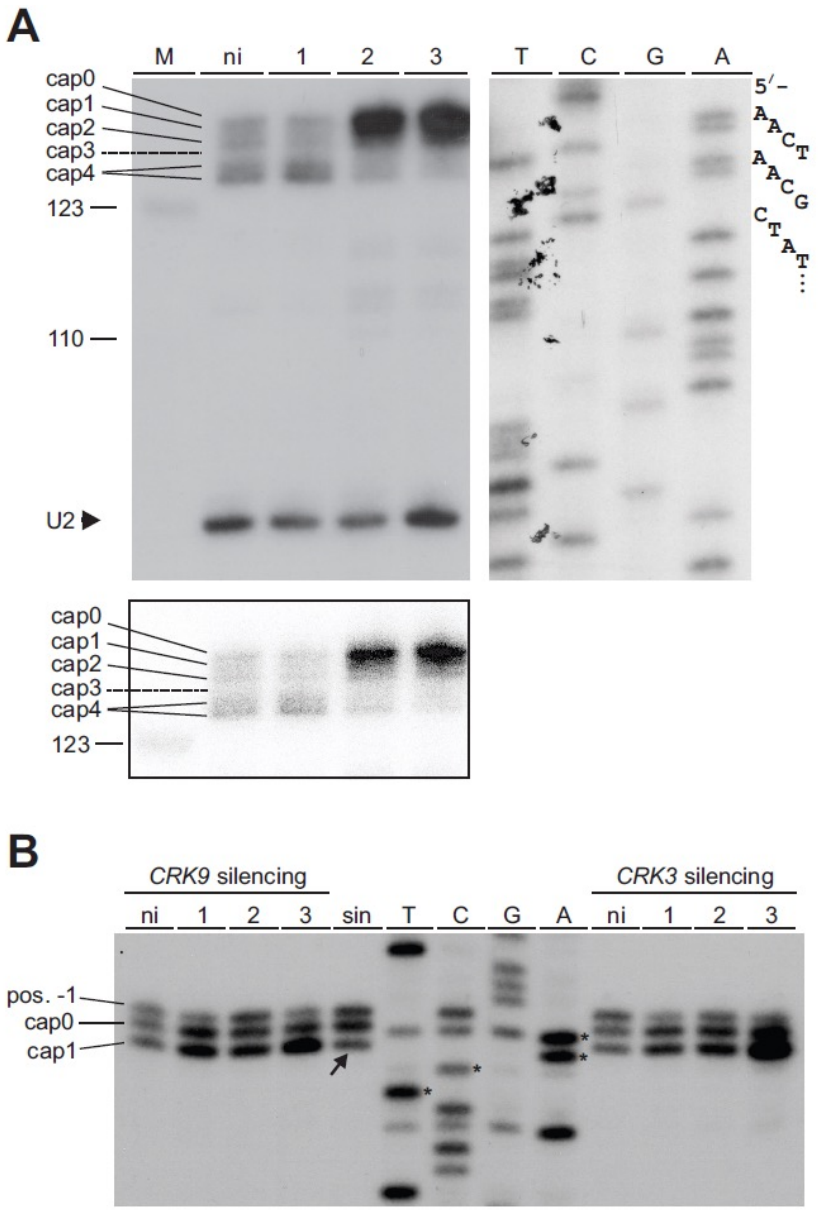



**Figure 3.4 *CRK9* silencing resulted in specific hypomethylation of SL RNA cap4.**

(**A**) Primer extension analysis with MMLV reverse transcriptase and reduced dNTP concentration (100 nM) of total RNA prepared from non-induced (ni) or *CRK9*-silenced cells using $^{32}$P-5′ end-labeled oligonucleotides SLf and U2f that are complementary to SL RNA and U2 snRNA, respectively. The extension products were separated on a high resolution 6% sequencing gel and visualized by autoradiography. The lower exposure of the SL RNA extension products at the bottom reveals that *CRK9* silencing led to accumulation of larger extension products, particularly that of the cap0 product. For comparison, a SL RNA-specific sequencing ladder was generated by linear amplification sequencing using labeled SLf; the nucleotide specification for each lane corresponds to the coding strand sequence. The sequencing ladder was co-separated on the same gel but required a longer exposure.

(**B**) A corresponding analysis of the U1 snRNA cap1 in *CRK9*- and *CRK3*-silenced cells. The first four U1 nucleotides of the sequencing ladder are marked by asterisks and the partial loss of cap1 in sinefungin-treated cells (sin) is denoted by the arrow.

To support this notion, we evaluated pseudouridylation of SL RNA at position 28 which is carried out by the SLA1 ribonucleoprotein particle (RNP) that comprises several proteins including the pseudouridine synthase CBF5 (54) and MTR1 (Zamudio et al., 2009b). We hypothesized that if MTR1 is not recruited to the SL RNA cap in *CRK9*-silenced cells then pseudouridylation may also be impaired because MTR1 and CBF5 belong to the same RNA-protein complex. However, the accumulating SL RNA in *CRK9*-silenced cells was clearly pseudouridylated (Attachment 2, Fig. S5) indicating that the SLA1 RNP can interact with SL RNA independent of CRK9-controlled RPB1 phosphorylations. As follows, our data do not unequivocally support the notion that RPB1 phosphorylation mediates cap4 formation. Testing this hypothesis is beyond the scope of this study because it requires the simultaneous knockdown of the two *T. brucei RPB1* genes and the ectopic expression of a tagged, RNAi-resistant version of the large *RPB1* gene carrying mutations of phosphorylated CTD residues. Nevertheless, the results presented here strongly indicate that the unique methylations of cap4 that are essential for *trans* splicing depend on CRK9.

### 4.5 *CRK9* silencing does not affect $m^7G$ capping of SL RNA.

The previous demonstration that prematurely terminated SL RNA transcripts as short as 30 nt were capped with an $m^7G$ residue led to the conclusion that $m^7G$ capping of *T. brucei* SL RNA occurs cotranscriptionally (Mair et al., 2000b). It was therefore possible that *CRK9* silencing and the resulting CTD dephosphorylation abrogated $m^7G$ capping of the SL RNA. To analyze the SL RNA capping status in *CRK9*-silenced cells we performed immunoprecipitations with an antibody directed against a TMG cap that also recognizes the $m^7G$ cap but does not interact with an unmethylated guanosine cap or uncapped RNAs. As shown in figure 3.5, the SL RNA was efficiently precipitated from total RNA of non-induced cells as well as of cells in which *CRK9* was silenced for up to three days. In contrast, SL RNA from sinefungin-treated cells that lacked methylations and the uncapped U5 snRNA, which served as an internal negative control,

were not precipitated. These results clearly demonstrated that CRK9-dependent CTD phosphorylation is not required for SL RNA $m^7G$ capping indicating that a potential interaction between trypanosome RNA pol II and the capping enzyme is independent of RPB1 phosphorylation.

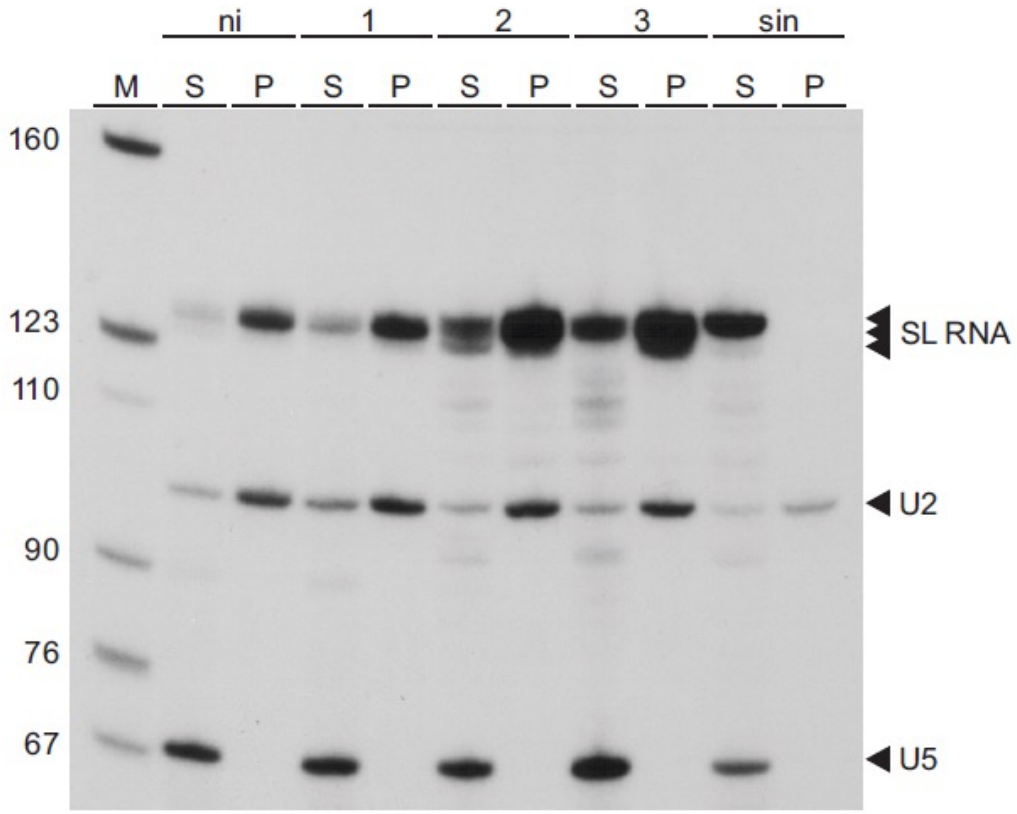



**Figure 3.5 *CRK9* silencing does not affect $m^7G$ capping of SL RNA.**

Total RNA preparations from non-induced cells (ni), cells in which *CRK9* was silenced for 1, 2 or 3 days, or cells which were treated with the methylation inhibitor sinefungin (sin) were immunoprecipitated with an antibody that recognizes both $m^{2,2,7}G$ and $m^7G$ caps. Equal amounts of supernatants (S) and immunoprecipitates (P) were analyzed by primer extension with oligonucleotides complementary to the SL RNA ($m^7G$ cap), the U2 snRNA ($m^{2,2,7}G$ cap), and the uncapped U5 snRNA.

**4.5 RPB1 phosphorylation and SL RNA cap4 formation depend on CRK9 kinase activity.**

Thus far, we have shown that inducible expression of dsRNA that targets the 5′-terminal CRK9 coding region led to an efficient knockdown of CRK9 mRNA and to RPB1 dephosphorylation, SL RNA cap4 hypomethylation, and the described RNA processing defects. To unequivocally demonstrate that these effects depended on CRK9 and its enzymatic activity and not to an off-target effect of the 500 bp-long dsRNA, we generated a new procyclic RNAi cell line in which we targeted the 3′ UTR of the CRK9 mRNA (Figure 3.6A). The corresponding knockdown led to a similar, rapid decline of cell numbers as observed previously and, as expected, the levels of mature and pre-mature α tubulin mRNA decreased and increased upon induction of dsRNA synthesis, respectively. The cap4 methylation defect and RPB1 dephosphorylation were similarly obvious (Figure 3.6A). Moreover, when we integrated a plasmid into an endogenous *CRK9* allele that fused an HA tag sequence and the RPA1 3′ UTR to the CRK9 coding region, we could rescue culture growth and all the defects (Figure 3.6B). The corresponding RNA analysis showed that endogenous *CRK9* mRNA was effectively knocked down whereas CRK9-HA mRNA was resistant to the RNAi response. Correspondingly CRK9-HA protein abundance remained unaffected upon doxycycline induction. These results unequivocally demonstrated that RPB1 phosphorylation and SL RNA cap4 formation are controlled by CRK9. In a final step, we wanted to make sure that the kinase activity of CRK9 is important for this control. In general, phosphorylation of the T loop threonine residue (T161 in human CDK1) leads to activation of a CDK/cyclin complex. The corresponding threonine residue in CRK9 of *T. brucei* had been mapped to position 533 (Hammarton, 2007). We therefore repeated the rescue experiment but this time we mutated the T533 to an alanine in the CRK9-HA protein (Figure 3.6C). RNAi induction showed a growth defect after 24 h similar to the other (non-rescued) *CRK9* silencing experiments. However, these cells recovered after 2-3 days to normal growth and this was observed with three independently obtained clonal cell lines.

Concomitant with recovery we reproducibly noticed a slightly higher migrating band of CRK9-HA$^{T553A}$ 3 days after induction (Figure 3.6C, arrow in CRK9-HA$^{T553A}$ immunoblot) suggesting that compensatory phosphorylations or other post-translational modifications led to an unconventional activation of CRK9-HA$^{T553A}$. Nevertheless, while only a minor perturbation of α tubulin (pre-)mRNA was apparent 3 days after induction, cap4 hypomethylation and RPB1 dephosphorylation were clearly detectable 2 days after induction. These data strongly indicated that CRK9 enzyme activity is required for RPB1 phosphorylation and SL RNA cap4 formation.

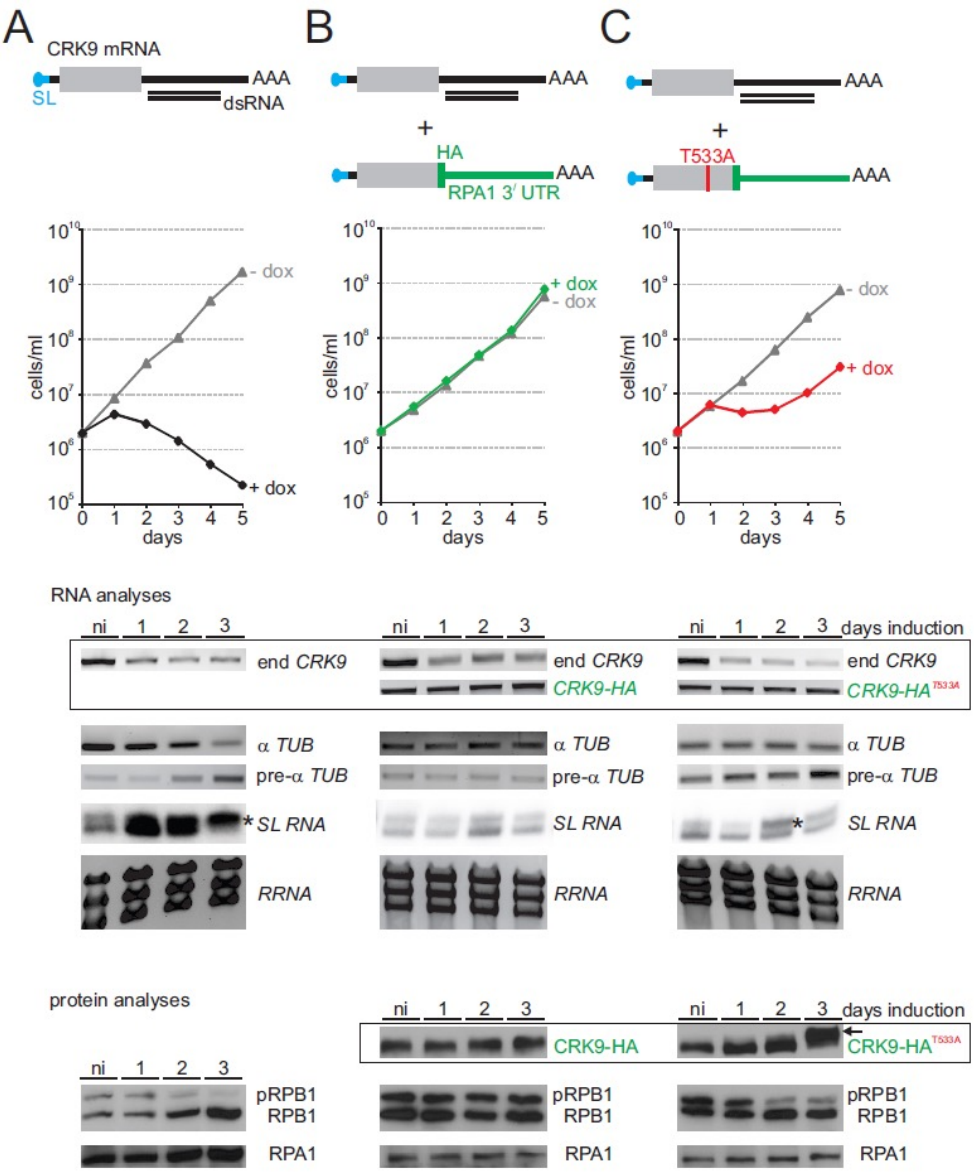



**Figure 3.6 RPB1 phosphorylation and SL RNA cap4 formation depend on CRK9 kinase activity.**

A *CRK9* RNAi parent cell line (**A**) was generated in which the inducibly expressed dsRNA targets the 3′ UTR of *CRK9* mRNA. This cell line was further modified by targeted integration of a plasmid into an endogenous allele that fused an HA tag sequence and the 3′UTR of the largest RNA pol I subunit RPA1 to the 3′ end of the gene (**B**). A third cell line was generated analogously except that the threonine codon at position 533 was mutated to an alanine codon (**C**). For each cell line, culture growth curves were obtained in the presence or absence of doxycycline and RNA and protein analyses were conducted with non-induced cells (ni) and with cells that were induced for 1, 2, or 3 days. Semi-quantitative RT-PCR analyses demonstrated efficient knockdown of endogenous (Castillo et al.) *CRK9* RNA and RNAi resistance of *CRK9-HA* RNA (boxed results). The *CRK9* silencing effect on α tubulin mRNA (*α Tub*) and pre-mRNA (*pre-αTub*) was assessed by semi-quantitative RT-PCR, and on SL RNA by primer extension assays. Detection of rRNA (*RRNA*) by ethidium bromide staining served as a control for RNA amount. Asterisks identify increased hypomethylated cap4 signals. Anti-HA and anti-RPB1 immunoblot analyses confirmed RNAi resistance of CRK9-HA (boxed) and revealed the effects of *CRK9* silencing on RBP1 phosphorylation. Detection of RPA1 served as a loading control. The arrow on the right identifies a slower migrating band of CRK9-HA$^{T553A}$ protein.

## 5. Discussion

We have analyzed the role in gene expression of four out of the eleven trypanosome CDKs. Our hypothesis was that, in line with phosphorylation of heptad repeats, this type of kinase phosphorylates the non-canonical RPB1 of trypanosomes and is of central importance for gene expression in this early-diverged eukaryote. Accordingly, we could demonstrate that silencing of trypanosome *CRK9* led to a loss of RPB1 phosphorylation. Since it was shown that the RPB1 CTD is phosphorylated (Chapman and Agabian, 1994; Das and Bellofatto, 2009) and phosphoproteomics exclusively detected phosphorylated residues in the CTD of the large RPB1 protein (Nett et al., 2009), it is likely that most if not all RPB1 phosphorylations are located in the CTD. It seems that all of these phosphorylations depend on CRK9 because silencing of its gene led to a loss of the pRPB1 band in immunoblots and to a concomitant increase of unphosphorylated RPB1 without the appearance of an intermediate band. Although we have not been able to show so far that CRK9 directly phosphorylates the CTD, our results are informative about the function of the non-canonical CTD. We found that reducing phosphorylated RPB1 to 30% on day 2 of CRK9 silencing, did not impact transcription of protein coding and *SLRNA* genes indicating a mechanistic difference in transcription initiation between trypanosomes and eukaryotes with a canonical CTD. In the latter, RNA pol II recruitment into the PIC involves the mediator complex which binds to hypophosphorylated RNA pol II. TFIIH/CDK7-mediated CTD phosphorylation disrupts this interaction, presumably facilitating promoter escape of the polymerase (Sogaard and Svejstrup, 2007). Trypanosomes do possess a mediator complex that is essential for *SLRNA* transcription, interacts with TFIIH, and is required for the formation of a transcription pre-initiation complex at the *SLRNA* promoter (Lee et al., 2010). However, whereas eukaryotic mediator consists of head, middle, tail and CDK8 modules, EM structure analysis of the trypanosome complex indicated that it is confined to the head module (Lee et al., 2010). Thus, the absence of a TFIIH kinase (Lee et al., 2009) and the presence of a minimal mediator complex suggest that, in trypanosomes, a mediator-CTD interaction does not occur and a

release of RNA pol II from mediator is not a prerequisite for promoter escape of the enzyme. Accordingly, tandem affinity purification of trypanosome mediator did not co-purify detectable amounts of RNA pol II subunits (Lee et al., 2010) and *vice versa* (Das et al., 2006; Devaux et al., 2006).

Furthermore, RPB1 dephosphorylation did not impact $m^7G$ capping of SL RNA. In higher eukaryotes, the capping enzyme binds to the CTD after $Ser^5$ of the heptad repeat motif is phosphorylated by CDK7 (Fabrega et al., 2003); capping then occurs co-transcriptionally. Since in trypanosomes, SL RNA capping does occur co-transcriptionally as well (Mair et al., 2000b), our results strongly suggest that the trypanosome capping enzyme interacts with the transcription machinery in a trypanosome-specific manner independent of CTD phosphorylation.

Nevertheless, *CRK9* silencing resulted in a striking defect in trypanosome gene expression, namely a block of *trans* splicing before the first transesterification reaction caused by hypomethylation of SL RNA cap4. Formation of the unique cap4 structure involves seven methylations on the first four nucleotides by three or more different methyltransferases. Thus, the most intriguing aspect of our work is the demonstration that cap4 formation is controlled by a single enzyme. Since the trypanosome cap4 structure is unique, essential for *trans* splicing (McNally and Agabian, 1992; Ullu and Tschudi, 1990), and important for translation (Zamudio et al., 2009b), identification of the methyltransferases has been a priority and, thus far, the three 2′-*O*-ribose methyltransferases MTR1, MTR2 (aka COM1, MT48) and MTR3 (aka MT57) which methylate positions 1, 2 and 3/4, respectively, have been identified (Arhin et al., 2006a; Arhin et al., 2006b; Hall and Ho, 2006; Zamudio et al., 2007; Zamudio et al., 2006). Surprisingly, however, elimination of individual *MTR* genes left trypanosomes viable and did not impair *trans* splicing. This functional redundancy of the known MTRs has dampened enthusiasm for these enzymes being potential chemotherapeutic targets. However, the dependence of a critical number of cap4 methylations on CRK9 alone reinstates cap4 formation as a target. CRK9

appears to be particularly promising in this respect. Firstly, CRK9 has an essential central role in trypanosome gene expression controlling RPB1 phosphorylation and SL RNA cap4 formation. Secondly, SL *trans* splicing does not occur in mammals and, thus, the block of SL RNA maturation is a parasite-specific defect. Thirdly, since SL RNA is consumed in *trans* splicing of every mRNA, partial inhibition of SL RNA biogenesis may already be lethal. This notion is backed up by independent observations that a reduction of SL RNA synthesis to ~30% is already lethal to cultured trypanosomes (Ruan et al., 2004; Schimanski et al., 2006). Fourthly, cyclin-dependent kinases in general are considered to be among the most promising drug targets due to their key function in cell proliferation (Malumbres and Barbacid, 2009). Accordingly, *CRK9* silencing was shown to affect trypanosome mitosis and cytokinesis (Gourguechon and Wang, 2009; Li, 2012).

How does CRK9 enable cap4 formation? A plausible explanation is that cap4 methylations occur co-transcriptionally requiring a phosphorylated CTD. This would be consistent with previous observations that cap4 methylations can be detected on newly synthesized, prematurely terminated SL RNA transcripts (Mair et al., 2000b). Part of our results support this scenario: *CRK9* silencing did not cause a detectable defect of cap1 formation in the RNA pol III-synthesized U1 snRNA suggesting that MTR1, the methyltransferase that forms cap1 in both U1 and SL RNA, remains functional during the *CRK9* knockdown. Thus, the loss of SL RNA cap1 may be explained by failed recruitment of MTR1 by dephosphorylated RNA pol II. Although MTRs did not co-purify with *T. brucei* RNA pol II (Das et al., 2006; Devaux et al., 2006), the interaction may have been lost due to rapid RPB1 dephosphorylation in extract. As our data show (Fig. 3.2A), the phosphorylated form of RPB1 was only detected in whole cell lysates and rapidly converted to the dephosphorylated form during extract preparation, possibly due to the presence of 15 putative homologs of the CTD phosphatase FCP1 in the *T. brucei* genome (www.TritrypDB.org).

On the other hand, there are findings which suggest that the *T. brucei* methyltransferases do not require RNA pol II for recruitment. *T. brucei* MTR1 was found to be in a complex with the SLA1 RNP that contains several subunits including the pseudouridine synthase CBF5, and which guides pseudouridylation of SL RNA position 28 (Zamudio et al., 2009b). While SL RNA pseudouridylation is not essential for trypanosome viability (Hury et al., 2009), CBF5 silencing affected cap4 formation indicating that the SLA1 RNP is important for MTR1 function (Barth et al., 2005). Accordingly, we found that pseudouridylation was not impaired during *CRK9* silencing which indicates that the SLA1 RNP can interact with the SL RNA independent of phosphorylated RNA pol II. Furthermore, MTR2 and MTR3 cap4 methylations depend on Sm protein binding to the SL RNA 3′ terminal domain (Mandelboim et al., 2003; Zeiner et al., 2004a), suggesting that the SL RNP possesses RNA pol II-independent determinants for MTR2 and MTR3 recruitment. However, these arguments are not conclusive and there are other findings which are more in line with co-transcriptional cap4 formation. Firstly, the interaction between SLA1 and SL RNA appears to be weak and transient since purification of the SLA1 RNP did not co-purify detectable amounts of SL RNA, indicating that base pairing alone may not be enough for SLA1 recruitment to the SL RNA (Zamudio et al., 2009b). Taken together, it will require a dedicated study to decipher whether CTD phosphorylation is directly linked to SL RNA cap4 formation or whether RPB1 phosphorylation and cap4 formation are independent processes.

Overall, our study has identified a cyclin-dependent kinase that is of central importance to trypanosome gene expression being essential for SL RNA maturation and RPB1 phosphorylation. Together with previous TFIIH and mediator characterizations, it has provided compelling evidence for RNA pol II transcription initiation and $m^7G$ capping functioning independently of a TFIIH-based, CDK7-mediated CTD phosphorylation event. Our data strongly suggest that the essential CTD function in RNA pol II transcription initiation has evolved after

the divergence of major protistan lineages because it appears to be coupled to the canonical CTD with heptad repeat motifs.

# Chapter IV

# The cyclin-dependent kinase CRK9, a new L-type cyclin and a CRK9-associated protein cooperate in facilitating spliced leader *trans* splicing in *Trypanosoma brucei*

## 1. Abstract

In eukaryotes, cyclin-dependent kinases (CDKs) control the cell cycle and critical steps in gene expression. The lethal parasite *Trypanosoma brucei,* member of the phylogenetic order Kinetoplastida, possesses eleven CDKs which, due to high sequence divergence, were generically termed CDC2-related kinases (CRKs). While several CRKs have been implied in the cell cycle, CRK9 is the first trypanosome CDK that was shown to control the unusual mode of gene expression found in kinetoplastids. In these organisms, protein-coding genes are arranged in tandem arrays which are transcribed polycistronically. Individual mRNAs are processed from precursor RNA by spliced leader (SL) *trans* splicing and polyadenylation. CRK9 ablation was lethal in cultured trypanosomes, causing a block of *trans* splicing at the first transesterification step. Additionally, *CRK9* silencing led to dephosphorylation of RNA polymerase II and to hypomethylation of the SL cap structure. Here, we tandem affinity-purified CRK9 and, among potential CRK9 substrates and modifying enzymes, discovered an unusual tripartite complex of CRK9, a new L-type cyclin (CYC12) and a protein, termed CRK9-associated protein (CRK9AP), that is only conserved among kinetoplastids. Silencing of either *CYC12* or *CRK9AP* reproduced the effects of depleting CRK9, identifying these proteins as functional partners of CRK9 *in vivo*. While mammalian cyclin L binds to CDK11, the CRK9 complex deviates substantially from that of CDK11, requiring CRK9AP for CRK9 complex integrity. Interference with this unusual CDK

rescued mice from lethal trypanosome infections, validating CRK9 as a potential chemotherapeutic target.

## 2. Introduction

*Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania* spp. are unicellular, vector borne, human parasites belonging to the early-diverged phylogenetic order Kinetoplastida. These parasites collectively affect millions of people worldwide primarily in developing countries, causing debilitating to fatal human diseases. There is no preventive vaccine for these diseases and the current medications are unsatisfactory due to poor efficacy, toxicity and developing drug resistance (Barrett and Croft, 2012). Trypanosomatid parasites share a unique mode of protein coding gene expression that is distinct from their hosts. A trypanosomatid genome is organized in large gene clusters, containing tandemly linked protein coding genes which are transcribed polycistronically. Individual mRNAs are resolved from pre-mRNA by spliced leader (SL) *trans* splicing and polyadenylation. In SL *trans* splicing, the SL, derived from the small nuclear SL RNA, is spliced onto the 5′ end of each mRNA (Gunzl, 2010; Michaeli, 2011). Like *cis* splicing, e.g. the removal of introns, SL *trans* splicing comprises two transesterifications, generating a Y-shaped structural intermediate after the first splicing step that is analogous to the lariat structure in *cis* splicing (Murphy et al., 1986; Sutton and Boothroyd, 1986). The 39 nt-long SL harbors an extensively modified cap structure, called cap4, that consists of a standard 7-methylguanosine cap nucleotide followed by four methylated nucleotides (Bangs et al., 1992). Cap4 is important for SL *trans* splicing (McNally and Agabian, 1992; Ullu and Tschudi, 1991) and for efficient translation (Zamudio et al., 2009a). Since each and every mRNA carries a SL, *trans* splicing is absolutely essential for the survival of kinetoplastids. In *T. brucei*, the cdc2-related kinase 9 (CRK9) appears to be of crucial importance for this unusual mode of gene expression since depletion of this cyclin-dependent kinase (CDK) led to a lethal block of *trans* splicing (Badjatia et al., 2013a).

CDKs are serine/threonine kinases that require association with a cyclin for enzymatic activity (Pines, 1995). CDKs were first identified as key regulators of cell cycle progression (Beach et al., 1982; Nasmyth and Reed, 1980), although now it is well established that there is a second group of CDK/cyclin complexes that have crucial functions in transcription and RNA splicing (Lim and Kaldis, 2013; Loyer et al., 2005). There are 13 CDKs in humans all of which share a conserved catalytic domain containing an ATP binding pocket, a PSTAIRE-like helix domain responsible for cyclin binding and an activating T-loop motif (Cao et al., 2014; Malumbres et al., 2009). While CDKs that are important for the cell cycle are regulated by their sequential binding to different cyclins, each of which are transcribed and degraded at defined cell cycle stages, the CDKs and cyclins that are involved in gene expression are constitutively expressed throughout the cell cycle, typically forming a single CDK-cyclin complex (Lim and Kaldis, 2013; Loyer et al., 2005). In metazoans, CDK7 has a dual role in both cell cycle and transcription. Uniquely, CDK7 forms a trimeric complex with Cyclin H and the RING finger domain protein MAT1 (ménage a trois 1) that phosphorylates and activates cell cycle CDKs in their T-loops and, therefore, has been termed the CDK-activating kinase (CAK). CAK is also part of the basal transcription factor TFIIH and, in this association, phosphorylates the carboxy-terminal domain of RPB1 during transcription initiation, promoting promoter clearance of RNA pol II (Akoulitchev et al., 1995; Dahmus, 1996; Serizawa et al., 1995; Shiekhattar et al., 1995) and recruitment of RNA processing factors including the capping enzyme (Rodriguez et al., 2000) reviewed in . The PITSLRE kinase CDK11 is another CDK that has a dual role (Loyer et al., 2005; Satyanarayana and Kaldis, 2009). In murine and human cells it is present as two isoforms, CDK11$^{p110}$ and CDK11$^{p58}$ (Cornelis et al., 2000; Gururajan et al., 1998; Li et al., 2004b; Xiang et al., 1994). While CDK11$^{p58}$ is implicated in mitosis, and is expressed due to a G2/M-specific internal ribosome entry site present in the CDK11$^{p110}$ mRNA (Cornelis et al., 2000; Li et al., 2004b); CDK11$^{p110}$ is constitutively expressed and involved in transcription and pre-mRNA splicing (Hu et al., 2003; Loyer et al., 1998; Trembley et al., 2004). Orthologs of CDK11 are

present in all metazoans and fission yeast but absent from budding yeast (Guo and Stiller, 2004; Liu and Kipreos, 2000).

*Trypanosoma brucei* harbors eleven CDKs that were generically termed CRK1-4 and CRK6-12 since their high sequence divergence has prevented orthologous assignments with CDKs of other eukaryotes. In addition, ten cyclins (CYC2-11) have been identified in *T. brucei* so far (Hammarton, 2007; Naula et al., 2005). Due to their crucial importance for gene expression and cell proliferation, CDKs are generally considered to be promising drug targets and, therefore, trypanosomatid CRKs have been a research focus. RNAi-mediated gene knockdowns identified CRK1-3, CRK6, CRK9 and CRK12 as essential CDKs in *T. brucei* (Badjatia et al., 2013a; Gourguechon and Wang, 2009; Jones et al., 2014; Mackey et al., 2011; Merritt and Stuart, 2013; Monnerat et al., 2013; Tu and Wang, 2004). Less is known about the specific functions of these important enzymes. CRK3 was shown to be the functional homolog of human CDK1; it partners with CYC6 (also referred to as Cyclin B2) and regulates mitosis (Hammarton et al., 2003; Li and Wang, 2003; Tu and Wang, 2004). Depletion of both CRK3 (Tu and Wang, 2004) and CYC6 (Hammarton et al., 2003; Li and Wang, 2003) results in a block of G2/M progression and cell death in the PF as well as BF. Interestingly, CRK3 also forms a pair with CYC2 (aka cyclin E1) *in vivo* (Van Hellemond et al., 2000). The knockdown of *CYC2* is lethal and causes a G1 arrest in cells (Hammarton et al., 2004; Li and Wang, 2003), but a function of CRK3 in G1 phase if any, is yet unknown. CRK1 is important for G1/S progression and interacts with CYC2 in a yeast two-hybrid screen but this interaction was not observed *in vivo* (Gourguechon et al., 2007; Tu and Wang, 2004; Van Hellemond et al., 2000). CRK2, CRK4 and CRK6 appear to have accessory functions in the cell cycle although their cyclin partners are unknown (Tu and Wang, 2005). Recently, CYC9 was shown to associate with CRK12 during a tandem affinity purification of CRK12. Both CYC9 and CRK12 were essential in BF, but gene silencing resulted in distinct phenotypes, suggesting that while CYC9 is important for cytokinesis, CRK12 has a role in endocytosis (Monnerat et al., 2013).

CRK9 is the first trypanosome CDK shown to have an essential role in gene expression. In addition, like mammalian CDK7 and CDK11, this CDK seems to function in the cell cycle, too, since ablation of CRK9 in insect-stage, procyclic form (PF) trypanosomes led to a defect in basal body segregation during cytokinesis (Gourguechon and Wang, 2009). CRK9's essential function in gene expression became apparent by RNA analysis of *CRK9*-silenced PFs and BFs both of which revealed a strong decline of mature mRNA and a concomitant increase of unspliced pre-mRNA. Similarly, while the *trans* splicing substrate SL RNA was dramatically increased in *CRK9*-silenced cells, the Y structure splicing intermediate exhibited a strong decline in abundance (Badjatia et al., 2013a). Moreover, introduction of an RNAi-resistant *CRK9* wild-type gene into these cells completely rescued this phenotype whereas introduction of an equivalent *CRK9* gene carrying a mutation in its T-loop sequence did not. Together, these results showed that CRK9 enzyme activity is crucial for the first step of splicing.

Interestingly, CRK9 ablation also caused dephosphorylation of RPB1 and hypomethylation of the SL cap, raising the possibility that CRK9 is a functional homolog of mammalian CDK7, directly phosphorylating the CTD of RPB1. However, biochemical and structural characterization of trypanosome TFIIH did not reveal a kinase sub-complex (Lee et al., 2010; Lee et al., 2009) and *CRK9* silencing did not lead to a RNA pol II-specific transcription defect (Badjatia et al., 2013a). Conversely, the *CRK9* silencing defects could be quantitatively reproduced by depleting a subunit of the spliceosomal PRP19 complex that, in other systems, was shown to be essential for spliceosome activation, indicating a potential role of CRK9 in the RNA processing branch of gene expression (Ambrosio et al., 2015). Thus, while its importance for trypanosome gene expression and viability has been clearly demonstrated, the nature of CRK9 remains undetermined.

Here we have characterized the enzyme complex and validated CRK9 enzyme activity as a potential drug target in the mouse. CRK9 forms a tripartite complex together with a novel, previously unannotated L-type cyclin that we termed CYC12 and a kinetoplastid-specific protein

that we call CRK9-associated protein (CRK9AP). *CYC12* and *CRK9AP* silencing reproduced the *CRK9* silencing phenotypes, identifying them as essential, functional partners of CRK9 *in vivo*. This is highly unusual as with the exception of CDK7, CDKs typically function in be reconstituted in active form in vitro and do not require additional protein partners for their kinase activity. Interestingly, CRK9AP appears to be important for the integrity of the complex as the knockdown of CRK9AP resulted in a concomitant loss of both CRK9 and CYC12.

## 3. Results

### 3.1 Characterization of the CRK9 enzyme complex

Since trypanosome CRK9 enzyme activity was required for spliced leader *trans* splicing, a parasite-specific step in the maturation of all mRNAs, we tagged and tandem affinity-purified the enzyme to identify its cyclin partner and other co-purifying proteins by liquid chromatography-tandem mass spectrometry (LC/MS/MS). By two consecutive transfections, we generated the clonal PF cell line TbC9ee, in which the hygromycin phosphotransferase gene (*HYG*$^{R}$) replaced one *CRK9* allele and targeted insertion of plasmid CRK9-PTP-NEO fused the sequence of the composite PTP tag, which encodes the protein C epitope (ProtC), a tobacco etch virus (TEV) protease cleavage site and a tandem protein A domain (ProtA), to the 3$^{/}$ end of the CRK9 coding region (Fig. 4.1A). To confirm that TbC9ee cells exclusively expressed tagged CRK9, we raised a specific anti-CRK9 immune serum in rats against a partial recombinant GST-tagged CRK9 protein that was expressed and purified from *Escherichia coli*. In an immunoblot of whole cell lysates, the immune serum detected a single protein band of ~100 kDa in wild-type trypanosomes which is slightly larger than the predicted CRK9 mass of 85.6 kDa (Fig. 4.1B). Accordingly to the molecular mass of 19 kDa for the PTP tag, in TbC9ee cells, this band was replaced by a ~120 kDa band that was also detected by a ProtA-specific probe, confirming exclusive expression of CRK9-PTP in TbC9ee cells. TbC9ee cells did not exhibit reduced

culture growth as compared to the wild-type cells. Additionally, the phosphorylation status of RPB1, which has been linked to CRK9 function, was comparable in wild-type and TbC9ee cells (Fig. 4.1B, bottom panel). Hence, we concluded that the PTP tag did not interfere with CRK9 function.

Facilitated by the PTP tag, we purified CRK9-PTP by IgG affinity chromatography, TEV protease cleavage and anti-ProtC immunoaffinity chromatography. An anti-ProtC immunoblot analysis revealed efficient capture of the tagged protein in all steps of the tandem affinity purification (Fig. 4.1C). Separation of the final eluate and, as controls, of the input material and the TEV protease fraction, by SDS-PAGE and staining of proteins by Sypro Ruby and Coomassie Blue, revealed a dominant band of ~105 kDa, representing CRK9-P (note that CRK9-PTP was reduced by ~15 kDa to CRK9-P by TEV protease cleavage), and, unexpectedly, many co-purified protein bands (Fig. 4.1D). Mass spectrometry of two independent CRK9-PTP purifications revealed a total of 162 proteins that were identified with an expect value smaller than 0.01 and at least by two unique peptides (Attachment 3, S1 Table). The majority of proteins, including those with very high Mascot Scores, were proteins of both ribosomal subunits. This massive co-purification of ribosomal proteins was surprising because CRK9 is a nuclear protein and not enriched in the nucleolus (Gourguechon and Wang, 2009), and previous PTP purifications of various nuclear complexes contained only few, low scoring ribosomal proteins. Since the functional association of CRK9 with ribosomal complexes remains unclear, we focused on non-ribosomal proteins. Table 4.1 lists the most significant identifications. While no annotated cyclin was detected in the purification, the protein with the second highest score is a new trypanosome cyclin (see below) which we termed CYC12 (Tb927.10.9160). Among the top scoring proteins are three dual specificity protein kinases capable of phosphorylating serines and threonines as well as tyrosines. While these kinases may be substrates of CRK9, it is equally possible that they modify CRK9 since phospho-proteomic studies revealed that residues threonine 533, serine 746, serine 750 and threonine

752 of CRK9 are phosphorylated in the cell (Nett et al., 2009; Urbaniak et al., 2013).The list also contains four proteins that are only conserved within the genus *Trypanosoma*, indicating that, in addition to CRK9's fundamental role in gene expression, there may be additional, trypanosome-specific functions associated with this kinase. Finally, as potential direct links to the *trans* splicing mechanism, polyubiquitin-binding protein 2 (PABP2) and the SR protein TSR1 were identified.

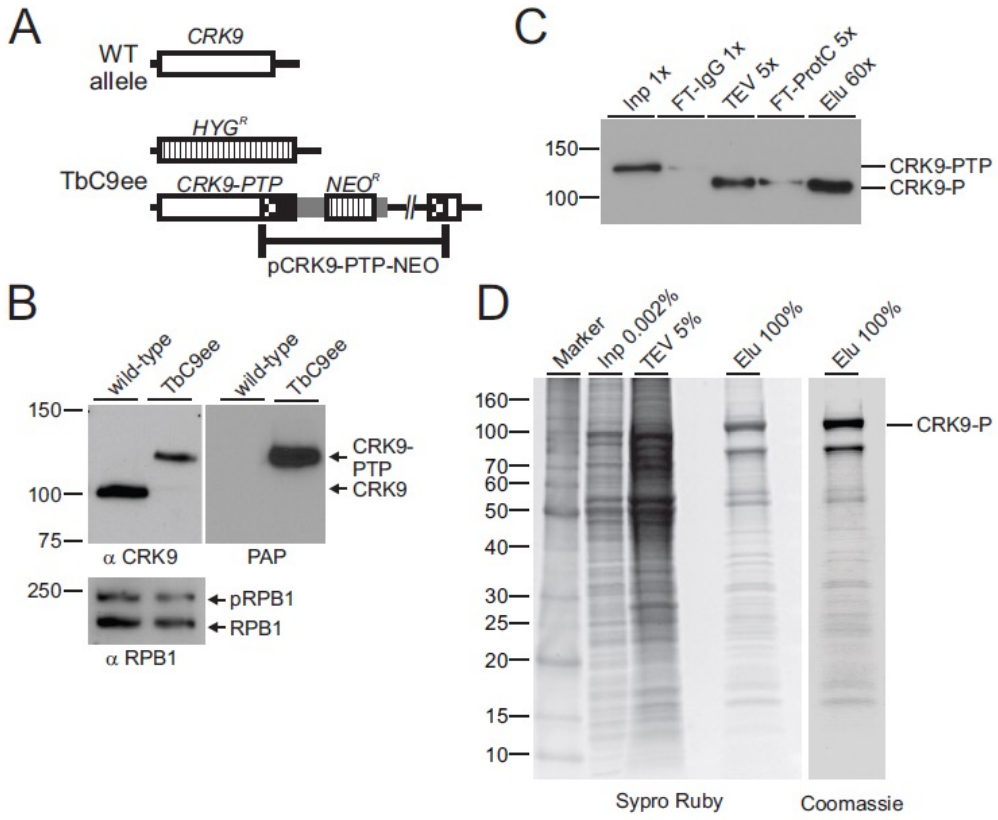



**Figure 4.1. Tandem affinity purification of CRK9.**

**(A)** Schematic depiction (not to scale) of the CRK9 locus in procyclic TbC9ee cells that exclusively express CRK9-PTP and no untagged CRK9. In these cells one wild-type CRK9 allele (open box) was knocked out by a hygromycin phosphotransferase gene (*HYG*$^R$, striped box). Integration of plasmid CRK9-PTP-NEO introduced the PTP sequence (black box) and the neomycin phosphotransferase (*NEO*$^R$) to the second allele. Smaller gray boxes indicate gene flanks for RNA processing signals and checkered boxes depict CRK9 sequences encoded in the plasmid.

**(B)** Immunoblot of whole cell lysates of wild-type trypanosomes and of TbC9ee cells, detecting CRK9 and CRK9-PTP, respectively, with the newly generated anti-CRK9 polyclonal immune serum. On the same blot, phosphorylated (p) and unphosphorylated RPB1 was detected as a loading control with a polyclonal antibody, and CRK9-PTP specifically with the peroxidase anti-peroxidase reagent (PAP) that binds to the ProtA domains of the PTP tag.

**(C)** Immunoblot monitoring of the CRK9-PTP purification detecting CRK9-PTP in crude extract (Inp) and the flowthrough of IgG affinity chromatography (FT-IgG), and CRK9-P in TEV protease eluate, the flowthrough of the anti-ProtC immunoaffinity chromatography (FT-ProtC) and the final eluate (Elu) with the monoclonal HCPC4 anti-ProtC antibody. The x-values indicate relative amounts analyzed.

**(D)** Protein analysis of the CRK9-PTP purification. Proteins of crude extract, the TEV eluate and the final eluate were separated on 10-20% SDS polyacrylamide gel and first stained with Sypro Ruby and, subsequently, with Coomassie blue (right panel). Marker sizes in kDa are indicated on the left.

**Table 4.1 Mass spectrometric identification of CRK9 co-purified proteins**

| Rank | Annotation [1] | Accession # [1] | $M_r$ (kDa) | Score [2] | % Coverage | emPAI [3] |
|---|---|---|---|---|---|---|
| 1 | CRK9 | Tb927.2.4510 | 85.5 | 13,418 | 73.9 | 10.3 |
| 2 | CYC12 | Tb927.10.9160 | 71.2 | 9,956 | 43.3 | 8.1 |
| 6 | protein kinase [4] | Tb927.7.3880 | 87.0 | 1,797 | 49.7 | 2.1 |
| 19 | hypothetical, conserved [5] | Tb927.9.13970 | 36.8 | 805 | 47.3 | 2.8 |
| 22 | protein kinase [4] | Tb927.10.350 | 58.6 | 795 | 35.5 | 1.5 |
| 24 | PABP2 | Tb927.9.10770 | 62.2 | 724 | 37.8 | 1.7 |
| 28 | protein kinase [4] | Tb927.3.1610 | 74.2 | 661 | 34.1 | 1.0 |
| 37 | CRK9AP | Tb927.3.4170 | 13.0 | 568 | 37.2 | 5.5 |
| 46 | hypothetical, conserved | Tb927.3.3740 | 69.1 | 525 | 34.4 | 0.8 |
| 50 | hypothetical, conserved [5] | Tb927.1.4680 | 37.3 | 474 | 36.3 | 1.2 |
| 59 | hypothetical, conserved | Tb927.10.11600 | 66.0 | 382 | 21.7 | 0.7 |
| 62 | MRB1-assoc. protein | Tb927.11.6320 | 53.2 | 353 | 29.3 | 0.6 |
| 69 | NRBD1 (RNA-binding) | Tb927.11.14000 | 28.8 | 299 | 37.8 | 1.7 |
| 73 | hypo. conserved [5] | Tb927.7.3080 | 47.3 | 285 | 31.9 | 1.0 |
| 74 | RACK1 (kinase receptor) | Tb927.11.11360 | 34.7 | 285 | 18.7 | 1.0 |
| 77 | hypothetical, conserved [5] | Tb927.4.3150 | 38.1 | 275 | 27.7 | 0.7 |
| 84 | PNO1 (pre-rRNA process.) | Tb927.9.11840 | 24.1 | 257 | 32.5 | 1.3 |
| 93 | TSR1, splicing factor | Tb927.8.900 | 37.4 | 226 | 22.1 | 0.6 |

List of CRK9-PTP co-purified proteins, identified by LC/MS/MS, that were not a common contaminant of previous tandem affinity purifications, not annotated a ribosomal protein, had a Mascot Score of at least 200 and an emPAI value equal or greater than 0.5. The rank number is according to the complete protein list in S1 Table. The CRK9 kinase complex subunits are highlighted in gray.

[1] Protein annotation and accession numbers are from the *T. b. brucei* 927 database at www.TriTrypDB.org, although proteins were identified through the *T. b. brucei* 427 Lister genome database whose annotation and assembly is not as complete.

[2] The Mascot Score corresponds to -10 x $LOG_{10}$(P), where P is the absolute probability that the observed match is a random event.

[3] The emPAI (**e**xponentially **m**odified **P**rotein **A**bundance **I**ndex) is a measure for the relative amount of an identified protein in the final eluate (Ishihama et al., 2005).

[4] Dual specificity kinases for tyrosine and serine/threonine phosphorylation.

[5] These proteins appear to be conserved only in the genus *Trypanosoma*.

To potentially separate the kinase complex from co-purifying proteins, we sedimented the eluate of a CRK9-PTP purification through a sucrose gradient by ultracentrifugation. While part of the enzyme and nearly all other co-purified proteins migrated to the bottom of the gradient, CRK9-P exhibited a strong sedimentation peak in fractions 9-11 (Fig. 4.2A). Protein bands of ~80, 37 and 15 kDa co-sedimented in these fractions. Since transcriptional CDKs of other systems are capable of autophosphorylating their T-loop motifs, we carried out a kinase assay with fractions 9, 11, 13, and 20 detecting disproportionately high CRK9-P phosphorylation in fractions 9 and 11, indicating that fractions 9-11 contain the active CDK9 complex (Fig. 4.2B). Mass spectrometry identified the protein of the 37 kDa band, which exhibited a slightly different sedimentation profile than the other two bands, as a putative ribosomal protein (Tb927.11.2050); it may be a direct interactor of CRK9 within the ribosome. The two other bands were non-ribosomal proteins: The 80 kDa band contained CYC12 whereas the 15 kDa band revealed a protein (Tb927.3.4170) whose conservation is conserved only among kinetoplastid organisms and which we termed CRK9-associated protein or CRK9AP (Attachment 3, S2 Figure). To enable verification of these protein identifications by reciprocal co-immunoprecipitation, we manipulated one *CYC12* allele in TbC9ee cells, fusing an HA tag C-terminally to CYC12, and raised a polyclonal immune serum in rats against recombinant, GST-tagged CRK9AP (Attachment 3, S3 Figure). As shown in figure 2C, precipitation of either CRK9-PTP, CYC12-HA or CRK9AP specifically co-precipitated the other two proteins, indicating the presence of a tripartite enzyme complex.

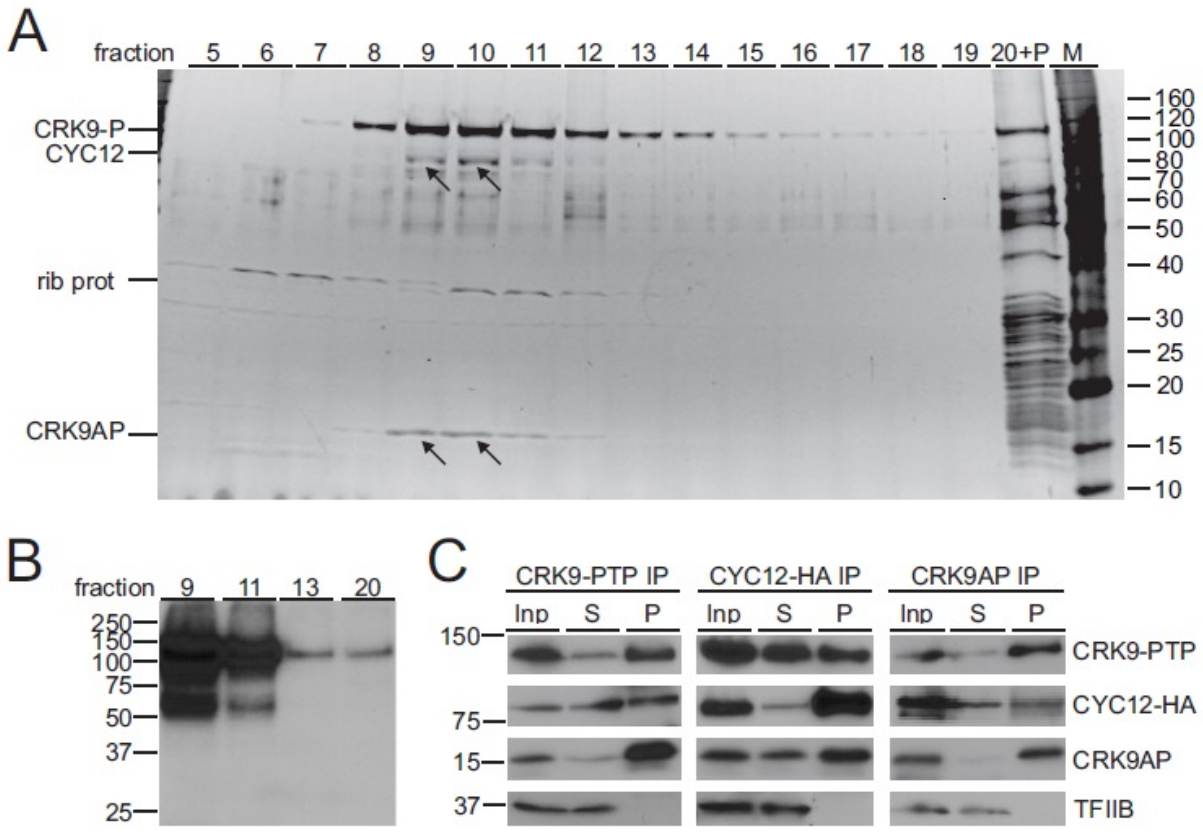



**Figure 4.2 CRK9 interacts and co-sediments with two unannotated proteins.**

**(A)** CRK9-PTP tandem affinity-purified material was sedimented through a 10-40% linear sucrose gradient by ultracentrifugation and fractionated into 20 aliquots from top to bottom. Proteins from each fraction were separated by SDS-PAGE and stained with sypro ruby. Protein bands were excised and identified by LC/MS/MS. Arrows point to the CYC12 and CRK9AP bands which co-sediment with CRK9 in fractions 9/10. The 35 kDa band with a peak in fractions 10/11 was found to be the putative ribosomal protein L5 (Tb927.9.15110/15150).

**(B)** Kinase assay with materials from indicated fractions suggest autophosphorylation of CRK9.

**(C)** Reciprocal co-IP assays of extracts prepared from a cell line in which CRK9 was exclusively PTP-tagged and an HA tag sequence was inserted at the 3'end of one CYC12 allele. The precipitate (P) was loaded at a fourfold excess relative to extract (Inp) and supernatant (S). Detection of the RNA pol II transcription factor TFIIB served as a negative precipitation control.

### 3.2 CYC12 is a new L-type cyclin

Mammalian cyclin L homologues exclusively partner with CDK11. There are two human cyclin L paralogs, L1 and L2, which share 60% identity. These proteins have a ~200 amino acid-long N-terminal composite CCL1 domain that comprises the highly conserved Cyclin_N domain, the less conserved Cyclin_C domain (Nugent et al. 1991) and additional sequence conservation, characteristic for transcriptional cyclins, around these domains. In addition, cyclin L is an SR-related protein that features a conserved, C-terminal, highly positively charged arginine/serine-rich RS domain with repetitive "SR" dipeptide motifs (Fig. 4.3A). A standard protein-protein BLAST search of the human proteome with the *T. brucei* CYC12 sequence returned L1 and L2 cyclins ($E=2e^{-09}$) but no other cyclins. A multiple sequence alignment of the CCL1 domain of cyclin L orthologs from model organisms and of kinetoplastid CYC12 sequences showed convincing sequence conservation across the whole domain, although, as expected, the Cyclin_C domain was less well conserved (Attachment 3, S4 Figure). Interestingly, both cyclin domains were disrupted by extensive kinetoplastid-specific sequence insertions which likely prevented recognition of CYC12 as a cyclin in trypanosomatid genome annotations and of the cyclin domains by the NCBI BLAST algorithm (Fig. 4.3A, Attachment 3, S4 figure). Analysis of the C-terminus of CYC12 revealed a strong reduction of SR dipeptide motifs when compared to the human cyclin L domain (21 versus 7). However, conserved among all kinetoplastid CYC12 orthologs, the C-terminal domain has a similarly high isoelectric point of ~12 as its counterparts in other eukaryotes.

To substantiate the notion that CYC12 is an L-type cyclin, we generated a phylogenetic tree with the sequence alignment of all human cyclin sequences and four representative kinetoplastid CYC12 sequences. To avoid a potential bias of the charged cyclin L/CYC12 C-terminal domain, we restricted the analysis to the CCL1 domain. As expected, CYC12 sequences unambiguously partitioned with the transcriptional cyclins and formed a branch with human L cyclins with a bootstrap value of 87% (Fig. 4.3B). A more extensive phylogenetic

analysis of complete sequence alignments of cyclins of model organisms and all known kinetoplastid CYC12 orthologs revealed a similar tree although the bootstrap value of the cyclin L/CYC12 cluster was reduced to 53% (Attachment 3, S5 Figure). Taken together, these data strongly indicate that kinetoplastid CYC12 represents a L-type cyclin.

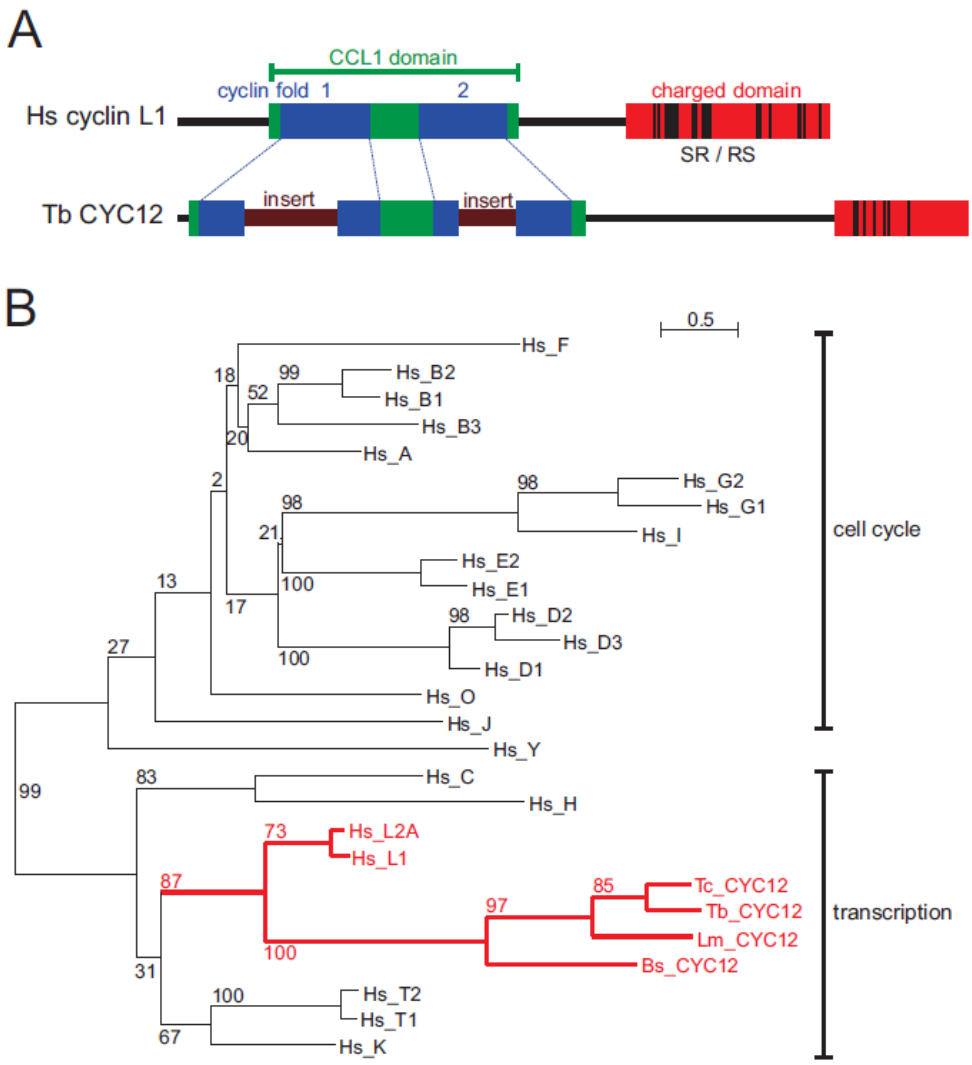



**Figure 4.3 Cyclin CYC12 is an L-type cyclin.**

**(A)** Schematic drawing to scale of the human L1 and *T. brucei* CYC12 cyclins. The two cyclin folds (blue) are embedded in the CCL1 domain (green). The charged RS domain (red) was defined by a hydrophilicity Kyte & Doolittle blot score of < -2. Black lines indicate SR or RS dipeptides. Both cyclin domains of CYC12 are disrupted by insertions.

**(B)** Phylogenetic tree, generated by the maximum likelihood algorithm and based on a multiple sequence alignment of the cyclin domains of human cyclins and of CYC12s from *T. brucei* (Tb), *T. cruzi* (Tc), *L. major* (Lm) and the bodonid *Bodo saltans* (Bs) (for accession numbers see Figure S5). Cyclins involved in the cell cycle and in transcriptional control are indicated. Bootstrap values are indicated in percentages and were derived from 1000 replicates. The common branch of human cyclins L and kinetoplastid CYC12s is drawn in red.

### 3.3 CYC12 and CRK9AP are functional partners of CRK9

To test whether CYC12 and CRK9AP are of similar importance as CRK9 for trypanosome gene expression, we generated PF and BF cell lines for conditional silencing of each gene. Since the culture growth defects upon gene silencing were very similar between the lines for each gene (data not shown), we analyzed the knockdown phenotype in detail in one of those cell lines. In PFs, *CYC12*- and *CRK9AP*-silencing strongly affected culture growth two days after induction of the gene knockdown. Although the defects were delayed by one day when compared to the *CRK9* knockdown (Badjatia et al., 2013a), trypanosome numbers diminished after the second day of induction (Figure 4.4A). The decline of PF trypanosome numbers was not as strong in *CYC12*-silenced cells as it was with *CRK9*- and *CRK9AP*-silenced cells, indicating a different dynamic of the lethal defect. In BFs, this variation of gene silencing was not observed and the knockdown of each of the three genes was even more lethal than in PFs. As was observed previously in the *CRK9* knockdown (Badjatia et al., 2013a), *CYC12* or *CRK9AP* silencing caused a culture growth defect within 24 hours of induction and, subsequently, cell numbers rapidly declined with no viable cells left on day 4 after induction (Attachment 3, S6 figure, (Badjatia et al., 2013a). Thus, we concluded that *CYC12* and *CRK9AP* are generally essential genes in *T. brucei*.

The *trans* splicing block upon *CRK9* silencing became apparent by a decrease of mature mRNA and the Y structure intermediate, the concomitant increase of unspliced pre-mRNA, and the accumulation of SL RNA with a hypomethylated cap structure (Badjatia et al., 2013a). Analyses of total RNA from *CYC12*- and *CRK9AP*-silenced cells revealed similar defects (Figure 4.4B). As analyzed for α tubulin and *RPB7* (RPB7 encodes a subunit of RNA pol II) by RT-PCR, the total coding RNA declined upon gene silencing whereas the abundance of unspliced RNA from these genes strongly increased during this period. Furthermore, primer extension assays of the same RNA preparations showed that, in both knockdowns, SL RNA increased, predominantly in its hypomethylated form, whereas the Y structure intermediate

declined during the time course experiments (Figure 4.4B). These experiments clearly indicated a block of *trans* splicing before the first transesterification step in both gene knockdowns. Analysis of α tubulin RNA in BFs revealed the same *trans* splicing defect (Attachment 3, S6 Figure).

Another specific effect of *CRK9* silencing was dephosphorylation of the RNA pol II subunit RPB1 (Badjatia et al., 2013a). To monitor the phosphorylation status of this RNA pol II subunit, we probed an immunoblot of whole cell lysates with a specific immune serum raised against the C-terminal domain of RPB1 (Badjatia et al., 2013a). As shown in Figure 4.4C, the signal for phosphorylated RPB1 clearly decreased upon both *CYC12*- and *CRK9A*P-silencing whereas the dephosphorylated signal concomitantly increased. Note that the disproportional increase of the unphosphorylated RPB1 signal likely is due to an increase of antibody-recognizable epitopes since the immune serum was raised against unphosphorylated protein.

Finally, as described previously, *CRK9* silencing leads to an atypical rounding up of cells (Badjatia et al., 2013a; Gourguechon and Wang, 2009). Virtually all PFs that survive 3 days of *CRK9* silencing exhibit a round shape (Attachment 3, S7 figure) reminiscent of FAT cells that were observed in the discovery of the RNA interference pathway upon tubulin gene knockdowns (Ngo). Silencing of *CYC12* and of *CRK9AP* resulted to the same extent in the same characteristic phenotype (Attachment 3, S7 figure). Rounding up of cells was also observed in BFs in all three gene knockdowns although the number of round cells detected after three days of induction was limited to 81%, 55%, and 50% (n = 50) in *CRK9*-, *CYC12*-, and *CRK9AP*-silenced cells, respectively. Possibly, the rapid onset of the deleterious effect in BFs prevented rounding up of all cultured cells.

Taken together, *CYC12* and *CRK9AP* silencing resulted in the same block of *trans* splicing and RPB1 dephosphorylation in both PFs and BFs as observed previously upon *CRK9* silencing. Moreover, all three gene knockdowns caused cells to round up, an atypical death phenotype. These data demonstrate that CRK9, CYC12 and CRK9AP are functional partners in

facilitating SL *trans* splicing in trypanosomes. Moreover, since this function requires an intact T-loop of CRK9 (Badjatia et al., 2013a), our data indicate that all three proteins are required *in vivo* for CRK9-dependent phosphorylation.

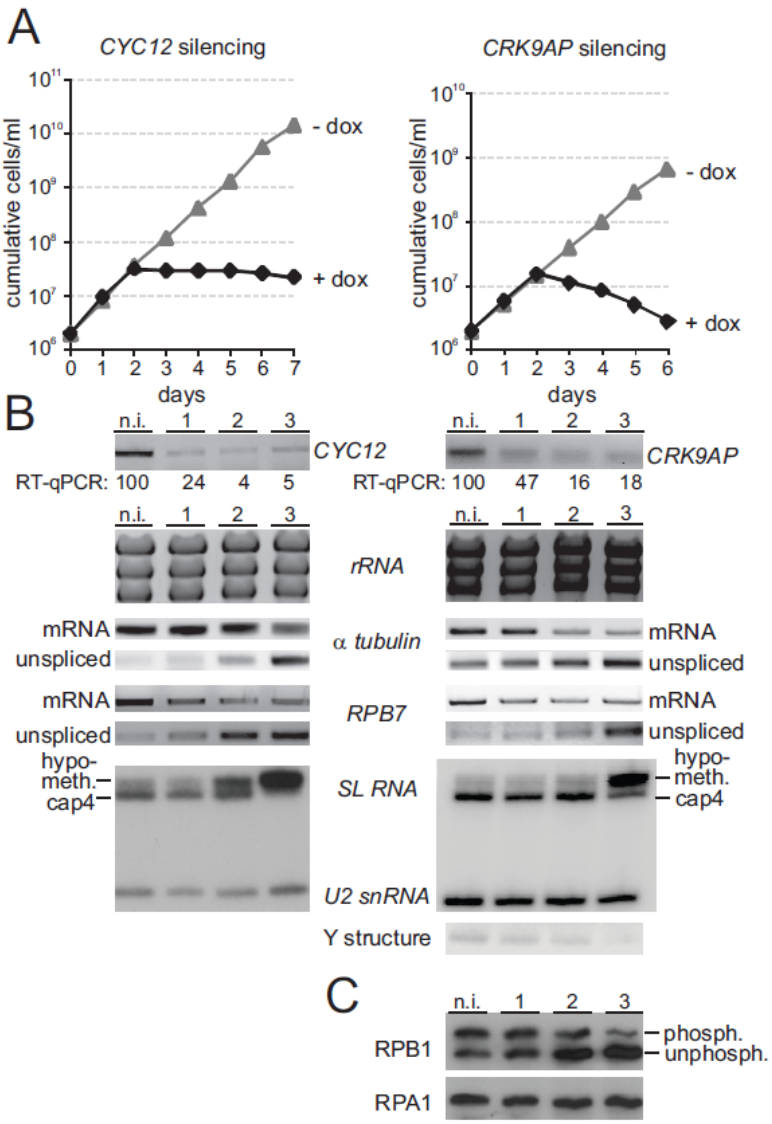



**Figure 4.4 CYC12 and CRK9AP are functional partners of CRK9.**

**(A)** Cumulative culture growth curves were obtained for *CYC12* and *CRK9AP* silencing in the absence and presence of doxycyline (dox), the gene knockdown-inducing compound. For each knockdown a representative growth curve is shown.

