Available online at www.sciencedirect.com
ScienceDirect
Molecular & Biochemical Parasitology 150 (2006) 201–210
MOLECULAR & BIOCHEMICAL PARASITOLOGY

Biochemical characterization of *Trypanosoma brucei* RNA polymerase II

Anish Das [a], Hong Li [b], Tong Liu [b], Vivian Bellofatto [a,*]

[a] *Department of Microbiology and Molecular Genetics, UMDNJ-New Jersey Medical School, International Center for Public Health, 225 Warren Street, Newark, NJ 07103, United States*
[b] *Department of Biochemistry, UMDNJ-New Jersey Medical School, 185 South Orange Avenue, Newark, NJ 07103, United States*

Received 22 February 2006; received in revised form 28 July 2006; accepted 1 August 2006
Available online 28 August 2006

Abstract

In *Trypanosoma brucei*, transcription by RNA polymerase II accounts for the expression of the spliced leader (SL) RNA and most protein coding mRNAs. To understand the regulation of RNA polymerase II transcription in these parasites, we have purified a transcriptionally active enzyme through affinity chromatography of its essential subunit, RPB4. The enzyme preparation is active in both promoter-independent and promoter-dependent in vitro transcription assays. Importantly, the enzyme is sensitive to α-amanitin inhibition, a hallmark of eukaryotic RNA polymerase II enzymes. Using mass spectrometric analysis we have identified the previously unobserved RPB12 subunit of *T. brucei* RNA polymerase II. TbRPB12 contains a conserved $CX_2CX_{10-15}CX_2C$ zinc binding motif that is characteristic of other eukaryotic RPB12 polypeptides. We also identified seven proteins that associate with *T. brucei* RNA polymerase II. While both bioinformatics and biochemical analysis have focused on the subunit structure of trypanosome RNA polymerases, this is the first study that reveals a functional RNA polymerase II enzyme.

Keywords: *Trypanosoma brucei*; RNA polymerase II; Transcription; Gene regulation; Proteomics

1. Introduction

Parasites in the genus *Trypanosoma* have complex life cycles that require differential protein expression. Protein-coding mRNAs are synthesized, processed, translated and recycled in stage-specific ways as parasites thrive in diverse environments. A unifying characteristic of all trypanosome mRNAs is a 5′ end leader sequence, called the spliced leader (SL), which provides a hypermethylated cap (cap 4) to mRNAs. Addition of this SL enables long polycistronic pre-mRNAs to be resolved into functional mRNAs having individual fates. In trypanosomes, the SL RNA genes and most of the protein coding genes, except those for the variant surface glycoprotein and the procyclins, are transcribed by an α-amanitin sensitive RNA polymerase II (RNA pol II) enzyme.

Our knowledge of RNA pol II transcription in trypanosomes is based almost exclusively on the transcription of the SL RNA gene. Along with RNA pol II, SL RNA gene transcription in *Trypanosoma brucei* requires additional polypeptides. tTBP (or TRF4), which is a protein closely related to the eukaryotic TATA-binding protein (TBP), transcription factor (TF) IIB [1,2] and the trypanosome version of the small nuclear RNA activating protein complex (tSNAPc) [3–5]. The tSNAP proteins were originally identified in the insect trypanosomatid *Leptomonas seymouri* as co-fractionating polypeptides that bound tightly to the *L. seymouri* SL RNA gene promoter.

Although RNA pol II transcribes the vast majority of the protein coding genes, very little is known about how this enzyme initiates and elongates polycistronic mRNAs. As pre-mRNA start sites are partially defined [6] in only one member of the Tri-Tryp parasite family (*T. brucei*, *Trypanosoma cruzi* and *Leishmania major*), it is critical to understand how RNA pol II is recruited to trypanosome chromosomes and accomplishes pre-mRNA initiation, elongation, termination and co-transcriptional *trans*-splicing [7].