**(B)** Analysis of total RNA prepared from non-induced cells and cells in which *CYC12* or *CRK9AP* were silenced for 1, 2 or 3 days. *CYC12* or *CRK9AP* mRNA as well as α tubulin and *RPB7* mRNA were analyzed by reverse transcription of oligo-dT and semi-quantitative PCR, whereas unspliced, pre-mRNA of α tubulin and *RPB7* were analyzed by reverse transcription of random hexamers and by PCR using an oligonucleotide upstream of the SL addition site. rRNA was visualized by ethidium bromide staining after separation in an agarose gel. SL RNA, U2 snRNA and the Y structure intermediate were detected by primer extension assays using a SL RNA and a U2 snRNA-specific primer in the same reactions.

**(C)** Anti-RPB1 immunoblot analysis of whole-cell lysates prepared from *CRK9AP*-silenced cells. Detection of the similar-sized RNA pol I subunit RPA1 served as a loading control.

### 3.4 *CRK9AP* silencing causes rapid co-loss of CRK9 and CYC12

A tripartite CRK9 enzyme complex would be highly unusual since, to our knowledge, eukaryotic CDK7 is the only CDK whose enzyme activity depends on three proteins. Thus, to understand the specific function of CRK9AP, we investigated the effect of CRK9AP ablation on the expression of its functional partners. While the *CRK9* and *CYC12* RNA abundances were unaffected by *CRK9AP* silencing in both PFs and BFs (data not shown), both proteins were rapidly lost upon CRK9AP depletion (Figure 4.5). This result indicates that CRK9AP is important for CRK9 complex assembly and/or integrity. Alternatively, CRK9AP may mediate nuclear import of the complex, although the amino acid sequence did not reveal a conventional nuclear localization signal (data not shown).

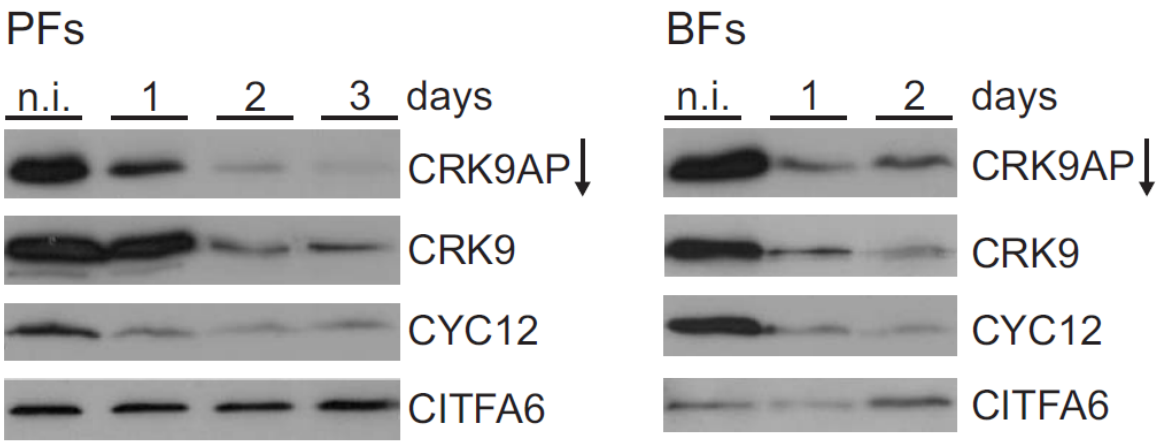



**Figure 4.5 CRK9AP depletion results in rapid co-loss of CRK9 and CYC12.** Immunoblot of whole cell lysates derived from non-induced (n.i.) and *CRK9AP*-silenced PF and BF trypanosomes. The arrow indicates the gene knockdown or *CRK9AP*. Detection of the class I transcription factor I subunit 6 (CITFA6) served as a loading control.

### 3.5 Validation of CRK9 as a potential chemotherapeutic target in the mouse model

Since CRK9 enzyme activity is essential for SL *trans* splicing in *T. brucei* (Badjatia et al., 2013a) and assembles into a unique complex with its functional CYC12 and CRK9AP partners, this CDK is a promising chemotherapeutic target for trypanosomatid parasites. Thus, we validated CRK9 as a drug target in the mouse host. We first used so-called single marker BF (smBF) trypanosomes, the parent line for conditional gene silencing experiments (Wirtz et al., 1999), to determine their lethal dose in intraperitoneal BALB/c mouse infections. Injection of 2 million smBFs was lethal to all mice tested within 3-5 days (data not shown). Accordingly, when we repeated the mouse infections with a cell line in which doxycycline induced the expression of CRK9 3$^{/}$ UTR dsRNA (Badjatia et al., 2013a), mice died on days 4 and 5 in the absence of doxycycline (n=15). However, when mice (n=16) obtained doxycycline in their drinking water, they all survived for two weeks, the time doxycycline was administered (Figure 4.6A). Interestingly, once doxycycline was removed from the drinking water, six out of sixteen mice survived for another month until the experiment was terminated, indicating that they were fully cured by the doxycycline treatment (data not shown). To control that this effect was due to CRK9 ablation and not an off-target effect, we introduced a *CRK9* transgene, which was resistant to RNAi due to a different 3/ UTR sequence, into the CRK9 knockdown cell line. As expected, these cells were not affected by doxycycline and were lethal to all mice in the absence (n=15) and presence (n=15) of doxycycline (Figure 4.6B). Thus, we concluded that inhibition of CRK9 is a valid strategy to combat trypanosome infections of mammalian hosts.

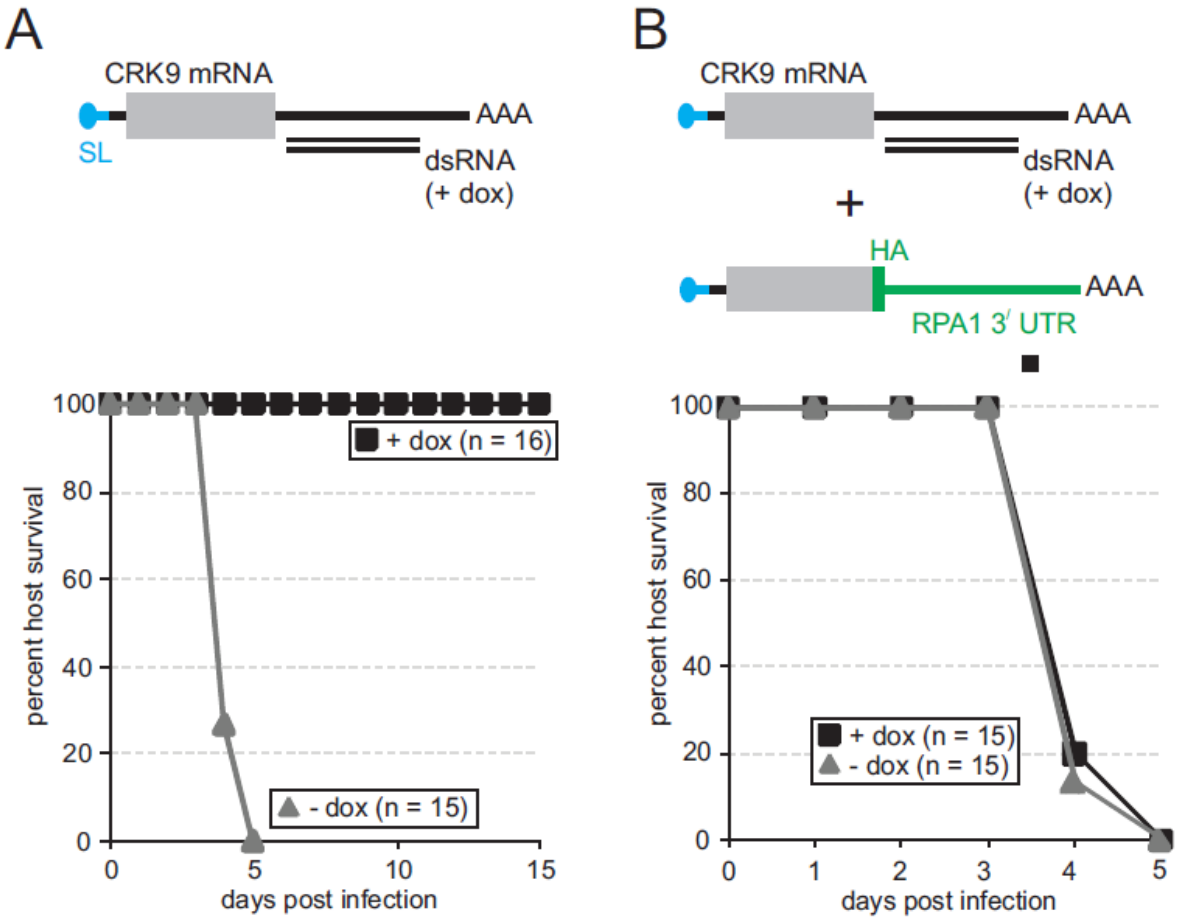



**Figure 4.6 Validation of CRK9 as a drug target in the mouse.**

**(A and B)** Two cell lines derived from the *T. brucei brucei* 427 smBF cell line were used for mouse infection studies. The line on the left **(A)** harbored a construct for conditional expression of dsRNA that targets the 3′ UTR of the *CRK9* mRNA. This line was further modified **(B)** by targeted integration of a plasmid into the endogenous *CRK9* locus that fused a functional HA tag sequence and the 3′ UTR of *RPA1* to the 3′ end of one *CRK9* allele, making the corresponding mRNA resistant to the RNAi response. As the survival graphs of infected mice show in the bottom panels, doxycycline treatment rescued every single mouse when CRK9 was depleted. This effect was completely abolished upon generating an RNAi-resistant *CRK9* gene in the same trypanosomes.

## 4. Discussion

CRK9 is the first CDK of trypanosomes known to be crucial for the parasite-specific path of gene expression. Ablation of CRK9 blocks SL trans splicing, leads to an accumulation of SL RNA with a hypomethylated cap structure and causes dephosphorylation of RPB1, the largest subunit of RNA pol II, indicating that CRK9 is a gene expression regulator of central importance (Badjatia et al., 2013a). Here we identified two un-annotated proteins that co-purified and co-sedimented with CRK9. Moreover, silencing of the respective genes was lethal and recapitulated the defects seen upon CRK9 ablation, identifying these two proteins as functional partners of CRK9 *in vivo* and suggesting that CRK9 may function in a tripartite enzyme complex. Multiple sequence alignments and phylogenetic evidence suggests that the larger protein is an L-type cyclin that was not recognized before due to unique sequence insertions in both its Cyclin_N and Cyclin_C domains. The smaller protein, CRK9AP, is conserved only among kinetoplastids without an established sequence motif.

CRK9 functionally partners with CYC12, which most likely is the kinetoplastid ortholog of human cyclin L. Interestingly, a standard protein-protein BLAST search of the human or *Drosophila melanogaster* proteome with the *T. brucei* CRK9 sequence returned various isoforms of CDK11 ($E=5e^{-43}$ & $2e^{-37}$ respectively). Since Cyclin L is a known functional partner of CDK11 (Loyer et al., 2008), this together with the phylogenetic clustering of CRK9 with CDK10/CDK11 in a previous study with a bootstrap of 45% (Monnerat et al., 2013), suggests that CRK9 is a divergent kinetoplastid ortholog of CDK11 kinase of higher eukaryotes. The members of CDK10/11 family have a conserved PIT/SSLRE-helix motif (Trembley et al., 2004) instead of the PSTAIRE-helix box that is characteristic of the cell cycle-related CDKs (Morgan and De Bondt, 1994) . This helix box is crucial for CDK-cyclin interaction and the activation of kinase subsequent to cyclin binding (Jeffrey et al., 1995; Morgan and De Bondt, 1994). Interestingly, while CRK9 appears to be related to CDK10/11 clade in phylogenetic analyses (Monnerat et al., 2013), all CRK9 orthologs from kinetoplastids have a conserved PPYLLRE (or

PPYMLRE for *Leishmania* species) motif instead of PITSLRE motif that is characteristic of CDK10/CDK11 branch.

CDK11$^{p110}$ isoform is essential for pre- mRNA splicing both *in vivo* and *in vitro* splicing assays (Choi et al., 2014; Hu et al., 2003). CDK11$^{p110}$/cyclin L associates with RNA pol II, CK2, transcription elongation factors as well as several splicing factors including PRP19 and the SR proteins: RNPS1, 9G8 and ASF/SF1 (Choi et al., 2014; Trembley et al., 2002; Trembley et al., 2003). In addition, 9G8, which is a crucial splicing factor (Cavaloc et al., 1994), was identified as a substrate for CDK11$^{p110}$ both *in vitro* and *in vivo* (Trembley et al., 2003). Our data clearly shows that CRK9 and the associated L-type cyclin, CYC12 are crucial for *trans* splicing in *T. brucei* and this suggests some functional overlap between trypanosome CRK9 and metazoan CDK11. Consistent with this several SR proteins including TSR1 that functions in splicing in trypanosomes and the trypanosome ortholog of ASF/SF1 were co-purified with CRK9. Since the function of SR proteins in splicing is typically regulated by phosphorylations (Shepard and Hertel, 2009), it will be interesting to find out if one of these co-purified SR proteins is a substrate of CRK9. Interestingly, a recent study shows that CDK11 homolog in *S. pombe* regulates transcription by phosphorylating mediator subunits (Drogat et al., 2012). These phosphorylations determined the assembly of the Mediator complex, an essential PIC component. This function of CDK11 appears to be unique to budding yeast, since it is either absent or yet to be identified for CDK11 from other eukaryotes. Nonetheless, all characterized CDK11 orthologs are implicated in the regulation of gene expression but during evolution they seem to have acquired specialized roles in gene expression in different organisms.

CRK9 is an unusual CDK since unlike most other CDKs that require only cyclin as a functional partner, CRK9's specific function seems to be dependent on both CYC12 and CRK9AP and it consistently associates with both proteins during purifications. To the best of our knowledge, CDK7 is the only other CDK that forms a *bona fide* tripartite complex called the CAK complex, which is crucial for both gene expression and the cell cycle (Fisher, 2005). Could

CRK9/CYC12/CRK9AP trimeric complex represent a divergent CAK complex? Similar to CAK's function in phosphorylation of RPB1's CTD, we show that CRK9 as well as CYC12 and CRK9AP are essential for RPB1 phosphorylations in *T. brucei.* However, while CAK mediated CTD phosphorylations are important for RNA pol II transcription, we did not observe a specific defect in RNA pol II transcription in CRK9-depleted cells. Moreover, CRK9 appears not to interact with TFIIH because our long list of CRK9 co-purified proteins did not reveal a single TFIIH subunit (Table S1), CRK9 was not discovered in eluates of several independent TFIIH tandem affinity purifications (Lecordier et al., 2007; Lee et al., 2010; Lee et al., 2009), and CRK9 was not found to occupy the *SLRNA* promoter *in vivo*, the only known *T. brucei* promoter that assembles a RNA pol II PIC (Badjatia et al., 2013a). Another function of CAK is to phosphorylate and activate the CDKs involved in cell cycle control (Fisher, 2005). If the trimeric CRK9 complex is important for activation of cell cycle-relevant CDKs of trypanosomes is yet to be tested. Interestingly, CRK9 does seem to function in the cell cycle, and was shown to be important for proper basal body segregation during cytokinesis (Gourguechon and Wang, 2009). However, further work is required to understand the role of CRK9 in cell cycle regulation in trypanosomes.

What is the functional role of CRK9AP in the trimeric CRK9 complex? In the CAK complex, the presence of MAT1 significantly stimulates the kinase activity of CDK7 towards CTD. MAT1 interacts directly with both Cyclin H and CDK7 (Busso et al., 2000), and functions as an assembly factor (Fisher et al., 1995; Tassan et al., 1995) Interestingly, in MAT1 deficient cells, the levels of CDK7 and cyclin H are significantly reduced (Korsisaari et al., 2002; Rossi et al., 2001; Sano et al., 2007). In this context, CRK9AP appears to function in a similar manner to MAT1, since presence of CRK9AP is crucial for the stability of both CRK9 and CYC12. However, while a N-terminal RING finger motif is a conserved feature of MAT1 proteins from different organisms (Busso et al., 2000), CRK9AP lacks any obvious functional motif and it is conserved only amongst kinetoplastids. Another group of small proteins that are implicated in

assembly of CDK-cyclin complexes is the Cip/Kip family of CDK inhibitor (CKI), which was shown to important for assembly and activation of CDK4/6 - cyclin D complexes (Cheng et al., 1999; LaBaer et al., 1997). Future work will determine if CRK9AP is a divergent ortholog of MAT1 or Cip/Kip proteins or it is a unique kinetoplastid-specific protein dedicated for the assembly/stability of CRK9 complex.

In summary, CRK9 assembles in a unique complex together with CYC12 and CRK9AP and regulates *trans* splicing and possibly cell cycle in trypanosomes. We also show that CRK9 is crucial for trypanosomes to establish infection in a mouse model. Since *trans* splicing is a parasite-specific process, and kinases including CDKs are considered suitable drug targets, CRK9 is an attractive drug target against trypanosomatid parasites. It is interesting to note that both CRK9 and CYC12 have large kinetoplastid specific insertions in their kinase and cyclin domain respectively. These insertions could be a result of the complex's specialized function in *trans* splicing and might mediate association with trypanosome specific splicing proteins. Nonetheless, the unique features of CRK9 complex may be exploited to selectively inhibit this CDK in a parasite-specific manner. In the future, identification of a substrate(s) of CRK9 would help to delineate the mechanism by which CRK9 affects gene expression. Potential substrate candidates, including TSR1, are listed in Table 1.

## 5. Methods

### 5.1 DNAs

For gene silencing of *CYC12* and *CRK9AP*, their coding regions from position 21 to position 520 and from position 1 to position 414, were integrated in stem-loop arrangements into the pT7-stl vector (Brandenburg et al., 2007) to reveal plasmids T7-CYC12-stl and T7-CRK9AP-stl, respectively. For HA-tagging of CYC12, 1140 bp of the 3′-terminal CYC12 coding sequence was inserted into the pC-HA-BLA vector (Lee et al., 2007), utilizing the vector's ApaI and NotI

restriction sites. The resulting vector, named pCYC12-HA-BLA, was linearized with BamHI before transfection. DNA oligonucleotides that were used in semi-quantitative and quantitative reverse transcription (RT)-PCR and in primer extension assays are specified in (Ambrosio et al., 2015; Badjatia et al., 2013a).

**5.2 Cells**

BF and PF *Trypanosoma brucei* cell culture was maintained, transfected, and cloned as described before (Lee et al., 2007; Park et al., 2011). Clonal cell lines for conditional *CYC12* and CRK9AP silencing experiments were generated by transfection of linearized and *RRNA* locus-targeted T7-CYC12-stl and T7-CRK9AP-stl constructs, respectively. For PF and BF lines, the constructs were transfected into 29-13 and smBF cells, respectively (Wirtz et al., 1999). dsRNA synthesis for gene knockdowns was induced with 2 μg/ml of doxycycline during RNAi experiments. Cells were counted and diluted daily to 2 × $10^6$ cells/ml for PF culture and to 2 × $10^5$ cells/ml for BF culture. CYC12 was fused C-terminally with an HA tag in *CRK9AP* PF and BF knockdown cell lines by transfection of pCYC12-HA-BLA.

**5.3 RNA analysis**

Semi-quantitative and quantitative RT-PCR was performed to determine relative amounts of various RNAs during gene silencing experiments. Total RNA was prepared from 8 x $10^7$ PF or 1 x $10^8$ BF cells using the TRIzol reagent (Invitrogen) or the hot-phenol method (Nguyen et al., 2007) respectively. Reverse transcription reactions were carried out with SuperScript II reverse transcriptase (Invitrogen) according to the manufacturer's specifications. Oligo-dT and random hexanucleotides (Roche) were used in reverse transcription for the analysis of mature mRNA and unspliced pre-mRNA, respectively. For each semi-quantitative PCR, the number of cycles for the linear amplification range was determined empirically. For RNA quantifications, cDNA preparations were analyzed by qPCR assays using the SsoFast EvaGreen Supermix (BioRad) on a CFX96 cycler (BioRad) according to the manufacturer's

recommendations. For each amplification, triplicate qPCR samples were analyzed. Oligonucleotide pairs that were used in qPCR reactions were evaluated for specificity by both agarose gel electrophoresis and melting curve analysis. Standard curves for oligonucleotide pairs were obtained from serial dilutions of non-induced cDNA samples and ranged in their coefficient of determination ($R^2$) value from 0.98 to 1.0. Samples were standardized with the 18S rDNA result from random hexamer-derived cDNA. The methylation status of SL RNA cap4 was analyzed by a modified primer extension as described previously (Badjatia et al., 2013a).

**5.4 Protein analysis**

The polyclonal anti-CRK9 and anti-CRK9AP immune sera were generated by immunization of rats (Schimanski et al., 2006) with N-terminal glutathione *S*-transferase fusion proteins, either comprising amino acids 100 to 300 of CRK9 or the entire coding region for CRK9AP that were expressed in *Escherichia coli* strain BL21Star(DE3) and purified by glutathione affinity chromatography (GE Healthcare). The CRK9 sequence was chosen due to its hydrophilic nature as determined by a hydrophilicity blot using the Kyte & Doolittle algorithm on the ProtScale server at http://web.expasy.org/protscale/.

CRK9 and CRK9AP were detected on immunoblots by 1:1000 dilution of the respective immune sera followed by a 1:5000 dilution of a monoclonal, peroxidase-labeled anti-rat IgG secondary antibody (Vector Laboratories). HA-tagged protein was detected with a commercial monoclonal rat anti-HA antibody (Roche), RPA1 with a polyclonal rabbit immune serum and CITFA6 with a polyclonal rat immune serum as described previously (Nguyen et al., 2012; Schimanski et al., 2003). Blots were developed with BM chemiluminescence blotting substrate (Roche) according to the manufacturer's protocol. PTP-tagged proteins were detected either with the peroxidase anti-peroxidase reagent (Sigma) or the monoclonal anti-ProtC antibody HPC4 (Roche).

Extract preparation and tandem affinity purification of PTP-tagged CRK9 was carried out according to the standard PTP purification protocol (Schimanski et al., 2005a). Purified proteins were separated on SDS–10 to 20% polyacrylamide gradient gels (BioRad) and stained either with Sypro ruby (BioRad) or with Coomassie blue (Gelcode Coomassie stain; Thermo Fisher Scientific). For the sedimentation analysis of the CRK9 complex, the final eluate of the CRK9 PTP purification was concentrated, dialyzed against E-80 buffer and sedimented in a linear 10-40% sucrose gradient as described before (Ambrosio et al., 2015). Briefly, the gradient was fractionated from top to bottom in twenty aliquots of 200µl each. Proteins from 150µl of each fraction were collected by a hydrophobic resin (StrataClean, Stratagene), and resuspended in SDS loading buffer for electrophoresis. The remaining 50µl aliquots were dialyzed against E-20 buffer (20 mM HEPES-KOH pH 7.7, 20 mM potassium glutamate, 20 mM potassium chloride, 3 mM magnesium chloride, 0.2 mM EDTA, 0.5 mM EGTA and 4 mM DTT) and used for *in vitro* kinase assays. Proteins that co-purified with CRK9-P were analyzed by LC/MS/MSfrom the gel lane of the final eluate by the Keck facility of Yale University as described previously (Ambrosio et al., 2015). Individual protein bands obtained after sucrose gradient sedimentation were analyzed equivalently.

**5.5 Phylogenetic analysis**

Amino acid sequences spanning the CCL1 domain (COG5333) of various human cyclins and of CYC12 from *T. brucei*, *T. cruzi*, *Leishmania major* and *Bodo saltans* were aligned using the multiple sequence alignment tool ClustalW at the EMBL European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/msa/muscle/). The alignment was uploaded onto the graphical user interface Seaview (Gouy et al., 2010) (http://pbil.univ-lyon1.fr/software/seaview.html) to perform the phylogenetic analysis. A maximum likelihood tree was generated employing the LG empirical matrix (Le and Gascuel, 2008) with optimized

invariable sites, substitution rate categories of 4, estimated gamma distribution and model equilibrium frequencies. Bootstrapping was performed with 1000 replicates.

**5.6 Mouse infections**

6-10 weeks old female BALB/c mice (The Jackson Laboratory, Bar Harbor, ME), were injected intraperitoneally with $2 \times 10^6$ cells in 0.1 ml of cold phosphate buffer saline (PBS) that was supplemented with 1% glucose. To ensure cell viability, the parasites and syringes were kept on ice prior to injections. To induce RNAi, mice were given 1 mg/ml doxycycline in their drinking water 2 days prior to injection and for up to 2 weeks post-injection. Drinking water was replaced daily with fresh doxycycline solution. Mice were monitored daily and sacrificed upon signs of distress Blood samples, randomly taken from sick mice, were investigated microscopically, revealing, in each case, parasitemia ranging from 0.8 to $1 \times 10^9$ trypanosomes per ml of blood.

# Chapter V

# The spliceosomal PRP19 complex of trypanosomes

## 1. Abstract

In trypanosomes, mRNAs are processed by spliced leader (SL) *trans* splicing, in which a capped SL, derived from SL RNA, is spliced onto the 5′ end of each mRNA. This process is mediated by the spliceosome, a large and dynamic RNA-protein machinery consisting of small nuclear ribonucleoproteins (snRNPs) and non-snRNP proteins. Due to early evolutionary divergence, the amino acid sequences of trypanosome splicing factors exhibit limited similarity to those of their eukaryotic orthologs making their bioinformatic identification challenging. Most of the ~60 protein components that have been characterized thus far are snRNP proteins because, in contrast to individual snRNPs, purification of intact spliceosomes has not been achieved yet. Here, we characterize the non-snRNP PRP19 complex of trypanosomes. We identified a complex that contained the core subunits PRP19, CDC5, PRL1, and SPF27, as well as PRP17, SKIP and PPIL1. Three of these proteins were newly annotated. The PRP19 complex was associated primarily with the activated spliceosome and, accordingly, *SPF27* silencing blocked the first splicing step. Interestingly, *SPF27* silencing caused an accumulation of SL RNA with a hypomethylated cap that closely resembled the defect observed previously upon depletion of the cyclin-dependent kinase CRK9, indicating that both proteins may function in spliceosome activation.

## 2. Introduction

Nuclear pre-mRNA splicing, in which introns are removed from exonic sequences, is an essential step for the expression of many eukaryotic genes. The splicing process, which occurs

by two consecutive transesterifications, is carried out by the spliceosome, a large and highly dynamic complex of the U1, U2, U4/U6 and U5 small nuclear ribonucleoproteins (snRNPs) and non-snRNP proteins (Smith et al., 2008; Wahl et al., 2009). After recognition of the splice sites and before the splicing process, a large spliceosomal complex is formed on the pre-mRNA that is known as the B complex, which contains all five snRNPs and many non-snRNP proteins. Subsequently, substantial structural and compositional alterations, including the exit of the U1 and U4 snRNAs and the exchange of protein components, leads to spliceosome activation. The first transesterification generates spliceosomal complex C which, after additional rearrangements, carries out the second splicing step. While it has recently been established that RNA catalyzes the splicing reaction (Fica et al., 2013), there are more than 200 spliceosome-associated proteins known in the human system several of which are indispensable for the splicing process (Hegele et al., 2012; Wahl et al., 2009). Among those is PRP19, a highly conserved splicing factor that is essential for the activation of the spliceosome (Chan et al., 2003; Makarova et al., 2004; Ohi and Gould, 2002). PRP19's role in RNA splicing is through a heteromeric protein complex known as the *Nineteen Complex* (NTC) in the budding yeast *Saccharomyces cerevisiae* (Tarn et al., 1994) and the PRP19/CDC5L complex in humans (Grote et al., 2010; Makarova et al., 2004). Although the protein composition of the human and yeast complexes differ, they share the subunits PRP19, CDC5L/Cef1p (human/yeast nomenclature), SPF27/Snt309, and PRL1/Prp46. In the human system, these subunits were shown to form a salt stable core complex (Grote et al., 2010). Additionally, there are 3-4 non-homologous proteins in each system that are considered *bona fide* subunits of these complexes (Chanarat and Strasser, 2013) which, from here on, will be generally referred to as "PRP19 complexes". Moreover, several other spliceosomal proteins were found to consistently co-purify with the PRP19 complex, although their association appears to be less stable than that of the *bona fide* subunits (Makarov et al., 2002; Ohi et al., 2002). In the human system, these proteins

have been termed “PRP19-related” proteins, and they include factors such as SKIP, SYF1, CRN, ISY1, etc. (Wahl et al., 2009).

Pre-mRNA splicing has been a focus in the protozoan parasites *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania* spp., which cause devastating human diseases, because maturation of each and every mRNA in these organisms requires spliced leader (SL) *trans* splicing (Gunzl, 2010; Michaeli, 2011; Preußer et al., 2012). Protein coding genes in these trypanosomatids are arranged in long tandem arrays that are polycistronically transcribed. Individual mRNAs are then processed from precursor RNA by SL *trans* splicing and polyadenylation. In *trans* splicing, the SL is derived from the 5′ end of the small nuclear SL RNA and spliced onto the 5′ end of each mRNA. The SL carries a so-called cap4 structure which comprises a standard $m^7G$ cap nucleotide and methylations of the first four SL nucleotides (Bangs et al., 1992). Cap4 formation appears to be a pre-requisite for *trans* splicing (McNally and Agabian, 1992; Ullu and Tschudi, 1991) although individual methylations are dispensable for the process (Arhin et al., 2006a; Arhin et al., 2006b; Zamudio et al., 2007; Zamudio et al., 2006). SL *trans* splicing is achieved by the same two transesterification reactions as in “*cis* splicing” (Murphy et al., 1986; Sutton and Boothroyd, 1986). Furthermore, dependence of *trans* splicing on U snRNAs (Tschudi and Ullu, 1990) and on orthologs of various spliceosomal factors (Liang et al., 2006; Luz Ambrósio et al., 2009; Tkacz et al., 2008; Tkacz et al., 2010) demonstrated that *trans* splicing in trypanosomes is carried out by a spliceosomal complex that is similar to its human and yeast counterparts. However, characterization of the trypanosomal spliceosome has been difficult because, so far, a larger spliceosomal complex could not be quantitatively purified from trypanosomatids. Furthermore, trypanosome protein sequences are highly divergent from their human and yeast counterparts, making bioinformatic annotation of proteins and their genes challenging in some cases. Nevertheless, tandem affinity purification of the *T. brucei* snRNP proteins SmD1 (Luz Ambrósio et al., 2009), SmB (Palfi et al., 2009), and

the *Leishmania tarentolae* snRNP proteins SmD3, LSm3 and U1A (Tkacz et al., 2010) led to the identification of several new splicing factors increasing the trypanosomatid repertoire of known spliceosomal proteins to ~60 (Gunzl, 2010; Preußer et al., 2012). Most of these proteins are snRNP proteins and relatively few non-snRNP proteins have been identified so far, suggesting that mainly individual snRNPs were purified and that many non-snRNP proteins remained undetected in these purifications (Gunzl, 2010). PRP19 is such a spliceosomal non-snRNP protein that was identified by sequence homology and shown to be essential for cell viability, *trans* splicing and removal of the *PAP* (poly A polymerase) intron in *T. brucei* (Tkacz et al., 2010). In addition, *PRP19* silencing led to hypomethylation of the SL RNA cap. Little is known, however, about PRP19 complex subunits. While CDC5 (Gunzl, 2010) and PRL1 (Tkacz et al., 2010) co-purified with splicing complexes and were then identified bioinformatically, an SPF27 ortholog has been elusive. Furthermore, only a few homologs of PRP19-related proteins have been identified so far (Gunzl, 2010). (Please note that since *T. brucei* is a parasite of humans, when possible, we prefer to use the nomenclature of the human system for trypanosome splicing factors).

Here we have tandem affinity-purified *T. brucei* PRP19 and mass spectrometrically identified 47 co-purifying proteins. Among those were 35 spliceosomal orthologs, which included the three newly annotated proteins SPF27, SKIP, and PPIL1, as well as six proteins which appear to be conserved only among trypanosomatids. A sedimentation analysis revealed a stable PRP19 complex of seven subunits that, although containing the four core subunits, deviated in protein composition from its human and yeast counterparts. As expected, PRP19 complex subunits were primarily associated with the activated spliceosome. Finally, silencing of *SPF27* affected SL RNA cap methylation similarly to the previous *PRP19* knockdown. However, as our data indicate, cap4 methylation defects may be a general phenomenon of blocking spliceosome activation in trypanosomes.

## 3. Materials and Methods

### 3.1 DNAs

pPRP19-PTP-NEO was generated by inserting 627 bp of the C-terminal *PRP19* coding region (position 901 to position 1527 relative to the translation initiation codon) into the pC-PTP-NEO tagging construct (Schimanski et al., 2005a) using the ApaI and NotI restriction sites. pCDC5-PTP-NEO, pSPF27-PTP-NEO, and pORC1-PTP-NEO were obtained analogously by cloning 519 bp (position 1663 to position 2181), 695 bp position -53 to position 642) and 326 bp (position 983 to position 1308) of the *CDC5*, *SPF27* and *ORC1* genes into pC-PTP-NEO, respectively. For pSPF27-HA-BLA, the 695 bp DNA fragment was cloned into pC-HA-BLA (Nguyen et al., 2007), again using the ApaI and NotI restriction sites. The conditional knockdown constructs pT7-SPF27-stl and pT7-ORC1-stl were obtained by cloning, respectively, 571 bp (position 7 to position 577) and 526 bp (position 7 to position 532) of the *SPF27* and *ORC1* coding regions in a stem-loop arrangement into the vector pT7-stl (Brandenburg et al., 2007). The constructs for conditional knockdown of *LSm2* and *CRK9* have been described previously (Badjatia et al., 2013a; Luz Ambrósio et al., 2009).

### 3.2 Cells

Cell culturing and stable transfection of procyclic *T. brucei brucei* strain 427 and of cell line 29-13 (Wirtz et al., 1999) was carried out as described previously (Laufer et al., 1999). After each transfection, cell lines were cloned by limiting dilution. Cell line TbP19ee was generated by inserting BmgBI-linearized pPRP19-PTP-NEO into one *PRP19* allele and deleting the remaining wild-type allele with a PCR product in which 100 bp of the *PRP19* 5′ and 3′ gene flanks were fused to the hygromycin phosphotransferase (*HYG*) coding region. Cell lines that express CDC5-PTP, SPF27-PTP or ORC1-PTP were generated by transfecting SnaBI-linearized pCDC5-PTP-NEO, BsmI-linearized pSPF27-PTP-NEO, and XhoI-linearized pORC1-PTP-NEO,

respectively. Cells that express both PRP19-PTP and SPF27-HA were obtained by two consecutive transfections, first with linearized pPRP19-PTP-NEO and then with BsmI-linearized pSPF27-HA-BLA. Cell lines for conditional gene knockdowns were generated by targeting the T7-stl constructs to the *RRNA* spacer region. For each transfection correct integration of DNA was verified by PCR with at least one oligonucleotide hybridizing outside the transfected DNA sequence. After selection, 427 cells with antibiotic resistance markers were grown in 40 µg/ml hygromycin, 40 µg/ml G418, and/or 10 µg/ml blasticidin whereas 29-13 cell lines were kept in 50 µg/ml hygromycin, 15 µg/ml G418, and 4 µg/ml phleomycin. Conditional gene silencing experiments were induced by adding doxycycline to a final concentration of 1 µg/ml. Cells were counted and diluted to a density of 2 x $10^6$ cells/ml daily.

**3.3 Protein analysis**

PTP-tagged proteins on immunoblots were detected either with the peroxidase anti-peroxidase reagent (Kartalou and Essigmann) or a monoclonal anti-protein C epitope antibody (Roche). Extract preparation and tandem affinity purification of PTP-tagged proteins were carried out according to the standard protocol (Schimanski et al., 2005a) Purified proteins were separated on SDS–10 to 20% polyacrylamide gradient gels (BioRad) and stained either with sypro ruby (BioRad) or with Coomassie blue (Gelcode Coomassie stain; Thermo Fisher Scientific). For the sedimentation analysis of PRP19 and CDC5 complexes, the final eluate of a standard tandem affinity purification (~1.8 ml) was vacuum concentrated to ~100 µl, dialyzed overnight against E-80 buffer (150 mM sucrose, 20 mM HEPES-KOH pH 7.7, 20 mM potassium glutamate, 80 mM potassium chloride, 3 mM magnesium chloride, 0.2 mM EDTA, 0.5 mM EGTA), loaded onto a 4 ml linear 10-40% sucrose gradient in E-80 buffer, and ultracentrifuged at 41,000 rpm in a Beckman Sw55Ti rotor for 19 h at 4 °C. Subsequently, the gradient was fractionated from top to bottom in 20 aliquots. Proteins in each aliquot were collected by a hydrophobic resin (StrataClean, Stratagene), and resuspended in SDS loading buffer for

electrophoresis. Proteins that co-purified with PRP19-P were analyzed from a gel lane of the final eluate by the Proteomic Core facility of the Rockefeller University. The gel lane was cut into multiple pieces and peptides, derived from trypsin-digested proteins, were separated by liquid chromatography and analyzed by tandem mass spectrometry (LC/MS/MS) using an Ultimate 3000 HPLC system (Dionex) and a nanospray LTQ Orbitrap XL (Thermo Scientific) mass spectrometer. Individual protein bands obtained after sucrose gradient sedimentation were analyzed equivalently by the Keck facility (Yale University). Proteins were identified using Mascot and NCBI non-redundant protein sequence database of eukaryotes, with carbamidomethyl (C) as static and oxidation (M) as variable modifications. Peptide Mass tolerance was set to ±25 ppm and fragment mass tolerance to ±0.8 Da. Peptides were considered as identified when their score was ≥22 and expectation values were ≤0.05.

**3.4 RNA analysis**

To analyze gene silencing effects on RNA abundances, total RNA was prepared from non-induced or doxycycline-induced cells by the hot phenol method as described previously (Nguyen et al., 2007). rRNA was separated by gel electrophoresis on reliant precast 1.25% SeaKem Gold agarose RNA gels (Lonza) and detected by ethidium bromide staining. For the analysis of mRNA and unspliced pre-mRNA, total RNA was first reverse transcribed with SuperScript II reverse transcriptase (Invitrogen) according to the manufacturer's protocol using, respectively, oligo(dT) and random hexamers as primers. Secondly, the cDNAs were subjected to semi-quantitative and quantitative (q)PCR analysis. Semi-quantitative PCR was performed using cycle numbers that were empirically determined to be within the linear amplification range for each oligonucleotide pair. The specificity of oligonucleotide pairs used in qPCR was first verified in analyzing their products by standard agarose gel electrophoresis. To further confirm the validity of the qPCR reactions, a melt curve analysis was included for each qPCR to ensure that only a single product was amplified. In addition, a linear regression analysis of a serial

dilution of input material was performed to make sure that the coefficient of determination ($r^2$) was within the 0.98 to 1.0 range.

Spliceosomal U snRNAs, SL RNA and, as a control, 7SL RNA were detected by primer extension assays using 10 µg of total RNA, 5′-$^{32}$P-end-labeled oligonucleotides and SuperScript II reverse transcriptase. Primers SL_PE and SL40, 7SL_PE, U1_PE, U2f, U4_PE, U5_PE, and U6_PE were used to detect SL RNA, 7SL RNA, U1, U2, U4, U5 and U6 snRNA, respectively. Primer extension products were separated on 8% polyacrylamide – 50% urea gels and visualized by autoradiography. The modified primer extension assay for characterizing the methylation status of the SL RNA cap and generation of the sequencing ladder was conducted with unmodified Moloney's murine leukemia virus (MMLV) reverse transcriptase (Invitrogen) and the Thermo Sequenase cycle sequencing kit (USB), as previously described (Badjatia et al., 2013a). For the analysis of PRP19 complex-associated U snRNAs, PTP-tagged proteins were precipitated by immunoglobulin G (Djikeng et al.) beads (GE Healthcare), and total RNA was prepared and analzyed by primer extension assays as previously published (Luz Ambrósio et al., 2009). DNA oligonucleotides that were used in primer extension assays and in competitive and [semi-]quantitative reverse transcription (RT-)PCR reactions are specified in Attachment 4, Supplementary Table S1.

### 3.5 Light microscopy

Immunofluorescence microscopy was performed as previously described (Luz Ambrósio et al., 2009). For the detection of PRP19-PTP, a rabbit polyclonal anti-ProtA immune serum (Kartalou and Essigmann) followed by an Alexa 594-conjugated anti-rabbit secondary antibody (Invitrogen) was used, while HA-tagged SPF27 was detected by a rat monoclonal anti-HA antibody (Roche) and an Alexa 488-conjugated anti-rat secondary antibody (Invitrogen). We excluded the possibility that anti-HA antibodies were non-specifically bound by ProtA by analyzing cells that expressed only PRP19-PTP with the antibody mix described above (data

not shown). Images were taken on a Zeiss Axiovert 200 microscope equipped with DAPI (4′,6′-diamidino-2-phenylindole), EGFP and Texas-Red filters using a 100x (1.3-numerical-aperture) oil immersion objective. Quantification of the co-localization was done using the Zeiss Axiovision 4.6.3.0 software program for co-localization.

## 4. Results

### 4.1 Tandem affinity purification of PRP19 and identification of co-purified proteins

For tandem affinity purification (Biswas et al.) of PRP19 (accession number Tb927.2.5240 at www.GeneDB.org or www.TriTrypDB.org) we generated the insect-stage, procyclic *T. brucei* cell line TbP19ee which expressed PRP19 with a C-terminal PTP tag (PRP19-PTP) and no untagged PRP19. The PTP tag is a composite tag consisting of a protein C epitope (ProtC) followed by a tobacco etch virus (TEV) protease cleavage site and tandem protein A domains (ProtA) (Schimanski et al., 2005a). The cell line was obtained by replacing one *PRP19* allele of wild-type procyclics with the coding region of the selectable marker hygromycin phosphotransferase (HYG) and by targeted integration of plasmid PRP19-PTP-NEO into the remaining *PRP19* allele, thereby fusing the PTP sequence to the *PRP19* gene (Figure 5.1A). Since TbP19ee cell cultures did not exhibit a growth defect (data not shown) and *PRP19* is an essential gene in *T. brucei* (Tkacz et al., 2010), we inferred that the tag did not disrupt PRP19 function. We then prepared extract from 5 x $10^{10}$ trypanosomes and performed PTP tandem affinity purification of PRP19-PTP. While the expected size of PRP19-PTP is 74 kDa, anti-ProtC immunoblotting of extract revealed a single band of ~80 kDa (Figure 5.1B, lane 1). PTP-based TAP involves IgG affinity chromatography through ProtA, release of the tagged protein by TEV protease cleavage, and recapture of the tagged protein by anti-ProtC immunoaffinity chromatography (Schimanski et al., 2005a). Immunoblot monitoring of the TAP procedure showed that both chromatography steps led to efficient depletion of the tagged

protein from extract and from TEV protease eluate (lanes 2 and 4) and that the final eluate contained a substantial amount of purified PRP19 (lane 5). Note, that removal of ProtA decreased the molecular mass of PRP19-P by ~15 kDa. The final eluate was then separated on a SDS-polyacrylamide gel and stained with Coomassie blue (Figure 5.1C). The profile of protein bands exhibited a massive protein enrichment in the 55-65 kDa range, which included PRP19-P, and three major protein bands with apparent sizes of ~85, 26 and 22 kDa. In addition, many bands of minor abundance and different size were detected which could be components of the spliceosome.

To identify co-purified proteins, we performed liquid chromatography-tandem mass spectrometry (LC/MS/MS) with the final TAP eluate that was not separated electrophoretically. By limiting the analysis to proteins that were identified by at least two unique peptides and excluding recurrent PTP purification contaminants, we identified 47 proteins (Table 5.1). Of those, 35 proteins were homologs of known spliceosomal proteins, two of the six proteins without annotation (*conserved hypothetical*) were previously co-purified with trypanosomatid splicing complexes (Luz Ambrósio et al., 2009; Palfi et al., 2009), and only six of the annotated proteins did not have a known splicing function. Moreover, the list contains several orthologs of human and yeast PRP19 complex subunits as well as PRP-related proteins (see below) confirming that the tagged PRP19 was functionally incorporated into RNA splicing complexes and indicating that the PRP19 purification was highly specific.

### 4.2 Identification of a stable T. brucei PRP19 complex of seven subunits

To determine the proteins that form a stable PRP19 complex, we sedimented PRP19-PTP-purified proteins through a linear sucrose gradient by ultracentrifugation. We took 20 fractions of the gradient from top to bottom, collected the proteins from each fraction with a hydrophobic resin and stained the proteins after electrophoretic separation in the gel with sypro ruby (Figure 5.1D). The major protein bands of the purification co-sedimented with peak signals

in fractions 14 and 15 which is between the ~230 kDa-large *T. brucei* SNAPc/TRF4/TFIIA transcription factor complex (Schimanski et al., 2005b) and the 444 kDa-large apoferritin marker. Furthermore, in fraction 20, major and minor bands were co-detected indicating that the purified material contained larger complexes that sedimented to the bottom of the gradient.

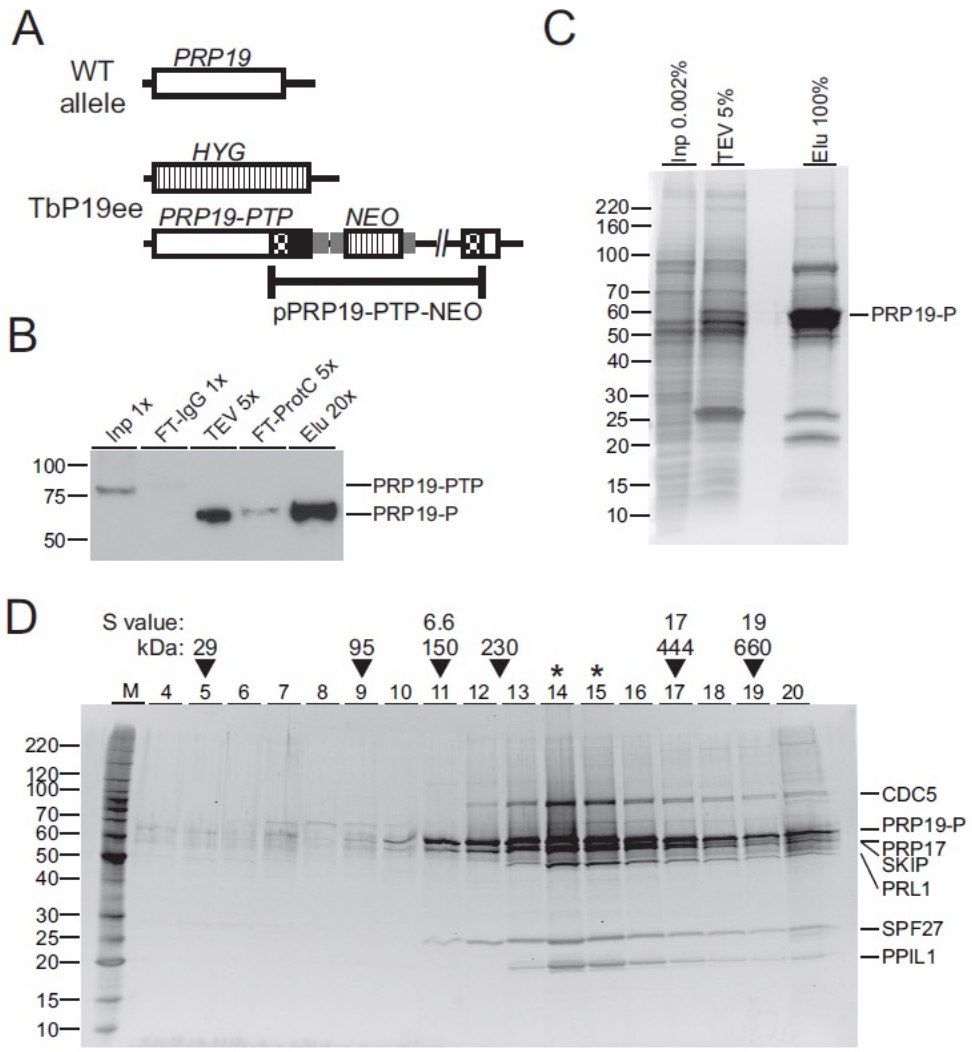



**Figure 5.1** ***Trypanosoma brucei*** **has a stable PRP19 complex of seven subunits A.**Schematic (not to scale) of a *PRP19* wild-type allele and the *PRP19* locus in procyclic TbP19ee cells that exclusively express PRP19-PTP and no untagged PRP19. PRP19, HYG/NEO coding regions and PTP tag are indicated by open, striped and black boxes, respectively. Gene flanks for RNA processing signals are drawn as smaller gray boxes. Checkered boxes indicate plasmid-derived portions of the PRP19 coding region. **B.** Immunoblot monitoring of the PRP19-PTP tandem affinity purification. PRP19-PTP and PRP19-P (after removal of the ProtA domains) were detected with the monoclonal anti-ProtC antibody HPC4 in crude extract (Inp), flowthrough of the IgG column (FT-IgG), TEV protease eluate (TEV), flowthrough of the anti-ProtC column (FT-ProtC) and the final eluate (Elu). x-Values indicate relative amounts loaded. **C.** Aliquots of extract, the TEV protease eluate and the final eluate were separated on a 10-20% SDS-polyacrylamide gradient gel and stained with Coomassie blue. Percentages indicate relative amounts analyzed. **D.** The eluate of a PRP19-PTP tandem affinity purification was sedimented through a 10-40% linear sucrose gradient by ultracentrifugation and 20 fractions were taken from top to bottom. Proteins from each fraction were separated on a 10-20% SDS-polyacrylamide gradient gel and stained with Sypro Ruby. Proteins specified on the right were identified by excision of individual protein bands and LC/MS/MS analysis. For comparison, sedimentations of *Taq* DNA polymerase (95 kDa), IgG (150 kDa, 6.6S), apoferritin (444 kDa, 17S) and thyroglobin (660 kDa, 19S) were analyzed in parallel gradients (arrowheads). In addition, sedimentation peaks from previous, comparable experiments of TEV protease (29 kDa) and the SNAPc/TRF4/TFIIA transcription factor complex (230 kDa) are indicated as well. Asterisks indicate the sedimentation peak.

## Table 5.1 Mass spectrometric identification of PRP19 co-purified proteins

| Annotation [1] | Accession # | $M_r$ (kDa) | Protein score | # Unique pept. | % Coverage | # Pept. |
|---|---|---|---|---|---|---|
| **PRP19** | Tb927.2.5240 | 54.2 | 65,296 | 32 | 75.1 | 39 |
| **CDC5** / Cef1 | Tb927.5.2060 | 80.1 | 31,713 | 57 | 82.0 | 64 |
| CDC40 / **PRP17** | Tb927.3.1930 | 52.8 | 9,959 | 33 | 83.7 | 37 |
| **SKIP** / Prp45 [2,3] | Tb927.9.5880 | 50.4 | 8,862 | 25 | 50.3 | 30 |
| **PRL1** / Prp46 | Tb927.10.10170 | 48.8 | 6,349 | 23 | 68.2 | 23 |
| **SPF27** / Snt309 [2,3] | Tb927.11.14150 | 24.5 | 5,775 | 16 | 74.8 | 23 |
| **PPIL1** [2] | Tb927.8.2090 | 21.7 | 5,615 | 11 | 67.0 | 13 |
| DHX38 / Prp16 | Tb927.10.7280 | 121.2 | 4,398 | 63 | 66.6 | 63 |
| U5-200K / Brr2 | Tb927.5.2290 | 249.3 | 3,936 | 67 | 37.2 | 67 |
| [U5-]PRP8 | Tb927.9.11110 | 276.8 | 3,811 | 79 | 36.0 | 79 |
| HSP73 (HSP70) | Tb927.11.11330 | 76.0 | 3,523 | 36 | 65.1 | 40 |
| CRN / Clf1, Syf3 | Tb927.10.9660 | 87.6 | 2,654 | 25 | 45.8 | 25 |
| CACTIN [3] | Tb927.11.11610 | 64.6 | 2,303 | 24 | 57.2 | 24 |
| U5-116K / Snu114 | Tb927.11.15430 | 105.4 | 2,165 | 33 | 43.8 | 33 |
| SYF1 | Tb927.5.1340 | 92.1 | 1,846 | 21 | 35.9 | 21 |
| KIAA1604 / Cwc22 | Tb927.11.10750 | 66.8 | 1,737 | 20 | 35.3 | 20 |
| Conserved hypoth. [3] | Tb927.8.5650 | 103.2 | 1,233 | 22 | 34.8 | 22 |
| Conserved hypoth. | Tb927.2.3400 | 36.6 | 852 | 10 | 32.9 | 10 |
| U5-CWC21 | Tb927.9.3480 | 20.4 | 693 | 10 | 55.2 | 10 |
| U2A′ (U2-40K) | Tb927.10.14360 | 36.5 | 662 | 15 | 65.3 | 15 |
| Conserved hypoth. [3] | Tb927.10.11230 | 34.7 | 567 | 11 | 42.7 | 11 |
| SmD1 | Tb927.7.3120 | 11.7 | 528 | 4 | 45.3 | 4 |
| LSm2 | Tb927.8.5180 | 13.3 | 507 | 5 | 44.8 | 5 |
| LSm8 | Tb927.3.1780 | 14.0 | 435 | 6 | 43.4 | 6 |
| VSG pseudogene | Tb927.5.4980 | 48.5 | 428 | 5 | 20.0 | 5 |
| aminopeptidase | Tb927.11.3570 | 98.0 | 408 | 9 | 15.5 | 9 |
| AD002 / **CWC15** | Tb927.10.11950 | 22.4 | 403 | 6 | 33.2 | 6 |
| U5-40K | Tb927.11.11150 | 35.0 | 398 | 5 | 18.7 | 5 |
| SmF | Tb927.9.10250 | 8.4 | 394 | 4 | 56.0 | 4 |
| PRP31 | Tb927.10.10700 | 39.6 | 371 | 4 | 13.2 | 4 |
| SSm2-1 (Sm15K) | Tb927.6.4340 | 12.8 | 342 | 3 | 33.3 | 3 |
| ISY1 | Tb927.8.1930 | 31.7 | 313 | 5 | 29.0 | 5 |
| SmD2 | Tb927.2.5850 | 12.5 | 282 | 4 | 35.1 | 4 |
| PABP2 (PABP1) | Tb927.9.10770 | 62.2 | 282 | 7 | 15.5 | 7 |
| Conserved hypoth. | Tb927.11.2960 | 12.0 | 280 | 5 | 40.7 | 5 |
| Actin A | Tb927.9.8850 | 41.9 | 266 | 3 | 11.4 | 3 |
| SSm2-2 (Sm16.5K) | Tb927.10.4950 | 14.7 | 255 | 4 | 31.3 | 4 |
| SmE | Tb927.6.2700 | 9.7 | 217 | 3 | 37.2 | 3 |
| SmB | Tb927.2.4540 | 12.3 | 169 | 5 | 40.4 | 5 |
| U2B″ | Tb927.3.3480 | 13.6 | 167 | 5 | 27.4 | 5 |
| Argonaute AGO1 | Tb927.10.10850 | 99.2 | 162 | 3 | 3.7 | 3 |
| ALBA3 | Tb927.4.2040 | 20.8 | 153 | 2 | 10.5 | 2 |
| Conserved hypoth. [3] | Tb927.4.3540 | 41.2 | 136 | 3 | 12.9 | 3 |
| Pumilo PUF2 | Tb927.10.12660 | 92.2 | 126 | 2 | 4.5 | 2 |
| LSm4 | Tb927.11.13960 | 14.2 | 122 | 2 | 15.0 | 2 |
| Conserved hypoth. [3] | Tb927.8.5840 | 21.5 | 118 | 2 | 10.7 | 2 |
| Cyclophilin (PPIase) | Tb927.9.11740 | 20.3 | 103 | 3 | 24.0 | 3 |

PRP19 co-purifying proteins, identified by LC/MS/MS, are ranked according to protein score. Listed proteins have a score of >100 and were identified by 2 or more unique peptides. Proteins that co-sedimented with PRP19 in linear sucrose gradients are highlighted in yellow, proteins denoted as PRP19 subunits or PRP19-related proteins in other systems in green, un-annotated proteins in blue, and [putative] spliceosome components in gray. Annotated proteins without a known role in RNA splicing were not highlighted. In addition, standard contaminants of trypanosome TAPs such as a/b tubulin, translation elongation factor, retrotransposon hot spot proteins, histones, low scoring ribosomal proteins, and chaperones were eliminated from this list, although it should be noted that prefoldin chaperones were particularly enriched in these purifications.

[1] Annotation is according to the human/yeast systems unless there is a trypanosome-specific name. Names in parentheses refer to alternative names published for *T. brucei*.

[2] Proteins that were annotated in this study.

[3] [putative] Spliceosomal proteins that co-purified in a trypanosome splicing complex for the first time.

To identify the proteins in the bands seen in fractions 14 and 15, gel slices were excised and their protein content analyzed by LC/MS/MS. We identified seven proteins which coincided with the seven proteins that obtained the highest protein score listed in Table 5.1. The largest protein was identified as trypanosome CDC5. The strong bands of 65 and 55 kD contained tagged PRP19-P and PRP17 (Palfi et al., 2009), respectively. The strong PRP19 signal is consistent with PRP19 forming a tetramer within a single PRP19 complex (Ohi et al., 2005). In contrast, PRP17 was not a component of previously characterized human and *S. cerevisiae* PRP19 complexes and, therefore, has not been considered a *bona fide* component of the PRP19 complex (Chanarat and Strasser, 2013; Wahl et al., 2009). On the other hand, systematic yeast-2-hybrid screens have identified PRP17 as a direct interactor of the PRP19 WD40 domain (Figure 5.2A) in human and yeast systems (Hegele et al., 2012; Ren et al., 2011). Moreover, in *Schizosaccharomyces pombe*, PRP17 was consistently detected in tandem affinity purifications of PRP19 complexes, and it was determined that there are four PRP17 molecules per complex (Ren et al., 2011). Our results corroborate these findings. In *T. brucei*, PRP17 co-purified and co-sedimented with PRP19 identifying it as a *bona fide* subunit of the trypanosome PRP19 complex. Moreover, the strong signals of the 65 and 55 kDa bands suggest that there is more than one PRP17 molecule present in the trypanosome complex. However, the strong 55 kDa signal appears to be not confined to PRP17 alone because we detected a second protein in this band. This protein is encoded in gene *Tb927.9.5880* which has been annotated as a "conserved, hypothetical protein" indicating that its sequence is conserved only among kinetoplastids. A standard BLAST search identified an internal SKIP/SNW domain strongly indicating that this protein is the ortholog of the spliceosomal Ski-binding protein (SKIP) and PRP45p of mammals and *S. cerevisiae*, respectively (Figs. 5.2A, B). Interestingly, in the human and yeast systems, SKIP/PRP45p is not considered to be a PRP19 complex subunit but a PRP19-related/associated protein (Fabrizio et al., 2009; Makarova et al., 2004).

The band of ~50 kDa has a lower strength signal and contained only a single protein, namely the trypanosome ortholog of human PRL1 and yeast PRP46p. It is considered a core component of the yeast and human PRP19 complexes (Fabrizio et al., 2009; Makarova et al., 2004). Trypanosomatid PRL1 had been identified previously, co-purifying with LSm3 in *Leishmania tarentolae* (Tkacz et al., 2010). Accordingly, *T. brucei* PRL1 shares clear sequence homology to human PRL1 ($E = e^{-80}$) including a C-terminal WD40 protein-protein interaction domain (Figure 5.2A).

The sixth protein, with an apparent size of 25 kDa, was found to be encoded by gene *Tb927.11.14150.* The corresponding amino acid sequence was conserved only among kinetoplastid organisms and, in contrast to the SKIP identification, did not reveal a conserved sequence motif. However, a key subunit of similar size of PRP19 complexes across systems is SPF27. To analyze whether this trypanosomatid protein could be a SPF27 ortholog, we performed a multiple sequence alignment of SPF27 orthologs from model organisms and of the kinetoplastid orthologs of Tb927.11.14150. As shown in Figure 5.2C, we found four short domains of sequence conservation between the two orthologous groups including a nearly invariant LPY motif near the N-terminus of the protein. Although this degree of sequence conservation is not strong, the findings that SPF27 is a *bona fide* PRP19 complex subunit in all systems and that Tb927.11.14150 was a major protein in the PRP19 purification co-regimenting with the PRP19 complex subunits argue that this protein is trypanosome SPF27.

Analysis of the smallest band of 21 kDa revealed a protein that is encoded by gene *Tb927.8.2090* and had been annotated as a cyclophilin-type peptidyl-prolyl *cis-trans* isomerase in the genome data base because it harbors the conserved domain of this enzyme class (Figure 5.2A). The protein was identified previously co-purifying with SmB and annotated as cyclophilin CYP2 (Palfi et al., 2009). However, a search for the closest homolog of this protein in the human genome returned the PRP19-related factor PPIL1. A multiple sequence alignment of the

trypanosomatid orthologs of this protein and PPIL1 sequences from different model organisms showed sequence conservation along the whole protein (Attachment 4, Fig. S1). Moreover, human PPIL1 binds to a 21 amino acid-long region in the N-terminal part of SKIP (Wang et al., 2010). This region is highly conserved in trypanosome SKIP sequences including two key residues (Figure 5.2D). Further support for Tb927.8.2090 being a PPIL1 ortholog comes from a comparison with yeast genome databases. While *S. pombe* harbors PPIL1 and full-length SKIP orthologs, *S. cerevisiae* does not have a PPIL1 ortholog (spliceosome database; (Cvitkovic and Jurica, 2013) and its SKIP protein lacks the N-terminal domain including the PPIL1 binding site (Wang et al., 2010). Accordingly, a blastp search of the *S. pombe* genome database with the Tb927.8.2090-encoded amino acid sequence identified the PPIL1 ortholog Cyp1 as the most similar sequence with a low Expect Value ($E = 2e^{-25}$) whereas the same search of the *S. cerevisiae* database returned the mitochondrial peptidylprolyl isomerase CPR3 as the most similar sequence with a considerably higher Expect Value ($E = 9e^{-14}$). We therefore concluded that gene *Tb927.8.2090* encodes *T. brucei* PPIL1.

Finally, to confirm the characterization of the PRP19 complex, we generated a cell line that expressed CDC5 with a C-terminal PTP tag. Tandem affinity-purification of CDC5-PTP and sedimentation analysis revealed a protein complex with similar banding pattern and sedimentation property as the PRP19-purified complex (Attachment 4, Fig. S2). Accordingly, mass spectrometry of the excised protein bands revealed the same seven proteins as in the PRP19-PTP-purified complex. Thus, we have identified a stable PRP19 complex in *T. brucei* consisting of seven subunits.

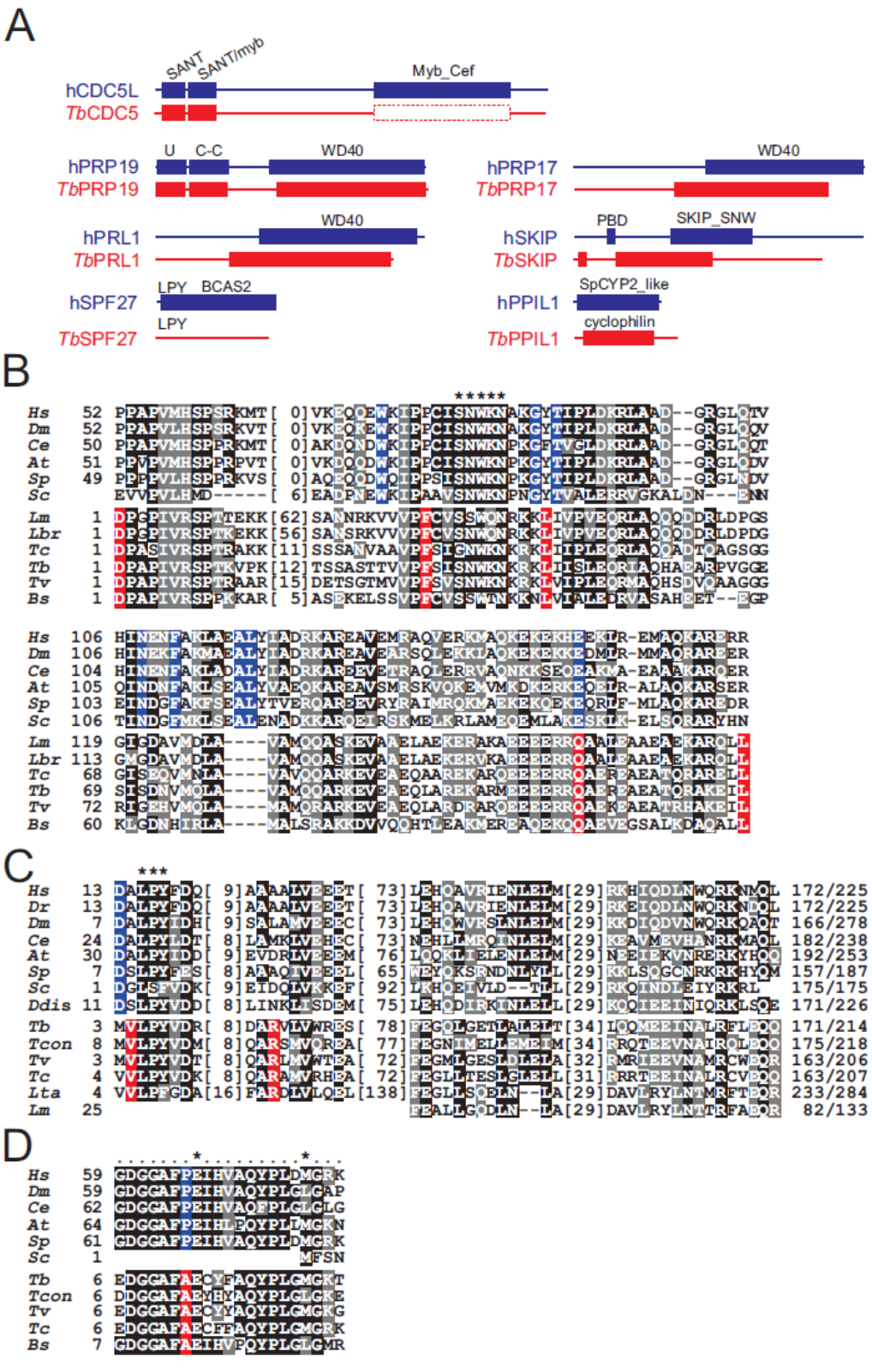

A
SANT
SANT/myb
Myb_Cef
hCDC5L
TbCDC5
U
C-C
WD40
hPRP19
TbPRP19
hPRP17
TbPRP17
hPRL1
TbPRL1
PBD
SKIP_SNW
hSKIP
TbSKIP
LPY
BCAS2
hSPF27
TbSPF27
SpCYP2_like
hPPIL1
cyclophilin
TbPPIL1
B
C
D

**Figure 5.2 Domain structure and sequence conservation in PRP19 complex subunits**

**A.** Domain structure drawn to scale of trypanosome PRP19 complex subunits and their human orthologues. The protein domains were identified by NCBI blastp searches of the Conserved Domain Database (Marchler-Bauer et al., 2011) (E-values are specified for *T. brucei* protein domains): CDC5: SANT domain (cd00167), E=1.03e$^{-08}$ ; SANT/myb-like domain (cd11659), E=8.07e$^{-10}$; the very limited sequence similarity between the human Myb_Cef domain (pfam11831) and the corresponding region of trypanosomatid CDC5 sequences is indicated by the dashed box. PRP19: U-Box (smart00504), E=1.06e$^{-14}$; PRP19 domain (coiled-coiled [C-C] protein interaction domain, pfam08606), E=4.87e$^{-20}$; WD40 (cd00200), E=5.77e$^{-18}$. PRL1: WD40 (cd00200), E=1.4e$^{-75}$. SPF27: while the LPY motif is present in *T. brucei* SPF27, the human BCAS2 domain (pfam05700) was not recognized. PRP17: WD40 (cd00200), E=2e$^{-37}$. SKIP: PPIL1-binding domain (PBD); SKIP_SNW domain (pfam02731), E=7.51e$^{-08}$. PPIL1: cyclophilin domain [cd00317], E= 2.18$^{e-27}$ (note that while the trypanosome sequence returned the general cyclophilin domain, the more specific SpCYP2_like cyclophilin domain (cd01922) was identified in the human ortholog.

**B.** Multiple sequence alignment of the SKIP/SNW domain of *Homo sapiens* (*Hs*, accession number NP_036377), *Drosophila melanogaster* (*Dm*, AGB95213), *Caenorhabditis elegans* (*Ce*, CAA98552), *Arabidopsis thaliana* (*At*, AEE35947), *Schizosaccharomyces pombe* (*Sp*, CAB41231), *Saccharomyces cerevisiae* (*Sc*, P28004), and the kinetoplastids *Leishmania major* (*Lm*, LmjF.15.1030), *Leishmania braziliensis* (*Lbr*, LbrM.15.1070), *Trypanosoma cruzi* (*Tc*, TcCLB.509445.20), *T. brucei* (*Tb*, Tb927.9.5880), *Trypanosoma vivax* (*Tv*, TvY486_0902160), and *Bodo saltans* (*Bs*, ACI16065). Since the N-terminal ~50 amino acids of the domain are only weakly conserved among kinetoplastids, they were omitted from the alignment. Numbers indicate the position within the SKIP/SNW domain. Positions with more than 50% similarity or identity are shaded in gray and black, respectively. Identical positions in model organisms without any conservation in kinetoplastids are shaded blue and identical position in kinetoplastids without conservation in model organisms are shaded in red. A hyphen indicates lack of an amino acid at this position. Numbers in parentheses specify lengths of non-conserved insertions. The highly conserved SNWKN motif is indicated by asterisks.

**C.** Corresponding alignment of four short conserved domains in SPF27 orthologues. Asterisks mark the highly conserved LPY motif which, among *Leishmania* species, is only present in *L. tarentolae* as LPF. Accession numbers: *Hs*, NP_005863; *Dr* (*Danio rerio*), NP_001007775; *Dm*, NP_651596; *Ce*, NP_498360; *At*, NP_566599; *Sp*, CAB57933; *Sc*, EEU07188; *Ddis* (*Dictyostelium discoideum*), XP_640072; *Tb*, Tb927.11.14150; *Tcon* (*Trypanosoma congolense*), TcIL3000.11.14450; *Tv*, TvY486_1114990; *Tc*, TcCLB.511727.110; *Lta* (*Leishmania tarentolae*), LtaP32.0970; *Lm*, LmjF.32.0900.

**D.** Multiple sequence alignment of the 21 amino acid-long PPIL1 binding domain in SKIP orthologues. Numbers indicate positions relative to the starting methionine. The two key residues for PPIL1 binding are marked by asterisks.

### 4.3 SPF27 silencing affects *trans* and *cis* splicing of pre-mRNA

Next, we wanted to verify for one co-purified subunit whether its depletion in cells had effects on cell viability, RNA splicing and SL cap4 formation which corresponded with the previous *PRP19* silencing experiment (Tkacz et al., 2010). We chose SPF27 since it was the least conserved subunit of the PRP19 complex. By stable integration of a tetracycline repressor-regulated *SPF27* stem-loop construct into procyclic 29-13 cells, which constitutively express the tetracycline repressor and T7 RNA polymerase (Wirtz et al., 1999), we generated a cell line for conditional *SPF27* silencing. Induction of the *SPF27* knockdown with the more stable tetracycline derivative doxycycline, which was added to the medium, stopped culture growth between one and two days of induction (Figure 5.3A). Subsequently, the cell density remained nearly constant until day 5 of the experiment strongly indicating that *SPF27* is an essential gene (Figure 5.3A). Interestingly, while the previous *PRP19* knockdown generated a similar growth defect (Tkacz et al., 2010), depletion of the essential snRNP proteins *LSm2* and *U5-Cwc21* led to rapidly declining cell numbers in our hands (Luz Ambrósio et al., 2009), suggesting that disabling the PRP19 complex affected trypanosomes differently than depletion of snRNP proteins. This difference appears not to be a consequence of inefficient gene knockdown because a semi-quantitative RT-PCR analysis showed that *SPF27* mRNA was rapidly and strongly reduced in the presence of doxycycline (Figure 5.3B). Analyzing total RNA prepared from non-induced cells and from cells in which *SPF27* was silenced for 1 to 3 days clearly revealed pre-mRNA splicing defects. While the amounts of the control RNAs, rRNA and the non-spliceosomal 7SL RNA, which is the RNA component of the signal recognition particle (Michaeli et al., 1992), did not change in these RNA preparations, semi-quantitative RT-PCR analysis clearly indicated that the abundance of mature α tubulin mRNA decreased upon *SPF27* silencing whereas, in random hexamer-derived cDNA of the same RNA preparations, the signal of unspliced α tubulin pre-mRNA, lacking the SL, clearly increased (Figure 5.3B). Similarly, a

competitive PCR of cDNA across the *PAP* intron showed a relative increase and decrease of unspliced and mature *PAP* RNA, respectively (Figure 5.3B). Together, these results revealed a splicing defect both in SL *trans* splicing (α tubulin mRNA) and in *cis* splicing (*PAP* mRNA).

To confirm that these changes in RNA abundances are specific to *SPF27* silencing and not due to pathways of death, we co-analyzed the knockdown of a non-spliceosomal gene that affected trypanosome culture growth similar to *SPF27* silencing. We chose the *ORC1* gene encoding subunit 1 of the origin recognition complex (Godoy et al., 2009) because, in our hands, the *ORC1* knockdown arrested culture growth after the second day of induction (Attachment 4, Fig. S3). The RNA analysis showed that, while *ORC1* mRNA abundance was very strongly reduced after 1 day of induction, the level of α tubulin mRNA was unaffected even after 3 days of *ORC1* silencing (Figure 5.3C). Together, these results confirmed that SPF27 is a spliceosomal protein.

Since the PRP19 complex is known to be important for spliceosome activation, we expected that *SPF27* silencing would affect the first step of splicing. In *trans* splicing such a defect should lead to an increase of the SL RNA splicing substrate and a decline of the Y structure intermediate that is formed after the first transesterification (Luz Ambrósio et al., 2009; Tkacz et al., 2010). To test this we conducted a primer extension assay with three radiolabeled oligonucleotides that specifically hybridize to the intronic sequence of the SL RNA and, as controls, to the U2 snRNA and the 7SL RNA. In the Y structure, the 5′ end of the intron is covalently linked to the branch site which forms a barrier for reverse transcriptase, leading to an intron-specific primer extension product. Indeed, in *SPF27*-silenced cells, primer extension showed a strong increase of SL RNA abundance on day 2 and 3 of induction and a concomitant reduction of the Y structure signal in these total RNA preparations, a phenotype that was not observed in *ORC1*-silenced cells (Figure 5.3D, left panel).

To substantiate these findings, we carried out quantitative RT-PCR analyses with three independently derived RNA preparations for each time point of *SPF27* silencing. On days two and three after induction, the amount of SL RNA increased ~6 to ~9-fold over the un-induced level, respectively (Figure 5.3D, right panel). These numbers were in accordance with those of a previous study in which silencing the gene of CDC2-related kinase 9 (*CRK9*) also resulted in a block of *trans* splicing before the first step (Badjatia et al., 2013a). In agreement with such a block, we found a ~3-fold increase of α tubulin pre-mRNA abundance in *SPF27*-silenced cells which coincided with a ~50% reduction of the mature mRNA level on day three of induction (Figure 5.3E, left panel; note that our analysis of mature mRNA did not discriminate between a correct mRNA and a polyadenylated, unspliced mRNA). Comparable results were obtained when we analyzed a second mRNA encoding the RPB7 subunit of RNA polymerase II (RNAPII; Figure 5.3E, middle panel). Finally, while the *PAP* mRNA level strongly decreased upon *SPF27* silencing as expected, a *PAP* intron-dependent qPCR detected only a small, insignificant increase of *PAP* pre-mRNA in induced cells (Figure 5.3E, right panel). Possibly, unspliced *PAP* mRNA is highly unstable, preventing an accumulation of this RNA in splicing-impaired cells.

In summary, the increase of SL RNA and unprocessed pre-mRNA levels paralleled by the concomitant decrease of the Y structure intermediate and mature mRNA abundances demonstrated that SPF27 depletion caused a block of the first *trans* splicing step, further indicating a role of the PRP19 complex for spliceosome activation in trypanosomes and corroborating SPF27 as a *bona fide* subunit of this complex.

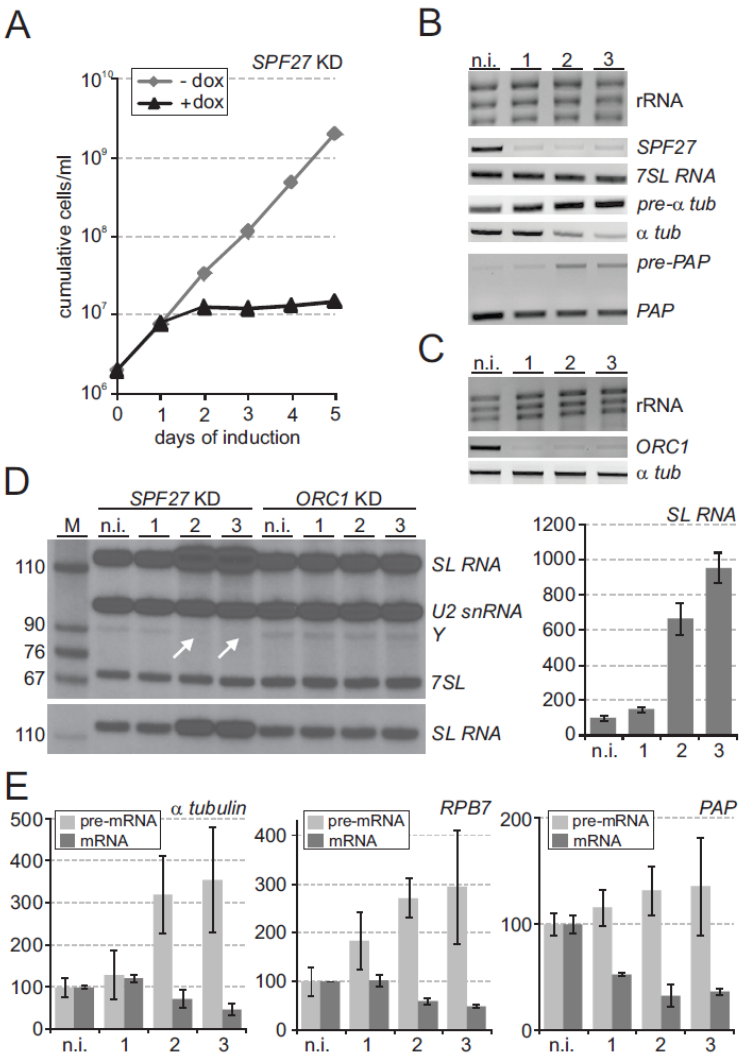



**Figure 5.3 *SPF27* silencing halted trypanosome proliferation and inhibited splicing of pre-mRNA.**

**A.** Procyclic trypanosome culture growth in the absence and presence of doxycycline, which induced *SPF27* silencing.

**B.** Analysis of total RNA prepared from non-induced cells (n.i.) and cells in which *SPF27* was silenced for 1, 2 or 3 days. rRNA was detected by ethidium bromide staining,7SL RNA and [pre-]mRNA of α tubulin (α *tub*) by semi-quantitative PCR and [pre-]mRNA of poly-A polymerase (PAP) by competitive PCR.

**C.** Corresponding analysis of rRNA, *ORC1* mRNA and α tubulin mRNA in *ORC1*-silenced cells.

**D.** Left panels, primer extension of total RNA prepared from *SPF27*- and *ORC1*-silenced cells with three $^{32}$P-5′-endlabeled oligonucleotides detecting SL RNA, U2 snRNA, the SL intron after the first step of SL *trans* splicing in the Y structure, and the 7SL RNA by denaturing PAGE and autoradiography. Arrows indicate the loss of the Y structure intermediate in *SPF27*-silenced cells. The lower panel is a lower exposure the SL RNA-specific product that better shows the increase of the SL RNA signal upon *SPF27* silencing. Right panel, RT-qPCR analysis of SL RNA abundance in *SPF27*-silenced cells. The amounts are relative to that of non-induced cells which was arbitrarily set to 100. Averages were derived from three independently conducted gene knockdown experiments.

**E.** Corresponding RT-qPCR analyses of pre-mRNA and of mature mRNA of α tubulin, RPB7 and PAP. For pre-mRNA analyses, random hexamer-derived cDNA was amplified with a sense oligonucleotide that hybridized upstream of the SL *trans* splice site (α tubulin and RPB7) or with an antisense oligonucleotide that hybridized within the *PAP* intron.

### 4.4 PRP19 and SPF27 co-localize in the nucleus

To further validate SPF27 as a component of the PRP19 complex, we examined if SPF27 co-localizes with PRP19 in trypanosomes. We generated a procyclic cell line in which PRP19 and SPF27 were C-terminally tagged with PTP and HA, respectively. Detection of both proteins by indirect immunofluorescence microscopy revealed that both proteins localized to the nucleus, exhibiting a speckle-like distribution that is characteristic of spliceosomal proteins, including trypanosome and human PRP19 (Figure 5.4) (Marechal et al., 2014; Spector and Lamond, 2011; Tkacz et al., 2010). Consistently, both proteins co-localized to a large extent (Pearson coefficient of 0.73, n = 53), supporting our findings that both proteins are subunits of the same spliceosomal protein complex. In comparison, PRP19 localization was previously found to be not as widespread in the nucleus as in our analysis but more focused at the so-called SL RNP factory in which SL RNA is synthesized (Tkacz et al., 2010). This discrepancy could be due to using different tags (PTP versus YFP), tagging different ends (C- versus N-terminus) or detecting PRP19 by indirect versus direct immunofluorescence microscopy.

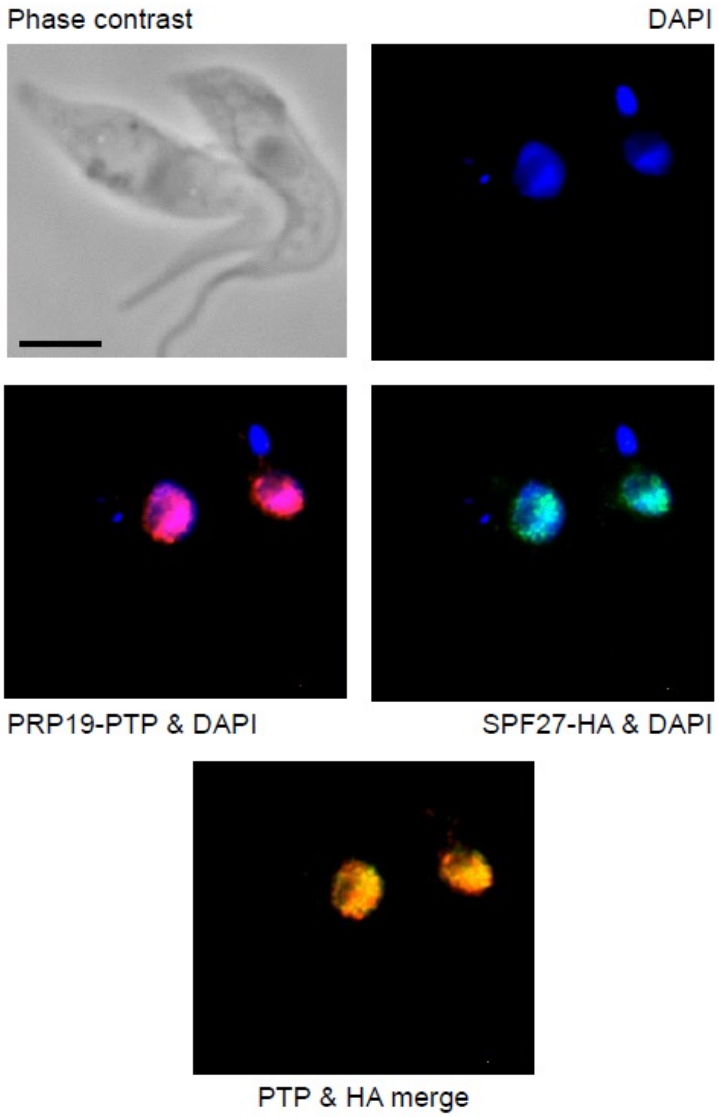



**Figure 5.4 PRP19 and SPF27 co-localize predominantly in nuclear speckles**

PRP19-PTP (red) was co-localized with SPF27-HA (green) in procyclic trypanosomes by indirect immunofluorescence. DNA detected by DAPI staining revealed the large nucleus and the smaller kinetoplast of trypanosomes. The black bar represents 10 µm.

### 4.5 SPF27 silencing leads to SL cap hypomethylation

SL cap4 methylations at positions 1, 2 and 3/4 are carried out by the methyl transferases MTR1 (Zamudio et al., 2007), MTR2 (also known as MT47 or COM1) (Arhin et al., 2006b; Hall and Ho, 2006) and MTR3 (also known as MT57) (Arhin et al., 2006a; Zamudio et al., 2006), respectively, and they occur sequentially from position 1 to 4 (Mair et al., 2000b) recently reviewed by (Sturm et al., 2012). Previously, Tkacz *et al*. (2010) reported that *PRP19* silencing led to a loss of cap3/4 and an accumulation of SL RNAs with cap2 indicating a functional link between PRP19 and MTR3. To test whether *SPF27* silencing led to a comparable defect, we performed a modified primer extension assay with MMLV reverse transcriptase and reduced ribonucleotide concentrations, both of which promote polymerization stops at methylated residues. We first extended oligonucleotide SL_PE which hybridizes to the single-stranded Sm binding region of the SL RNA (positions 111 to 130) and which we had used previously to show that *CRK9* silencing caused the loss of all cap4 methylations (Badjatia et al., 2013a). As shown in figure 5.5A, we found that the shorter extension product, indicative of SL RNA with a fully methylated cap4 structure, disappeared after 2 days of *SPF27* silencing and that the accumulated SL RNA signal was found almost exclusively in the top band instead, suggesting the predominant presence of SL RNA with a hypomethylated cap. Since a nearly congruent primer extension pattern was obtained when *CRK9* was silenced (Badjatia et al., 2013a), this result suggested a complete loss of cap4 methylations instead of a specific cap3/4 loss upon *SPF27* silencing. However, Tkacz *et al.* had used an oligonucleotide that hybridized further upstream of the SL RNA, precisely up to the GU splice site nucleotides of the SL RNA intron. Since it was possible that the use of different primers led to different extension results, we repeated the primer extension assay with oligonucleotide SL40 (Zamudio et al., 2006) that also hybridized up to GU splice site. Although the same assay conditions and RNA preparations were used, primer extension of SL40, in contrast to that of SL_PE, was different revealing cap4

products on days 2 and 3, and a cap2-specific product that was clearly detectable on accumulated SL RNA 2 days into the *SPF27* knockdown (Figure 5.5B). Possibly, the shorter extension of SL40 made reverse transcription more sensitive to methylated residues, resulting in a better detection of methylated and partially methylated SL RNA caps. Nonetheless, consistent with the SL_PE extension results, the cap4 product diminished over the time course, the cap2 product disappeared on day 3, and the accumulating SL RNA predominantly had a cap0. Thus, these findings did not reveal a specific cap3/4 defect but they clearly indicated that *SPF27* silencing led to an accumulation of SL RNA that is progressively hypomethylated with most SL RNA having a cap0 structure.

Since *SPF27* and *CRK9* silencing led to similarly strong accumulations of SL RNA and, according to the SL_PE primer extension results, to a similar loss of cap4 methylation, we co-analyzed total RNA from both gene knockdowns with primer SL40. In addition, we analyzed RNA from the *ORC1* knockdown as a negative control and repeated an *LSm2* silencing experiment which was shown previously to also affect the first step of *trans* splicing (Luz Ambrósio et al., 2009). As shown in figure 5.5C, depletion of the U6 snRNP-specific LSm2 protein caused only a mild increase of SL RNA that did not exhibit any methylation defect. This result is not due to ineffective gene silencing because *LSm2* silencing rapidly and strongly affected trypanosome proliferation in culture (Luz Ambrósio et al., 2009); data not shown). In contrast, the extent of SL RNA accumulation and the very strong cap0 signal were nearly identical in *SPF27* and *CRK9* knockdowns. The similarity of these phenotypes, despite the fact that *SPF27* silencing merely halted culture growth (Figure 5.1A) while *CRK9* silencing led to a strong decline of cell numbers (Badjatia *et al.*, 2013), suggests that SPF27 and CRK9 function in the same step of RNA splicing. *LSm2* silencing led to a loss of U6 snRNA (Luz Ambrósio et al., 2009; Tkacz et al., 2008) which should affect the formation of the B complex before

spliceosome activation. Therefore, strong SL RNA accumulation and cap hypomethylation may be specific consequences of interference with spliceosome activation.

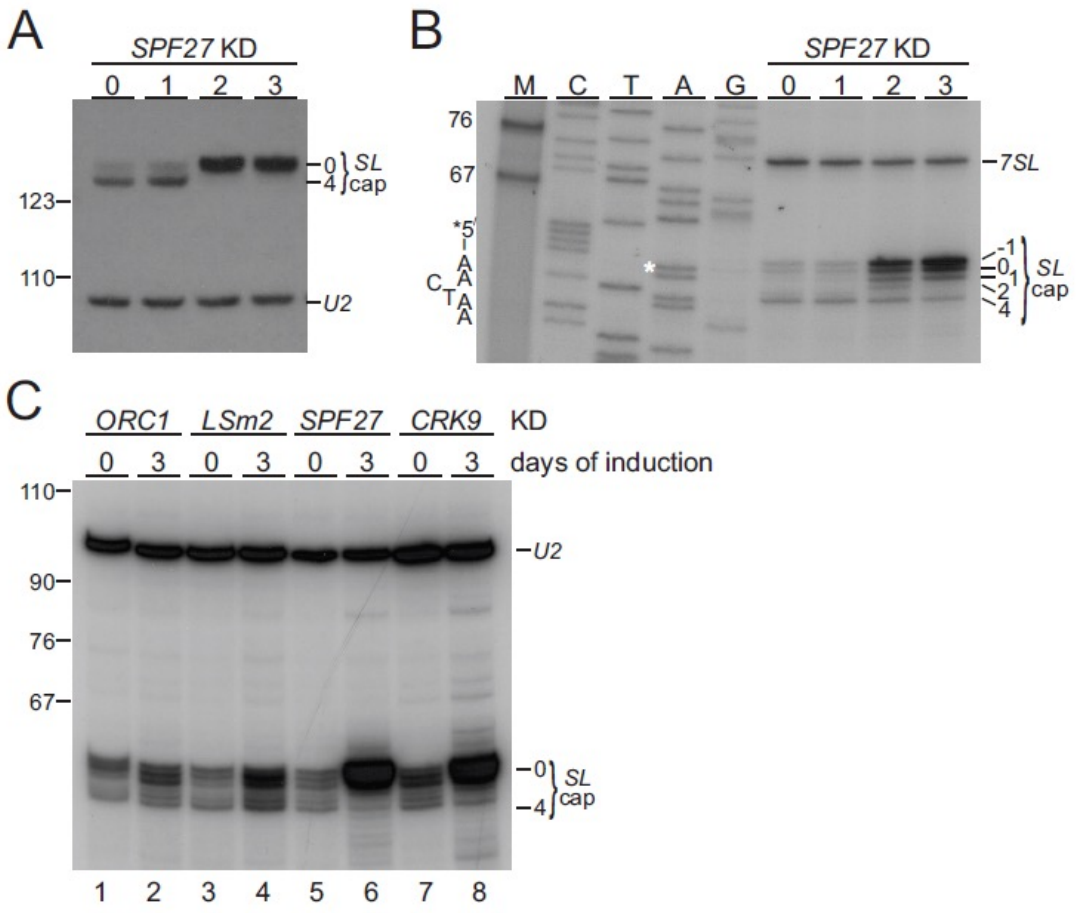



**Figure 5.5 Accumulated SL RNA in *SPF27*-silenced cells is hypomethylated**

**A.** SL RNA and, as a loading control, U2 snRNA primer extension products of total RNA, prepared from non-induced (n.i.) cells and cells in which *SPF27* was silenced for 1 to 3 days, were analyzed on a sequencing gel that separates SL RNA cap4–derived signals from signals of hypomethylated SL RNA caps (SL cap0). On the left, marker bands are indicated.

**B.** Corresponding primer extension analysis of the same RNA with oligonucleotide SL40 which hybridizes just downstream of the 5′ splice site in the SL RNA (position 40 to position 59 relative to SL RNA 5' end). For comparison, a sequencing ladder, which was obtained by linear amplification sequencing with the same oligonucleotide, was co-analyzed. Each lane is marked with the complementary base of the chain-terminating nucleotide used, so that the SL RNA sequence can be directly inferred from top to bottom. The starting A residue is marked with an asterisk. M, marker. On the left, the SL starting sequence 5′-AACTAA-3′ is indicated while specific cap products and the signal of the 7SL RNA, which was co-analyzed as a loading control, are specified on the right. Note that primer extension of SL40 generated a product with an extra nucleotide (position -1) that has been observed before and may be due to terminal transferase activity of reverse transcriptase (Zamudio et al., 2009a).

**C.** SL RNA primer extension analysis, using primer SL40, of total RNA which was prepared from uninduced cells or from cells in which *ORC1*, *LSm2*, *SPF27*, or *CRK9* was silenced for three days. Co-detection of U2 snRNA served as a loading control.

### 4.6 The trypanosome PRP19 complex is associated with U2, U5 and U6 snRNAs

In a final step, we wanted to see whether the trypanosome PRP19 complex is associated with a complex comparable to the activated spliceosome in the yeast and human systems, which is characterized by the presence of U2, U5 and U6 snRNAs and the absence of U4 and U1 snRNAs. For this analysis, we prepared extract from TbP19ee cells, and efficiently pulled down PRP19-PTP by IgG beads (Figure 5.6A). Total RNA, prepared from the precipitated material, was then analyzed by two distinct primer extension assays. In assay A, four oligonucleotides were used to detect SL RNA and the Y structure intermediate as well as U2, U4 and U6 snRNA, whereas assay B contained two oligonucleotides complementary to U1 and U5 snRNAs (Figure 5.6B). In a positive control precipitation of the PTP-tagged SmD1 core snRNP protein (Luz Ambrósio et al., 2009) all six snRNAs were efficiently detected by correctly sized primer extension products (lanes 1 and 2) while in a negative control precipitation of PTP-tagged ORC1 none of these RNAs co-precipitated (lanes 7 and 8). Since all PTP-mediated pull-downs were of similar efficiency (Figure 5.6A), these results confirmed the specificity of the primer extension assays. In comparison, the pull-down of PRP19 mainly co-precipitated U2 and U5 snRNA as well as the Y structure intermediate (Figure 5.6B, lanes 3 and 4). In addition, full length SL RNA and U6 snRNA were co-precipitated although with a reduction in relative signal strength whereas U4 and U1-specific products were hardly detectable. This snRNA pattern shows that the PRP19 complex is predominantly associated with a complex that resembles an activated spliceosome and it suggests that the trypanosome PRP19 complex assembles into the spliceosome after the discard of U1 and U4 snRNA has been initiated. Moreover, the clear detection of the Y structure in PRP19-precipitated material confirms that the trypanosome PRP19 remains associated with spliceosome after the first step of splicing. Since SmD1 binds directly to the SL RNA (Palfi et al., 2000) and PRP19 is associated with snRNAs only through the spliceosome, the reduction of SL RNA signal versus the positive control is expected.

Similarly, the reduction of the U6 signal is likely due to the presence of free U4/U6 snRNP in extract that is not assembled into the spliceosome and, therefore, is precipitated with SmD1 but not with a non-snRNP protein. When we repeated this experiment with two different cell lines that expressed either PTP-tagged CDC5 (lanes 6 and 7) or PTP-tagged SPF27 (lanes 9 and 10) the results were congruent with those for PRP19, further validating CDC5 and SPF27 as *bona fide* subunits of the trypanosome PRP19 complex. Overall, we found that only minor amounts of snRNPs co-precipitated with the PRP19 complex because, in contrast to the SmD1 precipitation in which snRNAs were efficiently co-depleted from the supernatant, the pull down of SPF27 did not detectably reduce the abundance of snRNAs in extract (Attachment 4, Fig. S4). This result is consistent with the finding that in trypanosomal extracts the majority of snRNAs are assembled in individual snRNPs rather than in large spliceosomes (Cross et al., 1991).

Finally, depletion of Sm and LSm proteins as well as of other snRNP proteins in trypanosomes led to a loss of the cognate U snRNA (Liu et al., 2004; Luz Ambrósio et al., 2009; Tkacz et al., 2008; Tkacz et al., 2010). In contrast, *SPF27* silencing for three days did not alter the abundance of spliceosomal snRNAs, verifying the non-snRNP nature of the trypanosome PRP19 complex (Attachment 4, Fig. S5). Thus, the effective co-precipitation of the U snRNAs with PRP19 complex components suggests that trypanosomes form splicing complexes that are comparable to the B and C complexes described in other organisms and that these complexes remained stable in our extracts.

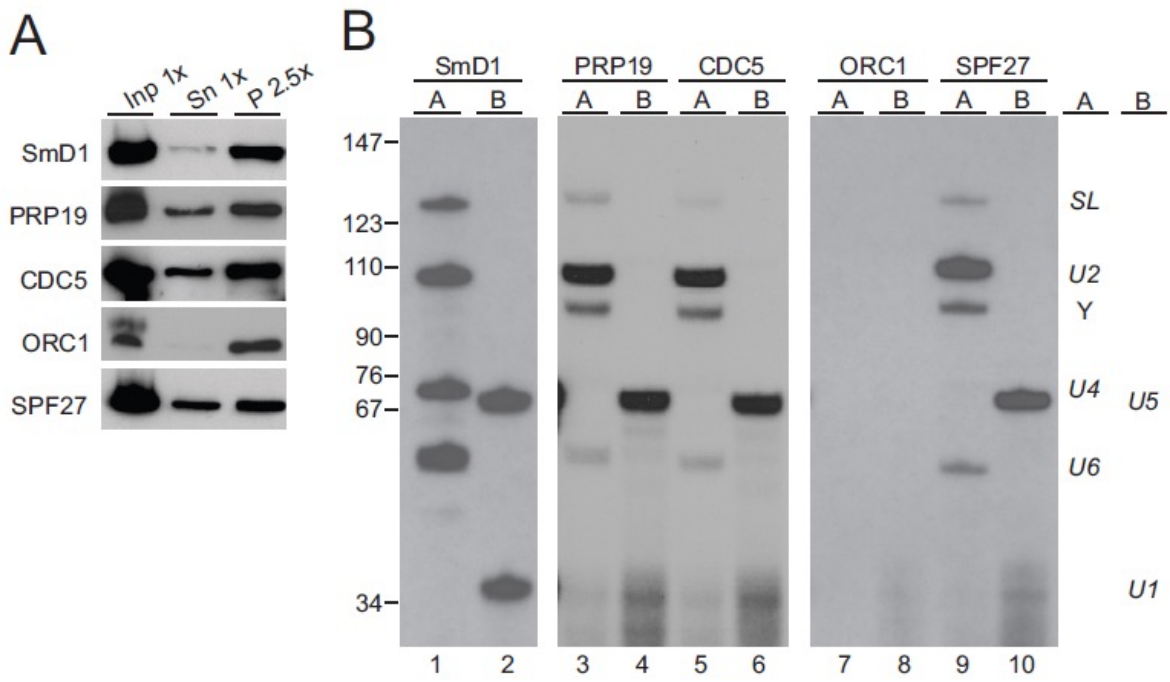



**Figure 5.6 The trypanosome PRP19 complex specifically associates with U2, U5 and U6 snRNAs**

**A.** Extract was prepared from procyclic cell lines that express SmD1, PRP19, CDC5, ORC1 or SPF27 with a functional, C-terminal PTP tag. PTP-tagged proteins were precipitated by IgG beads that bind the protein A domain of PTP and were detected with the peroxidase anti-peroxidase complex that is bound by protein A. Inp, input; Sn, supernatant; P, precipitate. x-Values indicate relative amounts analyzed.

**B.** Total RNA was prepared from the precipitates and analyzed by two primer extension reactions. In reaction A, oligonucleotides were used that are specific to SL RNA, U2 snRNA, U4 snRNA and U6 snRNA, whereas in reaction B, a U5- and a U1-specfic oligonucleotide were combined. The SmD1 and ORC1 pull-downs served as positive and negative controls, respectively. Note that the exposure time of the SmD1 analysis was much shorter than those of the other assays.

## 5. Discussion

In this study we have characterized the spliceosomal non-snRNP PRP19 complex of *T. brucei*. Since trypanosomes and the model organisms from yeast to humans belong to different phylogenetic lineages, e.g. the Excavata and Opisthokonta, respectively (Katz, 2012), a comparison of these complexes can distinguish between properties that are highly conserved over extended evolutionary time and those that vary between species. Accordingly, the subunits PRP19, CDC5, PRL1 and SPF27 are invariably present in the human, yeast and trypanosome complexes indicating that these four proteins form the conserved core of the complex (Table 5.2).

**Table 5.2 PRP19 complex subunits in *H. sapiens*, *T. brucei* and *S. cerevisiae***

| *H. sapiens* | *T. brucei* | *S. cerevisiae* |
|---|---|---|
| **PRP19** | **PRP19** | **Prp19** |
| **CDC5L** | **CDC5** | **Cef1** |
| **PRL1** | **PRL1** | **Prp46** |
| **SPF27** | **SPF27** | **Snt309** |
| AD002 | assoc. | assoc. (Cwc15) |
| CTNNB1 | n.k.h. | n.k.h. |
| HSP73 | assoc. | - |
| - | PRP17 | assoc. |
| assoc. | SKIP | assoc. |
| assoc. | PPIL1 | n.k.h. |
| assoc. | assoc. | Isy1 |
| assoc. | assoc. | Syf1 |
| assoc. | n.k.h. | Syf2 |
| assoc. | assoc. | Clf1 (Syf3) |

assoc. associated; proteins that were co-purified with PRP19 complexes but not considered *bona fide* subunits. In humans they are referred to as “PRP19-related” (Chanarat and Strasser, 2013; Wahl et al., 2009)

n.k.h. no known homologue

- protein was not co-purified

PRP17 is not a subunit of the human PRP19 complex and it is considered to be associated with the *S. cerevisiae* PRP19 complex. Trypanosome PRP17, however, stably co-purified and co-sedimented with the PRP19 complex independent of whether the complex was purified through PRP19 or CDC5 (Fig. 5.1D). This finding confirmed the association of *S. pombe* PRP17 and PRP19 (Ren et al., 2011), and it strongly suggested that PRP17 functions in the spliceosome through association with PRP19 across eukaryotes. In the model of the *S. pombe* PRP19 complex, a PRP19 tetramer binds four PRP17 molecules through interaction of the PRP17 N-terminus with the PRP19 WD40 domain (Ren et al., 2011). These domains are conserved in the trypanosome orthologs (Fig. 5.2A) and the strong protein bands that contained PRP19 and PRP17 (Fig. 5.1D) may indicate that this model does fit the trypanosome complex. However, our sedimentation analysis revealed a complex of ~350 kDa which is substantially smaller than the expected 675 kDa of a complex containing a PRP19 tetramer, four PRP17 molecules and monomers of each of the other subunits. Possibly, tandem purification and sedimentation led to a partial disruption of the PRP19 complex. Alternatively, we cannot rule out that there are two distinct PRP19 complexes of different composition that were not separated during sedimentation.

An interesting deviation from the yeast and human system was the finding that SKIP and PPIL1 were stably associated with trypanosome PRP19. Identification of these two proteins in trypanosomes implies an early evolutionary origin of the SKIP-PPIL1 sub-complex (Wang et al., 2010). Furthermore, although previous work demonstrated a direct interaction between the yeast SKIP ortholog and PRP46p (PRL1) *in vitro* (Albers et al., 2003), these proteins have not been considered *bona fide* subunits of the PRP19 complex and were classified as PRP19-related proteins (Chanarat and Strasser, 2013; Wahl et al., 2009). Since organization of PRP19-related proteins is in general not well understood, our finding suggests that the SKIP/PPIL1 sub-complex functions within the PRP19 complex independent of other PRP19-related proteins. This

seems to be at least true for trypanosomes since we recently succeeded in characterizing a distinct complex of PRP19-related proteins that does not contain PRP19 complex subunits including SKIP and PPIL1 (DL Ambrósio and A Günzl, unpublished data).

Another important aspect of our study was the discovery that depletion of SPF27 led to a strong accumulation of SL RNA with a hypomethylated cap. The extent of SL RNA accumulation and the pattern of cap hypomethylation closely resembled the effects seen upon *CRK9* silencing but not those of the LSm2 knockdown which also affected *trans* splicing (Fig. 5.5C) (Badjatia et al., 2013a; Luz Ambrósio et al., 2009). Thus, these findings suggested that CRK9 functions in spliceosome activation. Originally, our interest in CRK9 stemmed from its apparent role in the phosphorylation of RPB1, the largest subunit of RNAPII. In other systems, RPB1 phosphorylation is critically important for RNAPII transcription initiation and elongation (Buratowski, 2009). However, in trypanosomes the loss of RPB1 phosphorylation upon *CRK9* silencing did not lead to an RNAPII-specific transcription block but, instead, coincided with SL RNA cap hypomethylation and the *trans* splicing block. Since cap4 formation was thought to be co-transcriptional (Mair et al., 2000b) and cap4 methylations a pre-requisite for SL *trans* splicing (McNally and Agabian, 1992; Ullu and Tschudi, 1991), we had speculated that CRK9's primary substrate is RPB1, and that cap methylation and splicing defects are downstream consequences of RPB1 de-phosphorylation (Badjatia et al., 2013a). However, the possibility that CRK9 functions directly in spliceosome activation may point to a regulatory loop in trypanosomes in which the status of RNAPII phosphorylation is directly linked to the *trans* splicing mechanism. Accordingly, it will be interesting to analyze whether blocking spliceosome activation independent of CRK9 affects RNAPII phosphorylation.

Overall, this is the first characterization of a spliceosomal non-snRNP protein complex in trypanosomes. It is an important step towards understanding the spliceosome of *T. brucei* and related parasites. The human spliceosome is a validated anti-cancer drug with the first inhibitor

in clinical trial (Eskens et al., 2013; van Alphen et al., 2009). Given the special use of this RNA-protein machinery in trypanosome SL *trans* splicing, and considering the great evolutionary distance between trypanosomes and humans, a trypanosome-specific splicing inhibitor may wait to be discovered.

# Chapter VI

# Discussion and Future Directions

## 1. The conserved kinetoplastid XPB-R paralog is specialized in Nucleotide Excision Repair (NER)

The multi-subunit and multi-functional transcription factor TFIIH is an important component of the RNA pol II transcription pre-initiation complex (PIC) as well as the NER machinery in eukaryotes (Compe and Egly, 2012). The TFIIH core complex includes the DNA helicase XPB, which is essential for TFIIH's function in both transcription initiation and DNA repair. While model organisms such as yeast and humans have a single, bifunctional XPB that functions exclusively within TFIIH, kinetoplastids harbor two divergent XPB helicases. Our previous work showed that the larger of the two XPB paralogs is part of the trypanosome TFIIH complex that is crucial for *SLRNA* transcription (Lee et al., 2009; Lee et al., 2007). However, the functional significance of the smaller *XPB* paralog was unclear.

In this study, we functionally characterized the smaller XPB paralog that was termed XPB-R (R for repair) due to its essential role for NER in *T. brucei* (Badjatia et al., 2013b). We show that while XPB-R is dispensable of cell viability and transcription, it is important for NER in cells. Our results also indicate that XPB-R is not assembled in a TFIIH complex but that it does interact with the trypanosome p52 ortholog. p52 is a known regulator of XPB function in higher eukaryotes (Coin et al., 2007). Furthermore, gene silencing of *p52* but not of *XPD* rendered trypanosomes susceptible to UV light in a manner comparable to *XPB-R* silencing. This strongly indicates that the interaction of XPB-R with p52 is important for its function in NER. Consistent with our results, in a parallel study it was shown that RNAi of *XPB-R* increases the susceptibility of cells towards UV and cisplatin-induced DNA lesions that are repaired by NER (Machado et

al., 2014). Furthermore, this study also demonstrated that XPD and the larger XPB paralog do not function in DNA repair in trypanosome.

To the best of our knowledge, this is the first study that demonstrates a role of the XPB helicase in repair independent of the TFIIH complex. Our data support a model in which trypanosome TFIIH complex including the DNA helicases XPD and the larger XPB paralog is crucial for RNA pol II transcription but not required for DNA repair while XPB-R and its functional partner p52 are part of a TFIIH independent NER machinery (Figure 6.1). Since a similar XPB dualism is also found in other phylogenetically early-diverged eukaryotes such as ciliates and species of *Entamoeba,* it is possible that a repair-specialized XPB helicase is an ancient eukaryotic trait and that combining the repair and transcriptional functions in a single XPB helicase occurred later during evolution.

While our results clearly demonstrate that XPB-R is not essential for cell survival in *T brucei,* the *XPB-R*$^{-/-}$ knockout cell lines exhibited consistent slow growth when monitored for a period of over three months. This slow growth could be a result of accumulating DNA mutations in culture or it is possible that XPB-R has other functions in cells. Nonetheless, successful generation of clonal *XPB-R*$^{-/-}$ cell lines shows that *XPB-R* is not essential for culture growth of *T. brucei*. However, *XPB-R* could still be an important virulence factor in *T. brucei* for establishing successful infections in its host where growth conditions are very distinct from those in culture. This can be tested in a mouse model by injecting equal number of wild-type and *XPB-R*$^{-/-}$ parasites into mice and monitoring parasitemia and survival over a period of time. This will determine if XPB-R plays a role in virulence of the parasites.

Our results show that *XPB-R*$^{-/-}$ cells are more susceptible to NER inducing agents such as UV light and cisplatin as compared to the wild-type cells. This susceptibility is dose dependent. In order to further strengthen this finding, we can perform IF to determine if XPB-R localizes to the site of DNA damage specifically at DNA lesions induced by UV light and cisplatin. This can be done by colocalizing XPB-R and DNA lesions in cells that have been

exposed to UV or cisplatin. Commercially available monoclonal antibodies against 6-4 photoproducts and cyclobutane pyrimidine dimers (CPD) can be used to detect DNA lesions that are generated after UV exposure and anti-cisplatin antibody can be used to visualize cisplatin adducts. Since we have already established a clonal cell in which XPB-R is exclusively expressed as a fusion with the PTP tag, XPB-R and DNA lesions can be co-detected in this cell line using our established IF protocol. We would expect XPB-R to accumulate at the sites of DNA lesions after exposure to DNA damaging agents. This will be a strong support for our previous results that demonstrate XPB-R's involvement in NER.

Furthermore, IF can be used to determine if absence of *XPB-R* impedes the repair of Cisplatin- or UV-induced DNA lesions. DNA lesions can be visualized in wild type and *XPB-R*$^{-/-}$ cells at different time points after exposure to DNA damaging agents. We would expect to see a delay/loss in the repair of the lesions in *XPB-R*$^{-/-}$ cells. This will demonstrate a direct role of XPB-R in NER in trypanosomes.

Alternatively, a recently established more sensitive and quantitative NER assay termed Oligonucleotide Retrieval Assay (ORA**)** can be employed to compare repair efficiency in *XPB-R*$^{-/-}$ and wild-type cell lines (Shen et al., 2014). In this assay, 5′-biotinylated duplex DNA that contains a CPD adduct is transiently transfected into cells to serve as a substrate for NER. Thereafter, DNA is extracted from cells and the transfected oligonucleotide is isolated using capture by streptavidin beads. Finally, qPCR can be used to quantify the proportion of CPD lesions that were repaired *in vivo*. This sensitive assay can be employed to quantify the differences in NER efficiency in *XPB-R*$^{-/-}$ and wild-type cell lines.

A detailed analysis of NER in *T. brucei* showed that the trypanosome orthologs of CSB and XPG as well as XPB-R are crucial for the trypanosome NER pathway (Machado et al., 2014). In other eukaryotes, CSB is known to function in detection of DNA lesions specifically in the TC-NER pathway while XPC detects damage in the non-transcription coupled genomic NER pathway (Compe and Egly, 2012). XPG is one of the two endonucleases that mediate excision

of damaged DNA (Compe and Egly, 2012). Interestingly, while CSB seems to be important for NER in *T. brucei*, XPC appears to be dispensable. It is possible that trypanosome CSB detects DNA lesions in both TC-NER and genomic NER. However, since the crystal structure of archeal XPB revealed a DNA damage recognition domain, it can also be speculated that trypanosome XPB-R is capable of damage detection independent of XPC or CSB. In the archeon *Sulfolobus solfataricus*, a complex of XPB and Bax1 endonuclease (functional equivalent of XPG), can recognize, unwind and cleave NER substrates *in vitro* (Rouillon and White, 2010). If XPB-R and trypanosome XPG form a similar complex is not yet clear. This possibility can be tested by tandem affinity purification of XPG and/or by reciprocal co-immunoprecipitation analysis of XPB-R and XPG. Furthermore, NER efficiency of cellular extracts that are immmunodepleted of either XPB-R or XPG can be measured *in vitro* NER assays such as the ORA described above. Recombinant XPB-R and/or XPG can be added back to the extracts for reconstitution of the NER activity. Moreover, interaction of recombinant XPG with XPB-R can be examined in pull-down experiments. Lastly, it can be tested if recombinant XPB-R and XPG complex can recognize, open and cleave model NER substrates *in vitro* in a manner similar to repair by the archaeal XPB/Bax1 doublet.

Finally, the role of XPB-R in TC-NER, if any, can be tested in similar experiments as described above. TAP of CSB can potentially reveal novel factors that are important for TC-NER in *T. brucei.* Whether or not XPB-R interacts with CSB can be analyzed by reciprocal immunoprecipitations in extracts and *in vitro* pull-down experiments. Also, whether or not CSB specifically functions in TC-NER can be tested by measuring the NER activity of extracts depleted of CSB *in vitro.* However, since repair of model NER substrates is not coupled to transcription, depletion of CSB should not affect repair of *in vitro* substrates.

In conclusion, various experiments described above can help to further delineate the NER pathway in trypanosomes which is not well-characterized. It will also help us to understand the precise role of XPB-R in NER. These analyses can reveal unique features of trypanosome

NER machinery that are distinct from the NER machinery of the model organisms and they may elucidate the evolution of this important pathway in the eukaryotic kingdom.

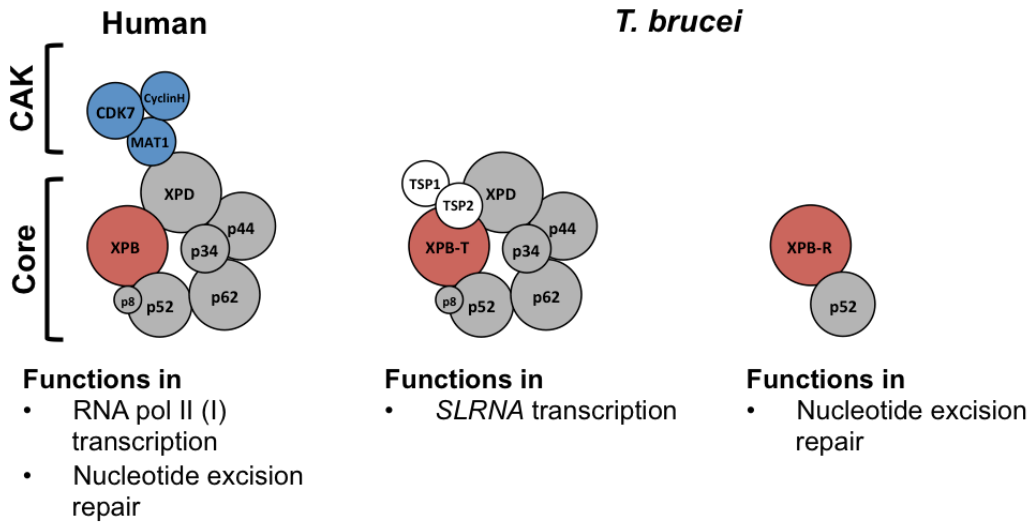



**Figure 6.1 A model for the comparison of TFIIH composition and function in human and *T. brucei*.**

The human TFIIH complex (model modified from (Compe and Egly, 2012)) consists of two structural and functional sub-complexes, namely the core complex, which includes the DNA helicases XPB and XPD, and the cyclin-dependent kinase (CDK)-activating kinase (CAK) complex. The XPB subunit of TFIIH has essential functions in both RNA pol II transcription and DNA repair. Trypanosomes contain two divergent paralogs of the XPB helicase of which only the larger termed XPB-T (T for transcription) is associated in a TFIIH complex together with orthologs of all core subunits as well as two trypanosome specific subunits called TSP1 and TSP2. *T. brucei* TFIIH is essential for *SLRNA* transcription by RNA pol II but lacks the CAK complex that is conserved amongst model eukaryotes. The smaller XPB paralog termed XPB-R (R for repair) in *T. brucei* is specialized in nucleotide excision repair (NER), and has no role in RNA pol II and RNA pol I transcription. It interacts with the p52 subunit of TFIIH, a known regulator of XPB in other systems, but does not appear to assemble in a TFIIH complex. This suggests a model in which trypanosomes possess a TFIIH, containing XPB-T, that is specialized in RNA pol II transcription while NER is carried out via a XPB-R based machinery.

## 2. Characterization of the unique CRK9 complex, which is crucial for SL *trans* splicing

Our biochemical and structural characterization of TFIIH revealed that trypanosome TFIIH lacks the CAK sub-complex (Lee et al., 2009; Lee et al., 2007; Rouillon and White, 2010), composed of CDK7, cyclin H and MAT1, which is typically conserved in all model eukaryotes. Given CDK7's essential role in the phosphorylation of the RPB1 CTD (Fisher, 2005), and the finding that the corresponding trypanosome CTD is phosphorylated as well (Chapman and Agabian, 1994; Das and Bellofatto, 2009), we analyzed the four most essential trypanosome CDKs, namely CRK1, CRK3, CRK7 and CRK9, in an attempt to identify the trypanosome CDK that is involved in CTD phosphorylation. Indeed, our study showed that ablation of CRK9 resulted in a loss of CTD phosphorylation. In addition, we found that this kinase is indispensable for the essential parasite-specific process of SL *trans* splicing in *T. brucei.* Silencing of *CRK9* led to the inhibition of SL *trans* splicing before the first transesterification step and it resulted in SL RNA cap hypomethylation and loss of RPB1 phosphorylation (Badjatia et al., 2013a).

Since trypanosomatids transcribe their protein coding genes in a polycistronic fashion, the processes of SL *trans* splicing and polyadenylation are required for resolving individual mRNAs from polycistronic pre-mRNA. Thus, SL *trans* splicing is a crucial RNA processing step for the maturation of each and every mRNA in these parasites. Accordingly, factors that are involved in SL *trans* splicing are typically essential for parasite survival. In this study, we found that *CRK9* is vital for parasite survival and that *CRK9* silencing led to arrest in culture growth just after one day of inducing the gene knockdown (Badjatia et al., 2013a). Furthermore, we showed that CRK9 is crucial for trypanosomes to establish infection in a mouse model (Chapter V). This is important since CDKs are considered to be suitable drug targets for dividing cells and several CDK inhibitors such as Flavopiridol have been tested or are in clinical trials for cancer therapy (Cicenas et al., 2014; Peyressatre et al., 2015). Since we have validated CRK9 as a

potential chemotherapeutic target for trypanosomes in a mouse model, characterization of the CRK9 complex is an important next step for the development of specific small molecule inhibitors of this enzyme in the future.

For this, we tandem affinity-purified CRK9 and sedimented the final eluate in a sucrose gradient. The sedimentation profile led to identification of a putative tripartite complex of CRK9 together with a previously unannotated cyclin, termed CYC12, and the novel protein CRK9AP. Expression silencing of both CYC12 and CRK9AP recapitulated the phenotypes of the *CRK9* knockdown, confirming that CYC12 and CRK9AP are functional partners of CRK9 *in vivo*.

The newly identified cyclin CYC12 is conserved amongst all kinetoplastids. Our bioinformatic and phylogenetic analyses strongly indicated that CYC12 is a L-type cyclin with a C-terminal arginine/serine-rich RS domain of high positive charge that is characteristic of L-type cyclins in other eukaryotes with the exception of *S. pombe* L-type cyclin that lacks this domain (Dickinson et al., 2002). Thus far, CYC12 is the only known 'transcriptional' cyclin of trypanosomes that is crucial for the parasite-specific mode of gene expression. Interestingly, kinetoplastid CYC12 deviates from known cyclins by large insertions in both of its cyclin domains, e.g. Cyclin_N and Cylin_C. It is likely that these sequence insertions prevented recognition of CYC12 as a cyclin in the annotation of the genome (Parsons et al., 2005). These insertions might reflect functionally important enzyme features that are kinetoplastid-specific . However, this remains to be tested.

Consistent with the identification of CYC12 as a putative L-type cyclin, a standard protein-protein BLAST search of the human or *Drosophila melanogaster* proteome with the *T. brucei* CRK9 sequence returned various isoforms of CDK11 ($E=5e^{-43}$ & $2e^{-37}$ respectively). Cyclin L is a known functional partner of CDK11 in model organisms (Loyer et al., 2008). $CDK11^{p110}$ isoform is implicated in the regulation of gene expression in eukaryotes, and was shown to be essential for pre-mRNA splicing in both *in vivo* and *in vitro* splicing assays (Choi et al., 2014; Hu et al., 2003). Since our results demonstrate CRK9's indispensable role in SL t*rans*

splicing, this finding together with the association of CRK9 with an L-type cyclin, strongly suggests that CRK9 is a divergent ortholog of CDK11.

Our study also identified a small 15 KDa protein termed CRK9AP that is required for *trans* splicing *in vivo*. CRK9AP appears to be kinetoplastid-specific with no recognizable orthologs in other systems and no identifiable functional motifs. Nonetheless, knockdown of *CRK9AP* mirrored the defects observed in *CRK9* and *CYC12* silencing. Our studies indicate that CRK9AP associates with CYC12 and CRK9 to form a stable tripartite complex. This is unique since with the exception of CDK7's CAK complex, CDKs function in heterodimeric complexes together with their cyclin partner. Silencing of CRK9AP leads to co-loss of CRK9 and CYC12 from cells strongly indicating that CRK9AP is important for stability and/or assembly of the complex. Interestingly, similar to the phenotype we observe upon CRK9AP silencing, in MAT1 deficient cells, the levels of CDK7 and cyclin H are significantly reduced (Korsisaari et al., 2002; Rossi et al., 2001; Sano et al., 2007). It is possible to speculate that CRK9AP's function in CRK9 complex is similar to that of MAT1 in CAK complex. However, it remains to be tested if CRK9AP is essential for CRK9's kinase activity *in vitro*. Previously, it has been shown that the C-terminal hydrophobic domain of MAT1 is sufficient for its association in the CAK complex and for activation of the kinase activity of CDK7 (Busso et al., 2000). It is interesting to note that CRK9AP also has a large hydrophobic region at the C-terminus. In the future, a mutational analysis of CRK9AP may help to decipher important domains/residues that mediate CRK9AP's interaction with CRK9 and CYC12.

Another family of small proteins that are implicated in assembly of CDK-cyclin complexes is the Cip/Kip family of CDK inhibitors (CKI). Although members of this protein family are typically considered as inhibitors of cell cycle-related CDKs, they were also shown to be important for the assembly and activation of CDK4/6 - cyclin D complexes (Cheng et al., 1999; LaBaer et al., 1997). However, the function of Cip/Kip family is restricted to CDKs involved in the regulation of the cell cycle and so far no study has demonstrated their role in

activation or inhibition of transcriptional CDKs. Nonetheless, future work will determine if CRK9AP is a divergent ortholog of MAT1 or Cip/Kip proteins or if it is a unique kinetoplastid-specific protein dedicated for the assembly/stability of CRK9 complex.

Our infection studies in mice with cells where *CRK9's* expression can be silenced by induction with doxycycline validated CRK9 as a suitable chemotherapeutic target. However, in order to screen for a large number of small molecule inhibitors of CRK9, a functional recombinant enzyme is required. Although our initial attempts to express full-length *CRK9* and *CYC12* in *E. coli* were not successful, recently we have effectively expressed *CRK9*, *CYC12* and *CRK9AP in vitro* in wheat-germ extracts. This is very promising as this system can allow us to tether the interactions of individual proteins with the others. Our preliminary results indicate that CRK9AP can interact with both CRK9 and CYC12 independently. Future pull-down experiments will determine if CRK9 can interact with CYC12 independent of CRK9AP or if CRK9AP functions as an assembly factor to mediate interaction between CYC12 and CRK9. Additionally, this system can be utilized to perform mutational analysis for identification of domains/regions of ALP or CYC12 that mediate interaction with CRK9 and each other. For example, pull down experiments after deletion of the RS domain of CYC12 and C-terminal hydrophobic region of ALP will determine if these regions are important for protein-protein interactions.

We do not yet know if CRK9 can assemble in a functional kinase complex in wheat-germ extracts. This can be tested by purifying the CRK9 complex from wheat-germ extract in which all three proteins are co-expressed and by subjecting the purified enzyme complex to autophosphorylation in kinase assays. If autophosphorylation of CRK9 is observed, then we can test if CRK9AP is essential for this phosphorylation. However, the CRK9 complex purified from wheat-germ may not be competent in autophosphorylation of CRK9. One possible reason for this could be that activation of CRK9 is dependent on post-translational modifications that are absent in wheat-germ extracts. Previous phospho-proteomic studies of *T. brucei* have identified

phosphorylations of CRK9 threonine 533 (T533), serine 746 (S746), S750 and T752 (Nett et al., 2009; Urbaniak et al., 2013). While our results showed that mutation of T533, which resides in the activating T-loop of the kinase, is deleterious to CRK9's function in cells, the significance of the remaining phosphorylations remains unknown. Phosphomimetic mutations of these four residues can determine if these phosphorylations are crucial or inhibitory for CRK9 kinase activity. Finally if we could establish a functional CRK9 complex from wheat germ extract, a small-scale screen of inhibitors can be performed which can give some leads for establishing a high-throughput screening (HTS) assay in the future.

In order to measure the activity of CRK9 *in vitro* kinase assays, we need to first identify the substrates(s) of CRK9 enzyme. Not only will this allow us to develop a robust *in vitro* kinase assay for CRK9 activity but it will also help delineate the mechanism by which CRK9 functions in SL *trans* splicing *in vivo*. A number of proteins were co-purified with CRK9 during TAP. Most significant proteins with high Mascot Scores are listed in Table 4.1. These are potential substrate candidates, including TSR1, which is a trypanosome SR protein shown to be important of splicing in *Trypanosoma brucei* (Gupta et al., 2014).

Expression of an analog-sensitive mutant of *CRK9* in cells is another powerful strategy to identify its targets *in vivo* (Gregan et al., 2007). Methionine at position 438 was identified as the gatekeeper residue of CRK9 by bioinformatic analysis (data not shown). The gatekeeper residue is a conserved bulky hydrophobic amino acid that resides in the ATP binding pocket of a kinase. Mutation of this amino acid to small residue such as alanine or glycine enlarges the ATP binding site such that the kinase can accommodate a hydrolysable ATP analog (e.g. *N*6-(benzyl)ATP) or specific cell-permeable inhibitors (e.g. 1-NA-PP1, 3-MB-PP1) (Gregan et al., 2007). These analogs or inhibitors can only bind to the mutant kinase and not to any other wild-type kinase. Thus, analog-sensitized CRK9 can be used to identify its substrates in two independent approaches. First, in an indirect approach, cells that exclusively express analog sensitive *CRK9* (*asCRK9*) will be grown in a medium containing cell permeable inhibitors such

as 1-NA-PP1 that specifically inhibits *as*CRK9. In a control experiment cells will be grown in medium lacking the inhibitor but containing amino acids labeled with heavy isotopes. Subsequently, equal number of cells will be harvested from both cultures, combined and analyzed by mass spectrometry to identify phosphoproteins in the sample. This SILAC based mass spectrometry approach will determine relative abundance of phosphoproteins in the two cell populations analyzed. In this approach we will be able to identify candidate proteins whose phosphorylation is directly or indirectly dependent on CRK9 kinase activity. In a second, more direct approach, cellular extract prepared from cells that express *asCRK9* will be subjected to an *in vitro* kinase assay with an ATP analog carrying a thio-phosphate group (N6B-ATPγS) (Blasius et al., 2011). Subsequently, proteins will be digested with trypsin and thio-phosphorylated peptides will be enriched with iodo-acetyl agarose beads (SulfoLink Coupling Resin, Thermo Scientific). Subsequently, the thio-phosphorylated peptides will be eluted with an oxidizing agent and analyzed by mass spectrometry. This approach will allow identification of direct CRK9 substrates. Commercially available antibodies against thiophosphate-ester (e.g. from Epitomics Inc., Burlingame, CA, USA) can be used in immunoblotting experiments to validate potential substrates or autophosphorylation of *as*CRK9.

In order to perform a HTS for specific CRK9 inhibitors we might need functional recombinant enzyme in large amounts. For this a Baculovirus-insect cell expression system can be established. Insect cell expression systems have been used in the past to purify active CDK-cyclin complexes in large quantities for structural analysis. Thus, this system might be useful to obtain functional CRK9 enzyme in amounts needed for HTS. Once a functional recombinant CRK9 enzyme can be obtained, it will be possible to establish a sensitive fluorescence-based HTS assay to screen for a wide range of (potential) kinase inhibitors. We will collaborate with drug screening centers to test for lead compounds. An important candidate is the Drug Discovery Unit in Dundee, UK which is led by Dr. Micheal Ferguson, a well-known trypanosomatid researcher.

In summary, different approaches described above can be used to further characterize the CRK9 complex, identify substrates of CRK9 and to discover specific small molecule inhibitors against CRK9. This study has significant medical implications for the treatment of trypanosomatid diseases since we have validated CRK9 as potential chemotherapeutic target. Furthermore, identification of CRK9 substrates *in vivo* will elucidate the mechanism by which CRK9 regulates the essential parasite-specific process of SL *trans* splicing and it may reveal other potential targets upstream or downstream of CRK9 function in this essential gene expression pathway.

## 3. Characterization of the trypanosome PRP19 complex

Having identified CRK9's crucial role in SL *trans* splicing, we wanted to further characterize the splicing machinery of trypanosomes. In recent years, biochemical characterizations of protein complexes and bioinformatics approaches have led to the identification of orthologs of almost all known spliceosomal snRNP proteins in trypanosomes. However a majority of non-snRNP splicing proteins remain unidentified (Gunzl, 2010; Preußer et al., 2012). In this study, we biochemically and functionally characterized the non-snRNP PRP19 complex of trypanosomes, which in other eukaryotes is essential for spliceosome activation (Ambrosio et al., 2015). By TAP and subsequent sedimentation of the purified PRP19 complex in a sucrose gradient and mass spectrometry of co-sedimented proteins, we identified a PRP19 complex that contained the orthologs of the known core subunits PRP19, CDC5, PRL1, and SPF27 that are also invariably present in yeast and human PRP19 complexes as well as PRP17, SKIP and PPIL1 orthologs which are not considered *bona fide* subunits of the PRP19 complex in the model organisms. This demonstrated that the trypanosome PRP19 complex consists of a conserved, albeit divergent core but, otherwise, deviates substantially in protein composition from its human and yeast counterparts. Moreover, this study is the first

characterization of a non-snRNP protein complex in trypanosomes. In the future, further biochemical characterizations, for example by TAP, will help to expand the repertoire of non-snRNP splicing factors in trypanosomes. These analyses may eventually lead to the identification of a *trans* splicing-specific protein that can be targeted in an antiparasitic strategy.

Recently, the human PRP19 complex was characterized by electron microscopy (EM) after TAP (Grote et al., 2010). This work revealed that the PRP19 complex has an asymmetric and elongated structure with a maximum dimension of ~20 nm and width of ~9 nm. Interestingly, the yeast PRP19 complex also has a similarly elongated morphology but a larger size with a side length of 40 nm (Ohi et al., 2005), indicating differences in the composition of the two complexes. In the future, we can perform EM analysis of the structure of the trypanosome PRP19 complex which might reveal unique structural features of the PRP19 complex in trypanosomatids.

Interestingly, gene silencing of components of the PRP19 complex led not only to defects in *trans* splicing but also to hypomethylation of cap4 and a loss of RPB1 phosphorylation. These phenotypes were very similar qualitatively and quantitatively to those we observed upon *CRK9* silencing. This suggests that CRK9 and PRP19 depletion affects *trans* splicing in a similar fashion and that cap4 hypomethylation and RPB1 dephosphorylation are a consequence of a block of *trans* splicing. Since PRP19 is important for activation of the spliceosome and our results strongly suggest that trypanosome PRP19 associates with the active spliceosome, it is possible that both PRP19 and CRK9 function in spliceosome activation. These findings also allude to a cross-talk between SL *trans* splicing and the status of RNA pol II phosphorylations. It will be interesting to decipher the signaling pathway that leads to the co-regulation of transcription, SL RNA biogenesis and *trans* splicing. Potential proteins involved in this pathway may have purified either with CRK9 (see above) or with the PRP19 complex.

Finally, our previous PTP purifications of SmD1 and PRP19 co-purified several so called ‘conserved hypothetical’ proteins. These are proteins that are conserved only amongst

trypanosomatids with no known ortholog in other systems. One of these conserved hypothetical proteins is encoded by a gene with the accession number Tb927.11.2960 and it was found to be enriched in the PRP19-PTP purification with a high protein score. *Tb927.11.2960* is an essential gene and silencing of its expression led to comparable defects in *trans* splicing and cap4 methylations as observed upon silencing of *PRP19* or *CRK9*. Our preliminary TAP analysis of this protein co-purified several proteins which, in the human system, have been termed PRP19-related proteins because they consistently co-purify with PRP19 subunits but are not stably associated with the PRP19 complex (Wahl et al., 2009). Most interestingly, a sedimentation analysis discovered a distinct complex of these proteins that has not been described in any other organisms and our initial analysis revealed that the complex enters the spliceosome before the PRP19 complex (data not shown). Thus, it is possible that the "PRP19-related" complex facilitates recruitment of the PRP19 complex into the spliceosome. This can be investigated by a conditional knockdown of genes encoding subunits of the newly identified complex and the parallel assessment of the snRNA profile that co-precipitates with PRP19. If the proposed functional relationship between these two distinct complexes can be established, this would provide new insight into spliceosome activation for the RNA splicing field in general. Overall, the spliceosome is difficult to investigate because it is such a huge and dynamic machinery. However, the methodology established in the laboratory and employed and further developed in this thesis project provides the means for a comprehensive analysis of the spliceosome and its signaling pathways to the other essential processes of gene expression.

## References


Abate, A.A., Pentimalli, F., Esposito, L., and Giordano, A. (2013). ATP-noncompetitive CDK inhibitors for cancer therapy: an overview. Expert opinion on investigational drugs *22*, 895-906.

Aboussekhra, A., Biggerstaff, M., Shivji, M.K., Vilpo, J.A., Moncollin, V., Podust, V.N., Protic, M., Hubscher, U., Egly, J.M., and Wood, R.D. (1995). Mammalian DNA nucleotide excision repair reconstituted with purified protein components. Cell *80*, 859-868.

Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., and Luhrmann, R. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. The EMBO journal *18*, 5789-5802.

Akhtar, M.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D., and Ansari, A.Z. (2009). TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. Mol Cell *34*, 387-393.

Aklilu, M., Kindler, H.L., Donehower, R.C., Mani, S., and Vokes, E.E. (2003). Phase II study of flavopiridol in patients with advanced colorectal cancer. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO *14*, 1270-1273.

Akoulitchev, S., Chuikov, S., and Reinberg, D. (2000). TFIIH is negatively regulated by cdk8-containing mediator complexes. Nature *407*, 102-106.

Akoulitchev, S., Makela, T.P., Weinberg, R.A., and Reinberg, D. (1995). Requirement for TFIIH kinase activity in transcription by RNA polymerase II. Nature *377*, 557-560.

Albers, M., Diment, A., Muraru, M., Russell, C.S., and Beggs, J.D. (2003). Identification and characterization of Prp45p and Prp46p, essential pre-mRNA splicing factors. Rna *9*, 138-150.

Alsford, S., duBois, K., Horn, D., and Field, M.C. (2012). Epigenetic mechanisms, nuclear architecture and the control of gene expression in trypanosomes. Expert Rev Mol Med *14*.

Alsford, S., Turner, D.J., Obado, S.O., Sanchez-Flores, A., Glover, L., Berriman, M., Hertz-Fowler, C., and Horn, D. (2011). High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. Genome Res *21*, 915-924.

Ambrosio, D.L., Badjatia, N., and Gunzl, A. (2015). The spliceosomal PRP19 complex of trypanosomes. Mol Microbiol *95*, 885-901.

Arhin, G.K., Li, H., Ullu, E., and Tschudi, C. (2006a). A protein related to the vaccinia virus cap-specific methyltransferase VP39 is involved in cap 4 modification in Trypanosoma brucei. Rna *12*, 53-62.

Arhin, G.K., Ullu, E., and Tschudi, C. (2006b). 2'-O-methylation of position 2 of the trypanosome spliced leader cap 4 is mediated by a 48 kDa protein related to vaccinia virus VP39. Mol Biochem Parasitol *147*, 137-139.

Assfalg, R., Lebedev, A., Gonzalez, O.G., Schelling, A., Koch, S., and Iben, S. (2012). TFIIH is an elongation factor of RNA polymerase I. Nucleic Acids Res *40*, 650-659.

Badjatia, N., Ambrósio, D.L., Lee, J.H., and Günzl, A. (2013a). Trypanosome cdc2-related kinase 9 controls spliced leader RNA cap4 methylation and phosphorylation of RNA polymerase II subunit RPB1. Mol Cell Biol *33*, 1965-1975.

Badjatia, N., Nguyen, T.N., Lee, J.H., and Gunzl, A. (2013b). Trypanosoma brucei harbours a divergent XPB helicase paralogue that is specialized in nucleotide excision repair and conserved among kinetoplastid organisms. Mol Microbiol *90*, 1293-1308.

Balasingham, S.V., Zegeye, E.D., Homberset, H., Rossi, M.L., Laerdahl, J.K., Bohr, V.A., and Tonjum, T. (2012). Enzymatic activities and DNA substrate specificity of Mycobacterium tuberculosis DNA helicase XPB. PLoS One *7*.

Bangs, J.D., Crain, P.F., Hashizume, T., McCloskey, J.A., and Boothroyd, J.C. (1992). Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. J Biol Chem *267*, 9805-9815.

Barrett, M.P., and Croft, S.L. (2012). Management of trypanosomiasis and leishmaniasis. British medical bulletin *104*, 175-196.

Barth, S., Hury, A., Liang, X.-H., and Michaeli, S. (2005). Elucidating the role of H/ACA-like RNAs in trans-splicing and rRNA processing via RNA interference silencing of the Trypanosoma brucei CBF5 pseudouridine synthase. J Biol Chem *280*, 34558-34568.

Beach, D., Durkacz, B., and Nurse, P. (1982). Functionally homologous cell cycle control genes in budding and fission yeast. Nature *300*, 706-709.

Bedez, F., Linard, B., Brochet, X., Ripp, R., Thompson, J.D., Moras, D., Lecompte, O., and Poch, O. (2013). Functional insights into the core-TFIIH from a comparative survey. Genomics *101*, 178-186.

Berglund, J.A., Abovich, N., and Rosbash, M. (1998). A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. Genes & development *12*, 858-867.

Berglund, J.A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. Cell *89*, 781-787.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D.C., Lennard, N.J., Caler, E., Hamlin, N.E., Haas, B.*, et al.* (2005). The genome of the African trypanosome Trypanosoma brucei. Science *309*, 416-422.

Berthet, C., Aleem, E., Coppola, V., Tessarollo, L., and Kaldis, P. (2003). Cdk2 knockout mice are viable. Current biology : CB *13*, 1775-1785.

Biswas, T., Pero, J.M., Joseph, C.G., and Tsodikov, O.V. (2009). DNA-dependent ATPase activity of bacterial XPB helicases. Biochemistry *48*, 2839-2848.

Blasius, M., Forment, J.V., Thakkar, N., Wagner, S.A., Choudhary, C., and Jackson, S.P. (2011). A phospho-proteomic screen identifies substrates of the checkpoint kinase Chk1. Genome biology *12*, R78.

Bonney, K.M. (2014). Chagas disease in the 21st century: a public health success or an emerging threat? Parasite *21*, 11.

Boothroyd, J.C., and Cross, G.A. (1982). Transcripts coding for variant surface glycoproteins of Trypanosoma brucei have a short, identical exon at their 5' end. Gene *20*, 281-289.

Brandenburg, J., Schimanski, B., Nogoceke, E., Nguyen, T.N., Padovan, J.C., Chait, B.T., Cross, G.A., and Günzl, A. (2007). Multifunctional class I transcription in *Trypanosoma brucei* depends on a novel protein complex. EMBO J *26*, 4856-4866.

Brem, R., Fernet, M., Chapot, B., and Hall, J. (2008). The methyl methanesulfonate induced S-phase delay in XRCC1-deficient cells requires ATM and ATR. DNA Repair (Amst) *7*, 849-857.

Bruzik, J.P., Van Doren, K., Hirsh, D., and Steitz, J.A. (1988). Trans splicing involves a novel form of small nuclear ribonucleoprotein particles. Nature *335*, 559-562.

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Mol Cell *36*, 541-546.

Burdette-Radoux, S., Tozer, R.G., Lohmann, R.C., Quirt, I., Ernst, D.S., Walsh, W., Wainman, N., Colevas, A.D., and Eisenhauer, E.A. (2004). Phase II trial of flavopiridol, a cyclin dependent kinase inhibitor, in untreated metastatic malignant melanoma. Investigational new drugs *22*, 315-322.

Busso, D., Keriel, A., Sandrock, B., Poterszman, A., Gileadi, O., and Egly, J.M. (2000). Distinct regions of MAT1 regulate cdk7 kinase and TFIIH transcription activities. J Biol Chem *275*, 22815-22823.

Campbell, D.A., Sturm, N.R., and Yu, M.C. (2000). Transcription of the kinetoplastid spliced leader RNA gene. Parasitology today *16*, 78-82.

Campbell, D.A., Thornton, D.A., and Boothroyd, J.C. (1984). Apparent discontinuous transcription of Trypanosoma brucei variant surface antigen genes. Nature *311*, 350-355.

Cao, L., Chen, F., Yang, X., Xu, W., Xie, J., and Yu, L. (2014). Phylogenetic analysis of CDK and cyclin proteins in premetazoan lineages. BMC evolutionary biology *14*, 10.

Castillo, E., Dea-Ayuela, M.A., Bolas-Fernandez, F., Rangel, M., and Gonzalez-Rosende, M.E. (2010). The kinetoplastid chemotherapy revisited: current drugs, recent advances and future perspectives. Current medicinal chemistry *17*, 4027-4051.

Cavaloc, Y., Popielarz, M., Fuchs, J.P., Gattoni, R., and Stevenin, J. (1994). Characterization and cloning of the human splicing factor 9G8: a novel 35 kDa factor of the serine/arginine protein family. The EMBO journal *13*, 2639-2649.

Chan, S.P., Kao, D.I., Tsai, W.Y., and Cheng, S.C. (2003). The Prp19p-associated complex in spliceosome activation. Science *302*, 279-282.

Chanarat, S., and Strasser, K. (2013). Splicing and beyond: the many faces of the Prp19 complex. Biochim Biophys Acta *1833*, 2126-2134.

Chapman, A.B., and Agabian, N. (1994). Trypanosoma brucei RNA polymerase II is phosphorylated in the absence of carboxyl-terminal domain heptapeptide repeats. J Biol Chem *269*, 4754-4760.

Chapman, R.D., Heidemann, M., Hintermair, C., and Eick, D. (2008). Molecular evolution of the RNA polymerase II CTD. Trends Genet *24*, 289-296.

Chappuis, F., Sundar, S., Hailu, A., Ghalib, H., Rijal, S., Peeling, R.W., Alvar, J., and Boelaert, M. (2007). Visceral leishmaniasis: what are the needs for diagnosis, treatment and control? Nature reviews Microbiology *5*, 873-882.

Cheng, M., Olivier, P., Diehl, J.A., Fero, M., Roussel, M.F., Roberts, J.M., and Sherr, C.J. (1999). The p21(Cip1) and p27(Kip1) CDK 'inhibitors' are essential activators of cyclin D-dependent kinases in murine fibroblasts. The EMBO journal *18*, 1571-1583.

Choi, H.H., Choi, H.K., Jung, S.Y., Hyle, J., Kim, B.J., Yoon, K., Cho, E.J., Youn, H.D., Lahti, J.M., Qin, J.*, et al.* (2014). CHK2 kinase promotes pre-mRNA splicing via phosphorylating CDK11(p110). Oncogene *33*, 108-115.

Cicenas, J., Kalyan, K., Sorokinas, A., Jatulyte, A., Valiunas, D., Kaupinis, A., and Valius, M. (2014). Highlights of the Latest Advances in Research on CDK Inhibitors. Cancers *6*, 2224-2242.

Coin, F., Marinoni, J.C., and Egly, J.M. (1998a). Mutations in XPD helicase prevent its interaction and regulation by p44, another subunit of TFIIH, resulting in Xeroderma pigmentosum (XP) and trichothiodystrophy (TTD) phenotypes. Pathologie-biologie *46*, 679-680.

Coin, F., Marinoni, J.C., Rodolfo, C., Fribourg, S., Pedrini, A.M., and Egly, J.M. (1998b). Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH. Nat Genet *20*, 184-188.

Coin, F., Oksenych, V., and Egly, J.-M. (2007). Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. Mol Cell *26*, 245-256.

Coin, F., Oksenych, V., Mocquet, V., Groh, S., Blattner, C., and Egly, J.M. (2008). Nucleotide excision repair driven by the dissociation of CAK from TFIIH. Mol Cell *31*, 9-20.

Coin, F., Proietti De Santis, L., Nardo, T., Zlobinskaya, O., Stefanini, M., and Egly, J.-M. (2006). p8/TTD-A as a repair-specific TFIIH subunit. Mol Cell *21*, 215-226.

Compe, E., and Egly, J.-M. (2012). TFIIH: when transcription met DNA repair. Nat Rev Mol Cell Biol *13*, 343-354.

Cornelis, S., Bruynooghe, Y., Denecker, G., Van Huffel, S., Tinton, S., and Beyaert, R. (2000). Identification and characterization of a novel cell cycle-regulated internal ribosome entry site. Molecular cell *5*, 597-605.

Crenshaw-Williams, K., and Bellofatto, V. (1999). In vivo transcriptional analysis of the spliced leader RNA gene in the trypanosomatid Leptomonas seymouri. Parasitology research *85*, 700-706.

Cross, M., Günzl, A., Palfi, Z., and Bindereif, A. (1991). Analysis of small nuclear ribonucleoproteins (RNPs) in Trypanosoma brucei: structural organization and protein components of the spliced leader RNP. MolCell Biol *11*, 5516-5526.

Cross, M., Wieland, B., Palfi, Z., Gunzl, A., Rothlisberger, U., Lahm, H.W., and Bindereif, A. (1993). The trans-spliceosomal U2 snRNP protein 40K of Trypanosoma brucei: cloning and analysis of functional domains reveals homology to a mammalian snRNP protein. The EMBO journal *12*, 1239-1248.

Cvitkovic, I., and Jurica, M.S. (2013). Spliceosome database: a tool for tracking components of the spliceosome. Nucleic Acids Res *41*, D132-D141.

Dacks, J.B., Walker, G., and Field, M.C. (2008). Implications of the new eukaryotic systematics for parasitologists. Parasitol Int *57*, 97-9104.

Dahmus, M.E. (1996). Reversible phosphorylation of the C-terminal domain of RNA polymerase II. The Journal of biological chemistry *271*, 19009-19012.

Das, A., Banday, M., and Bellofatto, V. (2008). RNA polymerase transcription machinery in trypanosomes. Eukaryot Cell *7*, 429-434.

Das, A., and Bellofatto, V. (2003). RNA polymerase II-dependent transcription in trypanosomes is associated with a SNAP complex-like transcription factor. Proc Natl Acad Sci U S A *100*, 80-85.

Das, A., and Bellofatto, V. (2009). The non-canonical CTD of RNAP-II is essential for productive RNA synthesis in Trypanosoma brucei. PLoS One *4*.

Das, A., Li, H., Liu, T., and Bellofatto, V. (2006). Biochemical characterization of Trypanosoma brucei RNA polymerase II. Mol Biochem Parasitol *150*, 201-210.

Das, A., Zhang, Q., Palenchar, J.B., Chatterjee, B., Cross, G.A., and Bellofatto, V. (2005). Trypanosomal TBP functions with the multisubunit transcription factor tSNAP to direct spliced-leader RNA gene expression. Mol Cell Biol *25*, 7314-7322.

Denker, J.A., Zuckerman, D.M., Maroney, P.A., and Nilsen, T.W. (2002). New components of the spliced leader RNP required for nematode trans-splicing. Nature *417*, 667-670.

Devaux, S., Lecordier, L., Uzureau, P., Walgraffe, D., Dierick, J.-F., Poelvoorde, P., Pays, E., and Vanhamme, L. (2006). Characterization of RNA polymerase II subunits of Trypanosoma brucei. Mol Biochem Parasitol *148*, 60-68.

Dickinson, L.A., Edgar, A.J., Ehley, J., and Gottesfeld, J.M. (2002). Cyclin L is an RS domain protein involved in pre-mRNA splicing. J Biol Chem *277*, 25465-25473.

Djikeng, A., Ferreira, L., D'Angelo, M., Dolezal, P., Lamb, T., Murta, S., Triggs, V., Ulbert, S., Villarino, A., Renzi, S.*, et al.* (2001). Characterization of a candidate Trypanosoma brucei U1 small nuclear RNA gene. Mol Biochem Parasitol *113*, 109-115.

Drogat, J., Migeot, V., Mommaerts, E., Mullier, C., Dieu, M., van Bakel, H., and Hermand, D. (2012). Cdk11-cyclinL controls the assembly of the RNA polymerase II mediator complex. Cell reports *2*, 1068-1076.

Dungan, J.M., Watkins, K.P., and Agabian, N. (1996). Evidence for the presence of a small U5-like RNA in active trans-spliceosomes of Trypanosoma brucei. The EMBO journal *15*, 4016-4029.

Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., and Murphy, S. (2007). Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. Science *318*, 1777-1779.

Egloff, S., Szczepaniak, S.A., Dienstbier, M., Taylor, A., Knight, S., and Murphy, S. (2010). The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. J Biol Chem *285*, 20564-20569.

Eskens, F.A., Ramos, F.J., Burger, H., O'Brien, J.P., Piera, A., de Jonge, M.J., Mizui, Y., Wiemer, E.A., Carreras, M.J., Baselga, J.*, et al.* (2013). Phase I pharmacokinetic and pharmacodynamic study of the first-in-class spliceosome inhibitor E7107 in patients with advanced solid tumors. Clinical cancer research : an official journal of the American Association for Cancer Research *19*, 6296-6304.

Evans, E., Moggs, J.G., Hwang, J.R., Egly, J.M., and Wood, R.D. (1997). Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. The EMBO journal *16*, 6559-6573.

Evans, T., Rosenthal, E.T., Youngblom, J., Distel, D., and Hunt, T. (1983). Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. Cell *33*, 389-396.

Fabrega, C., Shen, V., Shuman, S., and Lima, C.D. (2003). Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. Mol Cell *11*, 1549-1561.

Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Luhrmann, R. (2009). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. Mol Cell *36*, 593-608.

Fan, L., Arvai, A.S., Cooper, P.K., Iwai, S., Hanaoka, F., and Tainer, J.A. (2006). Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. Mol Cell *22*, 27-37.

Fantoni, A., Dare, A.O., and Tschudi, C. (1994). RNA polymerase III-mediated transcription of the trypanosome U2 small nuclear RNA gene is controlled by both intragenic and extragenic regulatory elements. Mol Cell Biol *14*, 2021-2028.

Feaver, W.J., Svejstrup, J.Q., Bardwell, L., Bardwell, A.J., Buratowski, S., Gulyas, K.D., Donahue, T.F., Friedberg, E.C., and Kornberg, R.D. (1993). Dual roles of a multiprotein complex from S. cerevisiae in transcription and DNA repair. Cell *75*, 1379-1387.

Fei, J., Kaczmarek, N., Luch, A., Glas, A., Carell, T., and Naegeli, H. (2011). Regulation of nucleotide excision repair by UV-DDB: prioritization of damage recognition to internucleosomal DNA. PLoS biology *9*, e1001183.

Fica, S.M., Tuttle, N., Novak, T., Li, N.S., Lu, J., Koodathingal, P., Dai, Q., Staley, J.P., and Piccirilli, J.A. (2013). RNA catalyses nuclear pre-mRNA splicing. Nature *503*, 229-234.

Fisher, R.P. (2005). Secrets of a double agent: CDK7 in cell-cycle control and transcription. J Cell Sci *118*, 5171-5180.

Fisher, R.P., Jin, P., Chamberlin, H.M., and Morgan, D.O. (1995). Alternative mechanisms of CAK assembly require an assembly factor or an activating kinase. Cell *83*, 47-57.

Fitch, M.E., Cross, I.V., Turner, S.J., Adimoolam, S., Lin, C.X., Williams, K.G., and Ford, J.M. (2003). The DDB2 nucleotide excision repair gene product p48 enhances global genomic repair in p53 deficient human fibroblasts. DNA Repair (Amst) *2*, 819-826.

Flores, O., Lu, H., Killeen, M., Greenblatt, J., Burton, Z.F., and Reinberg, D. (1991). The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. Proc Natl Acad Sci U S A *88*, 9999-10003.

Flores, O., Lu, H., and Reinberg, D. (1992). Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIH. J Biol Chem *267*, 2786-2793.

Franco, J.R., Simarro, P.P., Diarra, A., and Jannin, J.G. (2014). Epidemiology of human African trypanosomiasis. Clinical epidemiology *6*, 257-275.

Fregoso, M., Laine, J.-P., Aguilar-Fuentes, J., Mocquet, V., Reynaud, E., Coin, F., Egly, J.-M., and Zurita, M. (2007). DNA repair and transcriptional deficiencies caused by mutations in the Drosophila p52 subunit of TFIIH generate developmental defects and chromosome fragility. Mol Cell Biol *27*, 3640-3650.

Gaillard, P.H., Martini, E.M., Kaufman, P.D., Stillman, B., Moustacchi, E., and Almouzni, G. (1996). Chromatin assembly coupled to DNA repair: a new role for chromatin assembly factor I. Cell *86*, 887-896.

Ghazy, M.A., Brodie, S.A., Ammerman, M.L., Ziegler, L.M., and Ponticelli, A.S. (2004). Amino acid substitutions in yeast TFIIF confer upstream shifts in transcription initiation and altered interaction with RNA polymerase II. Mol Cell Biol *24*, 10975-10985.

Gilinger, G., and Bellofatto, V. (2001). Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. Nucleic Acids Res *29*, 1556-1564.

Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P., and Bentley, D.L. (2009). TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. Mol Cell Biol *29*, 5455-5464.

Godoy, P.D., Nogueira-Junior, L.A., Paes, L.S., Cornejo, A., Martins, R.M., Silber, A.M., Schenkman, S., and Elias, M.C. (2009). Trypanosome prereplication machinery contains a single functional orc1/cdc6 protein, which is typical of archaea. Eukaryot Cell *8*, 1592-1603.

Gourguechon, S., Savich, J.M., and Wang, C.C. (2007). The multiple roles of cyclin E1 in controlling cell cycle progression and cellular morphology of Trypanosoma brucei. J Mol Biol *368*, 939-950.

Gourguechon, S., and Wang, C.C. (2009). CRK9 contributes to regulation of mitosis and cytokinesis in the procyclic form of Trypanosoma brucei. BMC Cell Biol *10*, 68-68.

Gregan, J., Zhang, C., Rumpf, C., Cipak, L., Li, Z., Uluocak, P., Nasmyth, K., and Shokat, K.M. (2007). Construction of conditional analog-sensitive kinase alleles in the fission yeast Schizosaccharomyces pombe. Nature protocols *2*, 2996-3000.

Grondal, E.J., Evers, R., Kosubek, K., and Cornelissen, A.W. (1989). Characterization of the RNA polymerases of Trypanosoma brucei: trypanosomal mRNAs are composed of transcripts derived from both RNA polymerase II and III. The EMBO journal *8*, 3383-3389.

Grote, M., Wolf, E., Will, C.L., Lemm, I., Agafonov, D.E., Schomburg, A., Fischle, W., Urlaub, H., and Lührmann, R. (2010). Molecular architecture of the human Prp19/CDC5L complex. Mol Cell Biol *30*, 2105-2119.

Gunzl, A. (2010). The pre-mRNA splicing machinery of trypanosomes: complex or simplified? Eukaryot Cell *9*, 1159-1170.

Gunzl, A., Bruderer, T., Laufer, G., Schimanski, B., Tu, L.-C., Chung, H.-M., Lee, P.-T., and Lee, M.G.-S. (2003). RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in Trypanosoma brucei. Eukaryot Cell *2*, 542-551.

Gunzl, A., Cross, M., and Bindereif, A. (1992). Domain structure of U2 and U4/U6 small nuclear ribonucleoprotein particles from Trypanosoma brucei: identification of trans-spliceosomal specific RNA-protein interactions. Mol Cell Biol *12*, 468-479.

Gunzl, A., Ullu, E., Dorner, M., Fragoso, S.P., Hoffmann, K.F., Milner, J.D., Morita, Y., Nguu, E.K., Vanacova, S., Wunsch, S.*, et al.* (1997). Transcription of the Trypanosoma brucei spliced leader RNA gene is dependent only on the presence of upstream regulatory elements. Mol Biochem Parasitol *85*, 67-76.

Guo, Z., and Stiller, J.W. (2004). Comparative genomics of cyclin-dependent kinases suggest co-evolution of the RNAP II C-terminal domain and CTD-directed CDKs. BMC Genomics *5*, 69-69.

Gupta, S.K., Chikne, V., Eliaz, D., Tkacz, I.D., Naboishchikov, I., Carmi, S., Waldman Ben-Asher, H., and Michaeli, S. (2014). Two splicing factors carrying serine-arginine motifs, TSR1 and TSR1IP, regulate splicing, mRNA stability, and rRNA processing in Trypanosoma brucei. RNA biology *11*, 715-731.

Gururajan, R., Lahti, J.M., Grenet, J., Easton, J., Gruber, I., Ambros, P.F., and Kidd, V.J. (1998). Duplication of a genomic region containing the Cdc2L1-2 and MMP21-22 genes on human chromosome 1p36.3 and their linkage to D1Z2. Genome research *8*, 929-939.

Hall, M.P., and Ho, C.K. (2006). Functional characterization of a 48 kDa Trypanosoma brucei cap 2 RNA methyltransferase. Nucleic Acids Res *34*, 5594-5602.

Hammarton, T.C. (2007). Cell cycle regulation in Trypanosoma brucei. Mol Biochem Parasitol *153*, 1-8.

Hammarton, T.C., Clark, J., Douglas, F., Boshart, M., and Mottram, J.C. (2003). Stage-specific differences in cell cycle control in Trypanosoma brucei revealed by RNA interference of a mitotic cyclin. The Journal of biological chemistry *278*, 22877-22886.

Hammarton, T.C., Engstler, M., and Mottram, J.C. (2004). The Trypanosoma brucei cyclin, CYC2, is required for cell cycle progression through G1 phase and for maintenance of procyclic form cell morphology. The Journal of biological chemistry *279*, 24757-24764.

Hanawalt, P.C., and Spivak, G. (2008). Transcription-coupled DNA repair: two decades of progress and surprises. Nat Rev Mol Cell Biol *9*, 958-970.

Hannon, G.J., Maroney, P.A., and Nilsen, T.W. (1991). U small nuclear ribonucleoprotein requirements for nematode cis- and trans-splicing in vitro. J Biol Chem *266*, 22792-22795.

Hartree, D., and Bellofatto, V. (1995). Essential components of the mini-exon gene promoter in the trypanosomatid Leptomonas seymouri. Mol Biochem Parasitol *71*, 27-39.

Hartshorne, T., and Agabian, N. (1990). A new U2 RNA secondary structure provided by phylogenetic analysis of trypanosomatid U2 RNAs. Genes & development *4*, 2121-2131.

Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C.L., Pena, V., Lührmann, R., and Stelzl, U. (2012). Dynamic protein-protein interaction wiring of the human spliceosome. Mol Cell *45*, 567-580.

Henry, R.W., Sadowski, C.L., Kobayashi, R., and Hernandez, N. (1995). A TBP-TAF complex required for transcription of human snRNA genes by RNA polymerase II and III. Nature *374*, 653-656.

Hernandez, N. (2001). Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. The Journal of biological chemistry *276*, 26733-26736.

Ho, C.K., and Shuman, S. (2001). Trypanosoma brucei RNA triphosphatase. Antiprotozoal drug target and guide to eukaryotic phylogeny. J Biol Chem *276*, 46182-46186.

Hoogstraten, D., Bergink, S., Ng, J.M.Y., Verbiest, V.H.M., Luijsterburg, M.S., Geverts, B., Raams, A., Dinant, C., Hoeijmakers, J.H.J., Vermeulen, W.*, et al.* (2008). Versatile DNA damage detection by the global genome nucleotide excision repair protein XPC. J Cell Sci *121*, 2850-2859.

Horn, D. (2014). Antigenic variation in African trypanosomes. Mol Biochem Parasitol *195*, 123-129.

Hoskins, A.A., and Moore, M.J. (2012). The spliceosome: a flexible, reversible macromolecular machine. Trends in biochemical sciences *37*, 179-188.

Hsin, J.-P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes & development *26*, 2119-2137.

Hu, D., Mayeda, A., Trembley, J.H., Lahti, J.M., and Kidd, V.J. (2003). CDK11 complexes promote pre-mRNA splicing. J Biol Chem *278*, 8623-8629.

Hury, A., Goldshmidt, H., Tkacz, I.D., and Michaeli, S. (2009). Trypanosome spliced-leader-associated RNA (SLA1) localization and implications for spliced-leader RNA biogenesis. Eukaryot Cell *8*, 56-68.

Ibrahim, B.S., Kanneganti, N., Rieckhof, G.E., Das, A., Laurents, D.V., Palenchar, J.B., Bellofatto, V., and Wah, D.A. (2009). Structure of the C-terminal domain of transcription factor IIB from Trypanosoma brucei. Proc Natl Acad Sci U S A *106*, 13242-13247.

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein

amount in proteomics by the number of sequenced peptides per protein. Molecular & cellular proteomics : MCP *4*, 1265-1272.

Ivens, A.C., Peacock, C.S., Worthey, E.A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M.-A., Adlem, E., Aert, R.*, et al.* (2005). The genome of the kinetoplastid parasite, Leishmania major. Science *309*, 436-442.

Jae, N., Preusser, C., Kruger, T., Tkacz, I.D., Engstler, M., Michaeli, S., and Bindereif, A. (2011). snRNA-specific role of SMN in trypanosome snRNP biogenesis in vivo. RNA biology *8*, 90-100.

Jawdekar, G.W., and Henry, R.W. (2008). Transcriptional regulation of human small nuclear RNA genes. Biochimica et biophysica acta *1779*, 295-305.

Jawhari, A., Laine, J.-P., Dubaele, S., Lamour, V., Poterszman, A., Coin, F., Moras, D., and Egly, J.-M. (2002). p52 Mediates XPB function within the transcription/repair factor TFIIH. J Biol Chem *277*, 31761-31767.

Jeffrey, P.D., Russo, A.A., Polyak, K., Gibbs, E., Hurwitz, J., Massague, J., and Pavletich, N.P. (1995). Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. Nature *376*, 313-320.

Jones, N.G., Thomas, E.B., Brown, E., Dickens, N.J., Hammarton, T.C., and Mottram, J.C. (2014). Regulators of Trypanosoma brucei cell cycle progression and differentiation identified using a kinome-wide RNAi screen. PLoS pathogens *10*, e1003886.

Kamileri, I., Karakasilioti, I., and Garinis, G.A. (2012). Nucleotide excision repair: new tricks with old bricks. Trends Genet *28*, 566-573.

Karagiannis, J., and Balasubramanian, M.K. (2007). A cyclin-dependent kinase that promotes cytokinesis through modulating phosphorylation of the carboxy terminal domain of the RNA Pol II Rpb1p sub-unit. PLoS One *2*.

Kartalou, M., and Essigmann, J.M. (2001). Mechanisms of resistance to cisplatin. Mutat Res *478*, 23-43.

Katz, L.A. (2012). Origin and diversification of eukaryotes. Annual review of microbiology *66*, 411-427.

Kelly, S., Wickstead, B., and Gull, K. (2005). An in silico analysis of trypanosomatid RNA polymerases: insights into their unusual transcription. Biochem Soc Trans *33*, 1435-1437.

Kennedy, P.G. (2013). Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness). The Lancet Neurology *12*, 186-194.

Kim, M., Suh, H., Cho, E.-J., and Buratowski, S. (2009). Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. J Biol Chem *284*, 26421-26426.

Kolev, N.G., Franklin, J.B., Carmi, S., Shi, H., Michaeli, S., and Tschudi, C. (2010). The transcriptome of the human pathogen Trypanosoma brucei at single-nucleotide resolution. PLoS pathogens *6*, e1001090.

Konig, A., Schwartz, G.K., Mohammad, R.M., Al-Katib, A., and Gabrilove, J.L. (1997). The novel cyclin-dependent kinase inhibitor flavopiridol downregulates Bcl-2 and induces growth arrest and apoptosis in chronic B-cell leukemia lines. Blood *90*, 4307-4312.

Kooter, J.M., De Lange, T., and Borst, P. (1984). Discontinuous synthesis of mRNA in trypanosomes. The EMBO journal *3*, 2387-2392.

Korsisaari, N., Rossi, D.J., Paetau, A., Charnay, P., Henkemeyer, M., and Makela, T.P. (2002). Conditional ablation of the Mat1 subunit of TFIIH in Schwann cells provides evidence that Mat1 is not required for general transcription. J Cell Sci *115*, 4275-4284.

Kramer, A., and Utans, U. (1991). Three protein factors (SF1, SF3 and U2AF) function in pre-splicing complex formation in addition to snRNPs. The EMBO journal *10*, 1503-1509.

Krasikova, Y.S., Rechkunova, N.I., Maltseva, E.A., Petruseva, I.O., and Lavrik, O.I. (2010). Localization of xeroderma pigmentosum group A protein and replication protein A on damaged DNA in nucleotide excision repair. Nucleic Acids Res *38*, 8083-8094.

Kuhlman, T.C., Cho, H., Reinberg, D., and Hernandez, N. (1999). The general transcription factors IIA, IIB, IIF, and IIE are required for RNA polymerase II transcription from the human U1 small nuclear RNA promoter. Molecular and cellular biology *19*, 2130-2141.

LaBaer, J., Garrett, M.D., Stevenson, L.F., Slingerland, J.M., Sandhu, C., Chou, H.S., Fattaey, A., and Harlow, E. (1997). New functional activities for the p21 family of CDK inhibitors. Genes & development *11*, 847-862.

Laine, J.-P., and Egly, J.-M. (2006). When transcription and repair meet: a complex system. Trends Genet *22*, 430-436.

Laird, P.W., Kooter, J.M., Loosbroek, N., and Borst, P. (1985). Mature mRNAs of Trypanosoma brucei possess a 5' cap acquired by discontinuous RNA synthesis. Nucleic Acids Res *13*, 4253-4266.

Laird, P.W., ten Asbroek, A.L., and Borst, P. (1987). Controlled turnover and 3' trimming of the trans splicing precursor of Trypanosoma brucei. Nucleic Acids Res *15*, 10087-10103.

Lasda, E.L., and Blumenthal, T. (2011). Trans-splicing. Wiley interdisciplinary reviews RNA *2*, 417-434.

Laufer, G., and Gunzl, A. (2001). In-vitro competition analysis of procyclin gene and variant surface glycoprotein gene expression site transcription in Trypanosoma brucei. Mol Biochem Parasitol *113*, 55-65.

Laufer, G., Schaaf, G., Bollgonn, S., and Gunzl, A. (1999). In vitro analysis of alpha-amanitin-resistant transcription from the rRNA, procyclic acidic repetitive protein, and variant surface glycoprotein gene promoters in Trypanosoma brucei. Mol Cell Biol *19*, 5466-5473.

LeBowitz, J.H., Smith, H.Q., Rusche, L., and Beverley, S.M. (1993). Coupling of poly(A) site selection and trans-splicing in Leishmania. Genes & development *7*, 996-1007.

Lecordier, L., Devaux, S., Uzureau, P., Dierick, J.F., Walgraffe, D., Poelvoorde, P., Pays, E., and Vanhamme, L. (2007). Characterization of a TFIIH homologue from Trypanosoma brucei. Mol Microbiol *64*, 1164-1181.

Lee, J.H., Cai, G., Panigrahi, A.K., Dunham-Ems, S., Nguyen, T.N., Radolf, J.D., Asturias, F.J., and Gunzl, A. (2010). A TFIIH-associated mediator head is a basal factor of small nuclear spliced leader RNA gene transcription in early-diverged trypanosomes. Mol Cell Biol *30*, 5502-5513.

Lee, J.H., Jung, H.S., and Gunzl, A. (2009). Transcriptionally active TFIIH of the early-diverged eukaryote Trypanosoma brucei harbors two novel core subunits but not a cyclin-activating kinase complex. Nucleic Acids Res *37*, 3811-3820.

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

Lehmann, A.R. (2011). DNA polymerases and repair synthesis in NER in human cells. DNA Repair (Amst) *10*, 730-733.

Li, C., Harding, G.A., Parise, J., McNamara-Schroeder, K.J., and Stumph, W.E. (2004a). Architectural arrangement of cloned proximal sequence element-binding protein subunits on Drosophila U1 and U6 snRNA gene promoters. Molecular and cellular biology *24*, 1897-1906.

Li, T., Inoue, A., Lahti, J.M., and Kidd, V.J. (2004b). Failure to proliferate and mitotic arrest of CDK11(p110/p58)-null mutant mice at the blastocyst stage of embryonic cell development. Molecular and cellular biology *24*, 3188-3197.

Li, Z. (2012). Regulation of the cell division cycle in Trypanosoma brucei. Eukaryot Cell *11*, 1180-1190.

Li, Z., and Wang, C.C. (2003). A PHO80-like cyclin and a B-type cyclin control the cell cycle of the procyclic form of Trypanosoma brucei. J Biol Chem *278*, 20652-20658.

Liang, X.H., Liu, Q., Liu, L., Tschudi, C., and Michaeli, S. (2006). Analysis of spliceosomal complexes in Trypanosoma brucei and silencing of two splicing factors Prp31 and Prp43. MolBiochemParasitol *145*, 29-39.

Liang, X.H., Xu, Y.X., and Michaeli, S. (2002). The spliced leader-associated RNA is a trypanosome-specific sn(o) RNA that has the potential to guide pseudouridine formation on the SL RNA. Rna *8*, 237-246.

Lim, S., and Kaldis, P. (2013). Cdks, cyclins and CKIs: roles beyond cell cycle regulation. Development *140*, 3079-3093.

Lin, Y.C., Choi, W.S., and Gralla, J.D. (2005). TFIIH XPB mutants suggest a unified bacterial-like mechanism for promoter opening but not escape. Nature structural & molecular biology *12*, 603-607.

Liu, J., and Kipreos, E.T. (2000). Evolution of cyclin-dependent kinases (CDKs) and CDK-activating kinases (CAKs): differential conservation of CAKs in yeast and metazoa. Molecular biology and evolution *17*, 1061-1074.

Liu, Q., Liang, X.H., Uliel, S., Belahcen, M., Unger, R., and Michaeli, S. (2004). Identification and functional characterization of lsm proteins in Trypanosoma brucei. JBiolChem *279*, 18210-18219.

Liu, S., Rauhut, R., Vornlocher, H.P., and Luhrmann, R. (2006). The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. Rna *12*, 1418-1430.

Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W.H., Ranish, J., Shokat, K.M., and Hahn, S. (2009). Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. Mol Cell Biol *29*, 4852-4863.

Logan-Klumpler, F.J., De Silva, N., Boehme, U., Rogers, M.B., Velarde, G., McQuillan, J.A., Carver, T., Aslett, M., Olsen, C., Subramanian, S*., et al.* (2012). GeneDB--an annotation database for pathogens. Nucleic Acids Res *40*, 98-9108.

Loyer, P., Trembley, J.H., Grenet, J.A., Busson, A., Corlu, A., Zhao, W., Kocak, M., Kidd, V.J., and Lahti, J.M. (2008). Characterization of cyclin L1 and L2 interactions with CDK11 and splicing factors: influence of cyclin L isoforms on splice site selection. J Biol Chem *283*, 7721-7732.

Loyer, P., Trembley, J.H., Katona, R., Kidd, V.J., and Lahti, J.M. (2005). Role of CDK/cyclin complexes in transcription and RNA splicing. Cellular signalling *17*, 1033-1051.

Loyer, P., Trembley, J.H., Lahti, J.M., and Kidd, V.J. (1998). The RNP protein, RNPS1, associates with specific isoforms of the p34cdc2-related PITSLRE protein kinase in vivo. J Cell Sci *111 ( Pt 11)*, 1495-1506.

Lucke, S., Klockner, T., Palfi, Z., Boshart, M., and Bindereif, A. (1997). Trans mRNA splicing in trypanosomes: cloning and analysis of a PRP8-homologous gene from Trypanosoma brucei provides evidence for a U5-analogous RNP. The EMBO journal *16*, 4433-4440.

Luo, H., Gilinger, G., Mukherjee, D., and Bellofatto, V. (1999). Transcription initiation at the TATA-less spliced leader RNA gene promoter requires at least two DNA-binding proteins and a tripartite architecture that includes an initiator element. J Biol Chem *274*, 31947-31954.

Luz Ambrósio, D., Lee, J.H., Panigrahi, A.K., Nguyen, T.N., Cicarelli, R.M., and Günzl, A. (2009). Spliceosomal proteomics in *Trypanosoma brucei* reveal new RNA splicing factors. Eukaryot Cell *8*, 990-1000.

Machado, C.R., Vieira-da-Rocha, J.P., Mendes, I.C., Rajao, M.A., Marcello, L., Bitar, M., Drummond, M.G., Grynberg, P., Oliveira, D.A., Marques, C.*, et al.* (2014). Nucleotide excision repair in Trypanosoma brucei: specialization of transcription-coupled repair due to multigenic transcription. Mol Microbiol *92*, 756-776.

Mackey, Z.B., Koupparis, K., Nishino, M., and McKerrow, J.H. (2011). High-throughput analysis of an RNAi library identifies novel kinase targets in Trypanosoma brucei. Chemical biology & drug design *78*, 454-463.

Mair, G., Shi, H., Li, H., Djikeng, A., Aviles, H.O., Bishop, J.R., Falcone, F.H., Gavrilescu, C., Montgomery, J.L., Santori, M.I.*, et al.* (2000a). A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. Rna *6*, 163-169.

Mair, G., Ullu, E., and Tschudi, C. (2000b). Cotranscriptional cap 4 formation on the Trypanosoma brucei spliced leader RNA. J Biol Chem *275*, 28994-28999.

Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M., and Lührmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. Science *298*, 2205-2208.

Makarova, O.V., Makarov, E.M., Urlaub, H., Will, C.L., Gentzel, M., Wilm, M., and Lührmann, R. (2004). A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. EMBO J *23*, 2381-2391.

Malumbres, M., and Barbacid, M. (2001). To cycle or not to cycle: a critical decision in cancer. Nat Rev Cancer *1*, 222-231.

Malumbres, M., and Barbacid, M. (2009). Cell cycle, CDKs and cancer: a changing paradigm. Nat Rev Cancer *9*, 153-166.

Malumbres, M., Harlow, E., Hunt, T., Hunter, T., Lahti, J.M., Manning, G., Morgan, D.O., Tsai, L.H., and Wolgemuth, D.J. (2009). Cyclin-dependent kinases: a family portrait. Nature cell biology *11*, 1275-1276.

Malumbres, M., Sotillo, R., Santamaria, D., Galan, J., Cerezo, A., Ortega, S., Dubus, P., and Barbacid, M. (2004). Mammalian cells cycle without the D-type cyclin-dependent kinases Cdk4 and Cdk6. Cell *118*, 493-504.

Mandelboim, M., Barth, S., Biton, M., Liang, X.-h., and Michaeli, S. (2003). Silencing of Sm proteins in Trypanosoma brucei by RNA interference captured a novel cytoplasmic intermediate in spliced leader RNA biogenesis. J Biol Chem *278*, 51469-51478.

Mandelboim, M., Estrano, C.L., Tschudi, C., Ullu, E., and Michaeli, S. (2002). On the role of exon and intron sequences in trans-splicing utilization and cap 4 modification of the trypanosomatid Leptomonas collosoma SL RNA. J Biol Chem *277*, 35210-35218.

Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., Weese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R.*, et al.* (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res *39*, D225-D229.

Marechal, A., Li, J.M., Ji, X.Y., Wu, C.S., Yazinski, S.A., Nguyen, H.D., Liu, S., Jimenez, A.E., Jin, J., and Zou, L. (2014). PRP19 transforms into a sensor of RPA-ssDNA after DNA damage and drives ATR activation via a ubiquitin-mediated circuitry. Mol Cell *53*, 235-246.

Martin, J., Hunt, S.L., Dubus, P., Sotillo, R., Nehme-Pelluard, F., Magnuson, M.A., Parlow, A.F., Malumbres, M., Ortega, S., and Barbacid, M. (2003). Genetic rescue of Cdk4 null mice restores pancreatic beta-cell proliferation but not homeostatic cell number. Oncogene *22*, 5261-5269.

Martinez-Calvillo, S., Saxena, A., Green, A., Leland, A., and Myler, P.J. (2007). Characterization of the RNA polymerase II and III complexes in Leishmania major. Int J Parasitol *37*, 491-502.

Martinez-Calvillo, S., Vizuet-de-Rueda, J.C., Florencio-Martinez, L.E., Manning-Cela, R.G., and Figueroa-Angulo, E.E. (2010). Gene expression in trypanosomatid parasites. Journal of biomedicine & biotechnology *2010*, 525241.

Martinez-Calvillo, S., Yan, S., Nguyen, D., Fox, M., Stuart, K., and Myler, P.J. (2003). Transcription of Leishmania major Friedlin chromosome 1 initiates in both directions within a single region. Mol Cell *11*, 1291-1299.

Matera, A.G., and Wang, Z. (2014). A day in the life of the spliceosome. Nat Rev Mol Cell Biol *15*, 108-121.

Mathieu, N., Kaczmarek, N., and Naegeli, H. (2010). Strand- and site-specific DNA lesion demarcation by the xeroderma pigmentosum group D helicase. Proc Natl Acad Sci U S A *107*, 17545-17550.

Matthews, K.R., Tschudi, C., and Ullu, E. (1994). A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. Genes & development *8*, 491-501.

Maxon, M.E., Goodrich, J.A., and Tjian, R. (1994). Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. Genes & development *8*, 515-524.

McCall, L.I., and McKerrow, J.H. (2014). Determinants of disease phenotype in trypanosomatid parasites. Trends in parasitology *30*, 342-349.

McNally, K.P., and Agabian, N. (1992). Trypanosoma brucei spliced-leader RNA methylations are required for trans splicing in vivo. Mol Cell Biol *12*, 4844-4851.

Merritt, C., and Stuart, K. (2013). Identification of essential and non-essential protein kinases by a fusion PCR method for efficient production of transgenic Trypanosoma brucei. Mol Biochem Parasitol *190*, 44-49.

Mettus, R.V., and Rane, S.G. (2003). Characterization of the abnormal pancreatic development, reduced growth and infertility in Cdk4 mutant mice. Oncogene *22*, 8413-8421.

Michaeli, S. (2011). Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. FutureMicrobiol *6*, 459-474.

Michaeli, S. (2012). Spliced leader RNA silencing (SLS) - a programmed cell death pathway in Trypanosoma brucei that is induced upon ER stress. Parasit Vectors *5*, 107-107.

Michaeli, S., Podell, D., Agabian, N., and Ullu, E. (1992). The 7SL RNA homologue of *Trypanosoma brucei* is closely related to mammalian 7SL RNA. Mol Biochem Parasitol *51*, 55-64.

Michaeli, S., Roberts, T.G., Watkins, K.P., and Agabian, N. (1990). Isolation of distinct small ribonucleoprotein particles containing the spliced leader and U2 RNAs of Trypanosoma brucei. J Biol Chem *265*, 10582-10588.

Milhausen, M., Nelson, R.G., Sather, S., Selkirk, M., and Agabian, N. (1984). Identification of a small RNA containing the trypanosome spliced leader: a donor of shared 5' sequences of trypanosomatid mRNAs? Cell *38*, 721-729.

Mocquet, V., Kropachev, K., Kolbanovskiy, M., Kolbanovskiy, A., Tapias, A., Cai, Y., Broyde, S., Geacintov, N.E., and Egly, J.M. (2007). The human DNA repair factor XPC-HR23B distinguishes stereoisomeric benzo[a]pyrenyl-DNA lesions. The EMBO journal *26*, 2923-2932.

Monnerat, S., Almeida Costa, C.I., Forkert, A.C., Benz, C., Hamilton, A., Tetley, L., Burchmore, R., Novo, C., Mottram, J.C., and Hammarton, T.C. (2013). Identification and Functional Characterisation of CRK12:CYC9, a Novel Cyclin-Dependent Kinase (CDK)-Cyclin Complex in. PLoS One *8*, e67327.

Morgan, D.O., and De Bondt, H.L. (1994). Protein kinase regulation: insights from crystal structure analysis. Current opinion in cell biology *6*, 239-246.

Moser, J., Kool, H., Giakzidis, I., Caldecott, K., Mullenders, L.H., and Fousteri, M.I. (2007). Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III alpha in a cell-cycle-specific manner. Mol Cell *27*, 311-323.

Mouaikel, J., Narayanan, U., Verheggen, C., Matera, A.G., Bertrand, E., Tazi, J., and Bordonne, R. (2003). Interaction between the small-nuclear-RNA cap hypermethylase and the spinal muscular atrophy protein, survival of motor neuron. EMBO reports *4*, 616-622.

Murphy, W.J., Watkins, K.P., and Agabian, N. (1986). Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for trans splicing. Cell *47*, 517-525.

Murray, H.W., Berman, J.D., Davies, C.R., and Saravia, N.G. (2005). Advances in leishmaniasis. Lancet *366*, 1561-1577.

Naegeli, H., and Sugasawa, K. (2011). The xeroderma pigmentosum pathway: decision tree analysis of DNA quality. DNA Repair (Amst) *10*, 673-683.

Nakaar, V., Dare, A.O., Hong, D., Ullu, E., and Tschudi, C. (1994). Upstream tRNA genes are essential for expression of small nuclear and cytoplasmic RNA genes in trypanosomes. Mol Cell Biol *14*, 6736-6742.

Nakaar, V., Gunzl, A., Ullu, E., and Tschudi, C. (1997). Structure of the Trypanosoma brucei U6 snRNA gene promoter. Mol Biochem Parasitol *88*, 13-23.

Nasmyth, K.A., and Reed, S.I. (1980). Isolation of genes by complementation in yeast: molecular cloning of a cell-cycle gene. Proceedings of the National Academy of Sciences of the United States of America *77*, 2119-2123.

Naula, C., Parsons, M., and Mottram, J.C. (2005). Protein kinases as drug targets in trypanosomes and Leishmania. Biochimica et biophysica acta *1754*, 151-159.

Nett, I.R., Martin, D.M., Miranda-Saavedra, D., Lamont, D., Barber, J.D., Mehlert, A., and Ferguson, M.A. (2009). The phosphoproteome of bloodstream form Trypanosoma brucei, causative agent of African sleeping sickness. Mol Cell Proteomics *8*, 1527-1538.

Nguyen, T.N., Nguyen, B.N., Lee, J.H., Panigrahi, A.K., and Gunzl, A. (2012). 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 *11*, 1573-1581.

Nguyen, T.N., Schimanski, B., and Günzl, A. (2007). Active RNA polymerase I of *Trypanosoma brucei* harbors a novel subunit essential for transcription. MolCellBiol *27*, 6254-6263.

Nguyen, T.N., Schimanski, B., Zahn, A., Klumpp, B., and Gunzl, A. (2006). Purification of an eight subunit RNA polymerase I complex in Trypanosoma brucei. Mol Biochem Parasitol *149*, 27-37.

Nunes, M.C., Dones, W., Morillo, C.A., Encina, J.J., Ribeiro, A.L., and Council on Chagas Disease of the Interamerican Society of, C. (2013). Chagas disease: an overview of clinical and epidemiological aspects. Journal of the American College of Cardiology *62*, 767-776.

Nurse, P., and Thuriaux, P. (1980). Regulatory genes controlling mitosis in the fission yeast Schizosaccharomyces pombe. Genetics *96*, 627-637.

Nussbaum, K., Honek, J., Cadmus, C.M., and Efferth, T. (2010). Trypanosomatid parasites causing neglected diseases. Current medicinal chemistry *17*, 1594-1617.

Ogi, T., Limsirichaikul, S., Overmeer, R.M., Volker, M., Takenaka, K., Cloney, R., Nakazawa, Y., Niimi, A., Miki, Y., Jaspers, N.G*., et al.* (2010). Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. Mol Cell *37*, 714-727.

Ohi, M.D., and Gould, K.L. (2002). Characterization of interactions among the Cef1p-Prp19p-associated splicing complex. RNA *8*, 798-815.

Ohi, M.D., Link, A.J., Ren, L., Jennings, J.L., McDonald, W.H., and Gould, K.L. (2002). Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. Mol Cell Biol *22*, 2011-2024.

Ohi, M.D., Vander Kooi, C.W., Rosenberg, J.A., Ren, L., Hirsch, J.P., Chazin, W.J., Walz, T., and Gould, K.L. (2005). Structural and functional analysis of essential pre-mRNA splicing factor Prp19p. MolCell Biol *25*, 451-460.

Oksenych, V., Bernardes de Jesus, B., Zhovmer, A., Egly, J.-M., and Coin, F. (2009). Molecular insights into the recruitment of TFIIH to sites of DNA damage. The EMBO journal *28*, 2971-2980.

Oksenych, V., and Coin, F. (2010). The long unwinding road: XPB and XPD helicases in damaged DNA opening. Cell cycle *9*, 90-96.

Ortega, S., Prieto, I., Odajima, J., Martin, A., Dubus, P., Sotillo, R., Barbero, J.L., Malumbres, M., and Barbacid, M. (2003). Cyclin-dependent kinase 2 is essential for meiosis but not for mitotic cell division in mice. Nat Genet *35*, 25-31.

Palfi, Z., and Bindereif, A. (1992). Immunological characterization and intracellular localization of trans-spliceosomal small nuclear ribonucleoproteins in Trypanosoma brucei. J Biol Chem *267*, 20159-20163.

Palfi, Z., Gunzl, A., Cross, M., and Bindereif, A. (1991). Affinity purification of Trypanosoma brucei small nuclear ribonucleoproteins reveals common and specific protein components. Proc Natl Acad Sci U S A *88*, 9097-9101.

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

Palfi, Z., Lucke, S., Lahm, H.W., Lane, W.S., Kruft, V., Bragado-Nilsson, E., Seraphin, B., and Bindereif, A. (2000). The spliceosomal snRNP core complex of Trypanosoma brucei: cloning

and functional analysis reveals seven Sm protein constituents. ProcNatlAcadSciUSA *97*, 8967-8972.

Park, S.H., Nguyen, T.N., Kirkham, J.K., Lee, J.H., and Gunzl, A. (2011). Transcription by the multifunctional RNA polymerase I in Trypanosoma brucei functions independently of RPB7. Mol Biochem Parasitol *180*, 35-42.

Parsons, M., Worthey, E.A., Ward, P.N., and Mottram, J.C. (2005). Comparative analysis of the kinomes of three pathogenic trypanosomatids: Leishmania major, Trypanosoma brucei and Trypanosoma cruzi. BMC Genomics *6*, 127-127.

Passos-Silva, D.G., Rajao, M.A., Nascimento de Aguiar, P.H., Vieira-da-Rocha, J.P., Machado, C.R., and Furtado, C. (2010). Overview of DNA Repair in Trypanosoma cruzi, Trypanosoma brucei, and Leishmania major. J Nucleic Acids *2010*, 840768-840768.

Patterson, S., and Wyllie, S. (2014). Nitro drugs for the treatment of trypanosomatid diseases: past, present, and future prospects. Trends in parasitology *30*, 289-298.

Pays, E., Vanhollebeke, B., Uzureau, P., Lecordier, L., and Perez-Morga, D. (2014). The molecular arms race between African trypanosomes and humans. Nature reviews Microbiology *12*, 575-584.

Peyressatre, M., Prevel, C., Pellerano, M., and Morris, M.C. (2015). Targeting cyclin-dependent kinases in human cancers: from small molecules to Peptide inhibitors. Cancers *7*, 179-237.

Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. Genes & development *20*, 2922-2936.

Pines, J. (1995). Cyclins and cyclin-dependent kinases: a biochemical view. The Biochemical journal *308 ( Pt 3)*, 697-711.

Pitts, T.M., Davis, S.L., Eckhardt, S.G., and Bradshaw-Pierce, E.L. (2014). Targeting nuclear kinases in cancer: development of cell cycle kinase inhibitors. Pharmacology & therapeutics *142*, 258-269.

Preußer, C., Jae, N., Günzl, A., and Bindereif, A. (2012). Pre-mRNA splicing in *Trypanosoma brucei*: factors, Mechanisms, and Regulation. In RNA Metabolism in Trypanosomes, A. Bindereif, ed. (Springer Press), pp. 49-76.

Price, D.H., Sluder, A.E., and Greenleaf, A.L. (1989). Dynamic interaction between a Drosophila transcription factor and RNA polymerase II. Mol Cell Biol *9*, 1465-1475.

Qiu, H., Hu, C., and Hinnebusch, A.G. (2009). Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. Mol Cell *33*, 752-762.

Ramanathan, Y., Rajpara, S.M., Reza, S.M., Lees, E., Shuman, S., Mathews, M.B., and Pe'ery, T. (2001). Three RNA polymerase II carboxyl-terminal domain kinases display distinct substrate preferences. J Biol Chem *276*, 10913-10920.

Rane, S.G., Dubus, P., Mettus, R.V., Galbreath, E.J., Boden, G., Reddy, E.P., and Barbacid, M. (1999). Loss of Cdk4 expression causes insulin-deficient diabetes and Cdk4 activation results in beta-islet cell hyperplasia. Nat Genet *22*, 44-52.

Reed, S.I., Ferguson, J., and Groppe, J.C. (1982). Preliminary characterization of the transcriptional and translational products of the Saccharomyces cerevisiae cell division cycle gene CDC28. Mol Cell Biol *2*, 412-425.

Ren, L., McLean, J.R., Hazbun, T.R., Fields, S., Vander, K.C., Ohi, M.D., and Gould, K.L. (2011). Systematic two-hybrid and comparative proteomic analyses reveal novel yeast pre-mRNA splicing factors connected to Prp19. PLoS ONE *6*, e16719.

Richards, J.D., Cubeddu, L., Roberts, J., Liu, H., and White, M.F. (2008). The archaeal XPB protein is a ssDNA-dependent ATPase with a novel partner. J Mol Biol *376*, 634-644.

Rocha, A.A., Moretti, N.S., and Schenkman, S. (2014). Stress induces changes in the phosphorylation of Trypanosoma cruzi RNA polymerase II, affecting its association with chromatin and RNA processing. Eukaryot Cell *13*, 855-865.

Rodriguez, C.R., Cho, E.J., Keogh, M.C., Moore, C.L., Greenleaf, A.L., and Buratowski, S. (2000). Kin28, the TFIIH-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. Molecular and cellular biology *20*, 104-112.

Rossi, D.J., Londesborough, A., Korsisaari, N., Pihlak, A., Lehtonen, E., Henkemeyer, M., and Makela, T.P. (2001). Inability to enter S phase and defective RNA polymerase II CTD phosphorylation in mice lacking Mat1. The EMBO journal *20*, 2844-2856.

Rouillon, C., and White, M.F. (2010). The XBP-Bax1 helicase-nuclease complex unwinds and cleaves DNA: implications for eukaryal and archaeal nucleotide excision repair. J Biol Chem *285*, 11013-11022.

Rouillon, C., and White, M.F. (2011). The evolution and mechanisms of nucleotide excision repair proteins. Res Microbiol *162*, 19-26.

Roy, R., Adamczewski, J.P., Seroz, T., Vermeulen, W., Tassan, J.P., Schaeffer, L., Nigg, E.A., Hoeijmakers, J.H., and Egly, J.M. (1994). The MO15 cell cycle kinase is associated with the TFIIH transcription-DNA repair factor. Cell *79*, 1093-1101.

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

Ruan, J.P., Shen, S., Ullu, E., and Tschudi, C. (2007). Evidence for a capping enzyme with specificity for the trypanosome spliced leader RNA. Mol Biochem Parasitol *156*, 246-254.

Saito, R.M., Elgort, M.G., and Campbell, D.A. (1994). A conserved upstream element is essential for transcription of the Leishmania tarentolae mini-exon gene. The EMBO journal *13*, 5460-5469.

Sander, B., Golas, M.M., Makarov, E.M., Brahms, H., Kastner, B., Luhrmann, R., and Stark, H. (2006). Organization of core spliceosomal components U5 snRNA loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomicroscopy. Mol Cell *24*, 267-278.

Sano, M., Izumi, Y., Helenius, K., Asakura, M., Rossi, D.J., Xie, M., Taffet, G., Hu, L., Pautler, R.G., Wilson, C.R.*, et al.* (2007). Menage-a-trois 1 is critical for the transcriptional function of PPARgamma coactivator 1. Cell metabolism *5*, 129-142.

Satyanarayana, A., and Kaldis, P. (2009). Mammalian cell-cycle regulation: several Cdks, numerous cyclins and diverse compensatory mechanisms. Oncogene *28*, 2925-2939.

Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J.H., Chambon, P., and Egly, J.M. (1993). DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. Science *260*, 58-63.

Schimanski, B., Brandenburg, J., Nguyen, T.N., Caimano, M.J., and Gunzl, A. (2006). A TFIIB-like protein is indispensable for spliced leader RNA gene transcription in Trypanosoma brucei. Nucleic Acids Res *34*, 1676-1684.

Schimanski, B., Klumpp, B., Laufer, G., Marhofer, R.J., Selzer, P.M., and Gunzl, A. (2003). The second largest subunit of Trypanosoma brucei's multifunctional RNA polymerase I has a unique N-terminal extension domain. Mol Biochem Parasitol *126*, 193-200.

Schimanski, B., Nguyen, T.N., and Gunzl, A. (2005a). Highly efficient tandem affinity purification of trypanosome protein complexes based on a novel epitope combination. Eukaryot Cell *4*, 1942-1950.

Schimanski, B., Nguyen, T.N., and Günzl, A. (2005b). Characterization of a multisubunit transcription factor complex essential for spliced-leader RNA gene transcription in *Trypanosoma brucei*. MolCell Biol *25*, 7303-7313.

Schwartz, G.K., Ilson, D., Saltz, L., O'Reilly, E., Tong, W., Maslak, P., Werner, J., Perkins, P., Stoltz, M., and Kelsen, D. (2001). Phase II study of the cyclin-dependent kinase inhibitor flavopiridol administered to patients with advanced gastric carcinoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *19*, 1985-1992.

Seipelt, R.L., Zheng, B., Asuru, A., and Rymond, B.C. (1999). U1 snRNA is cleaved by RNase III and processed through an Sm site-dependent pathway. Nucleic Acids Res *27*, 587-595.

Senderowicz, A.M., Headlee, D., Stinson, S.F., Lush, R.M., Kalil, N., Villalba, L., Hill, K., Steinberg, S.M., Figg, W.D., Tompkins, A.*, et al.* (1998). Phase I trial of continuous infusion flavopiridol, a novel cyclin-dependent kinase inhibitor, in patients with refractory neoplasms. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *16*, 2986-2999.

Serizawa, H., Makela, T.P., Conaway, J.W., Conaway, R.C., Weinberg, R.A., and Young, R.A. (1995). Association of Cdk-activating kinase subunits with transcription factor TFIIH. Nature *374*, 280-282.

Shapiro, G.I. (2006). Cyclin-dependent kinase pathways as targets for cancer treatment. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *24*, 1770-1783.

Shapiro, G.I., Koestner, D.A., Matranga, C.B., and Rollins, B.J. (1999). Flavopiridol induces cell cycle arrest and p53-independent apoptosis in non-small cell lung cancer cell lines. Clinical cancer research : an official journal of the American Association for Cancer Research *5*, 2925-2938.

Shapiro, G.I., Supko, J.G., Patterson, A., Lynch, C., Lucca, J., Zacarola, P.F., Muzikansky, A., Wright, J.J., Lynch, T.J., Jr., and Rollins, B.J. (2001). A phase II trial of the cyclin-dependent kinase inhibitor flavopiridol in patients with previously untreated stage IV non-small cell lung cancer. Clinical cancer research : an official journal of the American Association for Cancer Research *7*, 1590-1599.

Sheader, K., te Vruchte, D., and Rudenko, G. (2004). Bloodstream form-specific up-regulation of silent vsg expression sites and procyclin in Trypanosoma brucei after inhibition of DNA synthesis or DNA damage. J Biol Chem *279*, 13363-13374.

Shen, J.C., Fox, E.J., Ahn, E.H., and Loeb, L.A. (2014). A rapid assay for measuring nucleotide excision repair by oligonucleotide retrieval. Scientific reports *4*, 4894.

Shepard, P.J., and Hertel, K.J. (2009). The SR protein family. Genome biology *10*, 242.

Shiekhattar, R., Mermelstein, F., Fisher, R.P., Drapkin, R., Dynlacht, B., Wessling, H.C., Morgan, D.O., and Reinberg, D. (1995). Cdk-activating kinase complex is a component of human transcription factor TFIIH. Nature *374*, 283-287.

Shlomai, J. (2004). The structure and replication of kinetoplast DNA. Current molecular medicine *4*, 623-647.

Siegel, T.N., Gunasekera, K., Cross, G.A., and Ochsenreiter, T. (2011). Gene expression in Trypanosoma brucei: lessons from high-throughput RNA sequencing. Trends in parasitology *27*, 434-441.

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

Sikorski, T.W., and Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. Current opinion in cell biology *21*, 344-351.

Smith, D.J., Query, C.C., and Konarska, M.M. (2008). "Nought may endure but mutability": spliceosome dynamics and the regulation of splicing. Mol Cell *30*, 657-666.

Sogaard, T.M., and Svejstrup, J.Q. (2007). Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. J Biol Chem *282*, 14113-14120.

Spector, D.L., and Lamond, A.I. (2011). Nuclear speckles. Cold Spring Harbor perspectives in biology *3*.

Staresincic, L., Fagbemi, A.F., Enzlin, J.H., Gourdin, A.M., Wijgers, N., Dunand-Sauthier, I., Giglia-Mari, G., Clarkson, S.G., Vermeulen, W., and Scharer, O.D. (2009). Coordination of dual incision and repair synthesis in human nucleotide excision repair. The EMBO journal *28*, 1111-1120.

Stechmann, A., and Cavalier-Smith, T. (2002). Rooting the eukaryote tree by using a derived gene fusion. Science *297*, 89-91.

Stewart, M., Haile, S., Jha, B.A., Cristodero, M., Li, C.-H., and Clayton, C. (2010). Processing of a phosphoglycerate kinase reporter mRNA in Trypanosoma brucei is not coupled to transcription by RNA polymerase II. Mol Biochem Parasitol *172*, 99-9106.

Sturm, N.R., Fleischmann, J., and Campbell, D.A. (1998). Efficient trans-splicing of mutated spliced leader exons in Leishmania tarentolae. J Biol Chem *273*, 18689-18692.

Sturm, N.R., Yu, M.C., and Campbell, D.A. (1999). Transcription termination and 3'-End processing of the spliced leader RNA in kinetoplastids. Mol Cell Biol *19*, 1595-1604.

Sturm, N.R., Zamudio, J.R., and Campbell, D.A. (2012). SL RNA biogenesis in kinetoplastids: A long and winding road. In RNA Metabolism in Trypanosomes, A. Bindereif, ed. (Springer Press), pp. 29-47.

Sugasawa, K., Okamoto, T., Shimizu, Y., Masutani, C., Iwai, S., and Hanaoka, F. (2001). A multistep damage recognition mechanism for global genomic nucleotide excision repair. Genes & development *15*, 507-521.

Sutton, R.E., and Boothroyd, J.C. (1986). Evidence for trans splicing in trypanosomes. Cell *47*, 527-535.

Takagi, Y., Sindkar, S., Ekonomidis, D., Hall, M.P., and Ho, C.K. (2007). Trypanosoma brucei encodes a bifunctional capping enzyme essential for cap 4 formation on the spliced leader RNA. J Biol Chem *282*, 15995-16005.

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol *56*, 564-577.

Tan, A.R., Headlee, D., Messmann, R., Sausville, E.A., Arbuck, S.G., Murgo, A.J., Melillo, G., Zhai, S., Figg, W.D., Swain, S.M.*, et al.* (2002). Phase I clinical and pharmacokinetic study of flavopiridol administered as a daily 1-hour infusion in patients with advanced neoplasms. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *20*, 4074-4082.

Tantin, D. (1998). RNA polymerase II elongation complexes containing the Cockayne syndrome group B protein interact with a molecular complex containing the transcription factor IIH components xeroderma pigmentosum B and p62. J Biol Chem *273*, 27794-27799.

Tarn, W.Y., Hsu, C.H., Huang, K.T., Chen, H.R., Kao, H.Y., Lee, K.R., and Cheng, S.C. (1994). Functional association of essential splicing factor(s) with PRP19 in a protein complex. EMBO J *13*, 2421-2431.

Tassan, J.P., Jaquenoud, M., Fry, A.M., Frutiger, S., Hughes, G.J., and Nigg, E.A. (1995). In vitro assembly of a functional human CDK7-cyclin H complex requires MAT1, a novel 36 kDa RING finger protein. The EMBO journal *14*, 5608-5617.

Thoma, F. (1999). Light and dark in chromatin repair: repair of UV-induced DNA lesions by photolyase and nucleotide excision repair. The EMBO journal *18*, 6585-6598.

Thomas, J.P., Tutsch, K.D., Cleary, J.F., Bailey, H.H., Arzoomanian, R., Alberti, D., Simon, K., Feierabend, C., Binger, K., Marnocha, R.*, et al.* (2002). Phase I clinical and pharmacokinetic trial of the cyclin-dependent kinase inhibitor flavopiridol. Cancer chemotherapy and pharmacology *50*, 465-472.

Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. Critical reviews in biochemistry and molecular biology *41*, 105-178.

Tietjen, J.R., Zhang, D.W., Rodriguez-Molina, J.B., White, B.E., Akhtar, M.S., Heidemann, M., Li, X., Chapman, R.D., Shokat, K., Keles, S.*, et al.* (2010). Chemical-genomic dissection of the CTD code. Nature structural & molecular biology *17*, 1154-1161.

Tirode, F., Busso, D., Coin, F., and Egly, J.M. (1999). Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. Mol Cell *3*, 87-95.

Tkacz, I.D., Cohen, S., Salmon-Divon, M., and Michaeli, S. (2008). Identification of the heptameric Lsm complex that binds U6 snRNA in Trypanosoma brucei. Mol Biochem Parasitol *160*, 22-31.

Tkacz, I.D., Gupta, S.K., Volkov, V., Romano, M., Haham, T., Tulinski, P., Lebenthal, I., and Michaeli, S. (2010). Analysis of spliceosomal proteins in Trypanosomatids reveals novel functions in mRNA processing. J Biol Chem *285*, 27982-27999.

Tkacz, I.D., Lustig, Y., Stern, M.Z., Biton, M., Salmon-Divon, M., Das, A., Bellofatto, V., and Michaeli, S. (2007). Identification of novel snRNA-specific Sm proteins that bind selectively to U2 and U4 snRNAs in Trypanosoma brucei. Rna *13*, 30-43.

Trembley, J.H., Hu, D., Hsu, L.C., Yeung, C.Y., Slaughter, C., Lahti, J.M., and Kidd, V.J. (2002). PITSLRE p110 protein kinases associate with transcription complexes and affect their activity. J Biol Chem *277*, 2589-2596.

Trembley, J.H., Hu, D., Slaughter, C.A., Lahti, J.M., and Kidd, V.J. (2003). Casein kinase 2 interacts with cyclin-dependent kinase 11 (CDK11) in vivo and phosphorylates both the RNA polymerase II carboxyl-terminal domain and CDK11 in vitro. J Biol Chem *278*, 2265-2270.

Trembley, J.H., Loyer, P., Hu, D., Li, T., Grenet, J., Lahti, J.M., and Kidd, V.J. (2004). Cyclin dependent kinase 11 in RNA transcription and splicing. Progress in nucleic acid research and molecular biology *77*, 263-288.

Truglio, J.J., Croteau, D.L., Van Houten, B., and Kisker, C. (2006). Prokaryotic nucleotide excision repair: the UvrABC system. Chem Rev *106*, 233-252.

Tschudi, C., and Ullu, E. (1990). Destruction of U2, U4, or U6 small nuclear RNA blocks trans splicing in trypanosome cells. Cell *61*, 459-466.

Tschudi, C., Williams, S.P., and Ullu, E. (1990). Conserved sequences in the U2 snRNA-encoding genes of Kinetoplastida do not include the putative branchpoint recognition region. Gene *91*, 71-77.

Tu, X., and Wang, C.C. (2004). The involvement of two cdc2-related kinases (CRKs) in Trypanosoma brucei cell cycle regulation and the distinctive stage-specific phenotypes caused by CRK3 depletion. J Biol Chem *279*, 20519-20528.

Tu, X., and Wang, C.C. (2005). Pairwise knockdowns of cdc2-related kinases (CRKs) in Trypanosoma brucei identified the CRKs for G1/S and G2/M transitions and demonstrated distinctive cytokinetic regulations between two developmental stages of the organism. Eukaryotic cell *4*, 755-764.

Ullu, E., and Tschudi, C. (1990). Permeable trypanosome cells as a model system for transcription and trans-splicing. Nucleic Acids Res *18*, 3319-3326.

Ullu, E., and Tschudi, C. (1991). Trans splicing in trypanosomes requires methylation of the 5' end of the spliced leader RNA. Proc Natl Acad Sci U S A *88*, 10074-10078.

Urbaniak, M.D., Martin, D.M., and Ferguson, M.A. (2013). Global quantitative SILAC phosphoproteomics reveals differential phosphorylation is widespread between the procyclic and bloodstream form lifecycle stages of Trypanosoma brucei. Journal of proteome research *12*, 2233-2244.

van Alphen, R.J., Wiemer, E.A., Burger, H., and Eskens, F.A. (2009). The spliceosome as target for anticancer treatment. BrJCancer *100*, 228-232.

van der Feltz, C., Anthony, K., Brilot, A., and Pomeranz Krummel, D.A. (2012). Architecture of the spliceosome. Biochemistry *51*, 3321-3333.

Van der Ploeg, L.H., Liu, A.Y., Michels, P.A., De Lange, T., Borst, P., Majumder, H.K., Weber, H., Veeneman, G.H., and Van Boom, J. (1982). RNA splicing is required to make the messenger RNA for a variant surface antigen in trypanosomes. Nucleic Acids Res *10*, 3591-3604.

Van Hellemond, J.J., Neuville, P., Schwarz, R.T., Matthews, K.R., and Mottram, J.C. (2000). Isolation of Trypanosoma brucei CYC2 and CYC3 cyclin genes by rescue of a yeast G(1) cyclin mutant. Functional characterization of CYC2. J Biol Chem *275*, 8315-8323.

Vermeulen, W., Bergmann, E., Auriol, J., Rademakers, S., Frit, P., Appeldoorn, E., Hoeijmakers, J.H., and Egly, J.M. (2000). Sublimiting concentration of TFIIH transcription/DNA repair factor causes TTD-A trichothiodystrophy disorder. Nat Genet *26*, 307-313.

Vidal, V.P., Verdone, L., Mayes, A.E., and Beggs, J.D. (1999). Characterization of U6 snRNA-protein interactions. Rna *5*, 1470-1481.

Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718.

Wang, P., Palfi, Z., Preusser, C., Lucke, S., Lane, W.S., Kambach, C., and Bindereif, A. (2006). Sm core variation in spliceosomal small nuclear ribonucleoproteins from Trypanosoma brucei. The EMBO journal *25*, 4513-4523.

Wang, Q.E., Zhu, Q., Wani, G., Chen, J., and Wani, A.A. (2004). UV radiation-induced XPC translocation within chromatin is mediated by damaged-DNA binding protein, DDB2. Carcinogenesis *25*, 1033-1043.

Wang, X., Zhang, S., Zhang, J., Huang, X., Xu, C., Wang, W., Liu, Z., Wu, J., and Shi, Y. (2010). A large intrinsically disordered region in SKIP and its disorder-order transition induced by PPIL1 binding revealed by NMR. J Biol Chem *285*, 4951-4963.

Watkins, K.P., Dungan, J.M., and Agabian, N. (1994). Identification of a small RNA that interacts with the 5' splice site of the Trypanosoma brucei spliced leader RNA in vivo. Cell *T6*, 171-182.

Weber, A., Chung, H.-J., Springer, E., Heitzmann, D., and Warth, R. (2010). The TFIIH subunit p89 (XPB) localizes to the centrosome during mitosis. Cell Oncol *32*, 121-130.

Welburn, S.C., Macleod, E., Figarella, K., and Duzensko, M. (2006). Programmed cell death in African trypanosomes. Parasitology *132 Suppl*, 7-7.

Will, C.L., and Luhrmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. Current opinion in cell biology *13*, 290-301.

Wirtz, E., Hartmann, C., and Clayton, C. (1994). Gene expression mediated by bacteriophage T3 and T7 RNA polymerases in transgenic trypanosomes. Nucleic Acids Res *22*, 3887-3894.

Wirtz, E., Leal, S., Ochatt, C., and Cross, G.A. (1999). A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in Trypanosoma brucei. Mol Biochem Parasitol *99*, 89-101.

Wu, S., Romfo, C.M., Nilsen, T.W., and Green, M.R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. Nature *402*, 832-835.

Xiang, J., Lahti, J.M., Grenet, J., Easton, J., and Kidd, V.J. (1994). Molecular cloning and expression of alternatively spliced PITSLRE protein kinase isoforms. The Journal of biological chemistry *269*, 15786-15794.

Xu, Y., Ben-Shlomo, H., and Michaeli, S. (1997). The U5 RNA of trypanosomes deviates from the canonical U5 RNA: the Leptomonas collosoma U5 RNA and its coding gene. Proc Natl Acad Sci U S A *94*, 8473-8478.

Yan, Q., Moreland, R.J., Conaway, J.W., and Conaway, R.C. (1999). Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. J Biol Chem *274*, 35668-35675.

Yoon, J.B., Murphy, S., Bai, L., Wang, Z., and Roeder, R.G. (1995). Proximal sequence element-binding transcription factor (PTF) is a multisubunit complex required for transcription of both RNA polymerase II- and RNA polymerase III-dependent small nuclear RNA genes. Molecular and cellular biology *15*, 2019-2027.

Yu, M.C., Sturm, N.R., Saito, R.M., Roberts, T.G., and Campbell, D.A. (1998). Single nucleotide resolution of promoter activity and protein binding for the Leishmania tarentolae spliced leader RNA gene. Mol Biochem Parasitol *94*, 265-281.

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

Zamudio, J.R., Mittra, B., Chattopadhyay, A., Wohlschlegel, J.A., Sturm, N.R., and Campbell, D.A. (2009b). Trypanosoma brucei spliced leader RNA maturation by the cap 1 2'-O-ribose methyltransferase and SLA1 H/ACA snoRNA pseudouridine synthase complex. Mol Cell Biol *29*, 1202-1211.

Zamudio, J.R., Mittra, B., Foldynova-Trantirkova, S., Zeiner, G.M., Lukes, J., Bujnicki, J.M., Sturm, N.R., and Campbell, D.A. (2007). The 2'-O-ribose methyltransferase for cap 1 of spliced leader RNA and U1 small nuclear RNA in Trypanosoma brucei. Mol Cell Biol *27*, 6084-6092.

Zamudio, J.R., Mittra, B., Zeiner, G.M., Feder, M., Bujnicki, J.M., Sturm, N.R., and Campbell, D.A. (2006). Complete cap 4 formation is not required for viability in Trypanosoma brucei. Eukaryot Cell *5*, 905-915.

Zeiner, G.M., Foldynova, S., Sturm, N.R., Lukes, J., and Campbell, D.A. (2004a). SmD1 is required for spliced leader RNA biogenesis. Eukaryot Cell *3*, 241-244.

Zeiner, G.M., Hitchcock, R.A., Sturm, N.R., and Campbell, D.A. (2004b). 3'-End polishing of the kinetoplastid spliced leader RNA is performed by SNIP, a 3'-->5' exonuclease with a Motley assortment of small RNA substrates. Mol Cell Biol *24*, 10390-10396.

Zhou, M., Halanski, M.A., Radonovich, M.F., Kashanchi, F., Peng, J., Price, D.H., and Brady, J.N. (2000). Tat modifies the activity of CDK9 to phosphorylate serine 5 of the RNA polymerase II carboxyl-terminal domain during human immunodeficiency virus type 1 transcription. Mol Cell Biol *20*, 5077-5086.

**Supporting information**

# *Trypanosoma brucei* harbors a divergent XPB helicase paralog that is specialized in nucleotide excision repair and conserved among kinetoplastid organisms

*Nitika Badjatia,[1] Tu N Nguyen, [1] Ju Huck Lee,[1,2] and Arthur Günzl [1,#]*

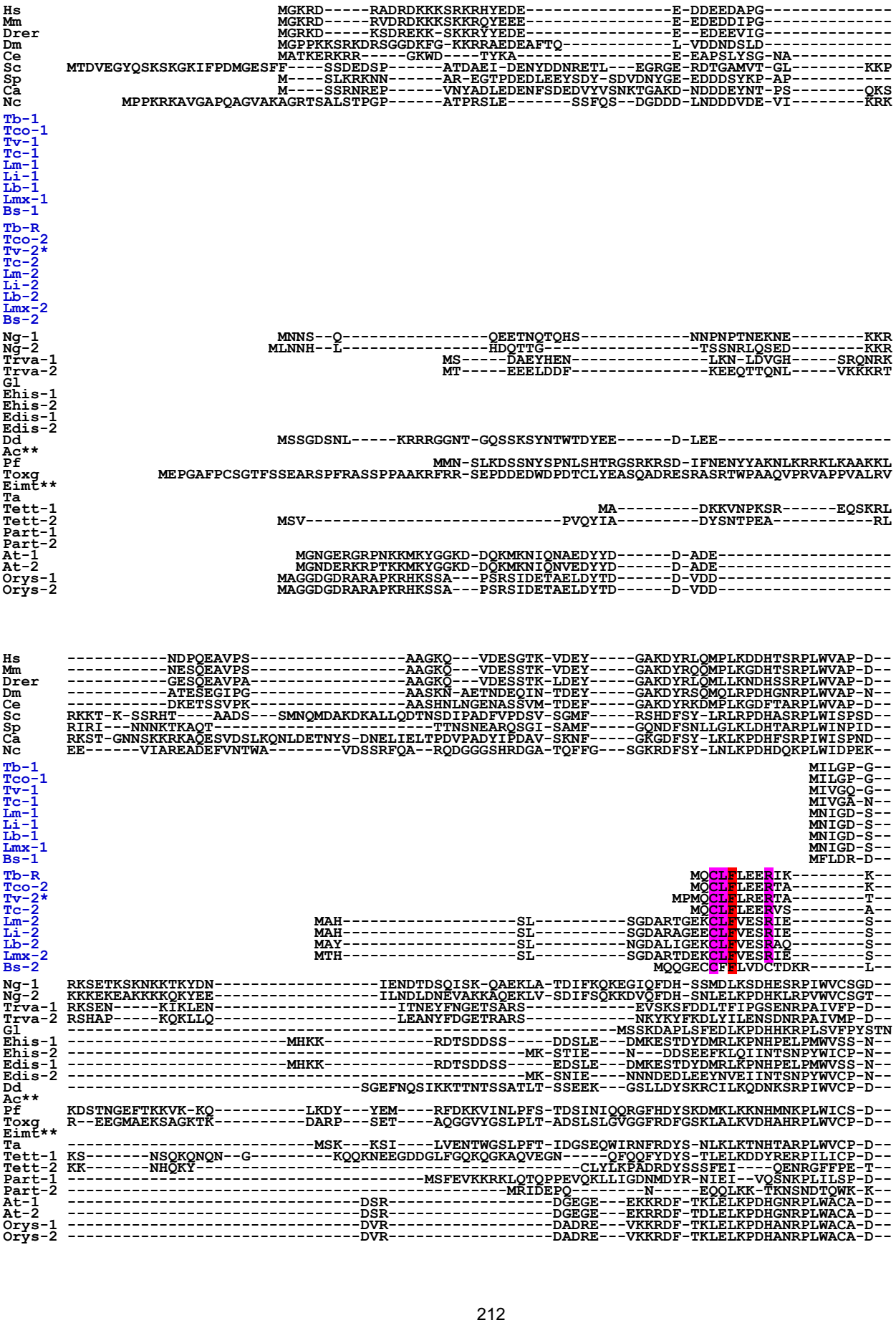
[1] Department of Genetics and Developmental Biology, University of Connecticut Health Center, 400 Farmington Avenue, Farmington, CT 06030-6403, USA

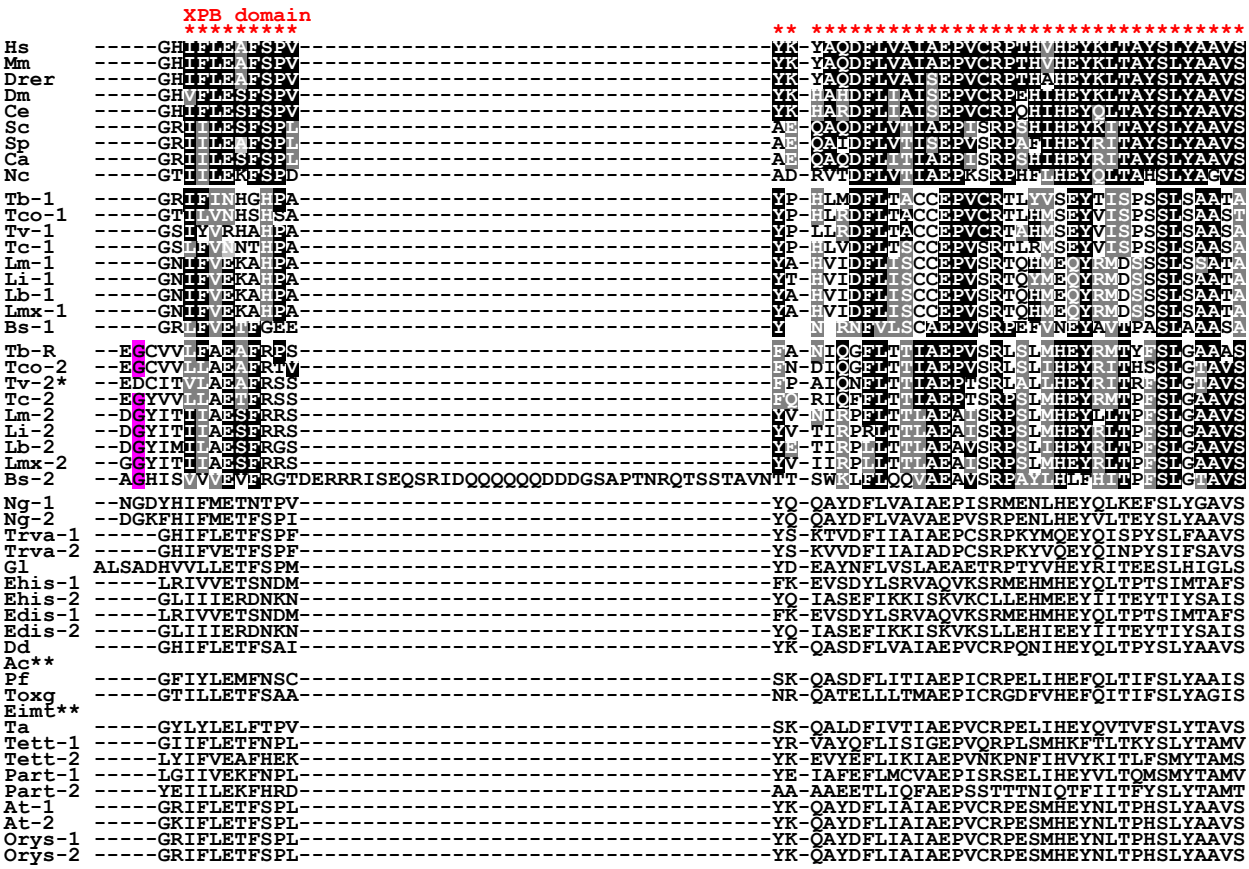
[2] Current address: Department of Microbiology & Molecular Genetics, University of Texas-Houston Medical School, 6431 Fannin Street, Houston, TX 77030, USA

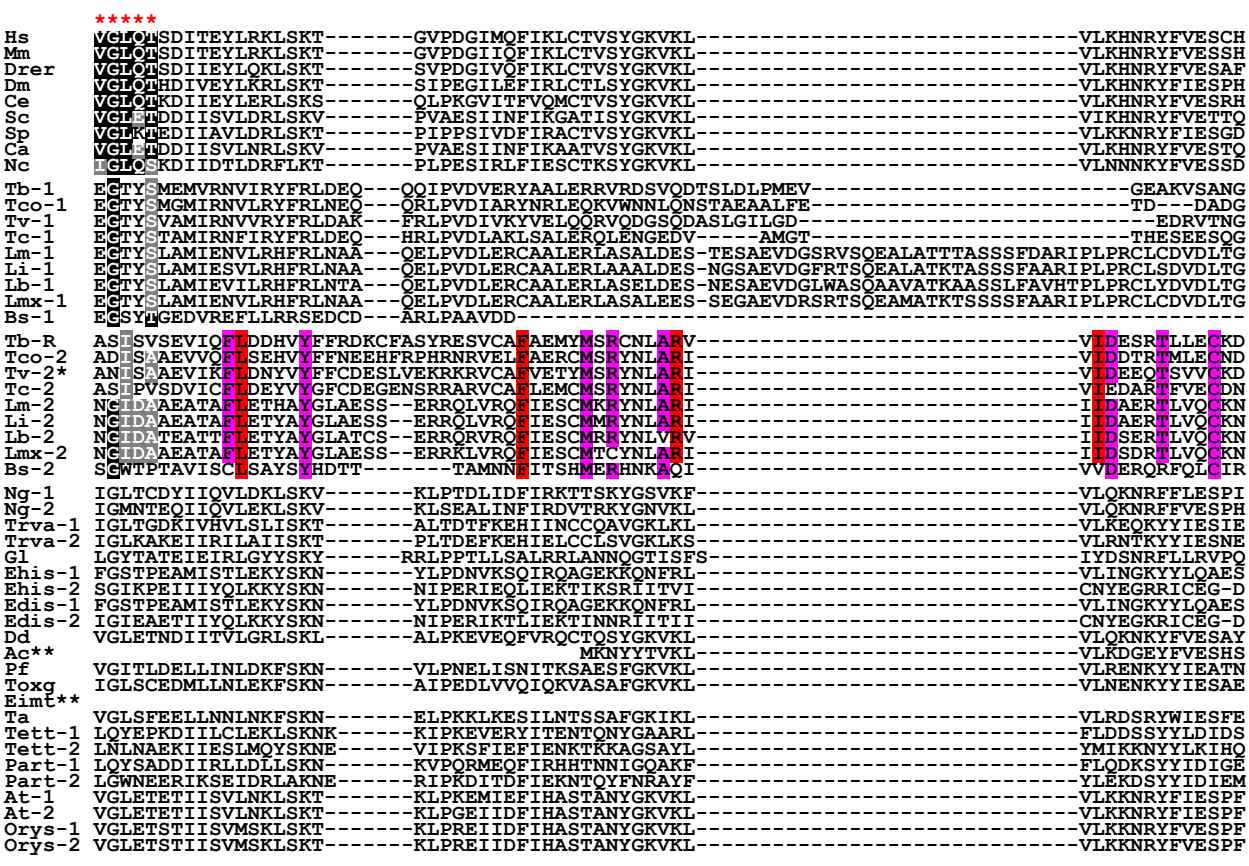
[#] For correspondence, Email: gunzl@uchc.edu; Tel: (860) 679-8878; Fax: (860) 679-8345.

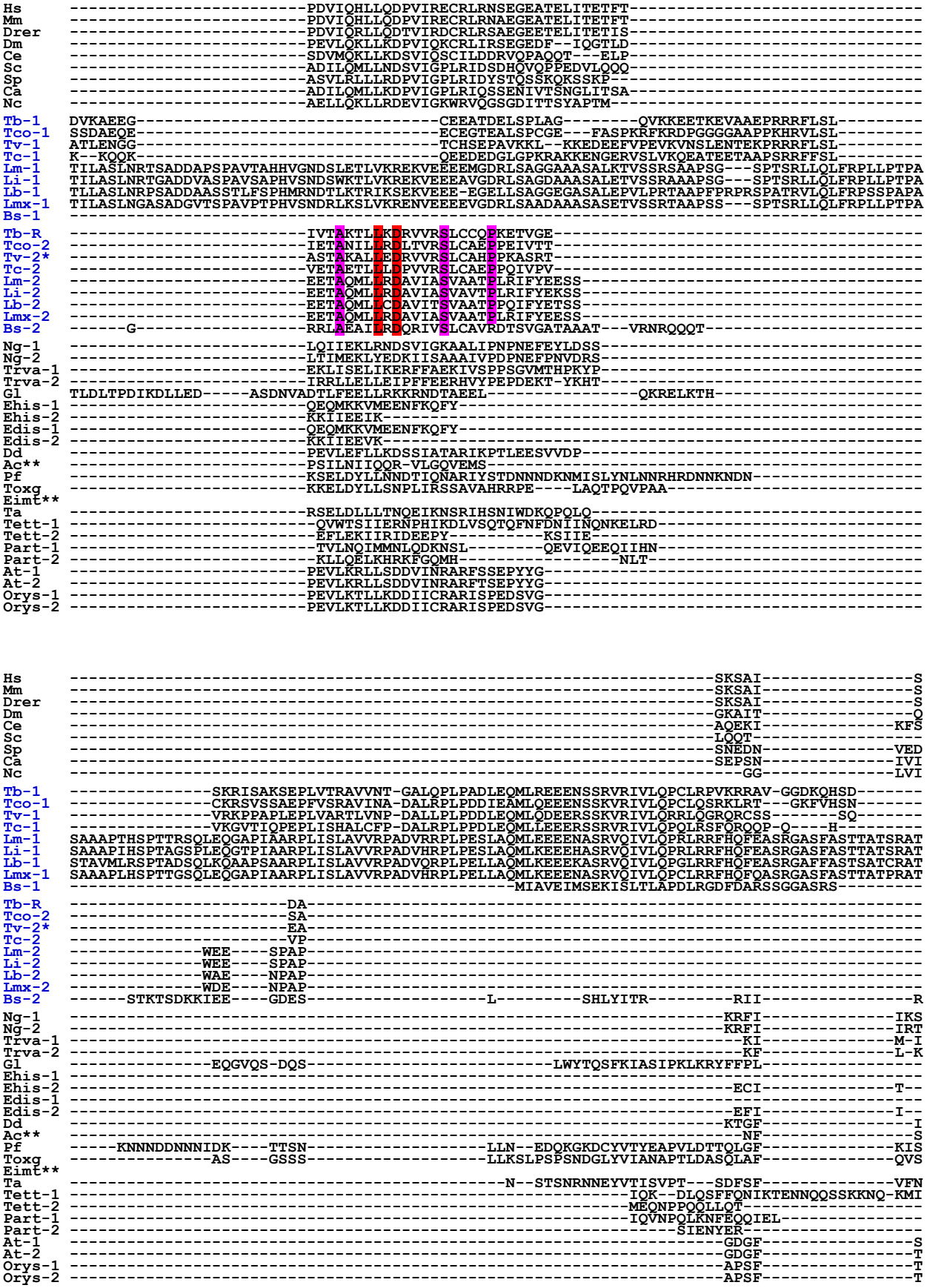
**Content**

CLUSTAL O(1.1.0) multiple sequence alignment



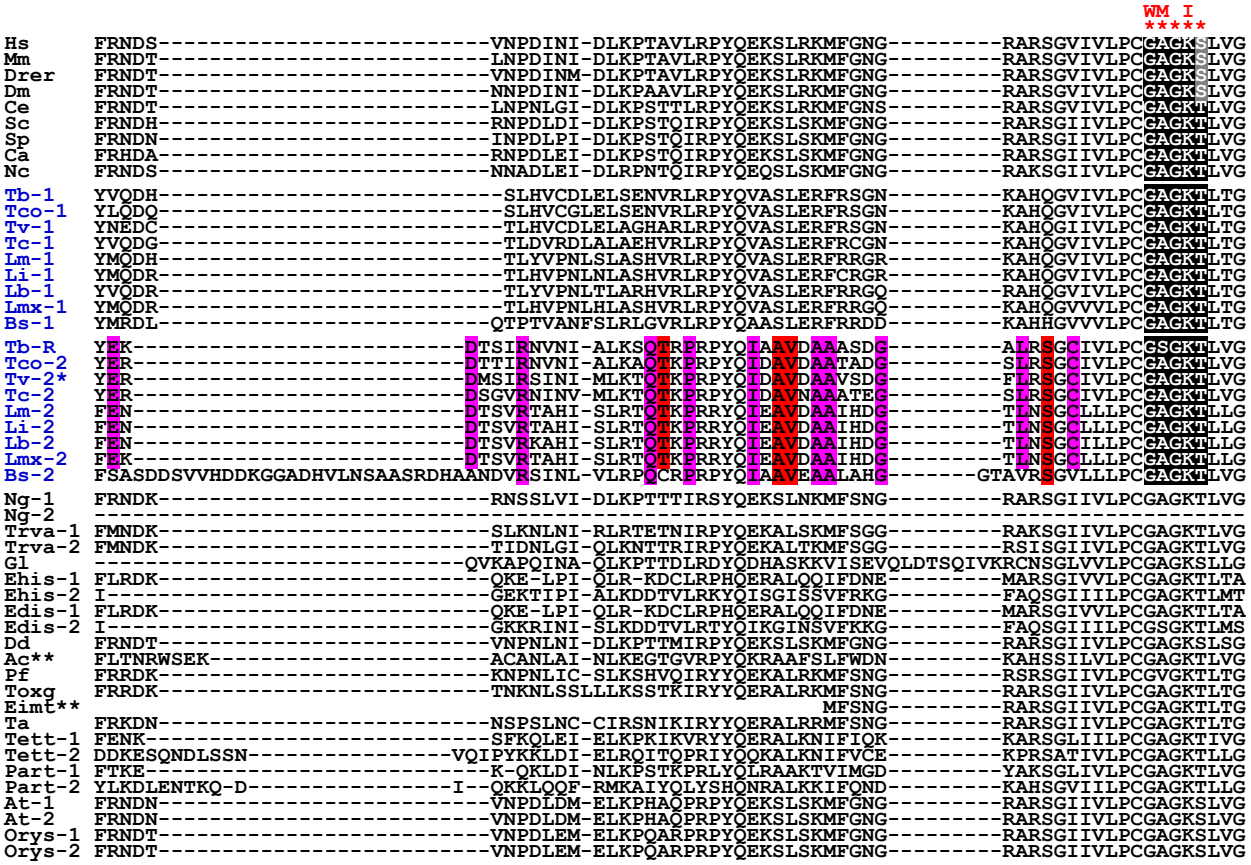

XPB_domain

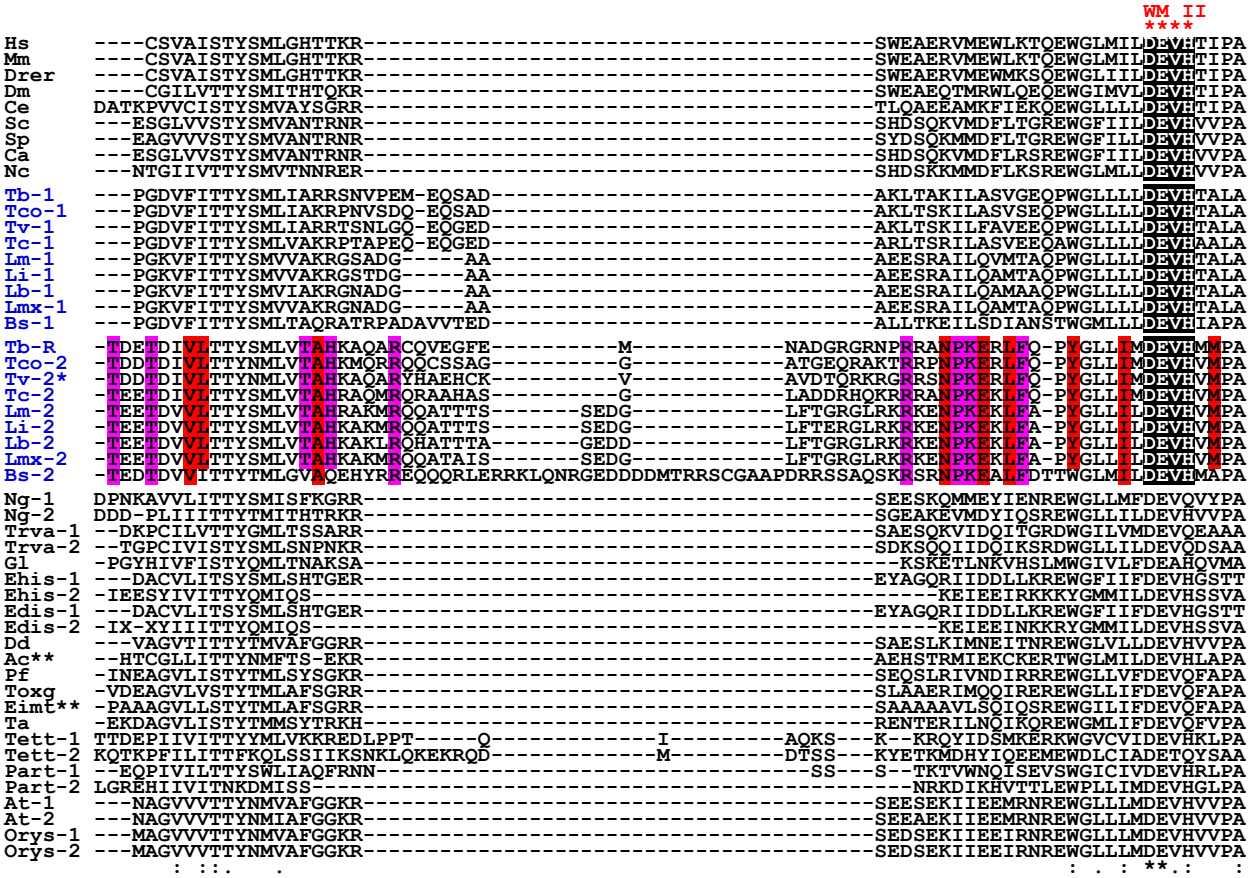


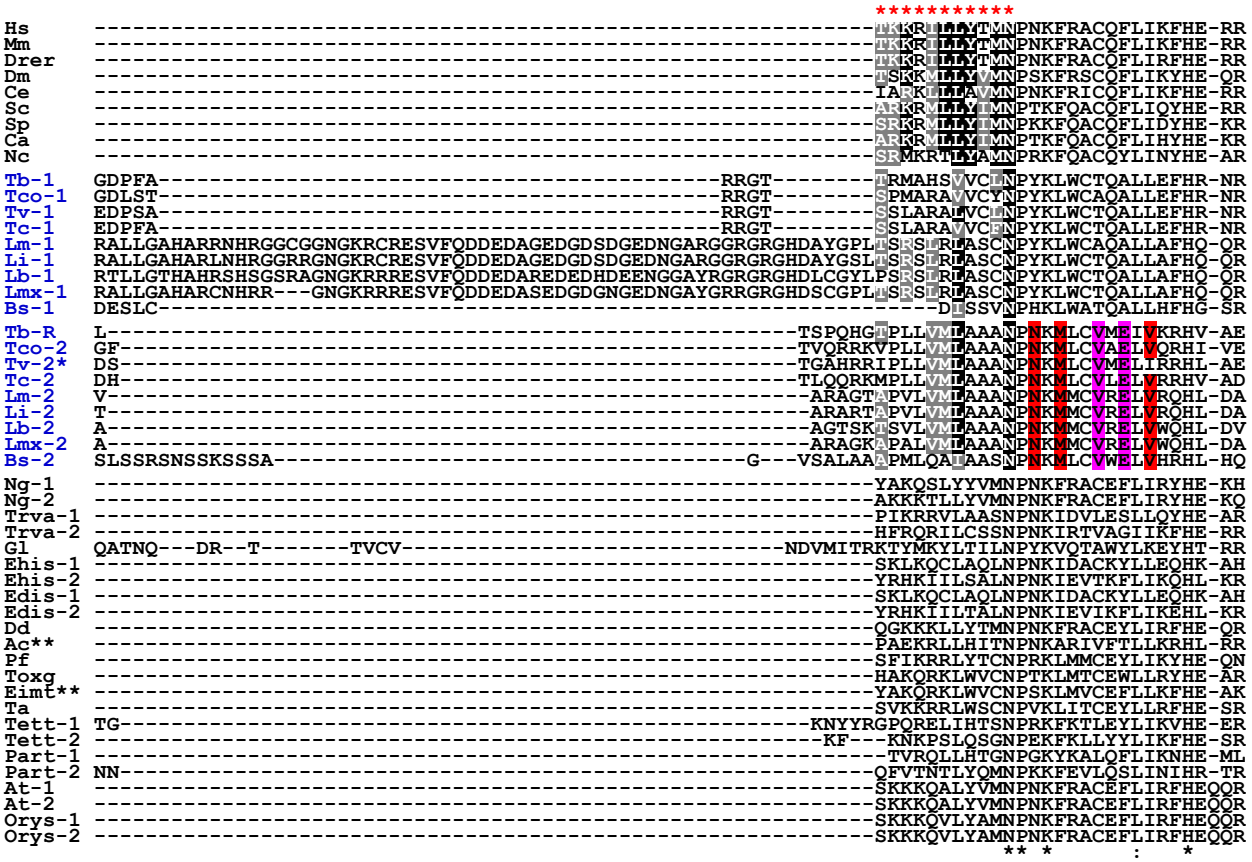


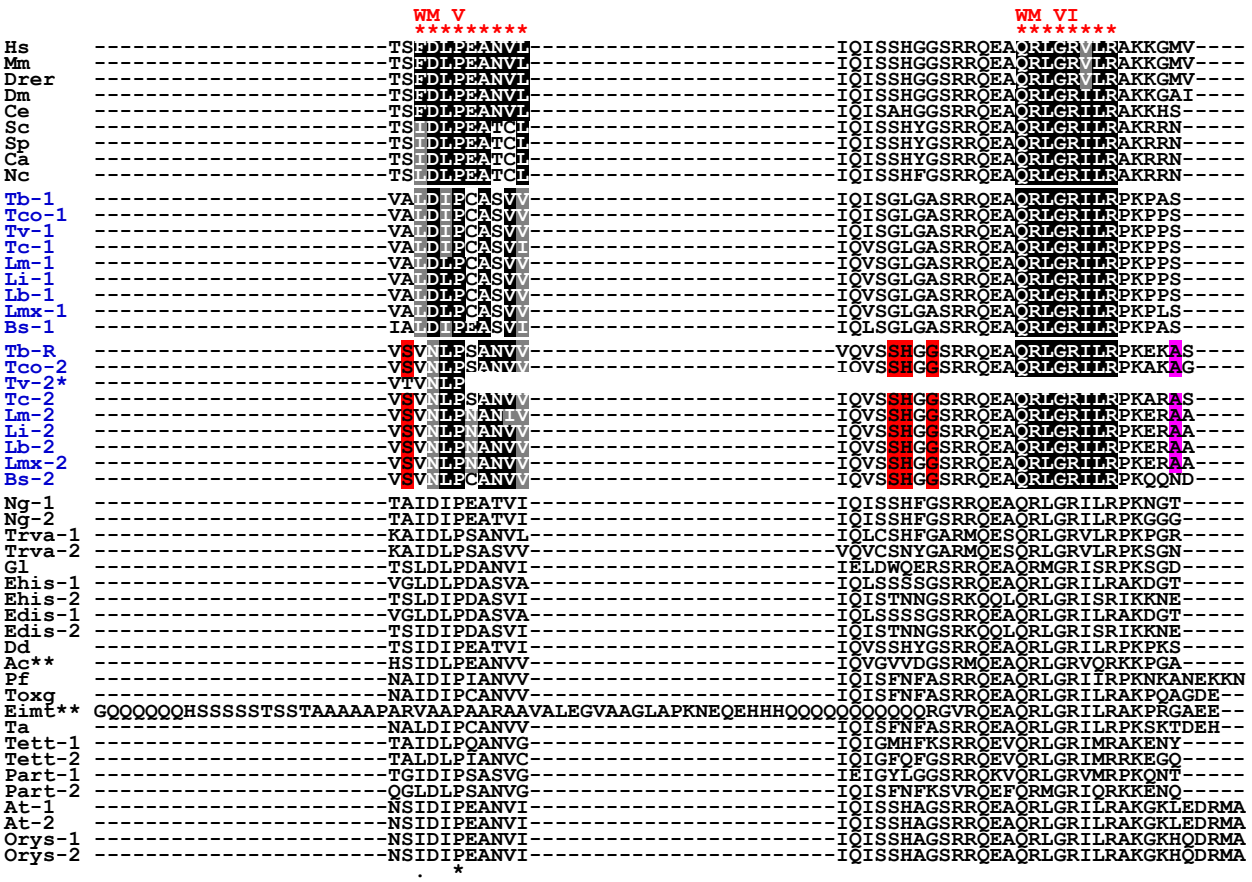
```
Hs      KTA-ESSG-------GP-STSRVT-DPQ----------GKSDIPMDLFDFYE---------QM--------------------------
Mm      KTAAEGSG-------GP-STSQGV-DAQ----------ATSDIPKDLFDFYE---------QM--------------------------
Drer    KSQQDN-G-------GP-SSSQPADGQR----------SGTQVPEDIFSYYE---------QM--------------------------
Dm      FGTKLPPGATD----KPTPDPAAAAGAD----------GTTAVPEDITDFYE---------KI--------------------------
Ce      HGNENEKGVE-----KDGAAADGTAAAA----------TDGKVPADIDEFYG---------KI--------------------------
Sc      ---A-------------G--KPAT----------------NVNPNDVEAVFS---------AVI--------------------------
Sp      KKDI-------------TNDSSKE----------------TAEKSSSDELFS---------AVV--------------------------
Ca      PGTK-------------KNDSTKE----------------DPANNGAEDIFA---------SVV--------------------------
Nc      PGTKDAAGVNQAGLNKSGNKKTAEDG-------------AADAAATNEADLYA---------AI--------------------------

Tb-1    ------EQ----------QCQRAEETKLAYFLTSPDRNHMEHLVSRLQDFLVPVLLHGTRRWVVSDVDRGV---------------EERS
Tco-1   ------FN------QLQRHQYQSDENKLAYFLTSPDRQRMEYLVSHLQEYLVPISLHGAIRWIMSDVDRGV---------------EERS
Tv-1    ------AH------NYQKQNESSEEEALAYFLVSHDRQHMEFLVEHLREYIVPVSLYGALRWVISDIDRGV---------------EERS
Tc-1    --------------HSQGQIHHEDEGRLVYFLVSQDRQQMEFLVSRLQEFLVPVSLHGVMRWVLSDVDRGV---------------EEKS
Lm-1    VSKER-NG------AATGALAERDDQQLFYFVQSRERAHLEHVLPALKEFLEPVVICGAERWIVSDVDRTPLGNTKGIGD------AERL
Li-1    VWKER-NG------AAAGALADRDDQQLLYFVQSRQRAHLEHVLPALKEFLEPVVICGAERWILSDVDRSPLGNTKDIGD------AERL
Lb-1    APKGR-NG------AATGTPADRDDQQLFYFVQSRQRVHLELVLPALKEFLKPVVICGVERWILSDVDRSPLMNTSDIKD------AERL
Lmx-1   VSKER-NG------AATGAVADQDDQQLFYFVQSRQRAHLEHVLPALKEFLEPVVICGAERWILSDIDRSPLGDTKGVSD------AERV
Bs-1    --------------NLFSSGVAPSSGNLAYFLQCSQKPLLEAFANEVKEFLDPYLLGGERAFVISDLVVSD---------------APRS

Tb-R    ----------------------------------------------------------------------------------------
Tco-2   ----------------------------------------------------------------------------------------
Tv-2*   ----------------------------------------------------------------------------------------
Tc-2    ----------------------------------------------------------------------------------------
Lm-2    ----------------------------------------------------------------------------------------
Li-2    ----------------------------------------------------------------------------------------
Lb-2    ----------------------------------------------------------------------------------------
Lmx-2   ----------------------------------------------------------------------------------------
Bs-2    APDEEEDE--------------------------------------------E-----------------------------------SHEMS

Ng-1    KSN--QVEAFQKQTLISTNSNSLSNT-------------MNEQHHVQKEMIG---------AIN--------------------------
Ng-2    RSKQNELEKFQKQALANTS------T-------------VNKGSKIQTEMIR---------AVN--------------------------
Trva-1  VDEKEAD------TIISGV-----------------------GQAAYYG----------SVQ--------------------------
Trva-2  VGKEELE------TIISGI-----------------------GESSYHG---------NVA--------------------------
Gl      ------------------------------------E---ALTSIRKKLKEMRDPLS--GTRSLIT-------------------IEEYL
Ehis-1  EGSAR-----------------------------------------------------------------------------------
Ehis-2  ----------------------------------------------------------------------------------------
Edis-1  EGSAR-----------------------------------------------------------------------------------
Edis-2  ----------------------------------------------------------------------------------------
Dd      INKEVVTGAQISGGLQANQS---------------------LDPVLKNDALS---------NLL--------------------------
Ac**    TTDE--------------------------------------------------------------------------------------
Pf      ESEKQLM---MEE-----------------------------------------------------------------------------
Toxg    ERDEETD---PAGLTGAGKSAEP--G--------------DTAPGAVTAVAG---------NLASPGASGD-GDPSETRKRPREEGRETS
Eimt**
Ta      RNHEAPE---SEE-----------------------------------------------------------------------------
Tett-1  VEDEQDDG-------------------------------------IPDKKN---------KLMKNGDSSI-ANY----------------
Tett-2  ----------------------------------------------------------------------------------------
Part-1  ----------------------------------------------------------------------------------------
Part-2  ----------------------------------------------------------------------------------------
At-1    VGRT-------SGELEAG-----------------------------------PG---------ELL--------------------------
At-2    IGKT-------SGELEAG-----------------------------------PG---------ELL--------------------------
Orys-1  ISKT-------AGEMASG-----------------------------------HE---------DLL--------------------------
Orys-2  ISKT-------AGEMASG-----------------------------------HE---------DLL--------------------------
```

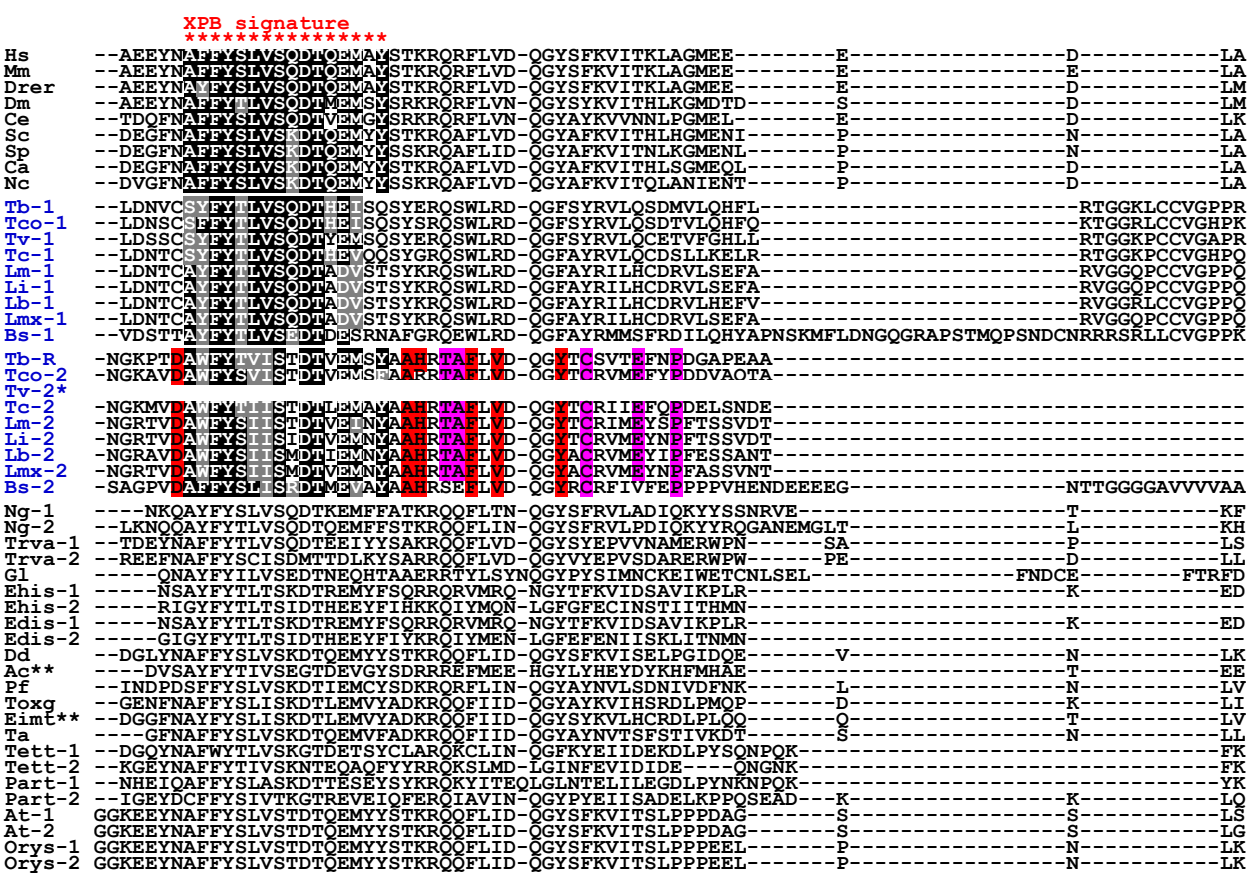
```
Hs      ------------------------------------DKDEEEE-E-----E--TQTVSFEVKQ---EMIEELQ-KRC-IH--LEYPLLAEYD
Mm      ------------------------------------DKDEEEE-E-----E--TQTVSFEVKQ---EMIEELQ-KRC-IC--LEYPLLAEYD
Drer    ------------------------------------DKEEEEE-E-----E--TQTVSFEIRQ---EMIEELQ-KRC-IQ--LEYPLLAEYD
Dm      ------------------------------------DKEEEDE-D-----EANLKTVSFEVAQ---EKIEVIQ-KRC-IE--IEHPLLAEYD
Ce      ------------------------------------DGDDEEDAE-----IRNLQLLTFEIKQ---ETIETVQ-KRC-IE--LEYPLLAEYD
Sc      ---------------------GG-------------DNEREEE-D-----DDIDAVHSFEIAN---ESVEVVK-KRC-QE--IDYPVLEEYD
Sp      ---------------------GL---------------QEE-E-----DDEDAVHLFEIKH---SSVETIK-KRC-AE--IDYPLLEEYD
Ca      ---------------------GN---------------RDD-D-----DDMDTIHSFEIAH---DSVEIVK-RRC-QD--IEYPVLEEYD
Nc      ----------------------------------NEEDD-E-----DDKDAVHAFEIPE---TAVEIVQ-RRC-LD--LGFPILEEYD

Tb-1    TAESGRAKTLRRLFEAPSSASGRSVASKSLT--NEGANGDGLGGGVGRRCTRIVYKSQVMD---GKMRNVRERLY-KE--LSVRADLFYD
Tco-1   TVESGRASVLRWLFDTPSRLPVFQCHGKEGAGATATPYSTGTGGGLG-RCTRIVYKSQVAD---GKMRNVRERLY-KQ--LGVRADLFYD
Tv-1    VVEDGRARTLRILFDTPSGQKSELQS-----------ANSSRMARSEKKYTRIVYKSQVMK---GRLRNVRERLF-KQ--LGVRVDLSYD
Tc-1    VVESGRAKTLRRLFDTPSVSFASAMD-------DVGSCATAAANSTGRKYSRVVYKSQVKD---GRLRNVRERLF-KE--LGVRADLFYD
Lm-1    ARDAGRPAVLRLLYEPPRVGTQYEKA-------------------EATPSASRFLYKCQVRD---GKLREVKECLF-SK--FDIRADCYYD
Li-1    ARDAGRPAVLRLLYESPRAGTRHEKA-------------------AATPSASRFLYKCQVRD---GKLREVKECLF-SK--FDIRADCYYD
Lb-1    TRDAGRPAVLRLLYEFPHVGTVSERT-------------------TTASTASRFVYKCQVRD---GKLREVKERLF-SK--FGIRADCYYD
Lmx-1   ARDAGRPAVLRLLYEPPRVGTRYEAA-------------------AATPSASCFVYKCRVRD---GKLREVKECLF-SK--FDIRADCYYD
Bs-1    ---SGSL--L-------------------------RERRLNGSSGPVTKQMVYKCRIRA---GCQRQIKEALY-QR---SIRVICTYD

Tb-R    ----------------------------------------------------CPHFLLKSR--AAARVVAER-C-VL--LGYPLQQQYE
Tco-2   ----------------------------------------------------RPYSLLRNR--AVTKVVAER-C-AA--LGYPLQQQYD
Tv-2*   ----------------------------------------------------CSF-LLRSR--AAARPVAER-C-VA--LGFPLQQQYD
Tc-2    ----------------------------------------------------HSLLVLRSR--AVARRVAER-C-VL--LGFPLQQQYD
Lm-2    ----------------------------------------------------YPSFLLQSR--AMSKVVAAQ-C-VV--LGLPIQQQYD
Li-2    ----------------------------------------------------YPSFLLQSR--AMSKVVAAQ-C-VV--LGLPIQQQYD
Lb-2    ----------------------------------------------------YPSFLLQSR--AMSKVVAAR-C-MV--LGFPIQQQYD
Lmx-2   ----------------------------------------------------YPSFLLQSR--VMSKVVAAQ-C-VV--LGLPIQQQYD
Bs-2    --SEAKNAAI-----------VA-----------DKEEHDDDG-------ELEEVQVLLQDH--AASRLVSDVLW-AH--HHLPICQVYH

Ng-1    ----------------------HS-----------NYEEEYEDDNDLDSEPHETLYSFEISS---KSIESVK-KQC-IV--IGHPTLDEYD
Ng-2    ----------------------QI-----------GDDEDDDADDDFDIS---------------------------------------
Trva-1  ----------------------RL-------------TEGFD-EIFE--VEKPKIYRFELKP---ETVRDVR-KHA-VD--NNLFLSDEYD
Trva-2  ----------------------RL-------------TEGID-DLLDTPLEKQTILRFQIKT---ESVREIR-QYA-VD--HNLFISDEYD
Gl      LLDEGIDAVR------------------------NEQVQRLQSLMAAEINTEAEMCEL--------KKIKEE-----------FAKAY-
Ehis-1  ------------------------------------------EVKESDELYRELDLDHHVFIIEVKQ---TSVFKLK-KKC-KK--KKVRVYEEYH
Ehis-2  ----------------------------------------------------INKNEYELK----SNIIVFK-KKC-QE--KGILINEKIE
Edis-1  ------------------------------------------EVKESDELYRELDLDHHVFIIEVKQ---TSVFKLK-KKC-KK--KKVRVYEEYH
Edis-2  ----------------------------------------------------INKKEYEIK----NNIIEFK-KKC-QE--KGILINEKIE
Dd      ------------------------------------EEEEEDTVN-----NSDQHFHSFEIDP---QQVEEVK-KRC-IQ--LDYPVLEEYD
Ac**    ----------------------------------------------ETGAQVLVYALPISTVQEVAEIK-KLL-LA--HNHPFIEEFD
Pf      ----------------------------------KKHANLNANENSTNSAEVYSFEVNC---DKIEEVK-QEA-LQ-TMQRPLLMEYD
Toxg    GAEEARGKAQVPL-ASPHASASP-----------TGEDGDSAKRKTMSAPTRQVFSFQVSS---DRVEEVK-RVS-VE-SMHRPLLNEYD
Eimt**
Ta      ------------------------------------TTTAKFAGRSDKTTTRQVFSFEVQQ---EKIEDLK-REA-LQ-TMRRPLVMEYD
Tett-1  --D------------DEDEEDSD-------------DGDDDFVYNFIMNNKTKRSFKRIQI-I---GDHFEVT-KAV-IN--CSVPLIQEYD
Tett-2  ----------------------Q-------------E-------ESRVWNFKFYRIKQ-K---HNLFQIT-EKI-IK--EGISLINEYH
Part-1  --K------------EADIEQMA-----------KEIDTFKDWLAQDKIQTIKFKRFRI-V---GDYFDVA-QAL-IR--SSVPLIQEYD
Part-2  ----------------QD-------------------------SLLNDN---KTFVKFKIKA---NEHFSVN-RELVED--FNVPVKQEFD
At-1    ------------------------------------NEAEFAAA-----AEEKETHSFEIDP---AQVENVK-QRC-LPNALNYPMLEEYD
At-2    ------------------------------------NEAELAAA-----AEEKETHSFEIDP---ALVENVK-QRC-LPNALNYPMLEEYD
Orys-1  ------------------------------------DGMELAAA-----TEDKETHSFEIDP---SQVENVK-QRC-LPNALNFPMLEEYD
Orys-2  ------------------------------------DGMELAAA-----TEDKETHSFEIDP---SQVENVK-QRC-LPNALNFPMLEEYD
```

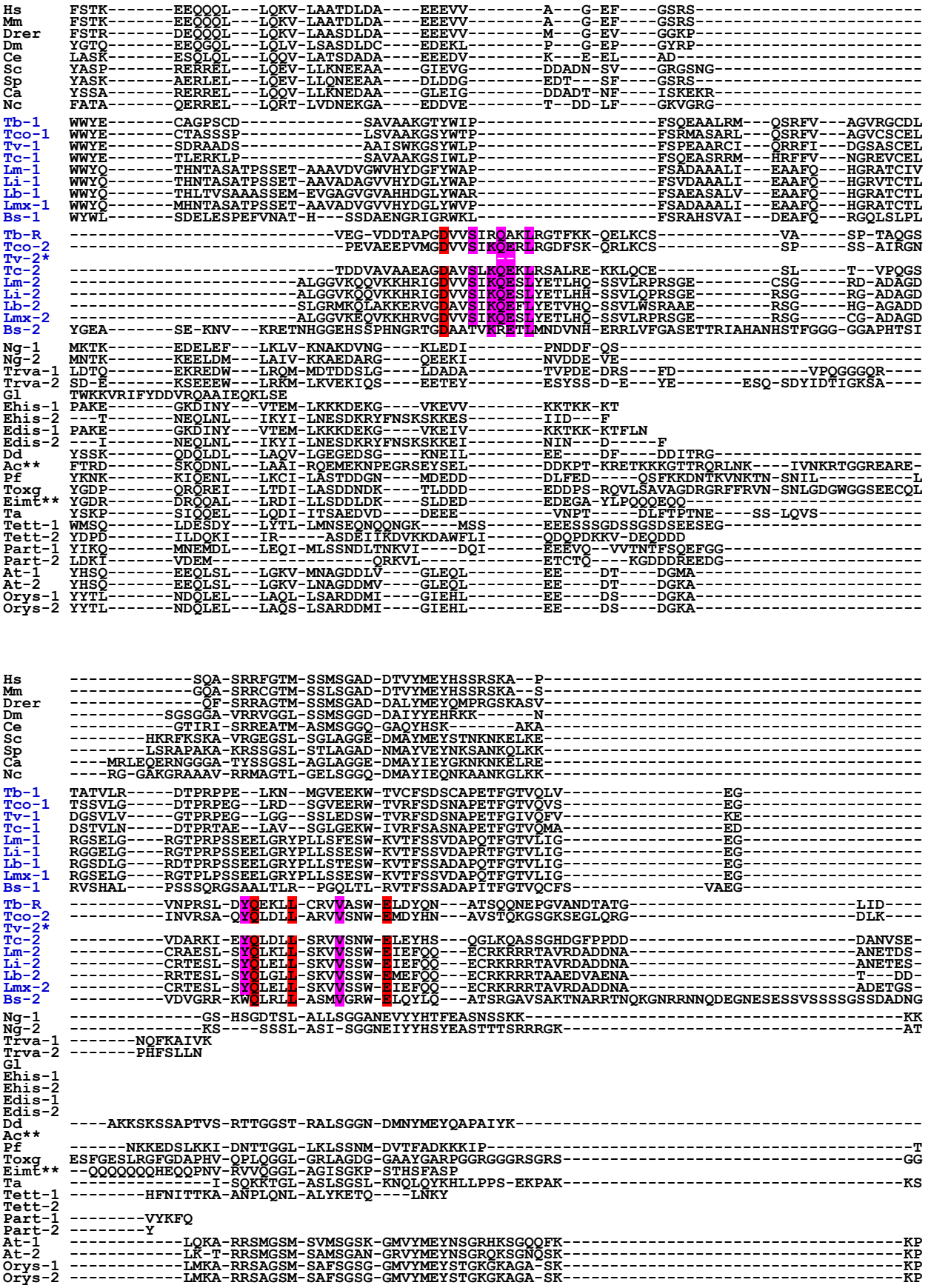

WM I
*****
Hs FRNDS----------VNPDINI-DLKPTAVLRPYQEKSLRKMFGNG---------RARSGVIVLPCGAGKSLVG
Mm FRNDT----------LNPDINI-DLKPTAVLRPYQEKSLRKMFGNG---------RARSGVIVLPCGAGKSLVG
Drer FRNDT----------VNPDINM-DLKPTAVLRPYQEKSLRKMFGNG---------RARSGVIVLPCGAGKSLVG
Dm FRNDT----------NNPDINI-DLKPAAVLRPYQEKSLRKMFGNG---------RARSGVIVLPCGAGKSLVG
Ce FRNDT----------LNPNLGI-DLKPSTTLRPYQEKSLRKMFGNS---------RARSGVIVLPCGAGKTLVG
Sc FRNDH----------RNPDLDI-DLKPSTQIRPYQEKSLSKMFGNG---------RARSGIIVLPCGAGKTLVG
Sp FRNDN----------INPDLPI-DLKPSTQIRPYQEKSLSKMFGNG---------RARSGIIVLPCGAGKTLVG
Ca FRHDA----------RNPDLEI-DLKPSTQIRPYQEKSLSKMFGNG---------RARSGIIVLPCGAGKTLVG
Nc FRNDS----------NNADLEI-DLRPNTQIRPYQEQSLSKMFGNG---------RAKSGIIVLPCGAGKTLVG
Tb-1 YVQDH----------SLHVCDLELSENVRLRPYQVASLERFRSGN---------KAHQGVIVLPCGAGKTLTG
Tco-1 YLQDQ----------SLHVCGLELSENVRLRPYQVASLERFRSGN---------KAHQGVIVLPCGAGKTLTG
Tv-1 YNEDC----------TLHVCDLELAGHARLRPYQVASLERFRSGN---------KAHQGIIVLPCGAGKTLTG
Tc-1 YVQDG----------TLDVRDLALAEHVRLRPYQVASLERFRCGN---------KAHQGVIVLPCGAGKTLTG
Lm-1 YMQDH----------TLYVPNLSLASHVRLRPYQVASLERFRRGR---------KAHQGVIVLPCGAGKTLTG
Li-1 YMQDR----------TLHVPNLNLASHVRLRPYQVASLERFCRGR---------KAHQGVIVLPCGAGKTLTG
Lb-1 YVQDR----------TLYVPNLTLARHVRLRPYQVASLERFRRGQ---------RAHQGVIVLPCGAGKTLTG
Lmx-1 YMQDR----------TLHVPNLHLASHVRLRPYQVASLERFRRGQ---------KAHQGVVVLPCGAGKTLTG
Bs-1 YMRDL----------QTPTVANFSLRLGVRLRPYQAASLERFRRDD---------KAHHGVVVLPCGAGKTLTG
Tb-R YEK----------DTSIRNVNI-ALKSQTRPRPYQIAAVDAAASDG---------ALRSGCIVLPCGSGKTLVG
Tco-2 YER----------DTTIRNVNI-ALKAQTKPRPYQIDAVDAATADG---------SLRSGCIVLPCGAGKTLVG
Tv-2* YER----------DMSIRSINI-MLKTQTKPRPYQIDAVDAAVSDG---------FLRSGCIVLPCGAGKTLVG
Tc-2 YER----------DSGVRNINV-MLKTQTKPRPYQIDAVNAAATEG---------SLRSGCIVLPCGAGKTLVG
Lm-2 FEN----------DTSVRTAHI-SLRTQTKPRRYQIEAVDAAIHDG---------TLNSGCLLLPCGAGKTLLG
Li-2 FEN----------DTSVRTAHI-SLRTQTKPRRYQIEAVDAAIHDG---------TLNSGCLLLPCGAGKTLLG
Lb-2 FEN----------DTSVRKAHI-SLRTQTKPRRYQIEAVDAAIHDG---------TLNSGCLLLPCGAGKTLLG
Lmx-2 FEK----------DTSVRTAHI-SLRTQTKPRRYQIEAVDAAIHDG---------TLNSGCLLLPCGAGKTLLG
Bs-2 FSASDDSVVHDDKGGADHVLNSAASRDHAANDVRSINL-VLRPQCRPRPYQIAAVEAALAHG-------GTAVRSGVLLLPCGAGKTLVG
Ng-1 FRNDK----------RNSSLVI-DLKPTTTIRSYQEKSLNKMFSNG---------RARSGIIVLPCGAGKTLVG
Ng-2 ----------
Trva-1 FMNDK----------SLKNLNI-RLRTETNIRPYQEKALSKMFSGG---------RAKSGIIVLPCGAGKTLVG
Trva-2 FMNDK----------TIDNLGI-QLKNTTRIRPYQEKALTKMFSGG---------RSISGIIVLPCGAGKTLVG
Gl ----------QVKAPQINA-QLKPTTDLRDYQDHASKKVISEVQLDTSQIVKRCNSGLVVLPCGAGKSLLG
Ehis-1 FLRDK----------QKE-LPI-QLR-KDCLRPHQERALQQIFDNE---------MARSGIVVLPCGAGKTLTA
Ehis-2 I----------GEKTIPI-ALKDDTVLRKYQISGISSVFRKG---------FAQSGIIILPCGAGKTLMT
Edis-1 FLRDK----------QKE-LPI-QLR-KDCLRPHQERALQQIFDNE---------MARSGIVVLPCGAGKTLTA
Edis-2 I----------GKKRINI-SLKDDTVLRTYQIKGINSVFKKG---------FAQSGIIILPCGSGKTLMS
Dd FRNDT----------VNPNLNI-DLKPTTMIRPYQEKSLSKMFGNG---------RARSGIIVLPCGAGKSLSG
Ac** FLTNRWSEK----------ACANLAI-NLKEGTGVRPYQKRAAFSLFWDN---------KAHSSILVLPCGAGKTLVG
Pf FRRDK----------KNPNLIC-SLKSHVQIRYYQEKALRKMFSNG---------RSRSGIIVLPCGVGKTLTG
Toxg FRRDK----------TNKNLSSLLLKSSTKIRYYQERALRKMFSNG---------RARSGIIVLPCGAGKTLTG
Eimt** MFSNG---------RARSGIIVLPCGAGKTLTG
Ta FRKDN----------NSPSLNC-CIRSNIKIRYYQERALRRMFSNG---------RARSGIIVLPCGAGKTLTG
Tett-1 FENK----------SFKQLEI-ELKPKIKVRYYQERALKNIFIQK---------KARSGLIILPCGAGKTIVG
Tett-2 DDKESQNDLSSN----------VQIPYKKLDI-ELRQITQPRIYQQKALKNIFVCE---------KPRSATIVLPCGAGKTLLG
Part-1 FTKE----------K-QKLDI-NLKPSTKPRLYQLRAAKTVIMGD---------YAKSGLIVLPCGAGKTLVG
Part-2 YLKDLENTKQ-D----------I--QKKLQQF-RMKAIYQLYSHQNRALKKIFQND---------KAHSGVIILPCGAGKTLLG
At-1 FRNDN----------VNPDLDM-ELKPHAQPRPYQEKSLSKMFGNG---------RARSGIIVLPCGAGKSLVG
At-2 FRNDN----------VNPDLDM-ELKPHAQPRPYQEKSLSKMFGNG---------RARSGIIVLPCGAGKSLVG
Orys-1 FRNDT----------VNPDLEM-ELKPQARPRPYQEKSLSKMFGNG---------RARSGIIVLPCGAGKSLVG
Orys-2 FRNDT----------VNPDLEM-ELKPQARPRPYQEKSLSKMFGNG---------RARSGIIVLPCGAGKSLVG
WM IA
****
Hs VTAACTVRKRCLVLGNSAVSVEQWKAQFKMWSTIDD----------SQICRFTSDAKDK-PIG--
Mm VTAACTVRKRCLVLGNSAVSVEQWKAQFKMWSTIDD----------SQICRFTSDAKDK-PIG--
Drer VTAACTVRKRCLVLGNSSVSVEQWKAQFKMWSTIDD----------SQICRFTSDAKDK-PIG--
Dm VTACCTVRKRALVLCNSGVSVEQWKQQFKMWSTADD----------SMICRFTSEAKDK-PMG--
Ce VTAVTTVNKRCLVLANSNVSVEQWRAQFKLWSTIQD----------KQLVRFTREAKDP-APSGA
Sc ITAACTIKKSVIVLCTSSVSVMQWRQQFLQWCTLQP----------ENCAVFTSDNKEM-FQT--
Sp ITAACTIKKSVIVLCTSSVSVMQWRQQFLQWSNIKP----------DHIAVFTADHKER-FHS--
Ca ITAACTIRKSVIVLCTSSVSVMQWRQQFLQWCTIQP----------ENVAVFTSENKEM-FAS--
Nc ITAACTIKKGVIVLCTSSMSVVQWRQEFLKWSNINP----------DDIAVFTAESKNK-FSG--
Tb-1 IGAAATVKKRTIVMCINVMSVLQWQREFIRWTNLSE----------DQVTVCIADKKQM------
Tco-1 IGAAATVKKRTIVMCINVMSVLQWQREFLRWTNLTE----------EQVTVCISDKKQM------
Tv-1 IGAATIVKKRTIVMCINVMSVLQWQQEFVRWTNLTE----------DQVTVCISDKKQM------
Tc-1 IGAATILKKRTIVMCINVISVLQWQREFFRWTDLSE----------DQVTVCISDKKQM------
Lm-1 IGAAATMQTTTIVMCINHMSVFQWQREFLRWTNLAE----------EEVTVCTAKVKQR------
Li-1 IGAAATMQTTTIVMCINNMSVFQWQREFLRWTDLAE----------DEVTVCTAKVKQR------
Lb-1 IGAAATMQTTTIVMCINNMSVFQWKREFLRWTDLTE----------DEVTVCTAKVKQR------
Lmx-1 IGAAATMQTTTIVMCINNMSVFQWQREFLRWTDLAE----------DEVTVCTAKVKQR------
Bs-1 IAASSIVSKKTIVMCCNVLSVMQWRKEFLRWTDLRE----------DQVTVCVRGMKQR------
Tb-R IMLLCKVKKPTLILCAGGVSVEQWRNQILEFASVCAPANNEEGDPSNS-----T-----TGEKVRTAAVGAARISCLTGKQKDE-I----
Tco-2 VMLLCKVKKPTLILCTGGISVEQWRNQILEFASLSAPSGGEEGSSSTAC----A-----TSKRARPPAVGAARISCLTGKQKDE-I----
Tv-2* IMLLCKVKKPTLILCAGSVSVEQWKNQILEFSTLGCSLGNDENRSSSGS---GT-----ASGRQRTRIEGVARISCLTAKQKDD-I----
Tc-2 IMLLCKVKKPTLILCAGSVSVEQWKNQILEFASICAHDTDDEM----------A-----AAEKHRSRLEGAARIACLTGKQKDP-V----
Lm-2 IMLMCKVKKPTLVLCAGSVSVEQWKSQILDYATLETPPRTDDGDASDAH----------GGRPHQIKDGAARIACLTGKQKDP-I----
Li-2 IMLMCKVKKPTLVVCAGSVSVEQWKSQILDYATLEAPPPTDDGDAANAH----------SGRSHQIKNGAARIACLTGKQKDP-I----
Lb-2 IMLMCKVKKPTLVVCAGSVSVEQWRSQILDYATLEAPPPTDNDDAADTQ----------GGCSHQIKSGATRIACLTGKQKDP-I----
Lmx-2 IMLMCKVKKPTLVVCAGGVSVEQWKSQILDYAVLEAPPPTDDGDAADAH----------GGRPHQIKNGAARIACLTGKQKDP-I----
Bs-2 VLLAAVVKKRTIIVCAGSISVEQWKNQLLEWADLTPLDEEDTATTHNIAARGGRGGGRGGSRGGQGRAVGGDRVACLTGKHKDP-I----
Ng-1 VTAASTIKKHTLVLCINTVSVFQWKNQFRLWSTIDD----------DSILVHTSQQKAT-LPK--
Ng-2 -----SAVETFYLFEINPTECIEWKNQFKLWSTIDE----------NRVVVFTSQEKDA-LPK--
Trva-1 ITAVATINKPAVVVCNSVEPVKQWANQFNLWTNVPG----------GVVVVLTSENKQA-LP---
Trva-2 IAALATINKPTVIVCNNRLTVKQWYNQILQYATMDM----------KKIFLFSDTEKQA-LPQ--
Gl VACACRLGHSCIVVTNGNLSSKQWKSQFLQFSTVES----------DRVYIFSANKDEETEFI--
Ehis-1 IAACSKIKRSTIVLTHTTQSVFQWKEEFLKWSTIKP----------DAIKLCVSSKKEQ-LGD--
Ehis-2 IGIICTFQQPTIIICTTTISIEQWRDEIRKWSTIPL----------NKIKCFSSIIKET------
Edis-1 IAACSKIKRSTIVLTHTTQSVFQWKEEFLKWSTIKP----------DAIKLCVSSKKEQ-LGD--
Edis-2 IGIICKFQEPTIIICTTTISIEQWRDEIRKWTNIPL----------NKIKCFSSIIKEE------
Dd ITAACTVKKSILVLCTSAVSVEQWKYQFKLWSNIEE----------RQISKFTSDNKEK-ISN--
Ac** ITVSSIIKKCILVFCTSMMAVNQWREQFLKWSDIEP----------SRISRFISDTRKKGEWD--
Pf ITAASTIKKSALFLTTSAVAVEQWKKQFEDFTNIHP----------RHIRILTSDYKLD-LWP--
Toxg ITAACTLRKSVMILTTSAVAVDQWRKQFEEYTTIDS----------SRLLTLTAETKQT-LWP--
Eimt** ITAVCTLQRSAVVLTTSAVAVEQWKRQFLENSSIKE----------KEIMLLTAESKQQ-LWG--
Ta IVAACTVRKSIFVLTTSAVAVEQWIKQFLDFTNISR----------TNLISLTSEHKTD-LWD--
Tett-1 VIAIERIKQSTVIICDSDVSVDQWRDELERWTTINK----------NKIVRLTGRIVDE--WPTR
Tett-2 IIVTEKIKGNTIVICDIDTATKQWKNEFIRWTTVKP----------DRIKIYTGKTKNL--QE--
Part-1 VLCMSLIKSSTVIICDSNVSVEQWKREIEGYTTIDN----------RRIIRVTGFAKDK--WQG-
Part-2 INVALKIKRKTLIICDQVNDVYQWQKSFIKFTEMDK----------NNLAIILRTQQEV--PMAV
At-1 VSAAARIKKSCLCLATNAVSVDQWAFQFKLWSTIRD----------DQICRFTSDSKER-FRG--
At-2 VSAAARIKKSCLCLATNAVSVDQWAYQFKLWSTIKD----------DQICRFTSDSKER-FRG--
Orys-1 VSAACRIKKSCLCLATNAVSVDQWAFQFKLWSTIKD----------EHISRFTSDNKEK-FRG--
Orys-2 VSAACRIKKSCLCLATNAVSVDQWAFQFKLWSTIKD----------EHISRFTSDNKEK-FRG--


WM II
WM III
RED
ThM domain




WM V
WM VI


XPB signature



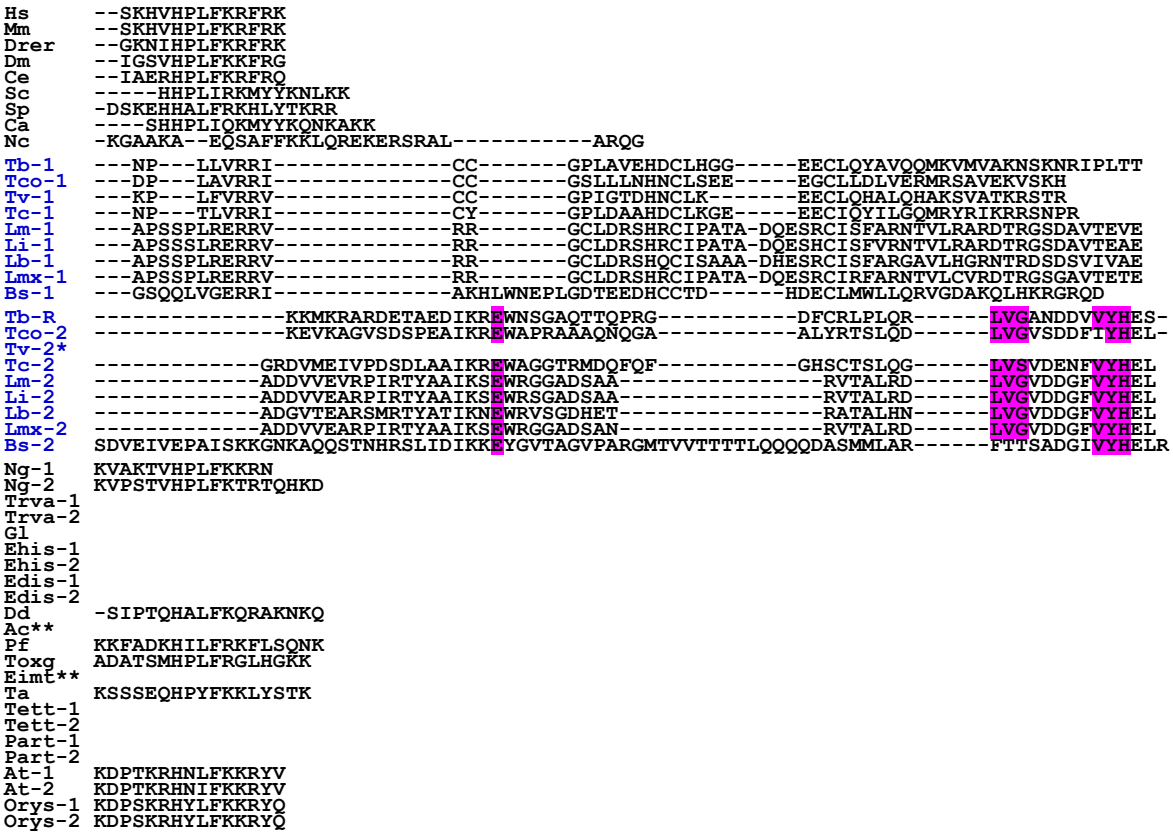



**Fig. S1.** Multiple sequence alignment of XPB amino acid sequences

The encoded protein sequences of specified *XPB* genes were aligned using the Clustal Omega server of the European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/services/web/toolform.ebi?tool=clustalo) at default parameters (Sievers *et al.*, 2011).

The XPB domain, Walker motifs (WM I, WM Ia, WM II-VI), RED residue loop, thumb-like domain (Thm) and XPB signature motif were highlighted in the members of the Opisthokonta supergroup, comprising the eukaryotic model organisms for TFIIH research (top nine sequences). If a position within the Opisthokonta was more than 50% identical or similar, the residue was highlighted in black or gray, respectively, and the highlighting was then applied to the kinetoplastid sequences (blue lettering) to demonstrate the conservation of these domains in the latter.

To detect putative DNA repair-specific sequence conservation in kinetoplastid XPB-R sequences, positions of XPB-R-specific identity (one mismatch allowed) that was not present in any of the kinetoplastid XPB sequences was highlighted in purple. If this position was also conserved in the Opisthokonta model organisms (more than 50% similarity) the identities were highlighted in red.

If a species contained two *XPB* sequences, the one with greater similarity to *T. brucei* XPB received the name affix 1 and the other sequence the affix 2.

Kinetoplastid accession numbers were derived from www.GeneDB.org. Ac, *Acanthamoeba castellanii* (XP_004356478); At-1/At-2, *Arabidopsis thaliana* (Q38861 / NP_568591); Bs-1/Bs-2 *Bodo saltans* (BS16690.1:pep / BS91250.1:pep); Ca, *Candida albicans* (EEQ43396); Ce, *Caenorhabditis elegans* (NP_499487); Dd, *Dictyostelium discoideum* (XP_647819); Dm, *Drosophila melanogaster* (NP_001137931); Drer, *Danio rerio* (Q7ZVV1); Edis-1/Edis-2, *Entamoeba dispar* (EDI_083940 / EDI_292260); Ehis-1/Ehis-2, *Entamoeba histolytica* (XP_649651 / XP_654948); Eimt, *Eimeria tenella* (ETH_00032995); Gl, *Giardia lamblia* (GL50803_16512); Hs, *Homo sapiens* (accession number P19447); Lb-1/Lb-2, *Leishmania braziliensis* (LbrM.29.0600 / LbrM.32.4160); Li-1/Li-2, *Leishmania infantum* (LinJ.29.0610 / LinJ.32.4070); Lm-1/Lm-2, *Leishmania major* (LmjF29.0590 / LmjF32.3920); Lmx-1/Lmx-2, *Leishmania mexicana* (LmxM.08_29.0590 / LmxM.31.3920); Mm, *Mus musculus* (NP_598419); Nc, *Neurospora crassa* (XP_957329); Ng-1/Ng-2, *Naegleria gruberi* (XP_002674455 / XP_002683056); Orys-1/Orys-2, *Oryza sativa* (EEE55219 / EEC71313.1); Part-1/Part-2, *Paramecium tetraurelia* (XP_001448651 / XP_001425941); Pf, *Plasmodium falciparum* (PF10_0369); Sc, *Saccharomyces cerevisiae* (CAY80363); Sp, *Schizosaccharomyces pombe* (NP_593474); Ta, *Theileria annulata* (XP_952567.1); Tb-1/Tb-R, *Trypanosoma brucei* XPB/XPB-R (Tb927.3.5100 / Tb927.11.16270); Tc-1/Tc-2, *Trypanosoma cruzi* (TcCLB.510149.50 / TcCLB.511527.20); Tco-1/Tco-2, *Trypanosoma congolense* (TcIL3000_3_3090 / TcIL3000.11.16180); Tett-1/Tett-2, *Tetrahymena thermophila* (XP_001032673 / XP_001470758); Toxg, *Toxoplasma gondii* (XP_002365647); Trva-1/Trva-2, *Trichomonas vaginalis* (XP_001582312 / XP_001326086); Tv-1/Tv-2, *Trypanosoma vivax* (TvY486_0304470 / TvY486_1117180).

* The *T. vivax* XPB-2 gene appears to be a partial gene, missing sequence for ~220 C-terminal amino acids

** The single *E. tenella* and *A. castellanii XPB* genes miss a large portion of the XPB N-terminus including the XPB domain, suggesting that the reading frames of these genes are incomplete.

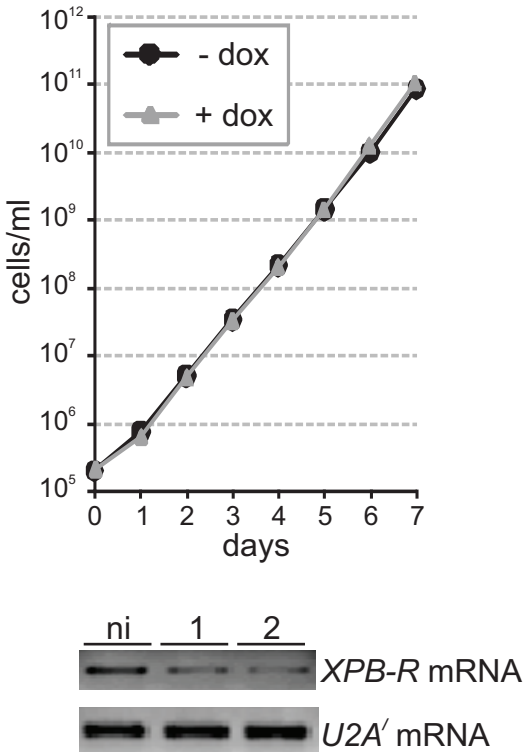



Fig. S2. *XPB-R* is not essential for trypanosome proliferation in culture.

Growth curve of a representative bloodstream form cell line in which doxycycline-induced *XPB-R* dsRNA synthesis led to RNAi-mediated *XPB-R* silencing. The lower panel shows a semi-quantitative PCR analysis of total RNA that was reverse transcribed with oligo-dT and prepared from non-induced cells (ni) and cells that were induced with doxycycline for one or two days. The spliceosomal *U2A'* coding sequence was amplified as a control.

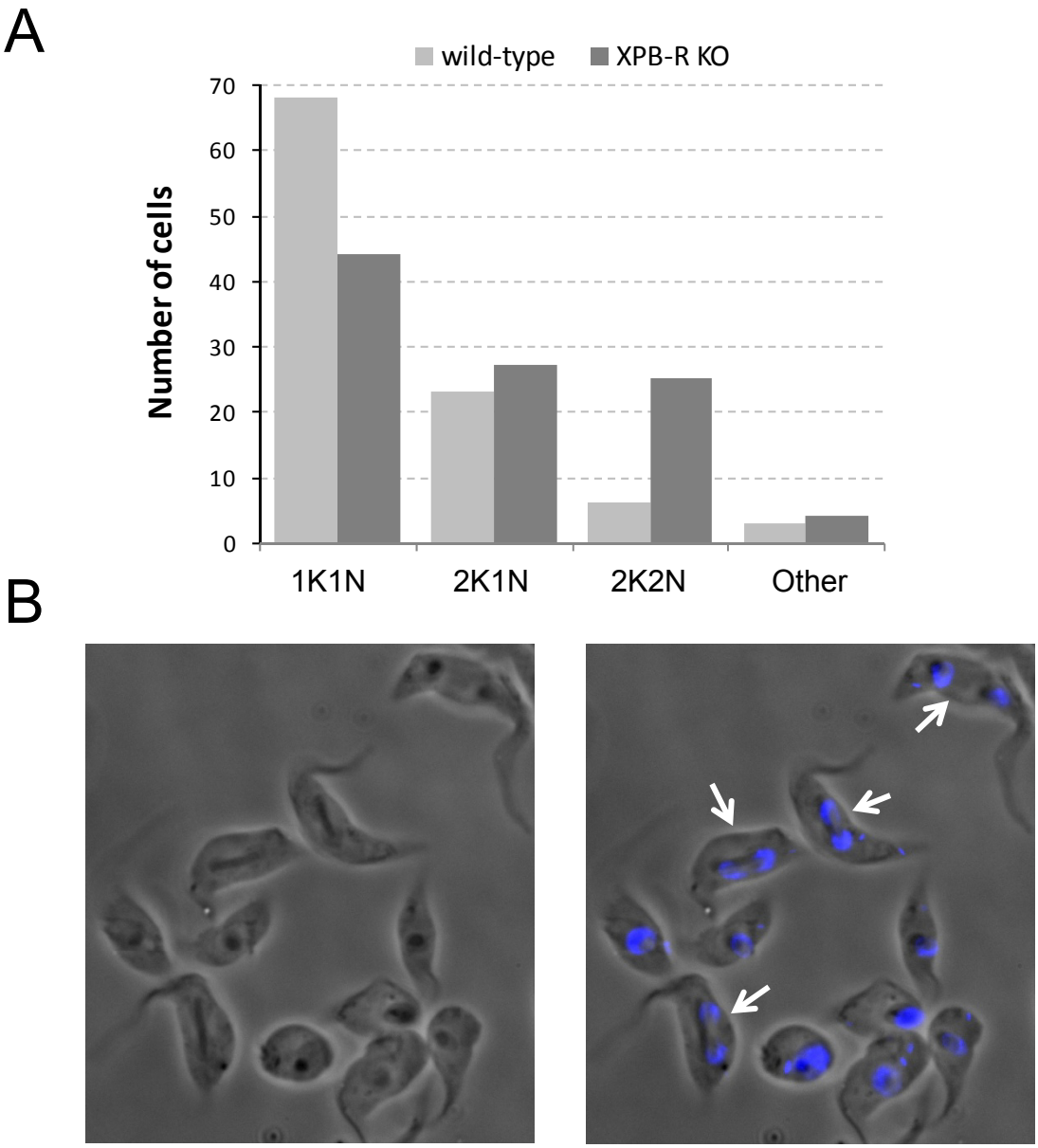



Fig. S3. *XPB-R* knockout appears to affect completion of mitosis

A. Diagram showing the number of wild-type procyclic (n=100) and *XPB-R* knockout cells (XPB-R KO; n= 100) that are in the 1K1N (1 kinetoplast and 1 nucleus), 2K1N, and 2K2N stages of the cell cycle. “Other” combines aberrant cells that have no nuclei (zoids) or more than two nuclei.

B. Image of XPB-R KO cells stained with DAPI showing an accumulation of 2K2N cells (white arrows).

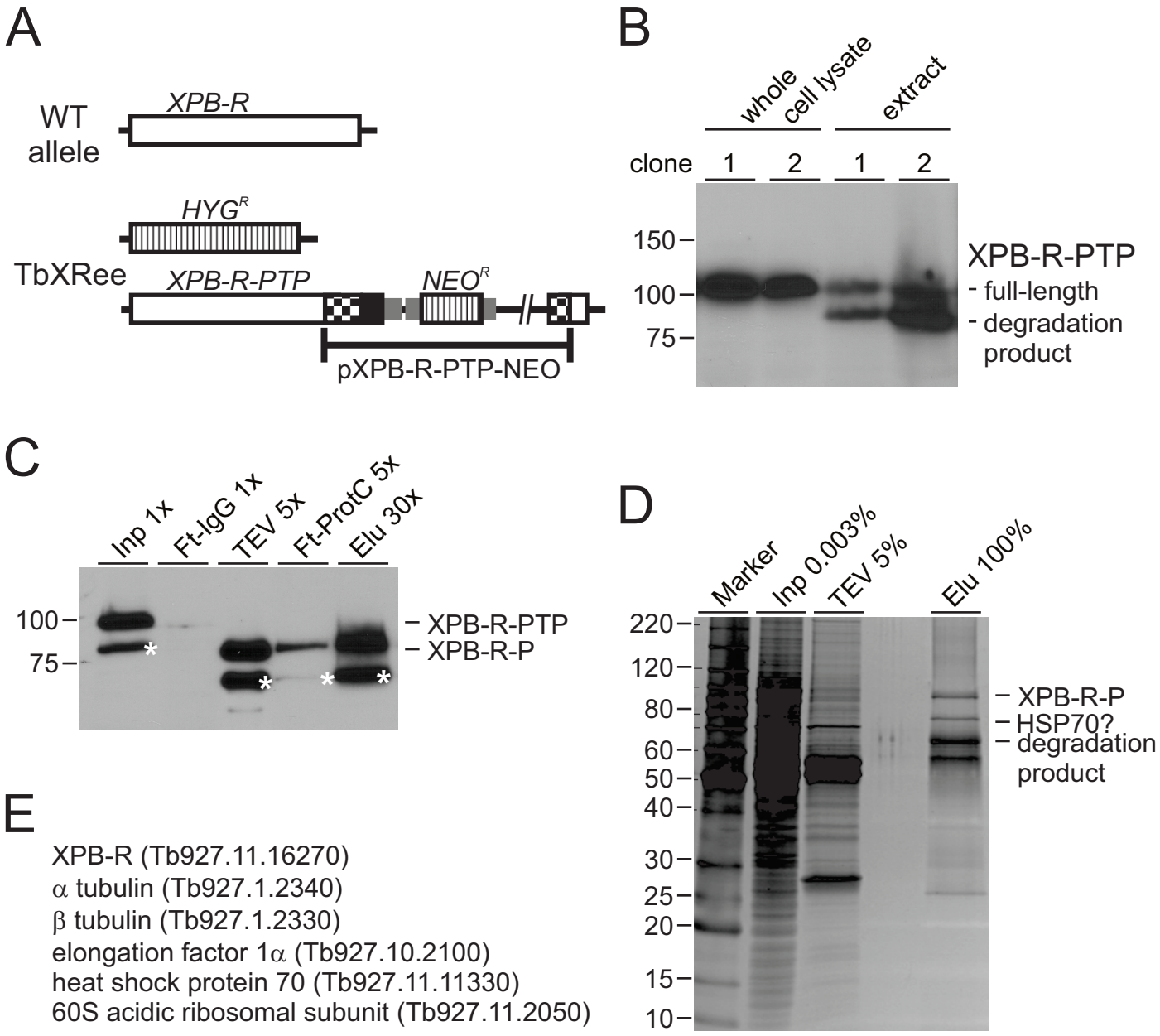



Fig. S4. PTP tagging and tandem affinity purification (TAP) of XPB-R

A. Schematic depiction (not to scale) of a *XPB-R* wild-type (WT) allele and of the *XPB-R* locus of procyclic TbXRee cells that exclusively express XPB-R-PTP and no wild-type, untagged XPB-R. XPB-R, $HYG^R$/$NEO^R$ coding regions and PTP tag are indicated by open, striped and black boxes, respectively. Gene flanks for RNA processing signals are drawn as smaller gray boxes.

B. Whole cell lysates and crude, cell-free extracts of two TbXRee cell lines were analyzed by immunoblotting with the monoclonal anti-Protein C epitope antibody HPC4 that recognizes the PTP tag.

*XPB-R-PTP is N-terminally degraded upon extract preparation.*

C. Immunoblot monitoring of XPB-R-PTP tandem affinity purification. Aliquots of crude extract (Inp), the flowthrough of the IgG affinity chromatography (Ft-IgG), the TEV protease elution (TEV), the flowthrough of the anti-ProtC affinity chromatography (Ft-ProtC), and the final eluate (Elu) were probed with the HPC4 antibody. The x-values indicate relative amounts analyzed. Note that during the TEV protease digest ~19 kDa of protein is removed from XPB-R-PTP resulting in XPB-R-P. Asterisks indicate the XPB-R degradation product.

D. Aliquots of input material (Inp), the TEV protease eluate (TEV) and the final eluate (Elu) were separated on a 10-20% SDS-polyacrylamide gradient gel and stained with sypro ruby. Percentages indicate relative amounts analyzed. Band assignments on the right are guided by apparent protein size.

E. List of proteins that were unambiguously identified by liquid chromatography/tandem mass spectrometry in the final TAP eluate.

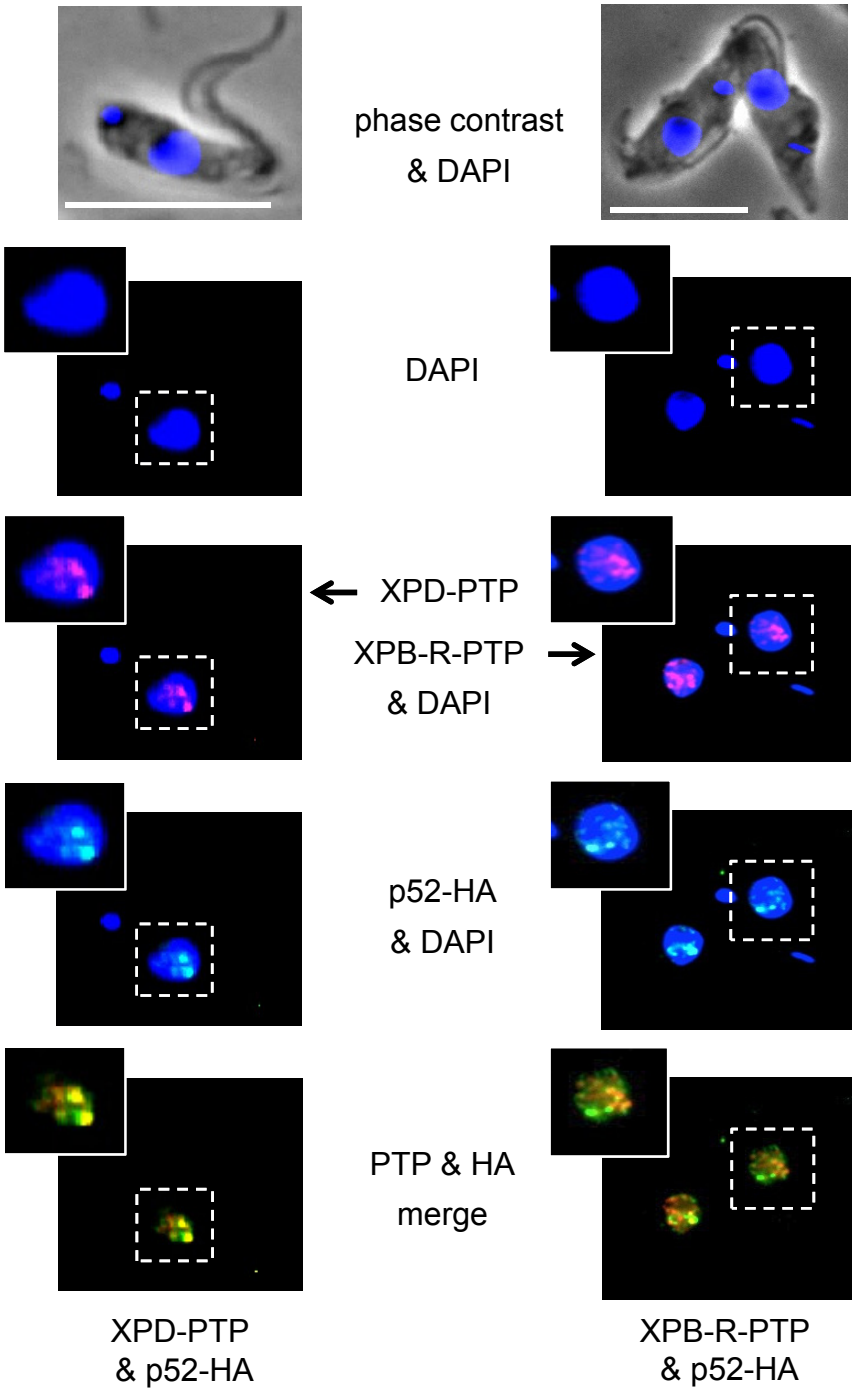



Fig. S5. XPB-R-PTP co-localizes with p52-HA outside putative *SLRNA* expression foci in the nucleus

p52-HA (green) was co-localized with XPD-PTP (red; left panels) or with XPB-R-PTP (red, right panels). White bars represent 10 μm.

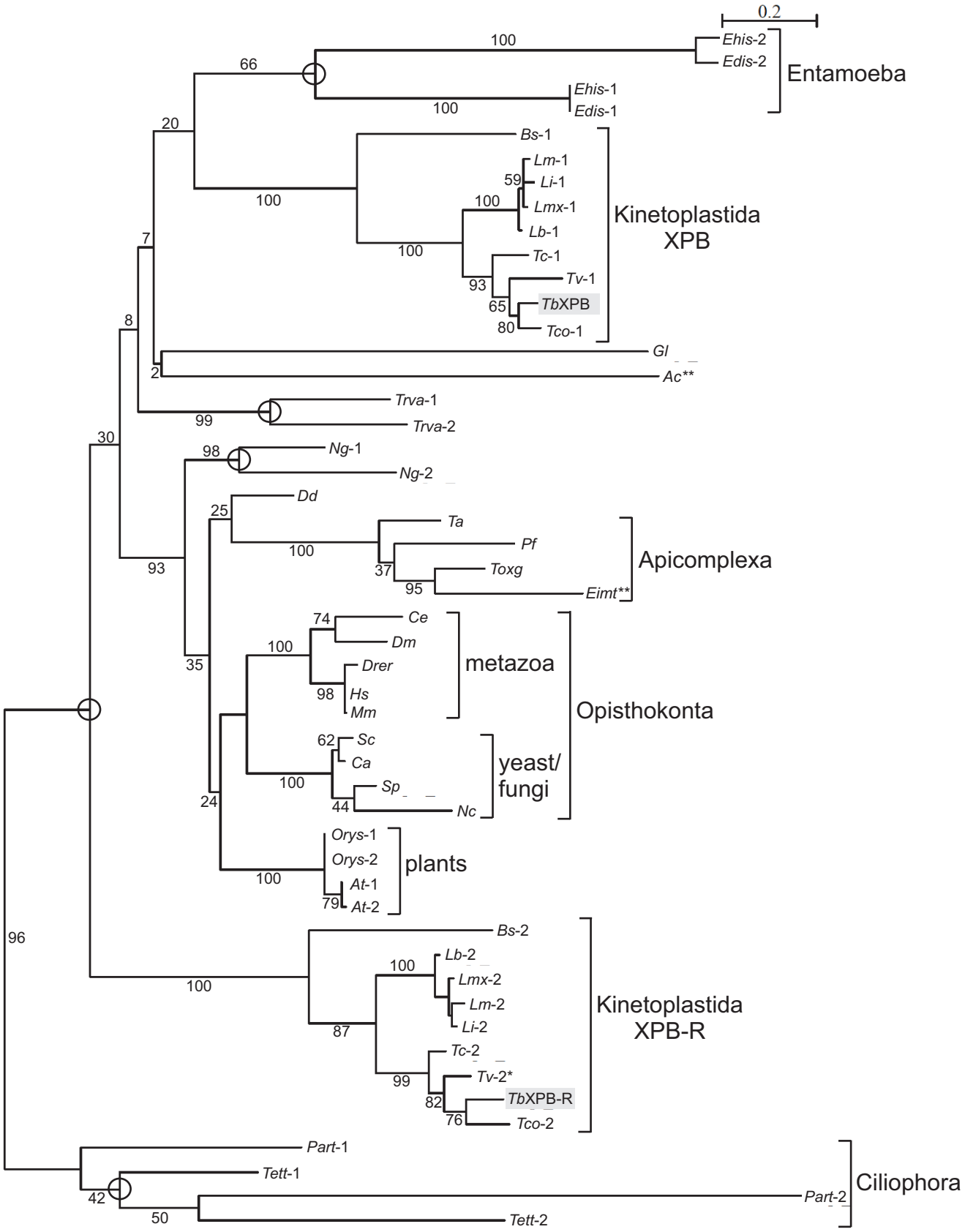

0.2
Ehis-2
Edis-2
Entamoeba
Ehis-1
Edis-1
Bs-1
Lm-1
Li-1
Lmx-1
Lb-1
Tc-1
Tv-1
TbXPB
Tco-1
Kinetoplastida
XPB
Gl
Ac**
Trva-1
Trva-2
Ng-1
Ng-2
Dd
Ta
Pf
Toxg
Eimt**
Apicomplexa
Ce
Dm
Drer
Hs
Mm
metazoa
Sc
Ca
Sp
Nc
yeast/
fungi
Opisthokonta
Orys-1
Orys-2
At-1
At-2
plants
Bs-2
Lb-2
Lmx-2
Lm-2
Li-2
Tc-2
Tv-2*
TbXPB-R
Tco-2
Kinetoplastida
XPB-R
Part-1
Tett-1
Part-2
Tett-2
Ciliophora

Fig. S6. Divergent *XPB* paralogs are present in several protistan taxa.

Multiple eukaryotic XPB amino acid sequences were aligned and phylogenetically analyzed using maximum likelihood. Bootstrap values were obtained by sampling a thousand replicates and are indicated as percentages. The scale bar indicates the number of amino acid substitutions per site. Open circles mark the splits of two divergent XPB paralogs within a taxon. If a species contained two *XPB* genes/sequences, the one with greater similarity to *T. brucei* XPB received the name affix 1 and the other sequence the affix 2. Accession numbers are specified in Fig. S1. *Ac, Acanthamoeba castellanii; At*-1/*At*-2, *Arabidopsis thaliana*; *Bs*-1/*Bs*-2, *Bodo saltans*; *Ca, Candida albicans; Ce, Caenorhabditis elegans; Dd*, *Dictyostelium discoideum; Dm, Drosophila melanogaster; Drer, Danio rerio; Eimt, Eimeria tenella; Gl*, *Giardia lamblia*; *Hs, Homo sapiens; Lb*-1/*Lb*-2, *Leishmania braziliensis*; *Li*-1/*Li*-2, *Leishmania infantum*; *Lm*-1/Lm-2, *Leishmania major*; *Lmx*-1/*Lmx*-2, *Leishmania mexicana*; *Mm, Mus musculus; Nc, Neurospora crassa; Ng*-1/*Ng*-2, *Naegleria gruberi*; *Orys*-1/*Orys*-2, *Oryza sativa*; *Part*-1/*Part*-2, *Paramecium tetraurelia*; *Pf, Plasmodium falciparum; Sc, Saccharomyces cerevisiae; Sp, Schizosaccharomyces pombe; Ta, Theileria annulata*; *Tb*-1/*Tb*-R, *Trypanosoma brucei*; *Tc*-1/*Tc*-2, *Trypanosoma cruzi*; *Tco*-1/*Tco*-2, *Trypanosoma congolense*; *Tett*-1/*Tett*-2, *Tetrahymena thermophila*; *Toxg, Toxoplasma gondii; Trva*-1/*Trva*-2, *Trichomonas vaginalis*; *Tv*-1/*Tv*-2, *Trypanosoma vivax.*

* The *T. vivax XPB-2* gene appears to be a partial gene, missing sequence for ~220 C-terminal amino acids

** The single *E. tenella* and *A. castellanii XPB* genes miss a large portion of the XPB N-terminus including the XPB domain, suggesting that the reading frames of these genes are incomplete.

*Method of Phylogenetic Analysis*

XPB amino acid sequences from 33 organisms representing the major eukaryotic supergroups (Dacks *et al.*, 2008) were aligned using the multiple sequence alignment tool MUSCLE at the EMBL European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/msa/muscle/). The alignment was uploaded onto the graphical user interface Seaview (Gouy *et al.*, 2010); http://pbil.univ-lyon1.fr/software/seaview.html) to drive the GBlocks program using the less stringent option for the selection of conserved sequence blocks that are likely to be aligned correctly and for the removal of poorly aligned positions (Talavera and Castresana, 2007). The best-fit evolution model for phylogenetic analysis was determined by using the ProtTest Server 2.4 (Abascal *et al.*, 2005); http://darwin.uvigo.es/software/prottest2_server.html). A maximum likelihood tree was generated employing the LG empirical matrix (Le and Gascuel, 2008) with estimated invariable sites, substitution rate categories of 4, estimated gamma distribution and model equilibrium frequencies.

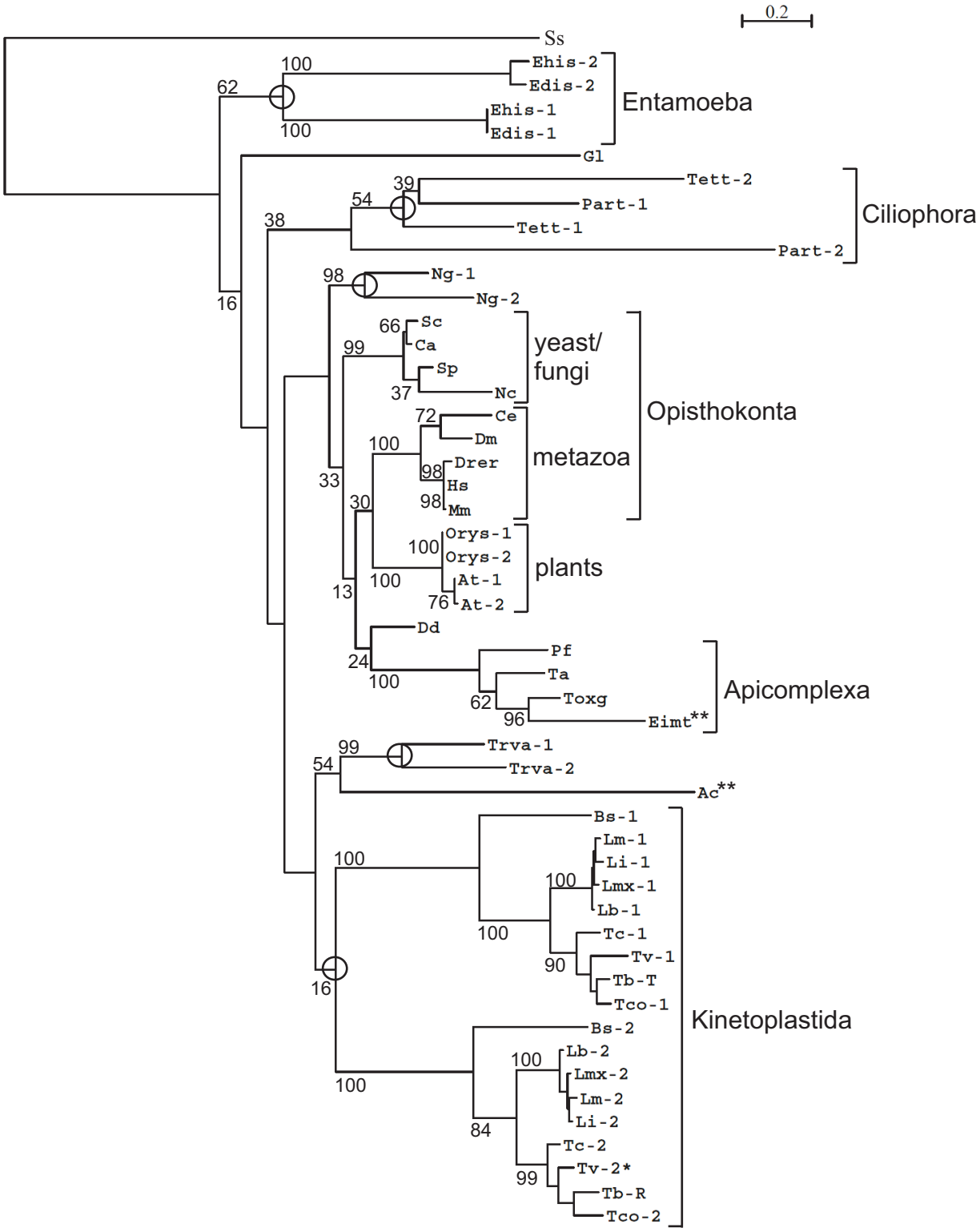



Fig. S7. XPB-derived phylogenetic tree with the archaeon *Sulfolobus solfataricus* as an outgroup

The phylogenetic analysis corresponds to that shown in figure 7 except that the XPB sequence from *Sulfolobus solfataricus* (Ss, accession number NP_342443) was included.

**Table S1. Oligonucleotide used in this study**

| pGene | Sequence | Usage |
|---|---|---|
| **XPB-R** | 5′-GGAGTGGAACTGCGAGGCCAGTAGTG-3′ / 5′-CCAAATCCACCTGAATACGGCGACG-3′ | Sense/antisense oligonucleotides that hybridize in the 5′ and 3′ XPB-R gene flanks, respectively; used to analyze correct integration of transfected DNA constructs that replace the wild-type *XPB-R* alleles. |
| | 5′-TAAGGGCCCAGCACACCGAACTGCATTCCTCG-3′/ 5′-TAACGGCCGCGATGACGTTGTCTATCATGAGAG T-3′ | Sense/antisense oligonucleotides that hybridize within the coding region of XPB-R; used for RT-PCR of XPB-R mRNA. |
| **U2A′** | 5′-TAGGGCCCTCCAAAAACGGCAGTGCCGATACTG C-3′ / 5′-TACGGCCGTGATGTGCGAGTCTTCTTTGTC CCTTTAGAAGC-3′ | Sense/antisense oligonucleotides that hybridize within the coding region of U2A′; used for RT-PCR of U2A′ mRNA. |
| ***SL RNA*** | 5′-CTACCGACACATTTCTGGC-3′ / 5′-GGTATGAGAA GCTCCCAGTAGCAGC-3′ | Sense/antisense oligonucleotides that hybridize in the *SLRNA* promoter region; used for PCR / qPCR amplifications in ChIP experiments. |
| | 5′-ATGGCTTATACGTGCTCGTTTCTCC-3′ / 5′-CACAT ATAGGCGCTTTAAAGTCTGCT-3′ | Sense/antisense oligonucleotides that hybridize in the *SLRNA* intergenic region; used for PCR / qPCR amplifications in ChIP experiments. |
| ***α-Tubulin*** | 5′-GTGCATTGAACGTGGATCTG-3′ / 5′-GCCTACCAC GAGCAACTCTC-3′ | Sense/antisense oligonucleotides that hybridize within the coding region of α-tubulin; used for PCR / qPCR amplifications in ChIP experiments. |
| ***p52*** | 5′-TAGGGCCCGCAGCCCCTTCAACAACAGTGG-3′ / 5′- TACGGCCGATCCGGAGTAATGAAATTGGAC-3′ | Sense/antisense oligonucleotides that hybridize within the coding region of p52; used for RT-PCR of p52 mRNA. |

## Supplemental References


Abascal,F., Zardoya,R., and Posada,D. (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**: 2104-2105.

Dacks,J.B., Walker,G., and Field,M.C. (2008) Implications of the new eukaryotic systematics for parasitologists. *Parasitol Int* **57**: 97-104.

Gouy,M., Guindon,S., and Gascuel,O. (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**: 221-224.

Le,S.Q., and Gascuel,O. (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* **25**: 1307-1320.

Sievers,F., Wilm,A., Dineen,D., Gibson,T.J., Karplus,K., Li,W. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**: 539.

Talavera,G., and Castresana,J. (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**: 564-577.

**Supplementary Information (Badjatia *et al.*)**



**Table S1** Accession numbers of relevant *T. brucei* genes

| Gene | Accession number [1] |
|---|---|
| *RPB1* | Tb927.4.5020, Tb927.8.7400 |
| *CRK1* | Tb927.10.1070 |
| *CRK3* | Tb927.10.4990 |
| *CRK7* | Tb927.7.1900 |
| *CRK9* | Tb927.2.4510 |
| *MTR1* | Tb927.10.7940 |
| *MTR2* | Tb927.11.4890 (previously Tb11.02.2500) |
| *MTR3* | Tb927.9.12040 (previously Tb09.211.3130) |
| *SNIP* | Tb927.8.3710 |

[1] Accession numbers are from the GeneDB (www.genedb.org)/TritrypDB (www.tritrypdb.org) genome data bases.

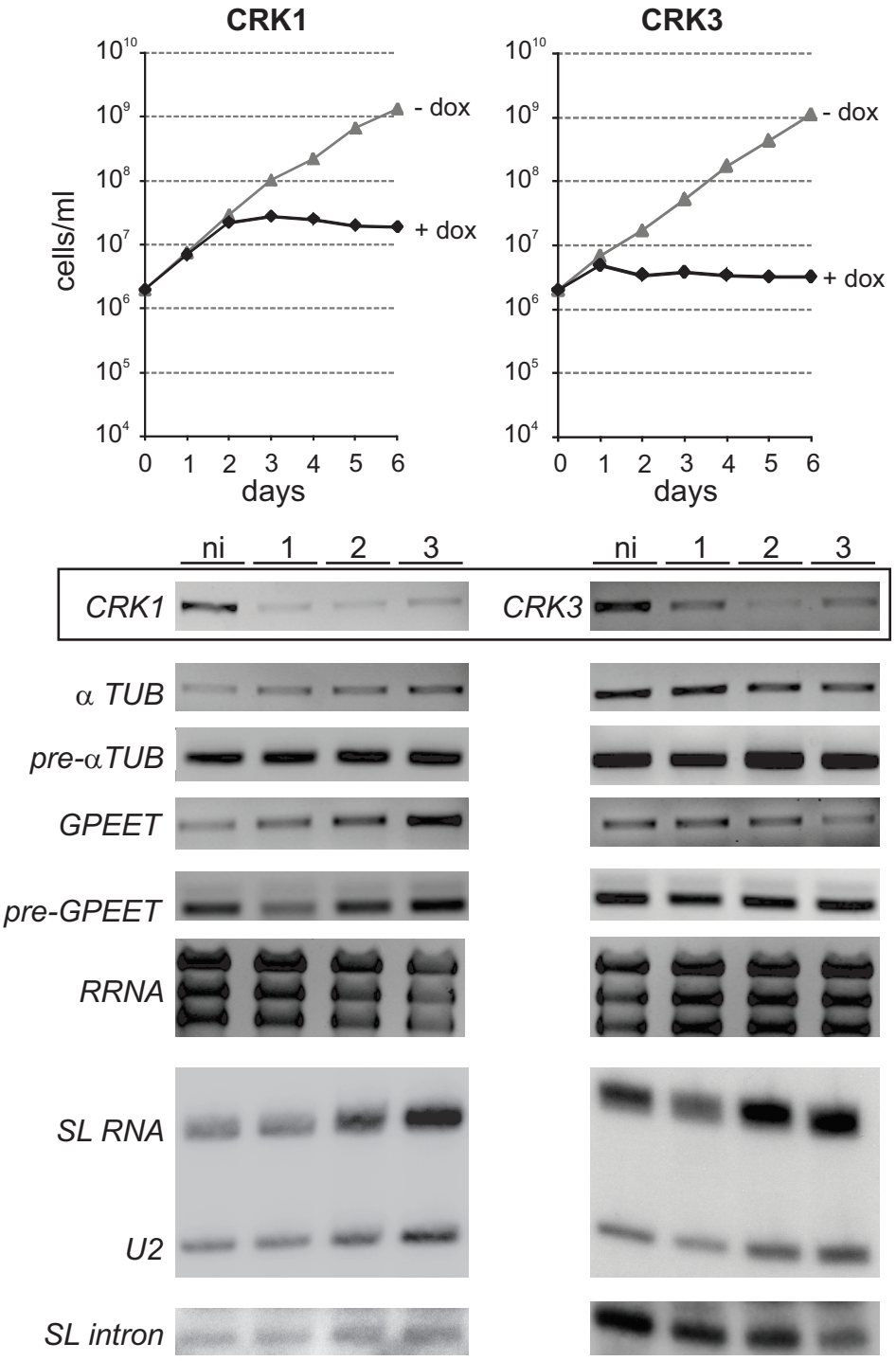



**FIG S1** *CRK1* and *CRK3* expression silencing. For each knockdown, representative growth curves from one of three independently derived clonal cell lines were obtained in the absence and presence of the dsRNA synthesis-inducing reagent doxycycline (-/+ dox). *CRK* mRNA levels were determined by semi-quantitative RT-PCR of total RNA in non-induced (ni) cells and in cells that were silenced for 1 to 3 days (boxed panels). In the same RNA preparations, the relative abundances of α tubulin (α TUB) and of *GPEET* procyclin mature and pre-mRNAs were determined by semi-quantitative RT-PCR, those of the large ribosomal RNAs by ethidium bromide staining, and those of the SL RNA, the U2 snRNA, and the SL intron by primer extension assays.

*CRK1* and *CRK3* silencing did not lead to a clear disturbance in gene expression patterns. In case of *CRK1* silencing, however, we did notice a reproducible, minor reduction in rRNA and compensatory increases in other RNAs suggesting a minor function of this kinase in the control of ribosome biogenesis.

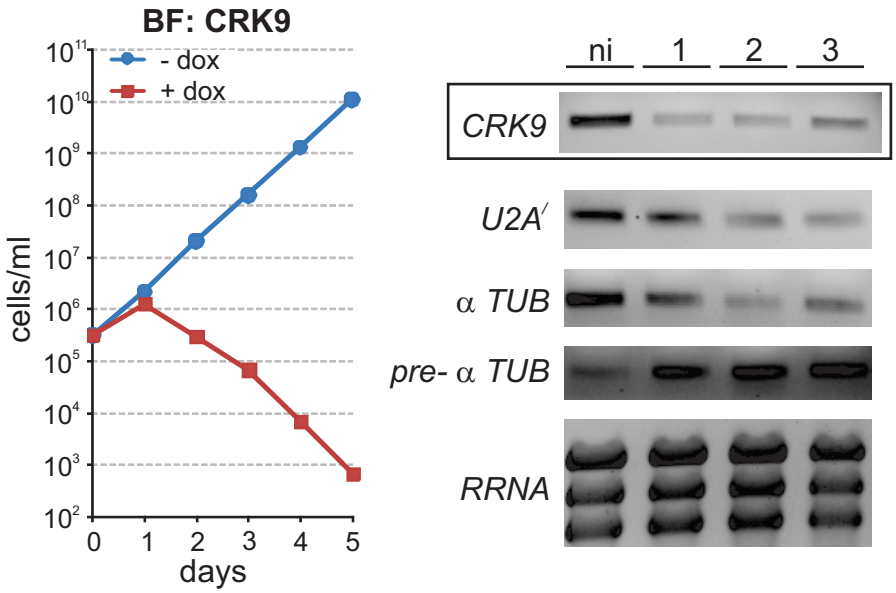



**FIG S2** *CRK9* silencing in bloodstream trypanosomes. A representative growth curve in the absence and presence of the dsRNA synthesis-inducing reagent doxycycline (-/+ dox) of one of two independently derived clonal cell lines. Mature mRNAs of CRK9, the spliceosomal protein U2A' and α tubulin (α TUB) as well as pre-mRNA of α tubulin (pre-α TUB) were analyzed by semi-quantitative reverse transcription (RT)-PCR in non-induced (ni) cells and cells that were silenced for 1, 2 or 3 days. The relative RNA abundances show the same trend as in procyclics: mRNA abundances decrease whereas the pre-mRNA level increases in *CRK9*-silenced bloodstream trypanosomes.

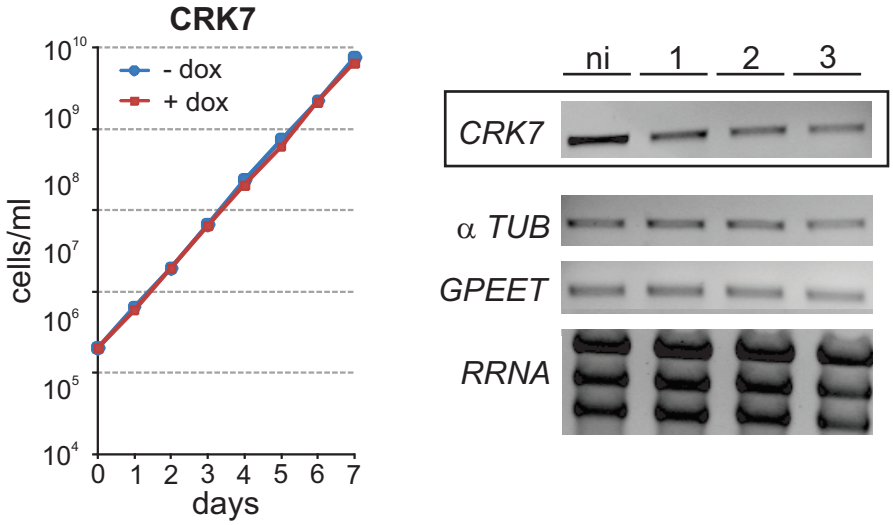



**FIG S3** *CRK7* silencing in procyclic trypanosomes. A representative growth curve of one of three independently derived clonal cell lines obtained in the absence and presence of doxycycline (-/+ dox). Abundances of CRK7, α tubulin (α TUB) and of *GPEET* procyclin mRNA were analyzed in total RNA prepared from non-induced (ni) cells and from cells that were silenced for 1, 2 or 3 days by semi-quantitative RT-PCR. The large ribosomal RNAs were detected by ethidium bromide staining.

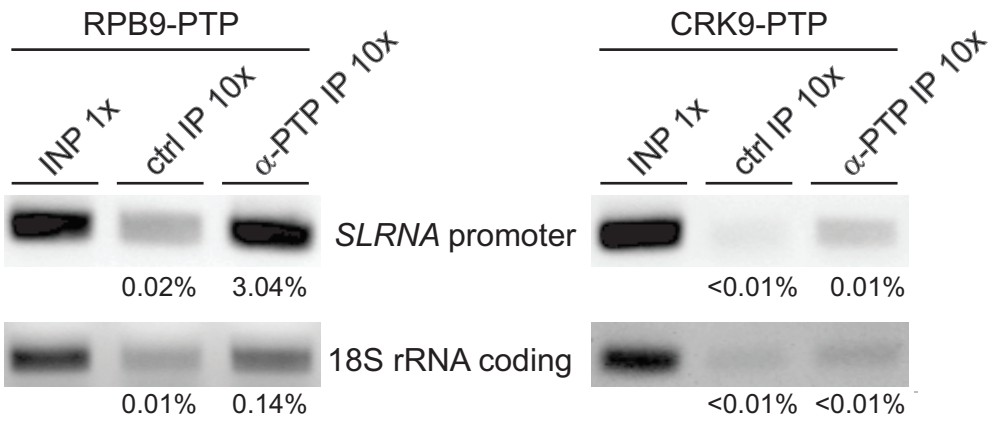



**FIG S4** Anti-CRK9-PTP chromatin immunoprecipitation (ChIP). ChIP experiments were carried out with two cell lines expressing either the RNA pol II subunit RPB9 or CRK9 with a functional fusion of the composite PTP tag. Chromatin was precipitated with a ChIP-grade rabbit polyclonal anti-ProtA antibody (α-PTP IP). In negative control immunoprecipitations (ctrl IP) a comparable, rabbit-derived non-specific antibody was used. Precipitated DNA was analyzed by amplification of the *SLRNA* promoter region and, as a negative control, of the 18S rRNA coding region. For comparison, DNA of total chromatin (INP) was analyzed. The percent precipitation values specified under each lane were determined by qPCR.

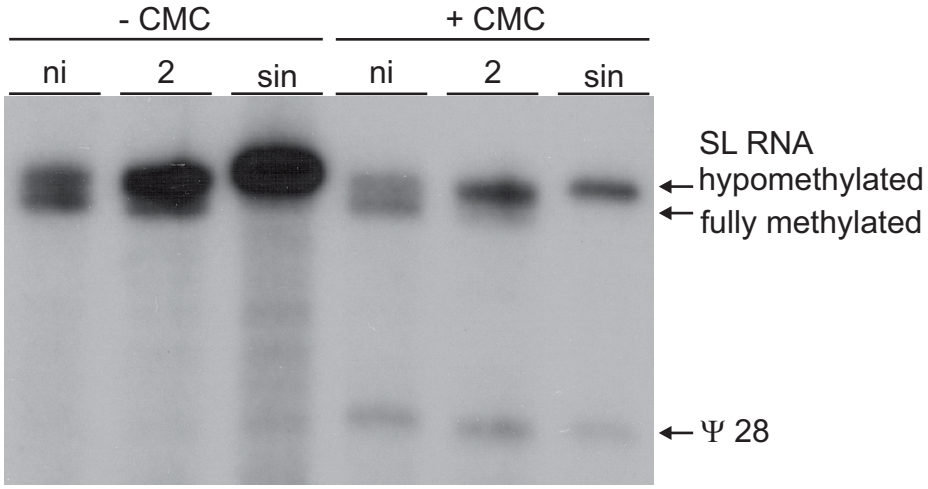



**FIG S5** *CRK9* silencing does not affect SL RNA pseudouridylation. Total RNA, prepared from non-induced cells (ni), from cells in which *CRK9* was silenced for 2 days and from cells which were incubated with the general methyltransferase inhibitor sinefungin (sin) for six hours, was analyzed by primer extension with 5′ radiolabeled oligonucleotide SLf. RNA was either mock incubated or treated with *N*-cyclohexyl-*N*′-β-(4-methylmorpholinium)ethyl-carbodiimide *p*-tosylate (-/+ CMC) according to a published protocol (Bakin and Ofengand, 1993, Biochemistry 32, 9754-9762) to create an extension stop before the pseudoruridylated residue at position 28 (Ψ 28). To account for the SL RNA increase in *CRK9*-silenced and sinefungin-treated cells the amount of RNA in these samples was reduced by one half.

**Table S2** List of oligonucleotides

| Gene | Name | Usage | Sequence (5′ – 3′) |
|---|---|---|---|
| *CRK1* | CRK1 ApaI | RT-PCR of *CRK1* mRNA | TAGGGCCCGTTACGAGCGGCTTCAGAAGATTG |
| | CRK1 NotI | RT-PCR of *CRK1* mRNA | TACGGCCGGAACTCGACAGAAAAGTATGGGTG |
| *CRK3* | CRK3 ApaI | RT-PCR of *CRK3* mRNA | TAGGGCCCCGGAATCTGGAAAGCGCGAGC |
| | CRK3 NotI | RT-PCR of *CRK3* mRNA | GATCGCGGCCGCTACAATGCTTGGGGCGTTAACC |
| *CRK7* | CRK7 ApaI | RT-PCR of *CRK7* mRNA | TAGGGCCCGACCGATTTGTCGGCGGTGCTCTG |
| | CRK7 EagI | RT-PCR of *CRK7* mRNA | TACGGCCGGTACTCGTTGAAATATGTATGAC |
| *CRK9* | CRK9 RNAi-S | RT-PCR of *CRK9* mRNA | GTTGTTAAGAAAGAAAATTGAGG |
| | CRK9 3′ RT | RT-(q)PCR of *CRK9* mRNA | CGCCGCCGCTGCCAACCCCACTC |
| | CRK9 qPCR 5′ | RT-qPCR of *CRK9* mRNA | GAGGAAGTGCAACAGCAGCG |
| | CRK9 UTR RNAi 5′ | RT-PCR of *CRK9* mRNA | GATAAGCTTACGCGTAGGATGGAAGGGGTTGTGAC |
| | CRK9 UTR RNAi 3′ | RT-PCR of *CRK9* mRNA | GATCTAGACTCGAGGCACTCGATACTTCAAAGAGC |
| | CRK9 Apa1 | RT-PCR of *CRK9* mRNA | TAGGGCCCGCTCTGTGAATGAATATGTTCCCG |
| | HA antisense | RT-PCR of *CRK9* mRNA | TTAAGCGTAGTCAGGTACGTCGTAAGGGTA |
| *α tubulin* | α-Tub Fw q1 | RT-(q)PCR of α tub. pre-mRNA | ACAGTTTCTGTACTATATTG |
| | α-Tub Rv q1 | RT-(q)PCR of α tub. pre-mRNA | GTGGATGCAGATAGCC |
| | α-Tub Fw q3 | RT-(q)PCR of α tub. mRNA | GTGCATTGAACGTGGATCTG |
| | α-Tub Rv q3 | RT-(q)PCR of  tub. mRNA | GAGAGTTGCTCGTGGTAGGC |
| *GPEET* | Pre-GPEET 5′ | RT-(q)PCR of *GPEET* pre-mRNA | CATGTTCTCGTGATCGCTGC |
| | Pre-GPEET 3′ | RT-(q)PCR of *GPEET* pre-mRNA | GCAGATAAAGGGAACGAGGTGC |
| | GPEET 5′ | RT-(q)PCR of *GPEET* mRNA | ATGGCACCTCGTTCCCTTTATCTGC |
| | GPEET 3′ | RT-PCR of *GPEET* mRNA | TTAGAATGCGGCAACGAGAGCAGC |
| | GPEET qPCR 3′ | RT-qPCR of *GPEET* mRNA | CCTTTGCCTCCCTTCACGATAAC |
| *18S rRNA* | 18SrRNAcoding5′ | RT-(q)PCR of 18s rRNA | TCATCAAACTGTGCCGATTAC |
| | 18SrRNAcoding3′ | RT-(q)PCR of 18s rRNA | CTATTGAAGCAATATCGG |
| *SLRNA* | SL_PE | Primer extension of SL RNA & RT-(q)PCR of *SL RNA* | CGACCCCACCTTCCAGATTC |
| | SL sense | RT-(q)PCR of *SL RNA* | ACAGTTTCTGTACTATATT |
| *U1 snRNA* | U1_PE | Primer extension of U1 snRNA | CACTCAAAGTTTACTGCA |
| *U2 snRNA* | U2_PE | Primer extension of U2 snRNA & RT-(q)PCR of *U2 snRNA* | ACAGGCAACAGTTTTGATCC |
| | U2 5′ | RT-(q)PCR of *U2 snRNA* | ATATCTTCTCGGCTATTTAGC |
| *U5 snRNA* | U5_PE | Primer extension of U5 snRNA | CCGCTCGAGGACACCCCAAAGTTT |

**Badjatia *et al.* – Supplemental material**

**S1 Table. CRK9-PTP co-purified proteins**

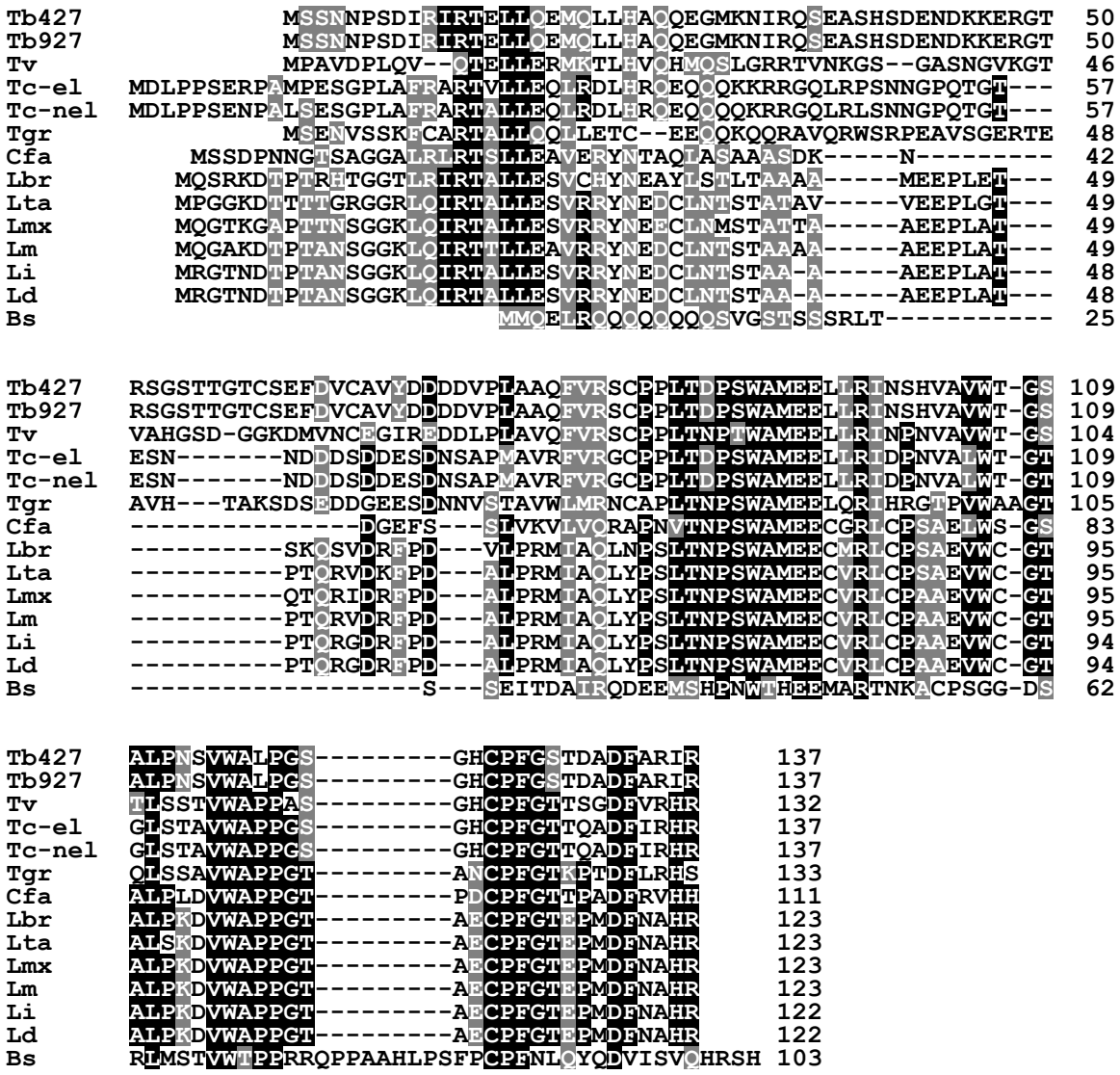
| Score | Expectation | Accession # | MW (Da) | % Coverage | emPAI | Annotation | Comment |
|---|---|---|---|---|---|---|---|
| 13418 | 0 | Tb927.2.4510 / Tb427.02.4510 * | 85,465 | 73.90 | 10.28 | CRK9 | |
| 9956 | 0 | Tb927.10.9160 / Tb427.10.9160 * | 71,208 | 43.30 | 8.09 | CYC12 | |
| 2551 | 0 | Tb927.1.2340 / Tb427.01.2340 * | 49,756 | 60.10 | 3.46 | alpha tubulin | |
| 2419 | 0 | Tb927.1.2330 / Tb11.v5.0469 * | 49,672 | 60.20 | 3.12 | beta tubulin | |
| 1897 | 0 | Tb927.11.11330 / Tb427tmp.01.3110 * | 76,002 | 52.70 | 3.12 | HSP70 | |
| 1797 | 0 | Tb927.7.3880 * | 87,018 | 49.70 | 2.13 | protein kinase | Domain PKc_DYRK_like (E=8.77e$^{-135}$) |
| 1358 | 0 | Tb11.v5.1046 * | 49,075 | 41.50 | 1.99 | elongation factor 1-alpha | |
| 1179 | 2.50E-114 | Tb927.3.5050 / Tb427.03.5050 * | 41,830 | 57.10 | 2.97 | 60S ribosomal protein L4 | |
| 1080 | 1.90E-104 | Tb927.10.3930 / Tb11.v5.0501 * | 29,404 | 52.70 | 4.35 | 40S ribosomal protein S3A, putative | |
| 1014 | 8.90E-98 | Tb927.10.14600 / Tb427.10.14600 * | 28,624 | 77.20 | 4.6 | 40S ribosomal protein S2, putative | |
| 1010 | 2.30E-97 | Tb927.11.3590 / Tb427tmp.02.1085 * | 30,611 | 63.60 | 6.49 | 40S ribosomal protein S4, putative | |
| 1006 | 4.60E-97 | Tb927.11.2050 * | 34,606 | 39.10 | 1.92 | 60S acidic ribosomal subunit protein, putative | |
| 991 | 1.80E-95 | Tb927.11.3590 / Tb11.v5.1059 * | 30,625 | 59.60 | 6.49 | 40S ribosomal protein S4, putative | |
| 965 | 6.40E-93 | Tb927.9.6070 / Tb427tmp.160.4450 * | 30,361 | 44.60 | 4.07 | 40S ribosomal protein S3, putative | |
| 936 | 5.00E-90 | Tb927.9.5690 / Tb427tmp.160.4200 * | 11,083 | 99.20 | 11.73 | 60S acidic ribosomal protein, putative | |
| 878 | 2.90E-84 | Tb927.8.6150 / Tb427.08.6150 | 25,004 | 35.20 | 2.72 | 40S ribosomal protein S8, putative | |
| 874 | 8.40E-84 | Tb927.7.1730 / Tb427.07.1730 * | 27,696 | 54.00 | 4.1 | 60S ribosomal protein L7, putative | |
| 817 | 4.40E-78 | Tb927.11.11820 / Tb427tmp.01.3675 | 16,212 | 61.70 | 8.6 | 40S ribosomal protein S17, putative | |
| 805 | 7.20E-77 | Tb927.9.13970 / Tb427tmp.211.4520 * | 36,766 | 47.30 | 2.84 | hypothetical conserved | Trypanosoma-specific, no human homolog |
| 800 | 1.90E-76 | Tb927.9.15110 / Tb427tmp.244.2730 * | 34,615 | 52.80 | 4.28 | 60S ribosomal protein L5, putative | |
| 796 | 5.50E-76 | Tb11.v5.0243 | 17,499 | 48.40 | 7.11 | 40S ribosomal protein S18, putative | |
| 795 | 6.30E-76 | Tb927.10.350 / Tb11.v5.0564 * | 58,564 | 35.50 | 1.5 | protein kinase, putative | Domain PKc_DYRK_like (E=1.77e$^{-129}$) |
| 742 | 1.30E-70 | Tb927.11.9710 / Tb427tmp.01.1470 | 24,601 | 60.80 | 4.28 | 60S ribosomal protein L10a, putative | |
| 724 | 7.60E-69 | Tb927.9.10770 / Tb427tmp.211.2150 | 62,195 | 37.80 | 1.71 | Polyadenylate-binding protein 2 | |
| 692 | 1.20E-65 | Tb927.11.6300 / Tb427tmp.02.4170 * | 21,372 | 44.80 | 2.15 | 40S ribosomal protein S5, putative | |
| 689 | 2.30E-65 | Tb927.11.10790 / Tb427tmp.01.2560 | 27,592 | 61.80 | 2.81 | 40S ribosomal protein SA, putative | |
| 683 | 1.10E-64 | Tb927.9.3920 / Tb427tmp.160.2490 * | 23,826 | 48.90 | 5.62 | ribosomal protein S7, putative | |
| 661 | 1.80E-62 | Tb927.3.1610 / Tb427.03.1610 | 74,199 | 34.10 | 0.95 | protein kinase, putative | Domain PKc_CLK (E=5.90e$^{-57}$) |
| 648 | 3.40E-61 | Tb927.9.8070 / Tb427tmp.211.0110 | 24,720 | 46.30 | 4.24 | 60S ribosomal protein L10, putative | |
| 647 | 4.00E-61 | Tb927.9.15360 / Tb427tmp.244.2630 * | 28,405 | 36.40 | 3.91 | 40S ribosomal protein S6, putative | |
| 641 | 1.50E-60 | Tb927.7.5180 / Tb427.07.5180 | 24,667 | 49.50 | 2.21 | 60S ribosomal protein L23a, putative | |
| 634 | 7.40E-60 | Tb927.10.190 / Tb427.10.190 * | 28,433 | 39.10 | 3.91 | 40S ribosomal protein S6, putative | |
| 632 | 1.40E-59 | Tb927.10.11390 / Tb427.10.11390 | 21,172 | 68.60 | 5.89 | 60S ribosomal protein L6, putative | |
| 605 | 5.90E-57 | Tb927.4.1100 / Tb427.04.1100 | 18,013 | 73.90 | 6.71 | ribosomal protein L21E (60S), putative | |
| 588 | 3.20E-55 | Tb927.7.2370 / Tb427.07.2370 | 19,952 | 45.80 | 3.19 | 40S ribosomal protein S15, putative | |
| 588 | 3.40E-55 | Tb927.10.3370 / Tb427.10.3370 | 11,528 | 63.00 | 7.16 | 60S acidic ribosomal protein P2, putative | |
| 568 | 3.30E-53 | Tb927.3.4170 / Tb427.03.4170 * | 13,001 | 37.20 | 5.5 | CRK9AP | |
| 567 | 4.30E-53 | Tb927.10.15290 / Tb427.10.15290 | 178,152 | 21.50 | 0.29 | hypothetical conserved | domain TBCC (E=1.13e$^{-09}$) Tubulin binding cofactor C |
| 555 | 6.70E-52 | Tb927.10.1100 / Tb427.10.1100 | 21,938 | 52.10 | 3.44 | 60S ribosomal protein L9, putative | |
| 541 | 1.80E-50 | Tb927.4.1790 / Tb427.04.1790 * | 54,402 | 45.10 | 1.49 | ribosomal protein L3, putative | |
| 539 | 2.30E-50 | Tb927.10.560 Tb427.10.560 | 20,075 | 38.90 | 3.16 | 40S ribosomal proteins S11, putative | |
| 535 | 6.10E-50 | Tb927.10.10970 / Tb427.10.10970 * | 80,780 | 31.60 | 0.76 | heat shock protein, putative | |
| 532 | 1.20E-49 | Tb927.10.6400 / Tb427.10.6400 | 59,466 | 37.40 | 1 | chaperonin HSP60, mitochondrial precursor | |
| 529 | 2.50E-49 | Tb927.9.11470 / Tb427tmp.211.2650 | 16,284 | 51.30 | 2.48 | 60S ribosomal protein L27a | |
| 527 | 3.80E-49 | Tb927.4.3550 / Tb427.04.3550 | 39,091 | 26.70 | 0.88 | 60S ribosomal protein L13a, putative | |
| 525 | 6.80E-49 | Tb927.3.3740 | 69,133 | 34.40 | 0.82 | hypothetical conserved | Zn finger (RING) protein, potential dsRNA-binding |
| 501 | 1.60E-46 | Tb927.1.3180 / Tb427.01.3180 | 20,047 | 39.00 | 3.16 | 40S ribosomal protein S11, putative | |
| 493 | 9.30E-46 | Tb927.7.1040 / Tb427.07.1040 | 16,987 | 57.00 | 5.82 | 40S ribosomal protein S16, putative | |
| 475 | 6.30E-44 | Tb927.9.11380 / Tb427tmp.211.2630 | 14,953 | 45.60 | 2.88 | 60S ribosomal protein L23, putative | |
| 474 | 7.80E-44 | Tb927.1.4680 / Tb427.01.4680 | 37,288 | 36.30 | 1.17 | hypothetical conserved | Trypanosoma-specific, no human homolog |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 469 | 2.40E-43 | Tb927.8.1110 / Tb427.08.1110 | 22,124 | 55.70 | 4.3 | 40S ribosomal protein S9, putative | |
| 465 | 6.20E-43 | Tb927.10.5360 Tb427.10.5360 | 19,274 | 56.50 | 3.4 | 40S ribosomal protein S10, putative | |
| 464 | 8.00E-43 | Tb927.10.7330 / Tb427.10.7330 | 15,674 | 29.50 | 1.83 | 40S ribosomal protein S24E, putative | |
| 454 | 8.50E-42 | Tb927.9.12200 / Tb427tmp.211.3270 | 21,651 | 40.10 | 2.11 | 60S ribosomal subunit protein L31, putative | |
| 418 | 3.40E-38 | Tb11.0290 | 15,503 | 58.20 | 1.85 | 40S ribosomal protein S14, putative | |
| 413 | 9.90E-38 | Tb927.9.7590 / Tb427tmp.160.5580 | 22,355 | 48.70 | 2.61 | 60S ribosomal protein L11, putative | |
| 406 | 5.30E-37 | Tb927.7.710 / Tb427.07.710 * | 70,168 | 17.10 | 0.42 | HSP70, putative | |
| 406 | 5.70E-37 | Tb927.11.16280 / Tb427tmp.01.7960 | 28,320 | 35.00 | 1.76 | 60S ribosomal protein L2, putative | |
| 382 | 1.10E-34 | Tb927.10.11600 / Tb427.10.11600 | 66,008 | 21.70 | 0.65 | hypothetical conserved | RS domain, Trypanosoma-specific, no human homolog |
| 378 | 3.30E-34 | Tb927.11.3510 / Tb427tmp.02.1000 * | 91,873 | 8.10 | 0.14 | hypothetical conserved | no human homolog |
| 367 | 4.30E-33 | Tb927.11.2990 / Tb427tmp.02.0490 | 46,419 | 20.60 | 0.09 | RNA-editing complex protein KREPB4 | |
| 353 | 1.10E-31 | Tb927.11.6320 / Tb427tmp.02.4190 | 53,247 | 29.30 | 0.59 | MRB1-associated protein | |
| 335 | 6.00E-30 | Tb927.11.14130 / Tb427tmp.01.5720 | 20,965 | 50.40 | 2.22 | ribosomal protein L18, putative | |
| 335 | 6.70E-30 | Tb927.6.2640 / Tb427.06.2640 | 57,920 | 28.30 | 0.9 | importin alpha subunit, putative (TbKap60) | |
| 334 | 7.70E-30 | Tb927.6.5120 / Tb427.06.5120 | 10,438 | 52.70 | 5.89 | 60S acidic ribosomal protein P2, putative | |
| 333 | 9.40E-30 | Tb927.10.5460 / Tb427.10.5460 | 14,586 | 40.00 | 3.04 | 60S ribosomal protein L24, putative | |
| 316 | 4.70E-28 | Tb927.9.5750 / Tb427tmp.160.4250 | 22,410 | 46.10 | 3.31 | tryparedoxin peroxidase | |
| 313 | 1.10E-27 | Tb927.10.15360 / Tb427.10.15360 | 192,223 | 11.00 | 0.16 | hypothetical conserved | DEAD-like helicase |
| 299 | 2.80E-26 | Tb927.11.14000 / Tb427tmp.01.5570 | 28,774 | 37.80 | 1.72 | RNA-binding protein NRBD1 | |
| 296 | 5.70E-26 | Tb927.10.1590 / Tb427.10.1590 | 12,367 | 47.10 | 8.96 | ribosomal protein L36, putative | |
| 295 | 6.00E-26 | Tb927.10.270 / Tb427.10.270 | 15,386 | 25.70 | 3.87 | 60S ribosomal protein L32, putative | |
| 295 | 6.60E-26 | Tb927.5.1820 / Tb427.05.1820 | 10,126 | 57.60 | 1.22 | 60S acidic ribosomal protein, putative | |
| 285 | 5.80E-25 | Tb927.7.3080 / Tb427.07.3080 | 47,288 | 31.90 | 1.01 | hypothetical conserved | Trypanosoma-specific, no human homolog |
| 285 | 6.90E-25 | Tb927.11.11360 / Tb427tmp.01.3170 | 34,668 | 18.70 | 1.04 | receptor for activated C kinase 1 (RACK1) | |
| 282 | 1.20E-24 | Tb927.11.6140 / Tb427tmp.02.4000 | 14,636 | 48.10 | 3 | 40S ribosomal protein S15A, putative | |
| 278 | 3.10E-24 | Tb927.2.820 * | 60,490 | 25.50 | 0.51 | retrotransposon hot spot protein, putative | |
| 275 | 6.60E-24 | Tb927.4.3150 / Tb427.04.3150 | 38,135 | 27.70 | 0.72 | hypothetical conserved | Trypanosoma-specific, no human homolog |
| 273 | 9.70E-24 | Tb927.10.9880 / Tb427.10.9880 | 21,858 | 27.20 | 1.11 | 60S ribosomal protein L18, putative | |
| 273 | 9.90E-24 | Tb927.11.11230 / Tb427tmp.01.3020 | 21,514 | 46.00 | 1.59 | 40S ribosomal protein L14, putative | |
| 270 | 2.20E-23 | Tb927.11.15880 / Tb427tmp.01.7535 | 15,572 | 37.40 | 5.25 | 60S ribosomal protein L27, putative | |
| 268 | 3.50E-23 | Tb927.11.6180 | 16,850 | 49.60 | 2.35 | 60S ribosomal protein L28, putative | |
| 262 | 1.30E-22 | Tb927.7.230 Tb427.07.230 | 11,163 | 59.10 | 1.94 | 40S ribosomal protein S33, putative | |
| 262 | 1.40E-22 | Tb927.6.4280 / Tb427.06.4280 * | 39,023 | 26.20 | 1.33 | glyceraldehyde 3-phosphate dehydrogenase | |
| 257 | 3.80E-22 | Tb927.9.11840 | 24,059 | 32.50 | 1.34 | hypothetical conserved | homolog of human pre-rRNA-processing protein PNO1 (E=9e$^{-65}$) |
| 254 | 8.70E-22 | Tb927.10.8430 / Tb427.10.8430 | 16,051 | 52.10 | 1.75 | 40S ribosomal protein S12, putative | |
| 252 | 1.40E-21 | Tb927.11.6790 / Tb427tmp.02.4620 | 88,968 | 27.00 | 0.45 | predicted WD40 protein | BOP1 |
| 251 | 1.60E-21 | Tb927.3.3310 / Tb427.03.3310 | 25,412 | 37.70 | 1.24 | 60S ribosomal protein L13, putative | |
| 249 | 2.70E-21 | Tb927.10.14580 / Tb427.10.14580 | 19,048 | 37.00 | 1.92 | 60S ribosomal protein L17, putative | |
| 245 | 6.70E-21 | Tb927.7.700 | 74,204 | 19.60 | 0.25 | hypothetical conserved | domain COG2319 (E=5.35e$^{-08}$; WD40 repeat) |
| 230 | 1.80E-19 | Tb927.4.1860 / Tb427.04.1860 | 18,834 | 47.90 | 1.95 | ribosomal protein S19, putative | |
| 229 | 2.80E-19 | Tb927.9.9950 / Tb427tmp.211.1480 | 55,884 | 18.50 | 0.34 | hypothetical conserved | no human homolog |
| 226 | 4.90E-19 | Tb927.5.4170 / Tb427.05.4170 | 11,135 | 47.40 | 3.28 | histone H4, putative | |
| 226 | 5.00E-19 | Tb927.8.900 / Tb427.08.900 | 37,395 | 22.10 | 0.55 | splicing factor TSR1 | |
| 224 | 7.30E-19 | Tb927.10.2890 / Tb11.v5.0650 | 46,563 | 26.50 | 0.43 | enolase | |
| 223 | 1.10E-18 | Tb927.9.13820 / Tb427tmp.211.4511 | 11,069 | 54.70 | 3.28 | kinetoplastid membrane protein KMP-11 | |
| 219 | 2.50E-18 | Tb927.4.2180 | 17,010 | 44.00 | 1.61 | 60S ribosomal protein L35a, putative | |
| 204 | 9.00E-17 | Tb927.11.9720 / Tb427tmp.01.1475 | 9,622 | 80.00 | 2.48 | 40S ribosomal protein S27, putative | |
| 199 | 2.70E-16 | Tb927.3.1370 / Tb427.03.1370 | 12,552 | 58.70 | 2.64 | 40S ribosomal protein S25, putative | |
| 198 | 3.30E-16 | Tb927.6.3740 / Tb427.06.3740 | 71,430 | 24.80 | 0.5 | HSP70, mitochondrial precursor, putative | |
| 197 | 4.50E-16 | Tb927.11.3340 / Tb427tmp.02.0840 | 26,640 | 21.90 | 0.59 | RNA-binding protein RBP34, putative | |
| 197 | 4.20E-16 | Tb927.8.6180 / Tb427.08.6180 | 16,444 | 40.80 | 1.7 | 60S ribosomal protein L26, putative | |
| 196 | 5.10E-16 | Tb927.8.1330 / Tb427.08.1330 * | 30,860 | 25.00 | 0.95 | 60S ribosomal protein L7a, putative | |
| 194 | 8.10E-16 | Tb927.11.10160 / Tb427tmp.01.1920 | 15,257 | 57.00 | 1.23 | 60S ribosomal protein L22, putative | |

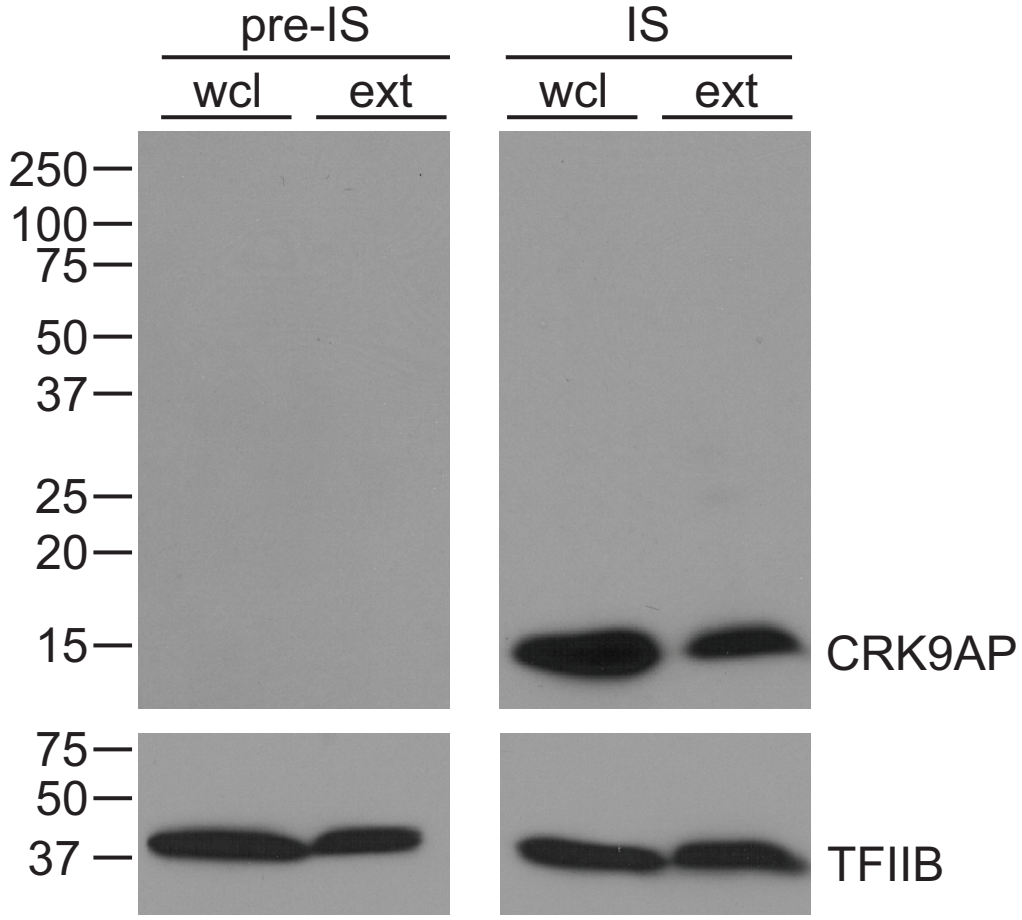
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 189 | 2.50E-15 | Tb927.2.5910 / Tb427.02.5910 | 17,308 | 48.90 | 2.25 | 40S ribosomal protein S13, putative | |
| 188 | 3.10E-15 | Tb927.8.760 / Tb427.08.760 | 35,022 | 15.10 | 0.26 | nucleolar RNA-binding protein Nopp44/46 | |
| 185 | 6.10E-15 | Tb927.10.1080 / Tb427.10.1080 | 15,919 | 33.90 | 1.16 | 40S ribosomal protein S23, putative | |
| 178 | 3.20E-14 | Tb927.9.8200 / Tb427tmp.211.0180 | 76,367 | 20.20 | 0.18 | hypothetical conserved | NOP7 |
| 171 | 1.60E-13 | Tb927.2.5130 | 43,530 | 31.10 | 0.33 | hypothetical conserved | Trypanosomatid-specific, no human homolog |
| 170 | 2.20E-13 | Tb927.10.2240 / Tb427.10.2240 | 66,436 | 21.40 | 0.21 | hypothetical conserved | Domain NTF2 (E=6.74e$^{-15}$) |
| 169 | 2.50E-13 | Tb927.10.220 / Tb427.10.220 | 10,464 | 43.10 | 1.16 | 60S ribosomal protein L37a, putative | |
| 169 | 2.30E-13 | Tb927.10.4740 / Tb427.10.4740 | 7,401 | 40.10 | 3.97 | nucleolar RNA-binding protein, putative | |
| 161 | 1.50E-12 | Tb927.11.10960 | 71,727 | 12.90 | 0.12 | hypothetical conserved | similar to human alpha-ketoglutarate-dependent dioxygenase (E=2e$^{-24}$) |
| 154 | 7.80E-12 | Tb927.11.4190 / Tb427tmp.02.1670 | 59,183 | 21.60 | 0.23 | hypothetical conserved | domain Proliferation-associated_protein_2G4 (E=4.66e$^{-58}$) |
| 152 | 1.20E-11 | Tb927.9.5150 / Tb427tmp.160.3670 * | 13,563 | 21.90 | 0.35 | ribosomal protein S6, putative | |
| 150 | 2.20E-11 | Tb927.11.8200 / Tb427tmp.01.0355 | 12,798 | 44.70 | 0.89 | ribosomal protein S26, putative | |
| 149 | 2.60E-11 | Tb927.2.6090 / Tb427.02.6090 | 12,446 | 42.40 | 2.68 | 60S ribosomal protein L44 | |
| 148 | 3.40E-11 | Tb927.10.10460 / Tb427.10.10460 | 12,562 | 34.20 | 1.63 | histone H2B, putative | |
| 144 | 8.60E-11 | Tb927.08.6030 / Tb427.08.6030 | 17,566 | 14.40 | 0.26 | 60S ribosomal protein L12, putative | |
| 137 | 4.20E-10 | Tb927.9.1810 / Tb427tmp.160.0700 | 15,090 | 42.30 | 2.85 | 60S ribosomal protein L35, putative | |
| 133 | 9.20E-10 | Tb927.9.8740 / Tb427tmp.211.0560 | 36,961 | 13.70 | 0.4 | RNA-binding protein DRBD3 | |
| 131 | 1.40E-09 | Tb927.4.4060 / Tb427.04.4060 | 104,023 | 13.30 | 0.17 | hypothetical conserved | putative chromosome segregation protein |
| 126 | 4.70E-09 | Tb927.7.2820 / Tb427.07.2820 | 14,245 | 26.30 | 1.36 | histone H2A, putative | |
| 125 | 6.20E-09 | Tb927.10.14700 | 37,678 | 14.90 | 0.39 | hypothetical conserved | no human homolog |
| 117 | 4.20E-08 | Tb927.11.15180 / Tb427tmp.01.6835 | 12,938 | 39.10 | 0.87 | hypothetical conserved | no human homolog |
| 115 | 6.00E-08 | Tb927.9.5320 | 55,412 | 16.70 | 0.35 | nucleolar RNA binding protein, putative | |
| 113 | 1.10E-07 | Tb927.10.3500 / Tb427.10.3500 | 96,597 | 5.00 | 0.09 | RNA-binding protein RBSR4, putative | |
| 110 | 1.90E-07 | Tb927.11.11330 / Tb427tmp.02.5450 | 71,403 | 10.80 | 0.19 | HSP70 | |
| 110 | 2.20E-07 | Tb927.9.11270 / Tb427tmp.211.2570 | 61,414 | 19.00 | 0.14 | T-complex protein 1, eta subunit, putative | |
| 108 | 3.30E-07 | Tb927.10.2370 / Tb427.10.2370 | 37,648 | 22.20 | 0.24 | La protein | |
| 107 | 4.50E-07 | Tb927.8.1720 | 68,077 | 6.60 | 0.06 | phosphatidylglycerophosphate synthase, putative | |
| 106 | 4.90E-07 | Tb927.1.180 / Tb427.01.180 | 94,247 | 6.70 | 0.09 | retrotransposon hot spot protein 1, putative | |
| 105 | 7.00E-07 | Tb927.8.3530 / Tb427.08.3530 | 37,810 | 34.00 | 0.39 | glycerol-3-phosphate dehydrogenase | |
| 105 | 7.00E-07 | Tb927.3.1920 / Tb427.03.1920 | 66,646 | 16.00 | 0.06 | NOT5 protein | |
| 103 | 0.0000011 | Tb927.1.2430 / Tb427.01.2430 | 14,741 | 45.50 | 0.73 | histone H3, putative | |
| 97 | 0.0000037 | Tb927.11.9550 / Tb427tmp.01.1310 | 37563 | 5 | 0.12 | replication factor C, subunit 4, putative | |
| 97 | 0.0000039 | Tb927.10.4110 / Tb427.10.4110 | 11545 | 45.7 | 1.01 | 60S ribosomal protein L30 | |
| 96 | 0.0000057 | Tb927.7.3680 / Tb427.07.3680 | 16972 | 31.8 | 0.62 | ubiquitin/ribosomal protein S27a, putative | |
| 95 | 0.0000061 | Tb927.3.3950 | 62209 | 9.7 | 0.07 | hypothetical conserved | no unambiguous human homolog |
| 95 | 0.0000067 | Tb927.11.8050 / Tb427tmp.01.0490 | 59922 | 12.5 | 0.15 | hypothetical conserved | putative SAS10 homolog (U3 processome complex, E=1.67e-$^{17}$) |
| 93 | 0.0000093 | Tb927.7.1780 / Tb427.07.1780 | 25956 | 16.5 | 0.17 | Adenine phosphoribosyltransferase, putative | |
| 92 | 0.000014 | H25N7.12 | 97751 | 18.7 | 0.09 | retrotransposon hot spot protein, RHS4 | |
| 91 | 0.000018 | Tb927.8.4660 / Tb427.08.4660 | 117829 | 7.9 | 0.07 | hypothetical conserved | putative homolog of human exportin-7 (E=3e$^{-23}$) |
| 90 | 0.000022 | Tb927.9.5730 / Tb427tmp.160.4240 | 47586 | 10.6 | 0.19 | nucleosome assembly protein-like protein | |
| 89 | 0.000023 | Tb927.11.10030 / Tb427tmp.01.1790 | 8217 | 62.9 | 3.3 | 60S ribosomal protein L29, putative | |
| 89 | 0.000024 | Tb927.8.3750 / Tb427.08.3750 | 54281 | 13.4 | 0.16 | nucleolar protein, putative | |
| 89 | 0.000028 | Tb927.10.4560 / Tb427.10.4560 | 94274 | 11.3 | 0.09 | elongation factor 2 | |
| 87 | 0.000039 | Tb927.2.1140 / Tb427.02.1140 | 52161 | 20.7 | 0.17 | retrotransposon hot spot protein, putative | |
| 86 | 0.000051 | Tb927.10.10360 / Tb427.10.10360 * | 367419 | 3.5 | 0.01 | microtubule-associated protein, putative | |
| 84 | 0.000083 | Tb927.3.3130 / Tb427.03.3130 | 177874 | 10.1 | 0.02 | hypothetical protein, conserved | no human homolog |
| 84 | 0.00009 | Tb927.11.16890 / Tb427tmp.01.8650 | 228855 | 11.4 | 0.02 | hypothetical protein, conserved | Tetratricopeptide repeat protein (TPR), putative |
| 83 | 0.000092 | Tb927.11.740 / Tb427tmp.03.0410 | 17809 | 8 | 0.26 | eukaryotic translation initiation factor 5a, putative | |
| 83 | 0.00011 | Tb927.10.13720 Tb427.10.13720 | 40636 | 6.5 | 0.11 | RNA-binding protein, putative (RBP29) | |
| 83 | 0.00011 | Tb09.v4.0126 | 55005 | 17.8 | 0.08 | variant surface glycoprotein, pseudogene | |
| 82 | 0.00013 | Tb927.2.560 / Tb427.02.560 | 95966 | 27.9 | 0.09 | retrotransposon hot spot protein, putative | |
| 81 | 0.00016 | Tb927.8.7530 / Tb427.08.7530 | 39779 | 8.8 | 0.11 | 3,2-trans-enoyl-CoA isomerase, putative | |
| 81 | 0.00016 | Tb927.10.15870 / Tb427.10.15870 | 26965 | 8.2 | 0.16 | RNA binding protein, putative | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 80 | 0.00022 | Tb927.1.1350 | 96521 | 9 | 0.04 | kinesin heavy chain, putative | |
| 79 | 0.00027 | Tb927.1.2100 / Tb427.01.2100 | 126075 | 5.6 | 0.03 | calpain-like cysteine peptidase, putative | |
| 78 | 0.0003 | Tb927.10.170 / Tb427.10.170 | 48363 | 8.6 | 0.09 | pseudouridine synthase, Cbf5p | |
| 76 | 0.00047 | Tb927.7.2980 / Tb427.07.2980 | 21229 | 12.4 | 0.47 | hypothetical protein, conserved | Nitroreductase 4 |
| 76 | 0.00057 | Tb927.10.14750 | 31653 | 30.9 | 0.3 | fibrillarin, putative | |
| 75 | 0.0006 | Tb927.6.5040 / Tb427.06.5040 | 24494 | 26 | 0.65 | ribosomal protein L15, putative | |
| 75 | 0.00065 | Tb11.v5.0537 | 70086 | 5.3 | 0.13 | polyubiquitin, putative | |
| 73 | 0.00094 | Tb927.2.2390 / Tb427.02.2390 | 16769 | 33.5 | 0.28 | hypothetical protein, conserved | no human homolog |
| 73 | 0.0011 | Tb927.7.5000 / Tb427.07.5000 | 29405 | 26.9 | 0.15 | 60S ribosomal protein L19, putative | |
| 72 | 0.0012 | Tb927.4.2040 / Tb427.04.2040 | 20796 | 39.7 | 0.48 | ALBA3 | |
| 71 | 0.0015 | Tb927.9.13650 / Tb427tmp.211.4460 | 20639 | 13.9 | 0.22 | ADP-ribosylation factor, putative | |
| 71 | 0.0015 | Tb927.2.100 | 94873 | 5.3 | 0.09 | retrotransposon hot spot protein 1, putative | |
| 70 | 0.0019 | Tb927.11.3120 / Tb427tmp.02.0620 | 74731 | 15.7 | 0.12 | nucleolar GTP-binding protein 1 | |
| 70 | 0.0021 | Tb927.4.3060 / Tb427.04.3060 | 17694 | 15.5 | 0.59 | hypothetical protein, conserved | no human homolog |
| 70 | 0.0022 | Tb927.6.2790 / Tb427.06.2790 | 36934 | 18.5 | 0.25 | L-threonine 3-dehydrogenase, putative | |
| 69 | 0.0027 | Tb927.8.4450 / Tb427.08.4450 | 47151 | 12.4 | 0.09 | RNA-binding protein, putative (RBP11) | |
| 67 | 0.0039 | Tb927.11.10280 / Tb427tmp.01.2050 | 131127 | 3.4 | 0.03 | zinc carboxypeptidase, putative | |
| 66 | 0.0054 | Tb927.10.9400 / Tb427.10.9400 | 31611 | 13.6 | 0.14 | SF1 (branch point binding protein) | |
| 65 | 0.0059 | Tb927.1.1700 / Tb427.01.1700 | 55121 | 11.8 | 0.16 | AATF protein, putative | |
| 65 | 0.0068 | Tb927.11.9590 / Tb427tmp.01.1350 | 48456 | 10.4 | 0.09 | S-adenosylhomocysteine hydrolase, putative | |
| 64 | 0.0089 | Tb927.10.8170 / Tb427.10.8170 | 144316 | 8 | 0.03 | nuclear pore complex protein (NUP155), putative | |
| 63 | 0.0093 | Tb927.10.15180 / Tb427.10.15180 * | 41221 | 5.6 | 0.11 | nucleosome assembly protein, putative | |
| 63 | 0.0094 | Tb11.v5.0366 | 120510 | 8.4 | 0.03 | hypothetical protein, conserved | putative chromosome segregation ATPase |

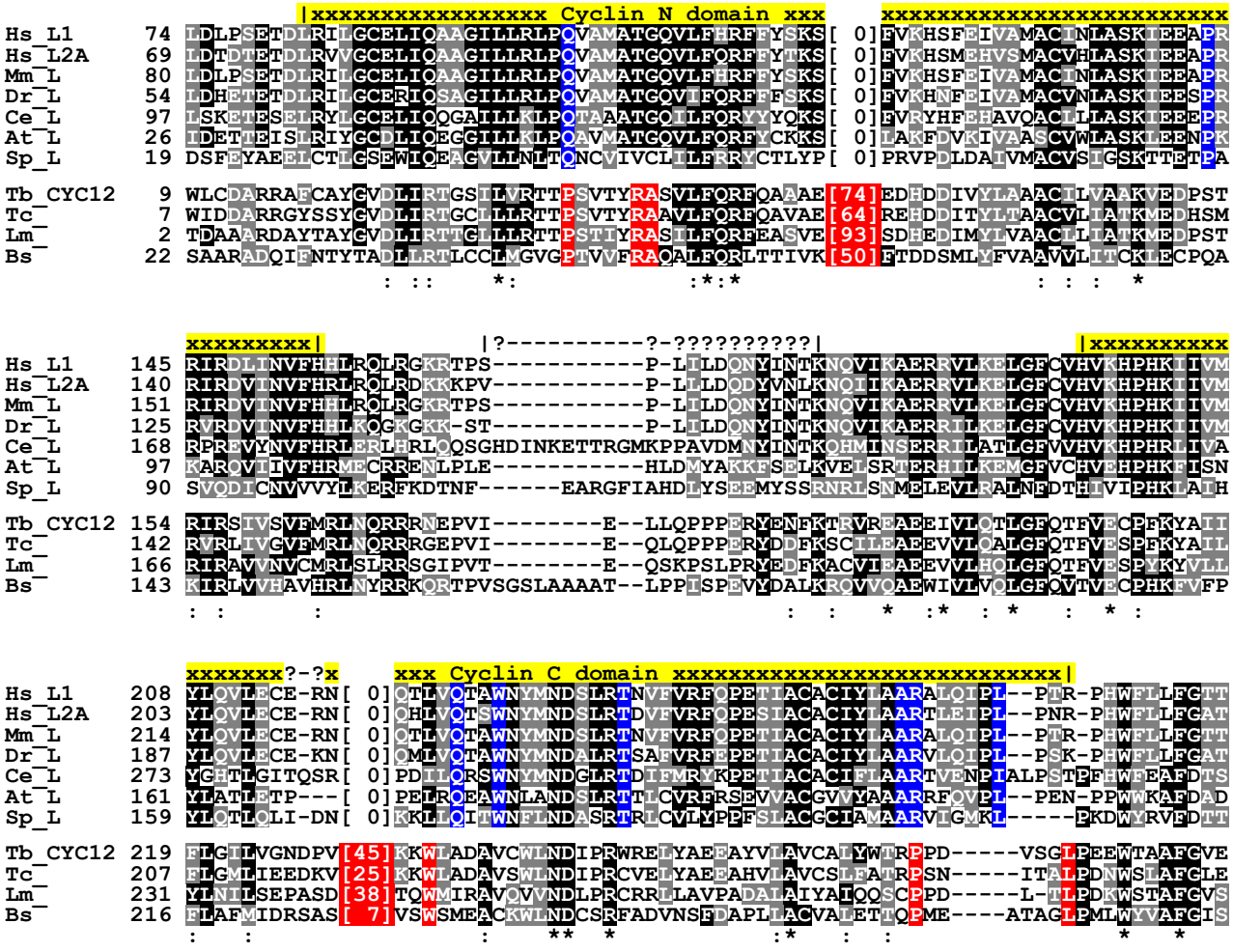
List of proteins that were identified in two CRK9-PTP tandem affinity purifications by LC/MS/MS. The list shows the protein IDs of the second, more comprehensive mass spectrometric analysis. All protein identifications with an Expect value smaller than 0.001 are listed and ranked according to their Mascot Scores. Proteins that were also identified in the first analysis are marked with an asterisk after their accession numbers. Peptides were identified in both the T. brucei brucei 927 and 427 genome databases at www.TriTrypDB.org. Annotations are according to the 927 database. Top-scoring and "hypothetical, conserved" proteins were analyzed by standard BLASTP searches of the Human Genome database. CRK9 complex subunits are highlighted in yellow, ribosomal proteins in green, non-ribosomal proteins in orange, and standard PTP purification contaminants in gray.



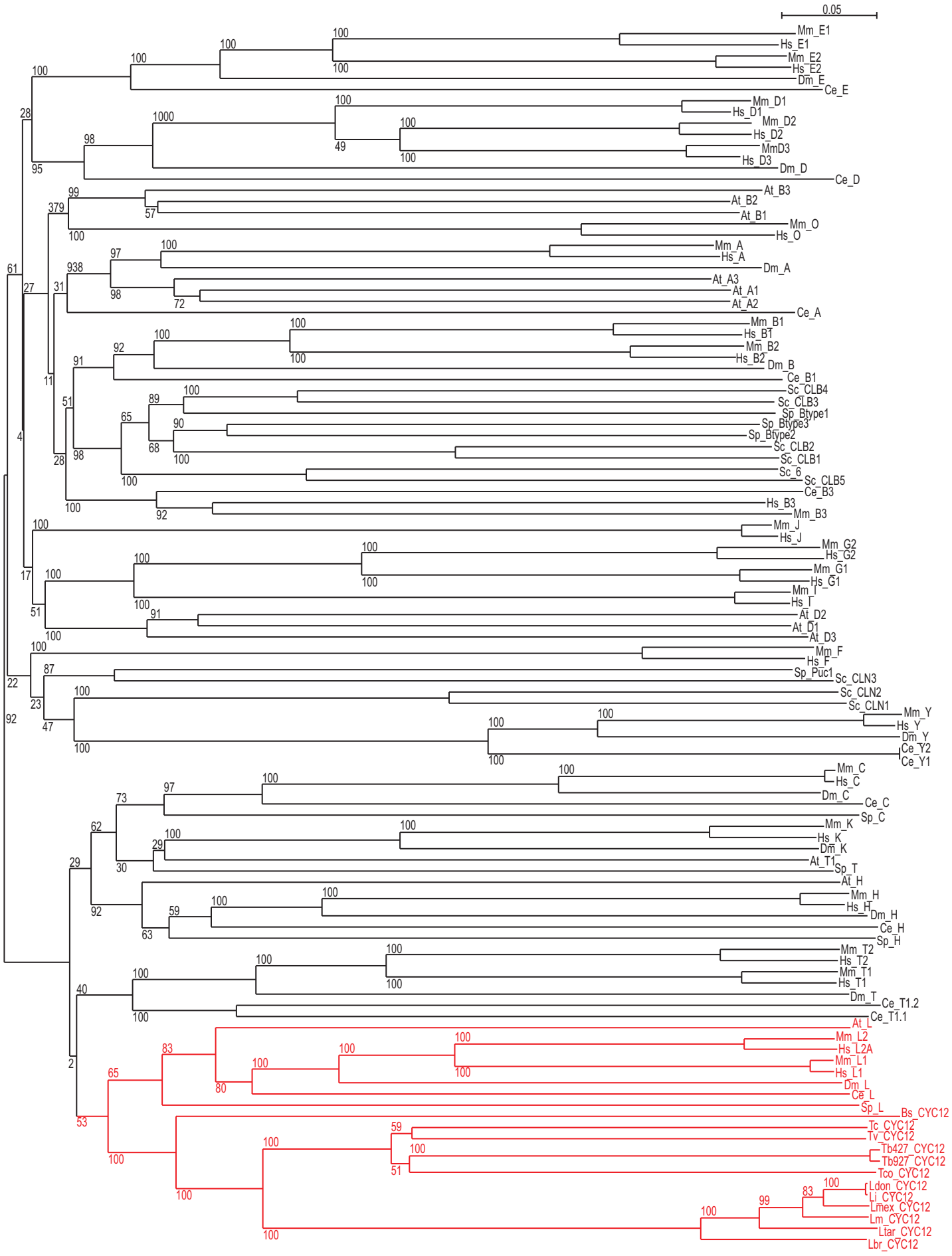

**S2 Figure. CRK9AP is conserved among kinetoplastid organisms.** Kinetoplastid CRK9AP sequences were aligned using the Clustal Omega server of the European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/services/web/toolform.ebi?tool=clustalo) at default para-meters [1]. Positions with more than 50% identity or similarity are highlighted in black or gray, respectively. Dashes indicate that a corresponding residue is missing. Sequences were obtained from www.TriTrypDB.org or www.GeneDB.org and comprise those of *T. brucei brucei* strains 427 (Tb427, accession number Tb427.03.4170) and 927 (Tb927, Tb927.3.4170), *Trypanosoma vivax* (Tv, TvY486_0303410), *Trypanosoma cruzi* CL Brener Esmeraldo-like (Tc-el; TcCLB.509669.80) and Non-Esmeraldo-like (Tc-nel; TcCLB.506175.30), *Trypanosoma grayi* (Tgr, Tgr.163.1040), *Crithidia fasciculata* (Cfa, CfaC1_25_1880), *Leishmania braziliensis* (Lbr, LbrM.29.1690), *Leishmania tarentolae* (Lta, LtaP29.1740), *Leishmania mexicana* (Lmx, LmxM.08_29.1585), *Leishmania major* (Lm, LmjF.29.1585), *Leishmania infantum* (Li, LinJ.29.1710), *Leishmania donovani* (Ld, LdBPK_291710.1), and the bodonid *Bodo saltans* (Bs, BS14910.1..pep).



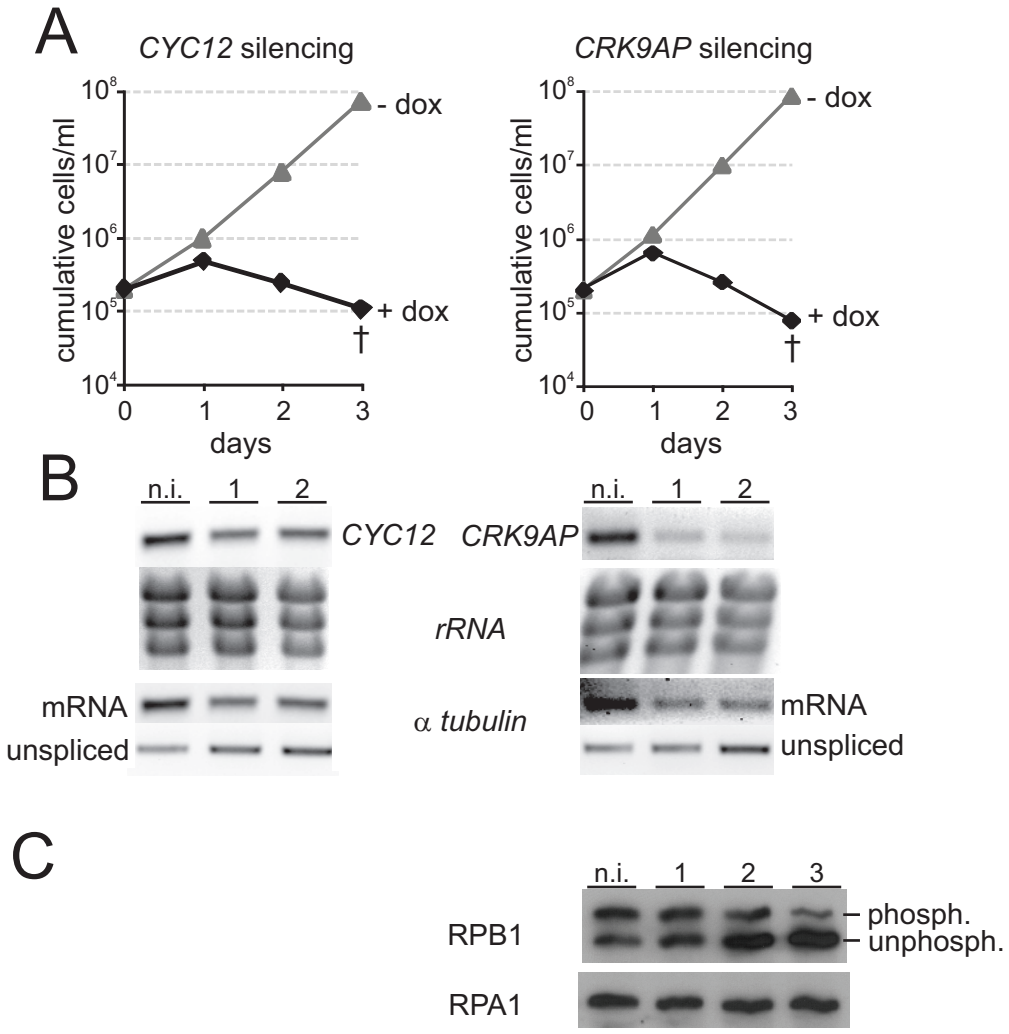

**S3 Figure. Generation of a highly specific rat anti-CRK9AP immune serum.** CRK9AP was expressed in *Escherichia coli* as a C-terminal fusion to glutathione S-transferase and purified from bacterial extract by glutathione affinity chromatography. By injecting the purified protein into the rat bloodstream, [pre-]immune serum was obtained according to a published protocol [1]. Pre-immune (pre-IS) and $\alpha$-CRK9AP immune sera (IS) were used to probe whole cell lysates (wcl) and crude extract (extr) of procyclic *Trypanosoma brucei brucei* strain 427. As a loading control, transcription factor TFIIB was detected on the same blots. Marker sizes in kDa are indicated on the left.



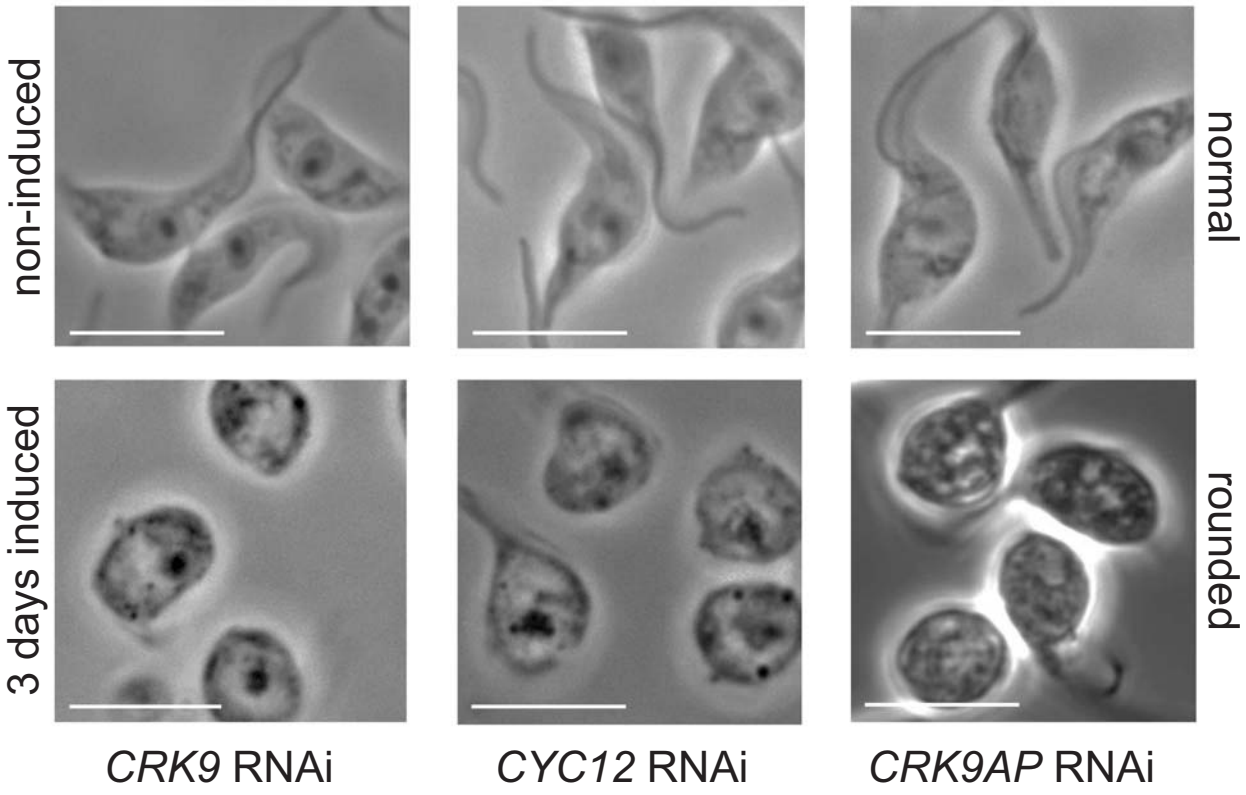

**S4 Figure. Multiple sequence alignment of the CCL1 domain of eukaryotic cyclin L and kinetoplastid CYC12 proteins.** Cyclin L and CYC12 sequences were aligned using the Clustal Omega server at default parameters. Shown are the sequences of the CCL1 domain (COG5333) as defined in human cyclin L1 according to the *Conserved Domain Database* [3]. Dashes indicate that a corresponding residue is missing. Numbers in parentheses specify number of residues without significant sequence similarity. Positions with more than 50% identity or similarity are highlighted in black or gray, respectively. Identical positions in model organisms without similarity in kinetoplastids were highlighted in blue and insertions or unique identical positions in kinetoplastids were highlighted in red. Stars and colons denote positions that are identical or similar in all sequences analyzed. Yellow highlighting indicates cyclin folds 1 and 2 within the CCL1 domain and question marks indicate positions of the human cyclin L1 sequence that were not recognized as part of the CCL1 domain. Cyclin L sequences are from Homo sapiens (Hs; L1, accession number NP_064703; L2A, NP_112199), *Mus musculus* (Mm, NP_064321), *Danio rerio* (Dr, NP_956034), *Caenorhabditis elegans* (Ce, NP_506007), *Arabidopsis thaliana* (At, NP_565622) and *Schizosaccharomyces pombe* (Sp, NP_593045). Kinetoplastid CYC12 sequences, obtained from www.TriTrypDB.org or www.GeneDB.org, were from *T. brucei* (Tb, accession number Tb927.10.9160), *Trypanosoma cruzi* (Tc, TcCLB.503525.20), *Leishmania major* (Lm, LmjF.36.5640), and the bodonid *Bodo saltans* (Bs, BS70770.1).


0.05
Mm_E1
Hs_E1
Mm_E2
Hs_E2
Dm_E
Ce_E
Mm_D1
Hs_D1
Mm_D2
Hs_D2
MmD3
Hs_D3
Dm_D
Ce_D
At_B3
At_B2
At_B1
Mm_O
Hs_O
Mm_A
Hs_A
Dm_A
At_A3
At_A1
At_A2
Ce_A
Mm_B1
Hs_B1
Mm_B2
Hs_B2
Dm_B
Ce_B1
Sc_CLB4
Sc_CLB3
Sp_Btype1
Sp_Btype3
Sp_Btype2
Sc_CLB2
Sc_CLB1
Sc_6
Sc_CLB5
Ce_B3
Hs_B3
Mm_B3
Mm_J
Hs_J
Mm_G2
Hs_G2
Mm_G1
Hs_G1
Mm_I
Hs_I
At_D2
At_D1
At_D3
Mm_F
Hs_F
Sp_Puc1
Sc_CLN3
Sc_CLN2
Sc_CLN1
Mm_Y
Hs_Y
Dm_Y
Ce_Y2
Ce_Y1
Mm_C
Hs_C
Dm_C
Ce_C
Sp_C
Mm_K
Hs_K
Dm_K
At_T1
Sp_T
At_H
Mm_H
Hs_H
Dm_H
Ce_H
Sp_H
Mm_T2
Hs_T2
Mm_T1
Hs_T1
Dm_T
Ce_T1.2
Ce_T1.1
At_L
Mm_L2
Hs_L2A
Mm_L1
Hs_L1
Dm_L
Ce_L
Sp_L
Bs_CYC12
Tc_CYC12
Tv_CYC12
Tb427_CYC12
Tb927_CYC12
Tco_CYC12
Ldon_CYC12
Li_CYC12
Lmex_CYC12
Lm_CYC12
Ltar_CYC12
Lbr_CYC12
cell cycle
transcription

**S5 Figure. CYC12 clusters phylogenetically with cyclin L of model organisms**. Amino acid sequences of cyclins from *Homo sapiens* (Hs), *Mus musculus* (Mm), *Drosophila melanogaster* (Dm), *Caenorhabditis elegans* (Ce), *Arabidopsis thaliana* (At), *Saccharomyces cerevisiae* (Sc), and *Schizosaccharomyces pombe* (Sp) as well as kinetoplastid CYC12 from *Trypanosoma brucei brucei* strains 427 (Tb427) and 927 (Tb927), *Trypanosoma congolense* (Tco), *Trypanosoma vivax* (Tv) *Trypanosoma cruzi* (Tc), *Leishmania major* (Lm), *Leishmania infantum* (Li), *Leishmania donovani* (Ldon), *Leishmania mexicana* (Lmex), *Leishmania tarentolae* (Ltar), *Leishmania braziliensis* (Lbr), and the bodonid *Bodo saltans* (Bs) were aligned using the Clustal Omega server at http://www.ebi.ac.uk [4]. The multiple sequence alignment was imported into the ClustalX software package [5] and phylogenetically analyzed using the neighborhood joining method. Bootstrap values were obtained by sampling a thousand replicates and are indicated as percentages. The node for the cyclin L/CYC12 cluster is drawn in red. Cyclin clusters of cell cycle-regulating and transcriptional CDKs, according to Ma *et al*. [6], are indicated. The cyclin sequences were obtained from the following accession numbers: Hs_A (CAA35986.1), Hs_B1 (CAO99273.1), Hs_B2 (AAI05087.1), Hs_B3 (CAC94915.1), Hs_C (AAH41123.1), Hs_D1 (AAH23620.1), Hs_D2 (CAA48493.1), Hs_D3 (AAA52137.1), Hs_E1 (AAH35498.1), Hs_E2 (AAC78145.1), Hs_F (AAB60342.1), Hs_G1 (AAC78145.1), Hs_G2 (AAC41978.1), Hs_H (AAA57006.1), Hs_I (AAF43786.1), Hs_J (AAH43175.1), Hs_K (AAH43175.1), Hs_L1 (AAH43175.1), Hs_L2A (Q96S94.1), Hs_O (NP_066970.3), Hs_T1 (AAC39664.1), Hs_T2 (AAW56073.1), Hs_Y (AAH94815.1), Mm_A (CAA81331.1), Mm_B1 (AAH85238.1), Mm_B2 (AAH08247.1), Mm_B3 (AAI38356.1), Mm_C (AAH03344.2), Mm_D1 (AAO13813.1), Mm_D2 (AAH49086.1), Mm_D3 (AAC53363.1), Mm_E1 (AAI38663.1), Mm_E1 (AAC80527.1) Mm_F (AAA63152.1), Mm_G1 (AAC42082.1), Mm_G2 (AAC32372.1), Mm_H (AAH38861.1), Mm_I (AAF43391.1), Mm_J (AAI20923.1), Mm_K (AAH27297.1), Mm_L1 (AAH94383.1), Mm_L2 (AAI32296.1), Mm_O (AAI47760.1), Mm_T1 (AAD13656.1), Mm_T2 (AAH54122.1), Mm_Y (NP_080760.2), Dm_A (NP_524030.2), Dm_B (AAF46904.1), Dm_C (CAA44720.1) Dm_D (NP_523355.2), Dm_E (AAF53477.1), Dm_H (NP_524207.1), Dm_K (AAN11146.1), Dm_L (putative, Dm_T (AAS64974.1), Dm_Y (AAF53122.1), Ce_A (AAA84393.1), Ce_B1 (Q10653.1), Ce_B3 (AAA84395.1), Ce_C (Q9TYP2.2), Ce_D (AAC35273.1), Ce_E (AAM78547.1), Ce_H (NP_494564.2), Ce_L (AAS64750.1), Ce_T1.1 (P34425.1), Ce_T1.2 (P34424.2), Ce_Y1 (NP_498857.2), Ce_Y2 (NP_498858.2), At_A1 (Q9C6Y3.1), At_A2 (AED93433.1), At_A3 (NP_199122.1), At_B1 (P30183.2), At_B2 (Q39068.2), At_B3 (NP_173083.3), At_D1 (P42751.3), At_D2 (P42752.3), At_D3 (P42753.3), At_H (BAB72144.1), At_L (Q8RWV3.2), At_T1 (NP_174775.1), Sc_CLB1 (CAA97112.1), Sc_CLB2 (CAA44195.1), Sc_CLB3 (CAA49201.1), Sc_CLB4 (CAA49202.1), Sc_CLB5 (AAA34503.1), Sc_6 (NP_011623.3), Sc_CLN1 (NP_013926.1), Sc_CLN2 (CAA97982.1) Sc_CLN3 (NP_009360.1), Sp_C (NP_595953.1), Sp_H (NP_595776.1), Sp_L (NP_593045.1), Sp_T (NP_596149.1) Sp_Puc1 (NP_596539.1), Sp_Btype1 (NP_588110.2), Sp_Btype2 (NP_595171.1), Sp_Btype3 (NP_593889.1), Tb927_CYC12 (Tb927.10.9160), Tb427_CYC12 (Tb427.10.9160), Tc_CYC12 (TcCLB.503525.20), Tv_CYC12 (TvY486_1009010), Tco_CYC12 (TcIL3000_10_7930), Lm_CYC12 (LmjF.36.5640), Ltar_CYC12 (LtaP36.5790), Lmex_CYC12 (LmxM.36.5640), Ldon_CYC12 (LdBPK_365890.1), Li_CYC12 (LinJ.36.5890), Lbr_CYC12 (LbrM.35.5920) and Bs_CYC12 (BS70770.1).




**S6 Figure. *CYC12* and *CRKAP* silencing in BFs.** (A) Growth curves of uninduced (- dox) BF cultures or cultures in which *CYC12* or *CRK9AP* was conditionally silenced by the addition of doxycyline to the medium (+ dox). Note that a second cell line for each gene knockdown was analyzed in parallel, producing nearly congruent growth curves (data not shown). The cross indicates that subsequent to day 3 of induction no intact cells were detected in cultures. (B) RNA analyses. Total RNA was prepared from [un-]induced cells and the relative amounts of mature CYC12/CRK9AP and α tubulin mRNA was determined by reverse transcription using an oligo-dT primer and semi-quantitative PCR, performed in the linear range of the amplification reaction. Unspliced, α tubulin pre-mRNA was analyzed by reverse transcription using random hexamers and PCR with an oligonucleotide that hybridized upstream of the α tubulin SL addition site. rRNA served as a loading control and was detected after RNA separation on agarose gel by ethidium bromide staining. (C) Both gene knockdowns led to dephosphorylation of the RNA pol II subunit RPB1. Immunoblot analysis of whole cell lysates of [un-]induced cells detecting RPB1 with an immune serum directed against the CTD. As a loading control, the similar sized protein RPA1, the largest subunit of RNA pol I, was detected on the same blot.




**S7 Figure. Depletion of CRK9, CYC12 or CRK9AP leads to round-up of cells.** Microscopic images of un-induced procyclic cells that look normal (top row) and of rounded cells from the same cell lines (bottom row) when CRK9, CYC12 or CRK9AP was depleted for three days. White scale bars correspond to 10 μm.

**Supplemental References**


1. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7: 539.

2. Schimanski B, Brandenburg J, Nguyen TN, Caimano MJ, Günzl A (2006) A TFIIB-like protein is indispensable for spliced leader RNA gene transcription in *Trypanosoma brucei*. Nucleic Acids Res 34: 1676-1684.

3. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, et al. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39: D225-D229.

4. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, et al. (2013) Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res 41: W597-600.

5. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948.

6. Ma Z, Wu Y, Jin J, Yan J, Kuang S, et al. (2013) Phylogenetic analysis reveals the evolution and diversification of cyclins in eukaryotes. Mol Phylogenet Evol 66: 1002-1010.

**Supporting information**

# The spliceosomal PRP19 complex of trypanosomes

***Daniela L. Ambrósio,[1,2] Nitika Badjatia,[1] and Arthur Günzl [1 #]***

[1] Department of Genetics and Genome Sciences, University of Connecticut Health Center, 400 Farmington Avenue, Farmington, CT 06030-6403, USA

[2] Current address: Department of Chemistry, Federal University of Lavras, Campus, Lavras, MG, 37200-000, Brazil

[#] For correspondence, Email: gunzl@uchc.edu; Tel: (860) 679-8878; Fax: (860) 679-8345.

**Content**

**Table S1.** List of oligonucleotides

| Purpose | Description | Sequence |
|---|---|---|
| RT-PCR | *SPF27* mRNA | 5′-AAGCTTACGCGTCCCATGGTGCTTCCTTAC-3′<br>5′-TCTAGACCCGAACGTTTCTCAGCC-3′ |
| RT-PCR | 7SL RNA | 5′-TTGCTCTGTAACCTTC-3′<br>5′-TCTACAGTGGCGACCTCAAC-3′ |
| RT-PCR | *ORC1* mRNA | 5′-AAGCTTACGCGTCGAAGGAGGGACAGTAAC-3′<br>5′-TCTAGAAAAACGCATACGGTAGAC-3′ |
| RT-[q]PCR | α tubulin pre-mRNA | 5′-GTAAGTGGTGGTGGCGTAAG-3′<br>5′-CAATGTGGATGCAGATAGCC-3′ |
| RT-[q]PCR | α tubulin mRNA | 5′-ACAGTTTCTGATCTATATTGATCTT-3′<br>5′-GAGAGTTGCTCGTGGTAGGC-3′ |
| Competitive RT-PCR | *PAP* [pre-]mRNA | 5′-GTGCAGCGGCACTCCCAAAAC-3′<br>5′-CGTTAAAACAGATGGACAAATC-3′ |
| RT-qPCR | SL RNA | 5′-ACAGTTTCTGTACTATATTG-3′<br>5′-CGACCCCACCTTCCAGATTC-3′ |
| RT-qPCR | *RPB7* pre-mRNA | 5′-CCACTCGAAGGAGTAGTTTTC-3′<br>5′-TTATGTGCACTTGCTGGTG-3′ |
| RT-qPCR | *RPB7* mRNA | 5′-CATGGGCCCGAGAGGAATATAAAAGTGGAGCCTC-3′<br>5′-ATTCTGATTTGTGCGGGC-3′ |
| RT-qPCR | *PAP* pre-mRNA (exon 1-intron) | 5′-GTGCAGCGGCACTCCCAAAAC-3′<br>5′-GGGATTAAGGAAAGAACTCAC-3′ |
| RT-qPCR | 18S rRNA | 5′-TCATCAAACTGTGCCGATTAC-3′<br>5′-CTATTGAAGCAATATCGG-3′ |
| Primer extension | SL_PE (SL RNA)<br>SL40 (SL RNA)<br>7SL_PE (7SL RNA)<br>U1_PE (U1 snRNA)<br>U2f (U2 snRNA)<br>U4_PE (U4 snRNA)<br>U5_PE (U5 snRNA)<br>U6_PE (U6 snRNA) | 5′-CGACCCCACCTTCCAGATTC-3′<br>5′-CTACTGGGAGCTTCTCATAC-3′<br>5′-GAACCCCCGCTTGTC-3′<br>5′-AGCACGGCGCTTTCGTGATG-3′<br>5′-ACAGGCAACAGTTTTGATCC-3′<br>5′-TACCGGATATAGTATTGCAC-3′<br>5′-GGGAGAGTGCTAATCTTCTC-3′<br>5′-GAACCCCCGCTTGTC-3′ |

The gene-specific sequence is underlined when oligonucleotides carry additional nucleotides.

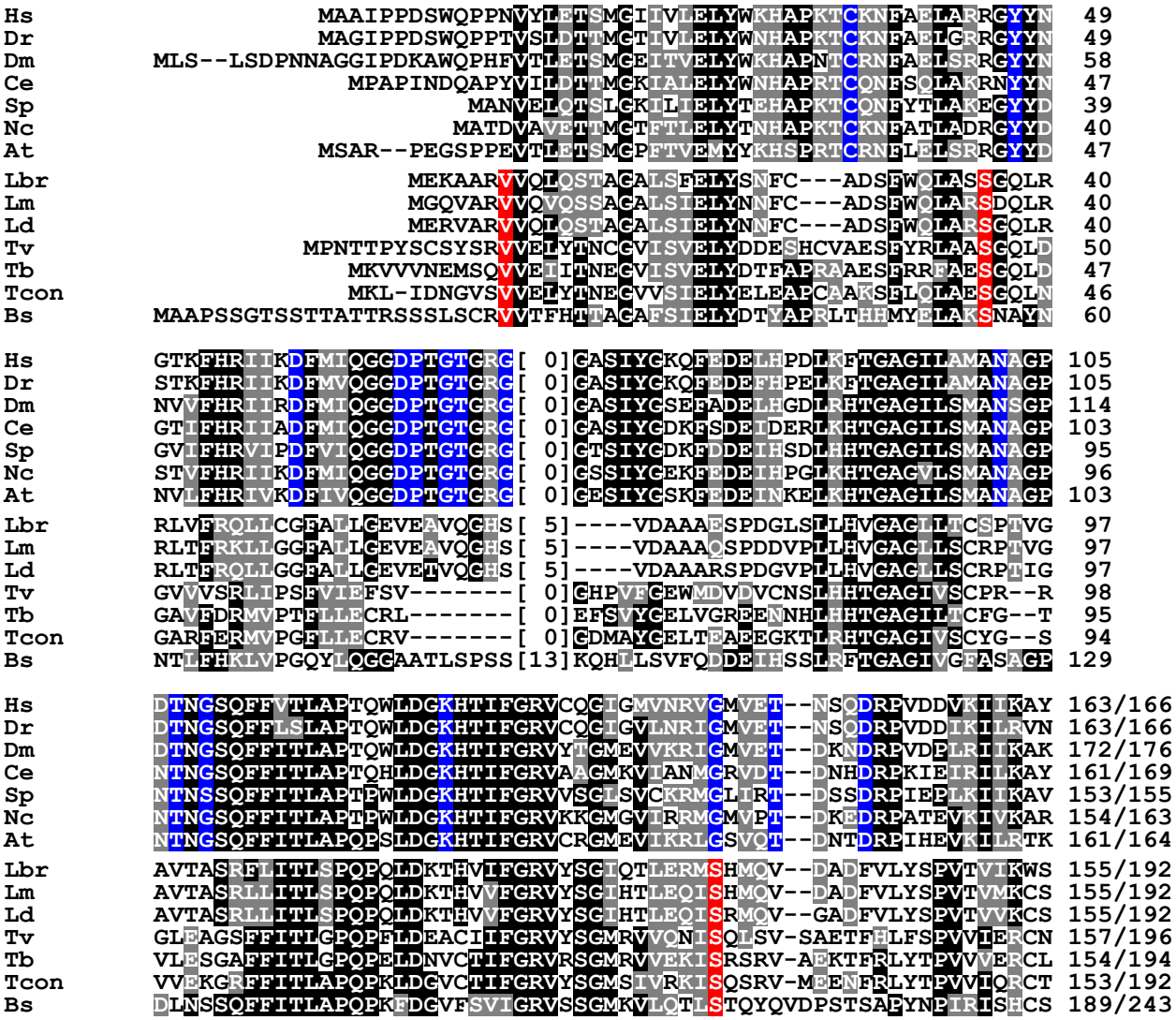


**Fig. S1.** Multiple sequence alignment of PPIL1 orthologues.

Clustal Omega (McWilliam *et al.*, 2013) was used at default parameters to align PPIL1 amino acid sequences from *Homo sapiens* (Hs, accession number NP_057143), *Danio rerio (Dr,* NP_001029350), *Drosophila melanogaster* (Dm, NP_523874), *Caenorhabditis elegans* (Ce, NP_501118), *Schizosaccharomyces pombe* (Sp, NP_593308), *Neurospora crassa* (Nc, XP_964739), *Arabidopsis thaliana* (At, NP_181157), and from the kinetoplastid species *Leishmania braziliensis* (Lbr, LbrM.23.0140), *Leishmania major* (Lm, LmjF.23.0125), *Leishmania donovani* (Ld, LdBPK_230140.1), *Trypanosoma vivax* (Tv, TvY486_0801580), *Trypanosoma brucei* (Tb, Tb927.8.2090), *Trypanosoma congolense* (Tc, TcIL3000_8_2090), and *Bodo saltans* (Bs, BS78785.1..pep). Positions with more than 50% similarity or identity are shaded in gray and black, respectively. Identical positions in model organisms without conservation in kinetoplastids are shaded blue and identical position is kinetoplastids without conservation in model organisms are shaded in red. A hyphen indicates lack of an amino acid at this position. Numbers in parentheses specify lengths of non-conserved insertions. *Saccharomyces cerevisiae* and *Trypanosoma cruzi* appear to lack a PPIL1 homolog.

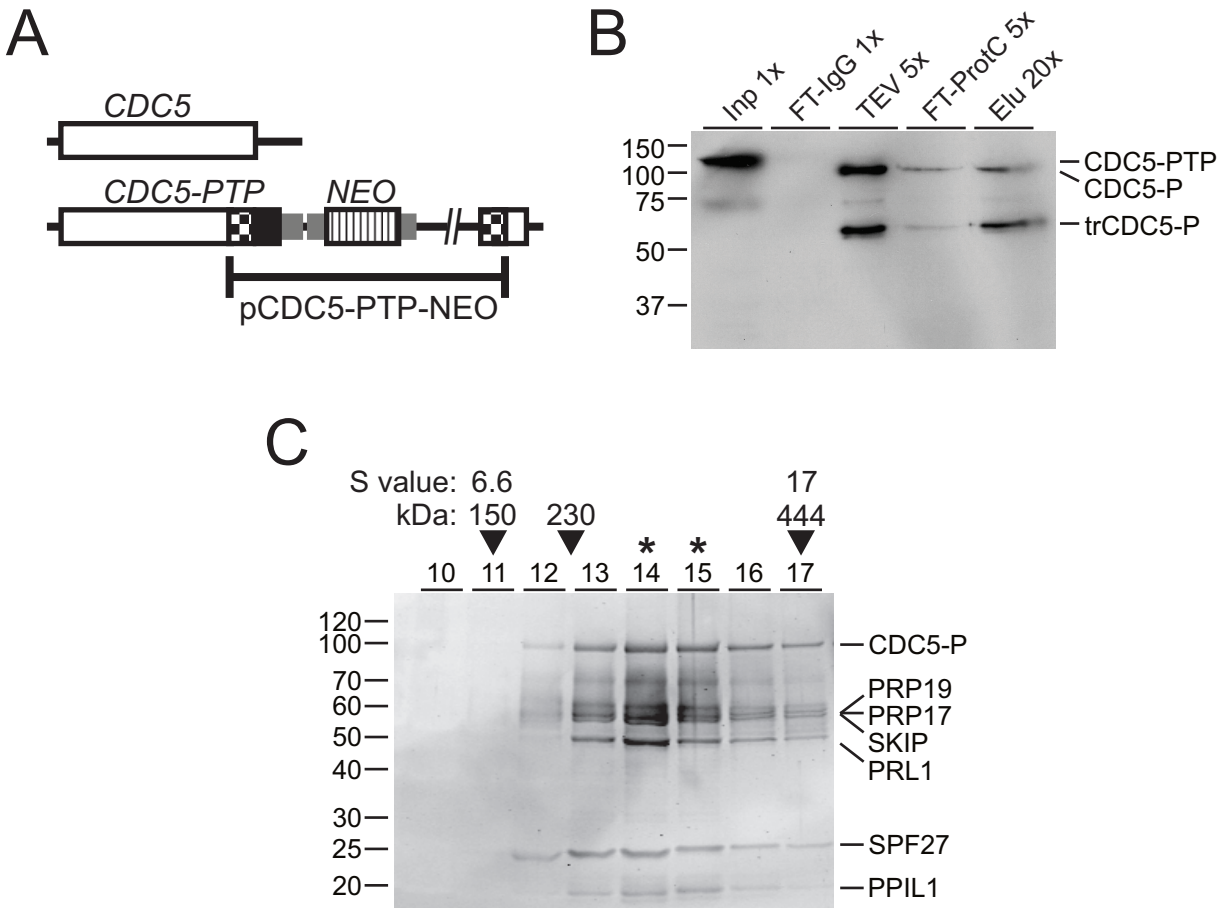



**Fig. S2.** PTP tagging and tandem affinity purification of CDC5 isolated the complete PRP19 complex

A. Schematic (not to scale) of the *CDC5* locus after integration of pCDC5-PTP-NEO, fusing the PTP coding sequence (black box) to an endogenous allele.

B. Immunoblot monitoring of the CDC5-PTP tandem affinity purification CDC5-PTP and CDC5-P (after removal of the ProtA domains) were detected with the anti-ProtC HPC4 antibody in crude extract (Inp), flowthrough of the IgG column (FT-IgG), TEV protease eluate (TEV), flowthrough of the anti-ProtC column (FT-ProtC) and the final eluate (Elu). x-Values indicate relative amounts loaded. Apparently, during extract preparation and purification, part of CDC5 was converted into a truncated form in which an N-terminal domain was cleaved off (trCDC5).

C. The final eluate of a CDC5-PTP tandem affinity purification was sedimented through a linear sucrose gradient, fractionated, and visualized by Sypro Ruby staining of an SDS-PAGE gel, the same as with PRP19 shown in figure 1D. Protein bands were excised and protein identities confirmed by LC/MS/MS. Asterisks indicate the sedimentation peak.

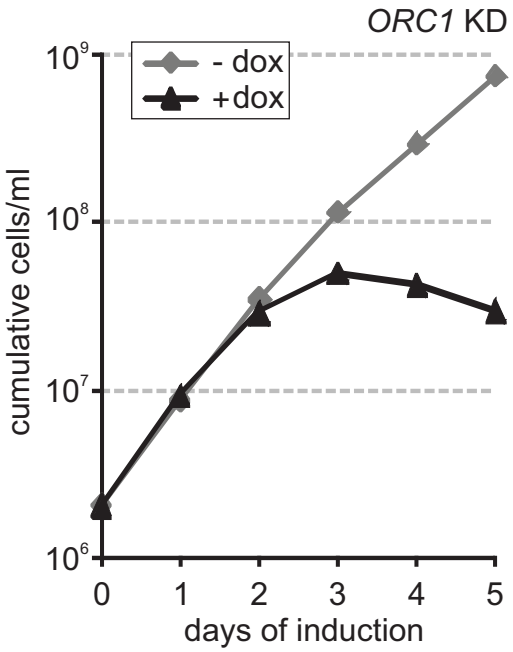



**Fig. S3.** *ORC1* silencing

Procyclic trypanosome culture growth in the absence and presence of doxycycline which induced *ORC1* silencing.

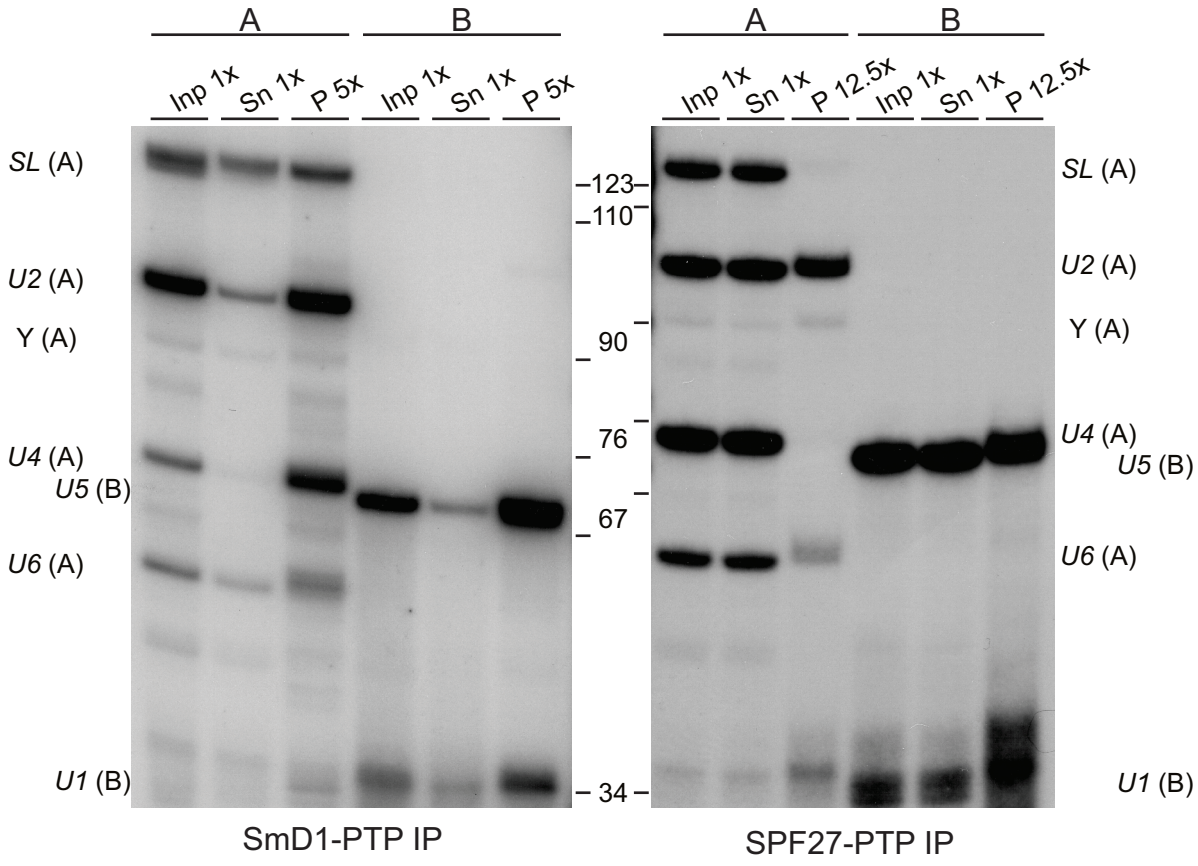



**Fig. S4.** Only a minor amount of spliceosomal U snRNAs are bound to the PRP19 complex in extract

Primer extension analysis of total RNA prepared from the SmD1-PTP and SPF27-PTP pull-downs documented in figure 6A. The analysis shows the snRNA profiles of crude extracts (Inp), supernatants (Sn) and precipitates (P) in reaction A which detects SL RNA, U2/U4/U6 snRNAs and the Y structure intermediate, and reaction B which detects U5/U1 snRNAs. x-Values indicate relative amounts loaded. Note that the high volume of the P sample for the SPF27 analysis (right panel) led to a slight upward shift of signals. Also note that the SmD1 and SPF27 reactions were separated on 6% and 8% polyacrylamide gels, respectively, leading to different spacing of primer extension products. Nevertheless, the results clearly show that SmD1 precipitation depleted U2, U4, U5, and U1 snRNAs and, to a lesser extent, SL RNA and U6 snRNA, whereas SPF27 precipitation did not lead to a detectable reduction of snRNAs. This indicates that, in extract, only a minor amount of snRNAs are stably associated with SPF27 in a spliceosomal complex.

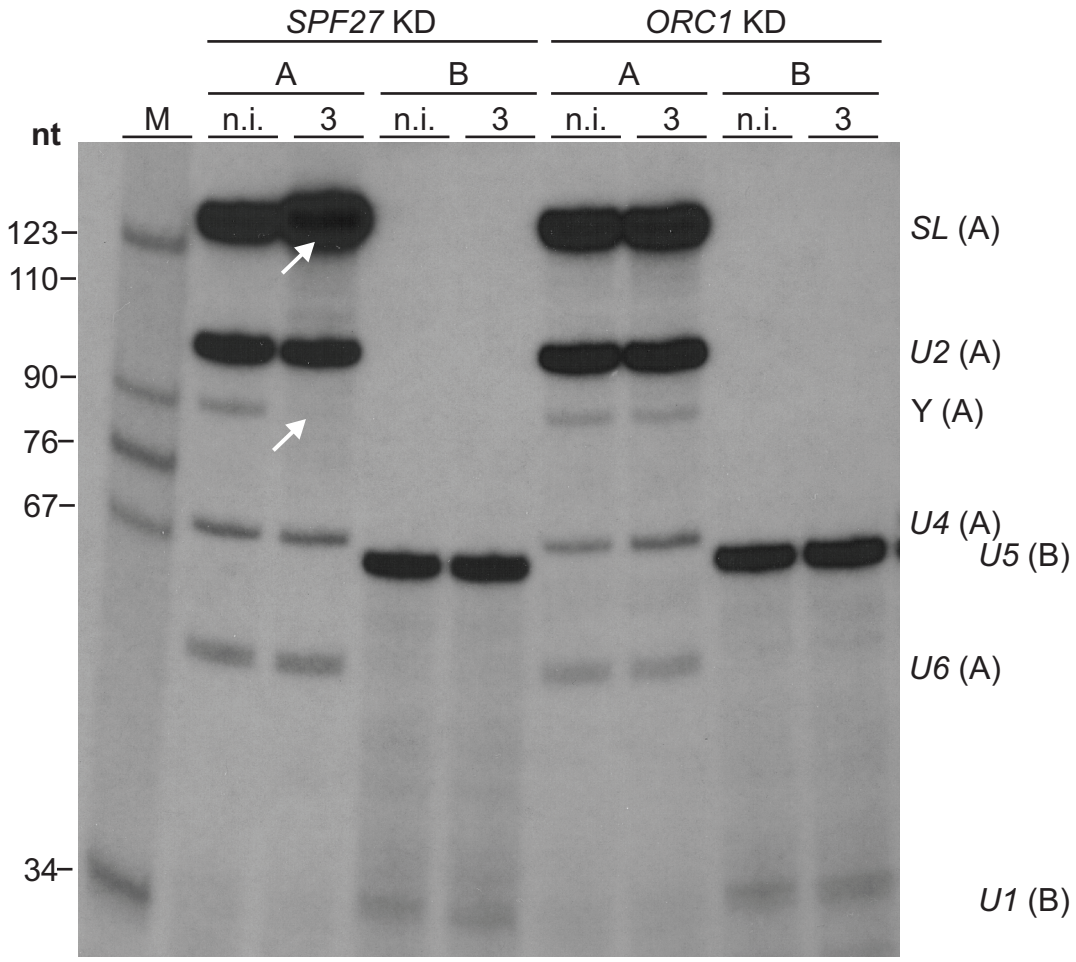



**Fig. S5.** Spliceosomal U snRNAs remain stable upon *TbSPF27* silencing

Total RNA was prepared from non-induced cells (n.i.) and from cells in which either *SPF27* or, as a control, *ORC1* was silenced (KD for knockdown) for 3 days. RNA abundances were determined by two primer extension assays. In assay A, $^{32}$P-5′-endlabeled oligonucleotides were combined that specifically hybridize to SL RNA and U2, U4 and U6 snRNAs. In assay B, U1- and U5-specific oligonucleotides were used. Primer extension products were separated on 8% polyacrylamide-50% urea gels and visualized by autoradiography. M, marker *Msp*I-digested pBR322. On the right, SL RNA and U snRNA products are specified. Y denotes a product of the SL RNA-specific oligonucleotide that is generated by the Y structure intermediate after the first step of splicing. The arrows point to increased and decreased signals of SL RNA and Y structure intermediate in RNA of *SPF27*-silenced cells, respectively. Other signal strengths remained comparable confirming that the trypanosome PRP19 complex is a non-snRNP component of the spliceosome.

## Supplemental Reference

McWilliam,H., Li,W., Uludag,M., Squizzato,S., Park,Y.M., Buso,N. *et al.* (2013) Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res* **41**: W597-W600.

Gene 556 (2015) 68–73

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

# Mono-allelic *VSG* expression by RNA polymerase I in *Trypanosoma brucei*: Expression site control from both ends?

Arthur Günzl *, Justin K. Kirkham, Tu N. Nguyen, Nitika Badjatia, Sung Hee Park

*Department of Genetics and Developmental Biology, University of Connecticut School of Medicine, 400 Farmington Avenue, Farmington, CT 06030-6403, USA*

ARTICLE INFO

*Article history:*
Received 2 September 2014
Received in revised form 22 September 2014
Accepted 23 September 2014
Available online 26 September 2014

*Keywords:*
RNA polymerase I
Mono-allelic gene expression
Trypanosome
Telomeric silencing

ABSTRACT

*Trypanosoma brucei* is a vector borne, lethal protistan parasite of humans and livestock in sub-Saharan Africa. Antigenic variation of its cell surface coat enables the parasite to evade adaptive immune responses and to live freely in the blood of its mammalian hosts. The coat consists of ten million copies of variant surface glycoprotein (VSG) that is expressed from a single *VSG* gene, drawn from a large repertoire and located near the telomere at one of fifteen so-called bloodstream expression sites (BESs). Thus, antigenic variation is achieved by switching to the expression of a different *VSG* gene. A BES is a tandem array of expression site-associated genes and a terminal VSG gene. It is polycistronically transcribed by a multifunctional RNA polymerase I (RNAPI) from a short promoter that is located 45–60 kb upstream of the *VSG* gene. The mechanism(s) restricting *VSG* expression to a single BES are not well understood. There is convincing evidence that epigenetic silencing and transcription attenuation play important roles. Furthermore, recent data indicated that there is regulation at the level of transcription initiation and that, surprisingly, the *VSG* mRNA appears to have a role in restricting *VSG* expression to a single gene. Here, we review BES expression regulation and propose a model in which telomere-directed, epigenetic BES silencing is opposed by BES promoter-directed, activated RNAPI transcription.



## 1. Introduction

The tsetse borne, unicellular parasite *Trypanosoma brucei*, which belongs to the phylogenetic order Kinetoplastida, is the only known organism that has evolved a multifunctional RNA polymerase I (RNAPI) system. This system is used to transcribe ribosomal gene units (*RRNA*) in the nucleolus, as in all eukaryotes, yet also to transcribe gene units that encode the parasite's major cell surface antigens (Günzl et al., 2003; Kooter and Borst, 1984). Trypanosomes have a unique mode of protein coding gene expression that allows them to utilize other RNA polymerases than RNAPII for the production of functional mRNA. In their genome, protein coding genes are arranged in long tandem arrays which are polycistronically transcribed. The precursor RNA is processed by spliced leader (SL) *trans* splicing and polyadenylation, resulting in mature, monocistronic mRNAs (Günzl, 2010; Michaeli, 2011; Preußer et al., 2012). Since in *trans* splicing the same capped leader sequence, derived from the SL RNA, is spliced onto the 5′ end of each mRNA, this process represents a post-transcriptional mode of capping that is decoupled from RNAPII transcription. Consequently, trypanosomes, in contrast to mammals (Grummt and Skinner, 1985), are able to use RNAPI to effectively and specifically express endogenous gene units that encode their major cell surface antigens (Rudenko et al., 1991; Zomerdijk et al., 1991a). This antigen, in mammalian-infective metacyclic and bloodstream form (BF) trypanosomes, is known as the variant surface glycoprotein (VSG), while the major cell surface antigen in insect-stage procyclic trypanosomes is procyclin.

*T. brucei* causes Human and Animal African Trypanosomiasis (also known as Sleeping Sickness and Nagana, respectively) throughout sub-Saharan Africa (Fevre et al., 2006). The parasite lives freely in the bloodstream of its mammalian host, evading the immune system by antigenic variation of its cell surface coat. The coat consists of ten million copies of the same VSG, shielding invariant membrane proteins from immune recognition (Schwede et al., 2011). *T. brucei* possesses roughly 2500 different *VSG* genes and pseudogenes (Cross et al., 2014), and periodic switching to the expression of an alternative *VSG* gene leads to antigenic variation. *VSG* genes are located in subtelomeric regions of 11 megabase, 5 intermediate-sized and ~100 minichromosomes covering, in total, ~30% of the genome (Ersfeld, 2011; Horn, 2014). However, the active *VSG* gene is invariably located next to the telomere within an expression site, with the coding region ending ~200–1800 bp upstream of the telomeric repeats. Metacyclic trypanosomes express a single *VSG*

*Abbreviations:* BES, bloodstream expression site; BF, bloodstream form [trypanosome]; CITFA, class I transcription factor A; DOT1B, disruptor of telomeric silencing; ESAG, expression-site associated gene; ESB, expression site body; FACT, "facilitates chromatin transcription" factor; NEO, neomycin phospho-transferase; NLP, nucleoplasmin-like protein; ORC1, origin recognition complex subunit 1; RNAP, RNA polymerase; RRNA, ribosomal RNA gene unit; SL, spliced leader; VSG, variant surface glycoprotein.

* Corresponding author at: Department of Genetics and Developmental Biology, University of Connecticut School of Medicine, 400 Farmington Avenue, Farmington, CT 06030-6403, USA.
*E-mail address:* gunzl@uchc.edu (A. Günzl).

http://dx.doi.org/10.1016/j.gene.2014.09.047
0378-1119/

monocistronically from one of five metacyclic expression sites in which the RNAPI promoter is located ~1–4 kb upstream of the coding region (Cross et al., 2014; Ginger et al., 2002; Kolev et al., 2012). Conversely, BFs express the active *VSG* from one of fifteen polycistronic "bloodstream expression sites" (BESs) which comprise a tandem array of typically 8–9 expression-site associated genes (ESAGs) and a terminal *VSG* gene (Fig. 1) (Hertz-Fowler et al., 2008). ESAGs appear to be important for the successful infection of the mammalian host since they encode a variant heterodimeric transferrin receptor (ESAG6 and ESAG7), whose varying affinity for transferrins of different host species is thought to expand the parasite's host range (Bitter et al., 1998). These also encode adenylate cyclases (ESAG4) that inhibit the innate immune system upon trypanosome lysis (Salmon et al., 2012).

The BES promoter resides 45–60 kb upstream of the telomere (Zomerdijk et al., 1990). It extends only 67 bp upstream of the transcription initiation site and comprises two short sequence elements (Pham et al., 1996; Vanhamme et al., 1995). Both elements are required for efficient binding of the multi-subunit class I transcription factor A (CITFA), which is essential for RNAPI transcription in the trypanosome (Brandenburg et al., 2007). The active BES is transcribed outside the nucleolus (Chaves et al., 1998), apparently in a small compartment termed the expression site body (ESB) (Navarro and Gull, 2001). In BFs the switch to the expression of another *VSG* occurs by two principal ways: either the active BES is silenced while one of the silent BESs is activated, or a DNA recombination event replaces the *VSG* gene in the active BES with a *VSG* gene from the repertoire.

Antigenic variation and mono-allelic *VSG* expression in *T. brucei* have been a research focus for decades. Several factors involved in BES silencing have been identified (see below) and BES silencing has been linked to DNA replication/ORC1 (Benmerzouga et al., 2013; Tiengwe et al., 2012), chromosome maintenance (Kim et al., 2013), and association of BESs with the nuclear lamina (DuBois et al., 2012). In addition, cohesin plays a critical role in maintaining the activated state of the BES during the cell cycle (Landeira et al., 2009). Recently, excellent and detailed reviews have addressed antigenic variation in trypanosomes and the biology of BES silencing (Alsford et al., 2012; Glover et al., 2013; Horn, 2014; Horn and McCulloch, 2010; Rudenko, 2010). Here we focus on the most recent findings of factors that appear to be directly involved in BES regulation, and propose a model in which BES-specific telomeric silencing is opposed by a mechanism that activates transcription initiation at the promoter of the active BES.

## 2. Telomeric silencing

The active *VSG* gene, independent of whether it resides in metacyclic or bloodstream expression sites, is invariably located near the telomere, indicating that the telomere has an essential function in regulating *VSG* expression. Accordingly, repression of RNAPI-mediated transcription by the telomere was directly demonstrated by integrating a plasmid with seeds for *de novo* telomere formation either at BESs or, internally, at *RRNA* loci (Glover and Horn, 2006). At the latter, tight repression extended only 2 kb upstream of telomeric repeats whereas, at inactive BESs, repression reached at least 5 kb in these experiments. The more extended repression of silent BESs was consistent with previous findings in which integration of RNAPI promoter-driven reporter cassettes at different positions of a silent BES were repressed, even when placed 14 kb upstream of the telomere (Horn and Cross, 1997). Several lines of evidence suggest that the pronounced silencing of BESs is dependent on the telomere. Depletion of the telomeric protein RAP1 led to de-repression of silent BESs, co-expression of multiple BES-encoded *VSG* genes, and the formation of additional extranucleolar RNAPI foci (Yang et al., 2009). Furthermore, depletion of the disruptor of telomeric silencing B (DOT1B), which methylates lysine 76 of trypanosome histone H3 (Janzen et al., 2006), similarly led to de-repression of silent BESs (Figueiredo et al., 2008). Direct evidence for repression of a BES from the telomere stems from a recent study in which induced expression of a *VSG* transgene, inserted into one of the *RRNA* loci, surprisingly led to a short-term, reversible attenuation of the active BES, indicating that *VSG* mRNA plays a direct role in the regulation of mono-allelic VSG expression (Batram et al., 2014). Interestingly, a time course experiment showed that this silencing of the active BES spread from the telomere towards the BES promoter in a DOT1B-dependent manner (Batram et al., 2014). Together, these data strongly indicated that BES silencing is directed by the telomere. Furthermore, it is likely that the *VSG* gene on silent BESs is protected from RNAPI transcription by more than one mechanism because *DOT1B* knockout cells could still shut down the active *VSG* gene upon ectopic *VSG* expression but were unable to attenuate expression of the remainder of the active BES (Batram et al., 2014).

## 3. BES transcription attenuation

Inactive BESs are completely silent only in regard to their telomere-proximal regions, including the terminal *VSG* gene. *VSG* mRNA from inactive BESs is $10^4$ to $10^5$-fold less abundant than that from the active BES (Figueiredo et al., 2008; Yang et al., 2009). Despite this strong difference, several observations have shown that transcription does initiate at silent BESs at a clearly detectable level. The first evidence came from a study in which insertion of a selectable marker gene 1 kb downstream of a "silent" BES promoter led to resistant parasites (Navarro and Cross, 1996). BES sequences are highly similar, especially at the promoter and in the proximal downstream region, differing from each other only by a few single nucleotide polymorphisms. However, the first genes within BESs are *ESAG7* and *ESAG6* which encode the heteromeric transferrin receptor and harbor short hypervariable regions that distinguish them from each other (Zomerdijk et al., 1991b). Analysis of *ESAG6* cDNA sequences, which on BESs are located ~5 kb downstream from the

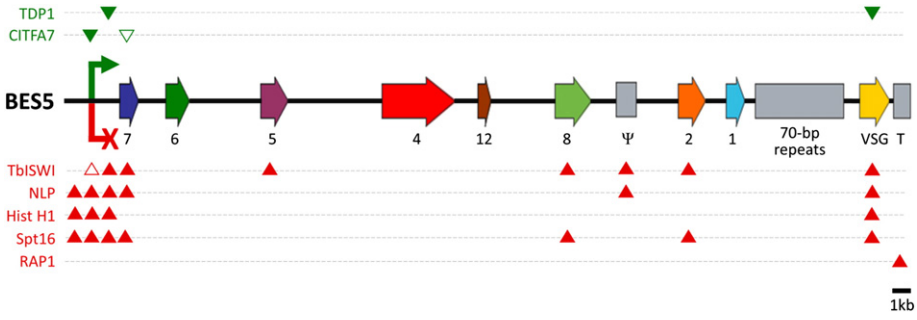



**Fig. 1.** Schematic outline of BES5 and interacting proteins. Depiction of BES5 (to scale) as a representative BES according to the published sequence (Hertz-Fowler et al., 2008). The diagram includes *ESAGs* (labeled 1, 2, 4–8 and 12), a *VSG* pseudogene (Ψ), 70 bp repeats preceding the terminal *VSG* gene, and the telomeric repeats (T). Note that some BESs have an additional promoter and an *ESAG10* gene ~14 kb upstream of the depicted promoter (not shown). The green arrow and red X represent the promoter when the BES is in the active state and silent state, respectively. Activating factors that are predominantly associated with the active BES are indicated above the diagram in green whereas factors which are implied in BES silencing are listed below the diagram in red. Filled and empty arrowheads indicate positive and negative ChIP results, respectively. Histone H1 and SPT16 associate predominately with silent sites whereas TbISWI and NLP were shown to interact equally with expression sites in both states.

promoter, revealed that 20% of the *ESAG6* mRNA in BFs was derived from various silent BESs whereas 80% stemmed from the active BES, demonstrating that, even in the absence of selective pressure, productive transcription did occur in the promoter-proximal domain of inactive BESs (Ansorge et al., 1999). Subsequently, a vast cDNA clone analysis along whole BESs showed that silent BESs contributed much more to the promoter-proximal cDNA pool than to pools of promoter-distant cDNAs, revealing that transcription that initiated at silent BESs was attenuated along the BES (Vanhamme et al., 2000). Recently, this approach was repeated with single cells, confirming that silent BESs are transcribed in their promoter-proximal region and that transcription was attenuated further downstream within a single trypanosome (Kassem et al., 2014). Finally, the demonstration that BES silencing spreads gradually from telomere to promoter and BES reactivation occurs gradually in the opposite direction (Batram et al., 2014) strongly supports the notion that transcription is attenuated at silent BESs.

## 4. Regulation of BES transcription initiation

Although the promoters of inactive BESs are not "silent", there is now convincing evidence that there is substantial regulation at BES promoters. Consistently, promoter-proximal RNA levels were found to be much higher from the active versus silent BESs. Thus, when the neomycin phosphotransferase gene (*NEO*) was inserted 1 kb downstream of the promoter of an inactive BES, it conferred parasite resistance to a low concentration of the drug G418 (1 μg/ml) while the same gene, when inserted at the identical position of an active BES, boosted resistance at least 100-fold (Navarro and Cross, 1996). The finding that in BFs 80% of *ESAG6* mRNA stemmed from the active BES suggested that there is at least a 50-fold stronger *ESAG6* expression from the active BES than from the average silent BES. When Yang et al. (2009) introduced a luciferase gene immediately downstream of the active or a silent BES promoter, the active BES produced 1500–4000-fold more light units than the silent BES.

More direct evidence for BES regulation at the level of transcription initiation came from the analysis of CITFA. CITFA consists of seven subunits, CITFA1-7, which are conserved only among kinetoplastid organisms, and the dynein light chain DYNLL1 (also known as LC8). Silencing of *CITFA1*, *CITFA2* and *CITFA7* was lethal to BFs grown in culture and strongly and specifically reduced the abundance of rRNA and *VSG* mRNA (Brandenburg et al., 2007; Nguyen et al., 2012; Park et al., 2014). Accordingly, depletion of CITFA2 from extract virtually abolished RNAPI transcription in vitro, as assayed by ~100 bp-long primer extension products, and the purified CITFA complex produced a specific gel shift with the BES promoter (Brandenburg et al., 2007). Moreover, a ChIP-seq analysis indicated that within a BES, CITFA7 occupancy was restricted to the promoter region (Nguyen et al., 2014). Together, these findings identified CITFA as a basal and general transcription initiation factor for RNAPI transcription in trypanosomes.

Interestingly, marking the active BES and a silent BES ~500 bp downstream of the transcription initiation site (Figueiredo et al., 2008) revealed that CITFA2 and CITFA7 predominantly occupied the promoter of the active BES relative to that of the marked silent BES, a phenotype that was maintained after consecutive in situ switches between the two marked sites (Nguyen et al., 2014). In accordance with CITFA's role as an RNAPI transcription initiation factor, higher CITFA occupancy at the active versus the silent BES promoter correlated with a ~70-fold higher abundance of promoter-proximal, unspliced RNA and a ~17-fold higher occupancy of the RNAPI-specific subunit RPB6z at the marker gene (Nguyen et al., 2014). Finally, *CITFA7* silencing led to a strong reduction of RNAPI occupancy and of promoter-proximal RNA levels, which directly demonstrated that CITFA binding to the promoter is required for high transcription rates in vivo (Nguyen et al., 2014). These data unequivocally showed that mono-allelic BES expression entails a mechanism that functions at the BES promoter, apparently limiting access of CITFA to silent BES promoters and/or ensuring maximal promoter occupancy of CITFA at the active BES.

It should be noted that this mechanism is not an "all or nothing"-mechanism because, in these experiments, the marked silent BES promoter was consistently occupied by CITFA above the level of negative control experiments. This finding is in accordance with promoter-proximal transcription occurring at silent BESs (see above) and it likely explains why hypersensitive DNase I sites in the promoter region, indicative of a bound transcription factor, were not restricted to the active BES but were also detected at a silent BES (Navarro and Cross, 1998). It appears that trypanosomes cannot completely shut down transcription initiation from silent BESs. Alternatively, low level transcription of the promoter-proximal part of silent BESs might serve a biological function. For instance, co-expression of different forms of the heteromeric transferrin receptor, e.g. *ESAG6* and *ESAG7*, could ensure initial survival in different mammalian hosts.

## 5. Factors involved in BES transcription regulation

There is strong evidence that inactive BESs are silenced epigenetically. Thus, while silent BESs have a nucleosomal structure, the active BES is largely depleted of nucleosomes (Figueiredo and Cross, 2010; Stanne and Rudenko, 2010). Direct evidence that nucleosomes are important for BES promoter silencing stems from depleting histone H3, which rapidly led to a ~11-fold de-repression of a GFP gene introduced downstream of the promoter of a silent BES (Alsford and Horn, 2012). In addition, CAF-1b, a replication-dependent histone chaperone, and the replication-independent chaperone ASF1A, were shown to be important for the inheritance and maintenance of the silenced state of BESs (Alsford and Horn, 2012). Interestingly, silencing the gene of either chaperone led to apparent nucleosome depletion and a de-repression of the promoter-proximal BES region. However, it did not affect expression of the corresponding *VSG* gene suggesting that nucleosomal structure is particularly important for the regulation of BES promoter activity.

Several chromatin remodeling and modifying proteins have been implicated in BES repression so far. The first epigenetic factor found to play a role in BES regulation was the chromatin remodeler TbISWI (Hughes et al., 2007). Depletion of this factor increased the mRNA abundance of a reporter gene inserted promoter-proximally into a silent BES up to 60-fold, whereas only a fivefold increase of the corresponding silent *VSG* mRNA was observed. TbISWI was found to occupy the entire length of both silent and active BESs, but was not enriched at BES promoters (Stanne et al., 2011). Although the specific function of TbISWI remains to be determined, these results suggest that TbISWI controls RNAPI transcription elongation rather than initiation.

Similar de-repression of a promoter-proximal reporter gene was observed when the histone deacetylase DAC3 (Wang et al., 2010), the linker histone H1 (Pena et al., 2014; Povelones et al., 2012), or the nucleoplasmin-like protein NLP (Narayanan et al., 2011) was depleted. The function of DAC3 appears to be promoter-specific since expression of the *VSG* gene in the marked BES was unaffected at both the mRNA and the protein level (Wang et al., 2010). However, direct association of DAC3 with BESs has not been demonstrated yet and it remains a possibility that DAC3's control of BES silencing is indirect.

The role of histone H1 in BES promoter repression has been more deeply investigated. Histone H1 is important for chromatin architecture and generally functions in chromatin condensation and transcription repression (Happel and Doenecke, 2009). Accordingly, co-silencing of the *T. brucei H1* multigene family opened up chromatin globally with the strongest effect on silent BES promoters (Pena et al., 2014). Metabolic labeling of nascent RNA then showed that histone H1 depletion resulted in an approximately six-fold higher promoter-proximal transcription rate at a silent BES, indicating that relaxation of the nucleosome structure in the promoter region led to an increase of the transcription initiation rate at the silent BES (Pena et al., 2014).

NLP is a ubiquitous nuclear protein and, accordingly, was found to be associated with all genomic loci analyzed, including the active and silent BESs (Narayanan et al., 2011). Despite this apparent general association with genomic DNA, NLP seems to be particularly important for BES promoter regulation. Depletion of NLP de-repressed a silent BES 45–65-fold, as measured by fluorescence derived from a promoter-proximal GFP gene. Moreover, *NLP* silencing also reduced promoter-proximal gene expression from the active BES about threefold (Narayanan et al., 2011). While it was speculated that NLP may have a dual function in BES silencing and in promoting processive transcription at the active BES (Narayanan et al., 2011), it is equally possible that loss of NLP enabled competition between silent and the active BES for the RNAPI transcription machinery. However, the specific function of NLP in BES regulation remains to be determined.

SPT16 is a subunit of the trypanosome FACT ("facilitates chromatin transcription") complex (Patrick et al., 2008) and appears to have a direct role in BES promoter silencing because it was found highly enriched at a silent BES promoter (Denninger et al., 2010). Accordingly, *SPT16* silencing increased promoter-proximal GFP expression from a silent BES up to 25-fold, yet de-repression did not extend to the *VSG* genes of silent BESs. However, the de-repression effect was strongly correlated with an arrest in the G2/early M cell cycle phase, raising the possibility that SPT16 does not generally facilitate BES repression in the bloodstream trypanosome. Moreover, SPT16 depletion strongly reduced *VSG* expression from the active BES, suggesting a separate BES-related function of SPT16 in facilitating processive RNAPI transcription. Overall, the specific function of FACT in the multifunctional RNAPI system remains unclear. While SPT16 has been co-purified with RNAPII of the related organism *Leishmania major* (Martinez-Calvillo et al., 2007), its association with *T. brucei* RNAPI remains to be shown.

The epigenetic factors discussed so far, including RAP1 and DOT1B (see Section 2, *Telomeric silencing*, above), function in BES silencing. The only such factor found to be important for efficient transcription of the active BES is the high mobility group protein TDP1, which belongs to a family of architectural chromatin proteins (Narayanan and Rudenko, 2013). Interestingly, TDP1 exhibited an inverse occupancy pattern to the core histone H3 at RNAPI-transcribed loci and was up to fivefold more abundant at the active BES promoter relative to a silent BES promoter. Accordingly, TDP1 depletion decreased the abundance of pre-rRNA and *VSG* mRNA from the active BES. In addition, TDP1, a nuclear protein, exhibited predominant localization to the nucleolus and the ESB, and its DNA association was found throughout the active BES and *RRNA* gene units. Thus, it appears that TDP1 facilitates high rates of processive RNAPI transcription required for trypanosome survival (Narayanan and Rudenko, 2013).

## 6. A model of BES regulation

It is difficult to integrate the data from BES de-repression studies because, for most factors, specific functions in BES silencing have not been determined yet. Nevertheless, recent data strongly indicated that BES regulation occurs at both ends of expression sites. RAP1 and DOT1B depletion studies have clearly shown that BESs are silenced by a telomere-directed mechanism. Moreover, the demonstration that BES silencing spreads from the telomere towards the promoter (Batram et al., 2014) strongly supports a telomere-directed BES silencing mechanism. However, it is unlikely that this is the only mechanism regulating monoallelic BES expression. If this was the case one would expect full activation of promoter-proximal transcription once telomeric silencing retreats beyond the promoter region, which should result in a leveling of the transcription rate between active and silent BESs (given the extremely high expression level of the active BES, it is unlikely that a trypanosome can support full activation of all fifteen BESs). However, in all cases reported, de-repression of silent BESs is, at best, moderate with the promoter-proximal expression level remaining manifold below that of the active BES. Furthermore, *RAP1* silencing did not strongly affect promoter-proximal transcription of de-repressed BESs and had only a minor influence on the high expression level of the active BES (Yang et al., 2009). Similarly, *DOT1B* silencing did not affect expression of the active BES at all (Figueiredo et al., 2008). These results strongly argue for the presence of a separate mechanism involved in BES regulation.

Transcription attenuation has been proposed to be that mechanism and, as discussed, there is clear evidence that it does occur on silent BESs (Kassem et al., 2014; Vanhamme et al., 2000). Moreover, some data suggested that transcription attenuation is caused by inefficient transcript processing (Vanhamme et al., 2000) rather than by repressive chromatin. However, recent data do not support this scenario. The finding that *RAP1* silencing caused gradual BES de-repression with the greatest effect on telomere-proximal genes, strongly indicated that transcription elongation on silent BESs is "*antagonized*" by telomere-directed spreading of repressive chromatin (Yang et al., 2009). Furthermore, upon removal of the apparent transcription elongation barrier, e.g. pronounced telomeric silencing, by RAP1 (Yang et al., 2009) or DOT1B (Figueiredo et al., 2008) depletion, promoter-proximal expression remained magnitudes below that of the active site, making it unlikely that transcription attenuation accounts for the strong difference in promoter-proximal transcription observed between active and silent BESs. Hence, transcription attenuation appears to be a consequence of epigenetic silencing rather than a regulatory mechanism, and seems to be in place to prevent the low level of transcription that does initiate at silent BESs from reaching the distally located *VSG* gene.

Based on the *RAP1* silencing results on BES de-repression, Yang et al. (2009) suggested that there has to be a mechanism functioning on the BES promoter that could explain the striking difference in promoter-proximal expression levels between the active and de-repressed/silent BESs. The strongest support for this idea stems from the demonstration that CITFA, which is absolutely required for RNAPI transcription, is predominantly associated with the active BES promoter, versus a silent site, strongly indicating that there is a mechanism in place that allows CITFA to preferentially interact with the active BES. In addition, the fact that depletion of several epigenetic factors increased promoter-proximal transcription with no or very little effect on the downstream *VSG* gene further supports the notion of a promoter-dependent regulatory mechanism.

Taking these data into account, we propose a model in which BESs are regulated by two opposing forces, namely telomere-directed epigenetic silencing acting on silent BESs and activated transcription initiation at the active BES (Fig. 2). In this model, the active BES promoter has unrestricted access to CITFA and RNAPI, allowing it to achieve the high transcription rate necessary for productive *VSG* expression. In addition, processive RNAPI transcription at the active BES is ensured by the presence of TDP1. At the same time, telomeric silencing at the active BES is impaired or pushed back so far that RNAPI transcription can extend productively past the *VSG* gene. In contrast, in this model, silent BES promoters are unable to recruit CITFA and RNAPI in sufficient amounts to allow for high transcription rates. In addition, a telomere-directed repressive epigenetic gradient spreading from the telomere into the BES causes transcription attenuation to prevent the low level of transcription, initiating at inactive BESs, from reaching the *VSG* gene.

How are BES promoters differentially regulated? An obvious barrier that could prevent CITFA from interacting with silent BES promoters and from recruiting RNAPI is a repressive chromatin structure at the promoter. The epigenetic factors whose depletion led to promoter-proximal BES de-repression, e.g. DAC3, histone H1 and NLP, may be important to build up a repressive chromatin structure at the promoter. If this is correct, depletion of these factors should lead to higher CITFA and RNAPI occupancies at de-repressed promoters. An alternative idea for the low promoter activity at "silent BESs" has been put forward in studies of CITFA. CITFA was found to be concentrated in both the nucleolus and the ESB (Nguyen et al., 2014), and it retained this localization even when its promoter-binding capability was impaired by

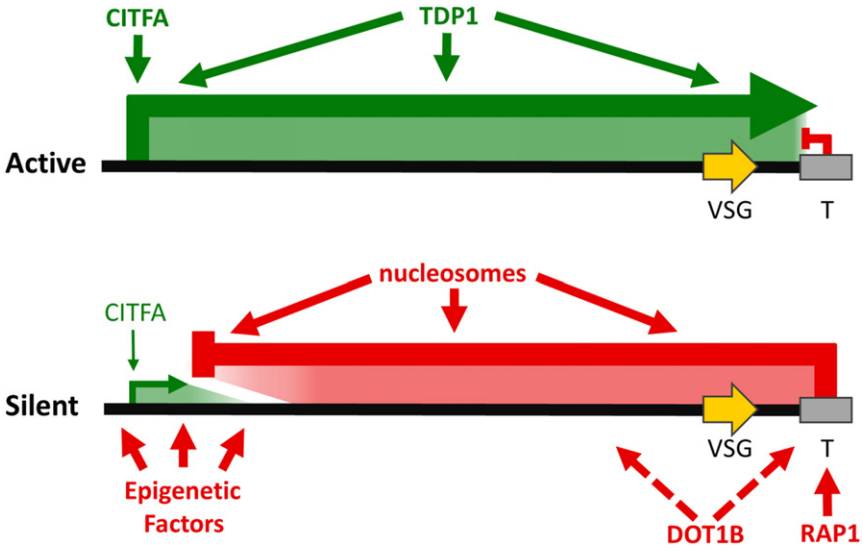



**Fig. 2.** Model of BES regulation in *T. brucei*. In the model of BES regulation, two opposing forces antagonize each other. The active BES is characterized by high transcription initiation rates and the lack of telomere-dependent epigenetic silencing, allowing unrestricted transcription elongation past the terminal VSG gene (green arrow). High processive transcription rates are facilitated by CITFA and TDP1. In silent BESs, low level RNAPI transcription initiation is opposed by BES-specific telomeric silencing that spreads towards the BES promoter causing transcription attenuation. This silencing depends on a nucleosomal structure, DOT1B and RAP1. RAP1 was shown to bind to telomeric repeats but the association of DOT1B with BESs remains to be determined (dotted line). Epigenetic factors may work together to build up a repressive chromatin structure at the promoter of silent BESs, preventing efficient binding of CITFA to its cognate DNA sequence elements. Alternatively, CITFA sequestration may limit the availability of the initiation factor for inactive BESs (not shown).

depletion of the essential subunit CITFA1 (Park et al., 2014). It was therefore suggested that sequestration of CITFA into the nucleolus and the ESB could restrict maximal RNAPI transcription to these compartments. This idea is in line with a previous study in which BFs were forced to co-express two BESs simultaneously by antibiotic selection. The two marked BESs were consistently detected in close spatial proximity (Chaves et al., 1999), as if they were competing for an essential expression factor. Furthermore, it may explain why de-repressed BES promoters remain much less active than the promoter from the active BES.

Finally, this model of two opposing forces is supported by the monitoring of the shut-down/reactivation of the active BES upon ectopic expression of *VSG* mRNA (Batram et al., 2014). In these experiments the active BES was gradually inactivated from the telomere towards the promoter, most likely by an active, telomere-directed process of repressive chromatin spreading, whereas the reactivation of the same BES occurred in the reverse direction, possibly by removal of nucleosomes by the transcription machinery. An important question emanating from this study is how does the ectopically expressed *VSG* mRNA cross-talk to the active BES? Although this question is beyond the scope of this article, it is tempting to speculate that the *VSG* mRNA sequestered an important factor for RNAPI transcription, allowing repressive chromatin to spread onto, and silence, the active BES.

## 7. Conclusion

Mono-allelic *VSG* expression in *T. brucei* differs from other allelic exclusion systems, such as *var* gene expression in *Plasmodium falciparum* (Guizetti and Scherf, 2013; Kirkman and Deitsch, 2012) or olfactory receptor expression in mammals (Magklara and Lomvardas, 2013) by the fact that the active *VSG* gene must be transcribed at an extremely high rate to enable rapidly proliferating trypanosomes to completely cover themselves with VSG. The careful measurement of RNA abundances and half lives indicated that the active *VSG* gene is transcribed at a 50-fold higher rate than a β tubulin gene (Ehlers et al., 1987). At the same time, *T. brucei* must ensure that *VSG* genes on other BESs are not expressed. The parasite achieves this balancing act apparently by restricting full RNAPI transcription initiation to the active BES and by shielding *VSG* genes on silent BESs by a telomere-dependent silencing mechanism that causes attenuation of RNAPI transcription.

## Acknowledgments

This work was supported by grants R01 AI059377 to A.G. and F30 AI110060 to J.K.K from the National Institute of Allergy and Infectious Diseases (NIAID) of the U.S. National Institutes of Health (NIH).

## References

Alsford, S., Horn, D., 2012. Cell-cycle-regulated control of VSG expression site silencing by histones and histone chaperones ASF1A and CAF-1b in *Trypanosoma brucei*. Nucleic Acids Res. 40, 10150–10160.
Alsford, S., duBois, K., Horn, D., Field, M.C., 2012. Epigenetic mechanisms, nuclear architecture and the control of gene expression in trypanosomes. Expert. Rev. Mol. Med. 14, e13.
Ansorge, I., Steverding, D., Melville, S., Hartmann, C., Clayton, C., 1999. Transcription of 'inactive' expression sites in African trypanosomes leads to expression of multiple transferrin receptor RNAs in bloodstream forms. Mol. Biochem. Parasitol. 101, 81–94.
Batram, C., Jones, N.G., Janzen, C.J., Markert, S.M., Engstler, M., 2014. Expression site attenuation mechanistically links antigenic variation and development in *Trypanosoma brucei*. Elife 3, e02324.
Benmerzouga, I., Concepcion-Acevedo, J., Kim, H.S., Vandoros, A.V., Cross, G.A., Klingbeil, M.M., Li, B., 2013. *Trypanosoma brucei* Orc1 is essential for nuclear DNA replication and affects both VSG silencing and VSG switching. Mol. Microbiol. 87, 196–210.
Bitter, W., Gerrits, H., Kieft, R., Borst, P., 1998. The role of transferrin-receptor variation in the host range of *Trypanosoma brucei*. Nature 391, 499–502.
Brandenburg, J., Schimanski, B., Nogoceke, E., Nguyen, T.N., Padovan, J.C., Chait, B.T., Cross, G.A., Günzl, A., 2007. Multifunctional class I transcription in *Trypanosoma brucei* depends on a novel protein complex. EMBO J. 26, 4856–4866.
Chaves, I., Zomerdijk, J., Dirks, M.A., Dirks, R.W., Raap, A.K., Borst, P., 1998. Subnuclear localization of the active variant surface glycoprotein gene expression site in *Trypanosoma brucei*. Proc. Natl. Acad. Sci. U. S. A. 95, 12328–12333.
Chaves, I., Rudenko, G., Dirks, M.A., Cross, M., Borst, P., 1999. Control of variant surface glycoprotein gene-expression sites in *Trypanosoma brucei*. EMBO J. 18, 4846–4855.
Cross, G.A., Kim, H.S., Wickstead, B., 2014. Capturing the variant surface glycoprotein repertoire (the VSGnome) of *Trypanosoma brucei* Lister 427. Mol. Biochem. Parasitol. 195, 59–73.
Denninger, V., Fullbrook, A., Bessat, M., Ersfeld, K., Rudenko, G., 2010. The FACT subunit TbSpt16 is involved in cell cycle specific control of VSG expression sites in *Trypanosoma brucei*. Mol. Microbiol. 78, 459–474.
DuBois, K.N., Alsford, S., Holden, J.M., Buisson, J., Swiderski, M., Bart, J.M., Ratushny, A.V., Wan, Y., Bastin, P., Barry, J.D., Navarro, M., Horn, D., Aitchison, J.D., Rout, M.P., Field, M.C., 2012. NUP-1 Is a large coiled-coil nucleoskeletal protein in trypanosomes with lamin-like functions. PLoS Biol. 10, e1001287.

Ehlers, B., Czichos, J., Overath, P., 1987. RNA turnover in *Trypanosoma brucei*. Mol. Cell. Biol. 7, 1242–1249.
Ersfeld, K., 2011. Nuclear architecture, genome and chromatin organisation in *Trypanosoma brucei*. Res. Microbiol. 162, 626–636.
Fevre, E.M., Picozzi, K., Jannin, J., Welburn, S.C., Maudlin, I., 2006. Human African trypanosomiasis: epidemiology and control. Adv. Parasitol. 61, 167–221.
Figueiredo, L.M., Cross, G.A., 2010. Nucleosomes are depleted at the VSG expression site transcribed by RNA polymerase I in African trypanosomes. Eukaryot. Cell 9, 148–154.
Figueiredo, L.M., Janzen, C.J., Cross, G.A., 2008. A histone methyltransferase modulates antigenic variation in African trypanosomes. PLoS Biol. 6, e161.
Ginger, M.L., Blundell, P.A., Lewis, A.M., Browitt, A., Günzl, A., Barry, J.D., 2002. *Ex vivo* and *in vitro* identification of a consensus promoter for VSG genes expressed by metacyclic-stage trypanosomes in the tsetse fly. Eukaryot. Cell 1, 1000–1009.
Glover, L., Horn, D., 2006. Repression of polymerase I-mediated gene expression at *Trypanosoma brucei* telomeres. EMBO Rep. 7, 93–99.
Glover, L., Hutchinson, S., Alsford, S., McCulloch, R., Field, M.C., Horn, D., 2013. Antigenic variation in African trypanosomes: the importance of chromosomal and nuclear context in VSG expression control. Cell. Microbiol. 15, 1984–1993.
Grummt, I., Skinner, J.A., 1985. Efficient transcription of a protein-coding gene from the RNA polymerase I promoter in transfected cells. Proc. Natl. Acad. Sci. U. S. A. 82, 722–726.
Guizetti, J., Scherf, A., 2013. Silence, activate, poise and switch! Mechanisms of antigenic variation in *Plasmodium falciparum*. Cell. Microbiol. 15, 718–726.
Günzl, A., 2010. The pre-mRNA splicing machinery of trypanosomes: complex or simplified? Eukaryot. Cell 9, 1159–1170.
Günzl, A., Bruderer, T., Laufer, G., Schimanski, B., Tu, L.C., Chung, H.M., Lee, P.T., Lee, M.G., 2003. RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. Eukaryot. Cell 2, 542–551.
Happel, N., Doenecke, D., 2009. Histone H1 and its isoforms: contribution to chromatin structure and function. Gene 431, 1–12.
Hertz-Fowler, C., Figueiredo, L.M., Quail, M.A., Becker, M., Jackson, A., Bason, N., Brooks, K., Churcher, C., Fahkro, S., Goodhead, I., Heath, P., Kartvelishvili, M., Mungall, K., Harris, D., Hauser, H., Sanders, M., Saunders, D., Seeger, K., Sharp, S., Taylor, J.E., Walker, D., White, B., Young, R., Cross, G.A., Rudenko, G., Barry, J.D., Louis, E.J., Berriman, M., 2008. Telomeric expression sites are highly conserved in *Trypanosoma brucei*. PLoS One 3, e3527.
Horn, D., 2014. Antigenic variation in African trypanosomes. Mol. Biochem. Parasitol. 195, 123–129.
Horn, D., Cross, G.A., 1997. Position-dependent and promoter-specific regulation of gene expression in *Trypanosoma brucei*. EMBO J. 16, 7422–7431.
Horn, D., McCulloch, R., 2010. Molecular mechanisms underlying the control of antigenic variation in African trypanosomes. Curr. Opin. Microbiol. 13, 700–705.
Hughes, K., Wand, M., Foulston, L., Young, R., Harley, K., Terry, S., Ersfeld, K., Rudenko, G., 2007. A novel ISWI is involved in VSG expression site downregulation in African trypanosomes. EMBO J. 26, 2400–2410.
Janzen, C.J., Hake, S.B., Lowell, J.E., Cross, G.A., 2006. Selective di- or trimethylation of histone H3 lysine 76 by two DOT1 homologs is important for cell cycle regulation in *Trypanosoma brucei*. Mol. Cell 23, 497–507.
Kassem, A., Pays, E., Vanhamme, L., 2014. Transcription is initiated on silent variant surface glycoprotein expression sites despite monoallelic expression in *Trypanosoma brucei*. Proc. Natl. Acad. Sci. U. S. A. 111, 8943–8948.
Kim, H.S., Park, S.H., Günzl, A., Cross, G.A., 2013. MCM-BP is required for repression of life-cycle specific genes transcribed by RNA polymerase I in the mammalian infectious form of *Trypanosoma brucei*. PLoS One 8, e57001.
Kirkman, L.A., Deitsch, K.W., 2012. Antigenic variation and the generation of diversity in malaria parasites. Curr. Opin. Microbiol. 15, 456–462.
Kolev, N.G., Ramey-Butler, K., Cross, G.A., Ullu, E., Tschudi, C., 2012. Developmental progression to infectivity in *Trypanosoma brucei* triggered by an RNA-binding protein. Science 338, 1352–1353.
Kooter, J.M., Borst, P., 1984. Alpha-amanitin-insensitive transcription of variant surface glycoprotein genes provides further evidence for discontinuous transcription in trypanosomes. Nucleic Acids Res. 12, 9457–9472.
Landeira, D., Bart, J.M., Van Tyne, D., Navarro, M., 2009. Cohesin regulates VSG monoallelic expression in trypanosomes. J. Cell Biol. 186, 243–254.
Magklara, A., Lomvardas, S., 2013. Stochastic gene expression in mammals: lessons from olfaction. Trends Cell Biol. 23, 449–456.
Martinez-Calvillo, S., Saxena, A., Green, A., Leland, A., Myler, P.J., 2007. Characterization of the RNA polymerase II and III complexes in *Leishmania major*. Int. J. Parasitol. 37, 491–502.
Michaeli, S., 2011. Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. Future Microbiol 6, 459–474.
Narayanan, M.S., Rudenko, G., 2013. TDP1 is an HMG chromatin protein facilitating RNA polymerase I transcription in African trypanosomes. Nucleic Acids Res. 41, 2981–2992.
Narayanan, M.S., Kushwaha, M., Ersfeld, K., Fullbrook, A., Stanne, T.M., Rudenko, G., 2011. NLP is a novel transcription regulator involved in VSG expression site control in *Trypanosoma brucei*. Nucleic Acids Res. 39, 2018–2031.
Navarro, M., Cross, G.A., 1996. DNA rearrangements associated with multiple consecutive directed antigenic switches in *Trypanosoma brucei*. Mol. Cell. Biol. 16, 3615–3625.
Navarro, M., Cross, G.A., 1998. In situ analysis of a variant surface glycoprotein expression-site promoter region in *Trypanosoma brucei*. Mol. Biochem. Parasitol. 94, 53–66.
Navarro, M., Gull, K., 2001. A pol I transcriptional body associated with VSG mono-allelic expression in *Trypanosoma brucei*. Nature 414, 759–763.
Nguyen, T.N., Nguyen, B.N., Lee, J.H., Panigrahi, A.K., Günzl, A., 2012. 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 11, 1573–1581.
Nguyen, T.N., Müller, L.S., Park, S.H., Siegel, T.N., Günzl, A., 2014. 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. 42, 3164–3176.
Park, S.H., Nguyen, B.N., Kirkham, J.K., Nguyen, T.N., Günzl, A., 2014. A new strategy of RNA interference that targets heterologous sequences reveals CITFA1 as an essential component of class I transcription factor A in *Trypanosoma brucei*. Eukaryot. Cell 13, 785–795.
Patrick, K.L., Luz, P.M., Ruan, J.P., Shi, H., Ullu, E., Tschudi, C., 2008. Genomic rearrangements and transcriptional analysis of the spliced leader-associated retrotransposon in RNA interference-deficient *Trypanosoma brucei*. Mol. Microbiol. 67, 435–447.
Pena, A.C., Pimentel, M.R., Manso, H., Vaz-Drago, R., Pinto-Neves, D., Aresta-Branco, F., Rijo-Ferreira, F., Guegan, F., Pedro Coelho, L., Carmo-Fonseca, M., Barbosa-Morais, N. L., Figueiredo, L.M., 2014. *Trypanosoma brucei* histone H1 inhibits RNA polymerase I transcription and is important for parasite fitness *in vivo*. Mol. Microbiol. 93, 645–663.
Pham, V.P., Qi, C.C., Gottesdiener, K.M., 1996. A detailed mutational analysis of the VSG gene expression site promoter. Mol. Biochem. Parasitol. 75, 241–254.
Povelones, M.L., Gluenz, E., Dembek, M., Gull, K., Rudenko, G., 2012. Histone H1 plays a role in heterochromatin formation and VSG expression site silencing in *Trypanosoma brucei*. PLoS Pathog. 8, e1003010.
Preußer, C., Jae, N., Günzl, A., Bindereif, A., 2012. Pre-mRNA splicing in *Trypanosoma brucei*: factors, mechanisms, and Regulation. In: Bindereif, A. (Ed.), RNA Metabolism in Trypanosomes. Springer Press, pp. 49–76.
Rudenko, G., 2010. Epigenetics and transcriptional control in African trypanosomes. Essays Biochem. 48, 201–219.
Rudenko, G., Chung, H.M., Pham, V.P., Van der Ploeg, L.H., 1991. RNA polymerase I can mediate expression of CAT and neo protein-coding genes in *Trypanosoma brucei*. EMBO J. 10, 3387–3397.
Salmon, D., Vanwalleghem, G., Morias, Y., Denoeud, J., Krumbholz, C., Lhomme, F., Bachmaier, S., Kador, M., Gossmann, J., Dias, F.B., De Muylder, G., Uzureau, P., Magez, S., Moser, M., De Baetselier, P., Van Den Abbeele, J., Beschin, A., Boshart, M., Pays, E., 2012. Adenylate cyclases of *Trypanosoma brucei* inhibit the innate immune response of the host. Science 337, 463–466.
Schwede, A., Jones, N., Engstler, M., Carrington, M., 2011. The VSG C-terminal domain is inaccessible to antibodies on live trypanosomes. Mol. Biochem. Parasitol. 175, 201–204.
Stanne, T.M., Rudenko, G., 2010. Active VSG expression sites in *Trypanosoma brucei* are depleted of nucleosomes. Eukaryot. Cell 9, 136–147.
Stanne, T.M., Kushwaha, M., Wand, M., Taylor, J.E., Rudenko, G., 2011. TbISWI regulates multiple polymerase I (Pol I)-transcribed loci and is present at Pol II transcription boundaries in *Trypanosoma brucei*. Eukaryot. Cell 10, 964–976.
Tiengwe, C., Marcello, L., Farr, H., Dickens, N., Kelly, S., Swiderski, M., Vaughan, D., Gull, K., Barry, J.D., Bell, S.D., McCulloch, R., 2012. Genome-wide analysis reveals extensive functional interaction between DNA replication initiation and transcription in the genome of *Trypanosoma brucei*. Cell Rep. 2, 185–197.
Vanhamme, L., Pays, A., Tebabi, P., Alexandre, S., Pays, E., 1995. Specific binding of proteins to the noncoding strand of a crucial element of the variant surface glycoprotein, procyclin, and ribosomal promoters of *Trypanosoma brucei*. Mol. Cell. Biol. 15, 5598–5606.
Vanhamme, L., Poelvoorde, P., Pays, A., Tebabi, P., Xong, H.V., Pays, E., 2000. Differential RNA elongation controls the variant surface glycoprotein gene expression sites of *Trypanosoma brucei*. Mol. Microbiol. 36, 328–340.
Wang, Q.P., Kawahara, T., Horn, D., 2010. Histone deacetylases play distinct roles in telomeric VSG expression site silencing in African trypanosomes. Mol. Microbiol. 77, 1237–1245.
Yang, X., Figueiredo, L.M., Espinal, A., Okubo, E., Li, B., 2009. RAP1 is essential for silencing telomeric variant surface glycoprotein genes in *Trypanosoma brucei*. Cell 137, 99–109.
Zomerdijk, J.C., Ouellette, M., ten Asbroek, A.L., Kieft, R., Bommer, A.M., Clayton, C.E., Borst, P., 1990. The promoter for a variant surface glycoprotein gene expression site in *Trypanosoma brucei*. EMBO J. 9, 2791–2801.
Zomerdijk, J.C., Kieft, R., Borst, P., 1991a. Efficient production of functional mRNA mediated by RNA polymerase I in *Trypanosoma brucei*. Nature 353, 772–775.
Zomerdijk, J.C., Kieft, R., Duyndam, M., Shiels, P.G., Borst, P., 1991b. Antigenic variation in *Trypanosoma brucei*: a telomeric expression site for variant-specific surface glycoprotein genes with novel features. Nucleic Acids Res. 19, 1359–1368.