The transcriptional paradigm for eukaryotes is that general transcription factors, gene-specific proteins and sequence-specific DNA elements work together to modify chromatin structure and recruit the transcriptional machinery to chromosomes. RNA pol II is positioned on the DNA in a preinitiation complex that undergoes a conformational change, allowing the template strand of the DNA to sit within the active site cleft of the enzyme

* Corresponding author. Tel.: +1 973 972 4483x2 4406; fax: +1 973 972 3644.
E-mail address: bellofat@umdnj.edu (V. Bellofatto).

0166-6851/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.molbiopara.2006.08.002

(reviewed in Ref. [8]). Transcription initiates with the synthesis of the first phosphodiester bond within the new RNA chain. The enzyme likely is released from the promoter region of the DNA after numerous short 'abortive' RNAs are produced, and the polymerase begins to productively initiate longer transcripts. Proteins that affect the switch from abortive to productive RNA synthesis, as well as factors involved in the process of elongation, splicing and termination of the RNA chains all appear to interact either directly or indirectly with the enzyme. The specific region of polymerase that recruits these myriad factors is the carboxyl terminal region of the largest subunit of RNA pol II (reviewed in Refs. [9,10]). Eukaryotic RNA pol II enzymes are composed of 12 polypeptides, designated RPB1-12. The largest polypeptide, RPB1, is a 220 kDa protein that contains seven domains (A–G), each of which are reminiscent of bacterial RNA polymerase subunit β' (rpoC), and an RNA pol II-specific carboxyl terminus, termed the carboxyl-terminal domain (CTD) [11]. The CTD, as defined in yeast and mammals, is a heptapeptide sequence (YSPTSPS), tandemly repeated numerous times (26–52 repeats). The Ser2 and Ser5 positions are phosphorylated in coordination with the status of RNA pol II during RNA synthesis [8].

Over a decade ago, *T. brucei* RPB1 (TbRPB1) was cloned based on its similarity to other eukaryotic RPB1 polypeptides and shown to exist in two closely related non-allelic forms. The *T. brucei* proteins are strikingly different from their yeast and mammalian orthologs in that they lacked a tandemly repeated heptapeptide sequence following their highly conserved A–G domains [12,13]. Recent completion of genome sequences for the Tri-Tryp organisms predicts that *T. brucei* RNA pol II likely possess the 12 subunits, characteristic of eukaryotic RNA pol II enzymes [14]. However, conformation of this prediction as well as RNA pol II transcriptional analysis in trypanosomes requires biochemical studies.

In this report, we describe the purification of a *T. brucei* RNA pol II that transcribes DNA in both a promoter-independent and promoter-specific fashion. This enzyme activity is inhibited by α-amanitin, the hallmark of RNA pol II. We also identify seven proteins that co-purify with *T. brucei* RNA pol II and may be important for its function in these parasites.

2. Materials and methods

2.1. Trypanosome cell lines, transfection and plasmid constructions

Procyclic (tsetse midgut form) parasites of the wild-type *T. brucei* Lister 427 strain were cultured and transfected as described previously [15]. One of the two RPB4 alleles was knocked out by inserting a PCR-generated DNA sequence consisting of the NEO marker (Neomycin resistance gene) flanked on one end by a 610 bp fragment and on the other end by a 540 bp fragment, consisting of the upstream and downstream regions, respectively, that surround the RPB4 open reading frame. Linear DNA (5 μg) was gel-purified and transfected into wild-type procyclic *T. brucei* cells. Transformants were selected with 15 μg/ml G418. To replace the second RPB4 allele with a tandem affinity (TAP)-tagged gene copy, the RPB4 open reading frame was PCR amplified from *T. brucei* strain 427 genomic DNA and fused with an in-frame carboxyl terminal TAP-tagged ORF. These sequences were inserted into a BLE (phleomycin resistance gene)-containing cassette consisting of the aldolase (ALD) 3′ UTR, followed by the actin (ACT) 5′ UTR and BLE marker, and followed by the RPB4 540 bp downstream region. Linear DNA (5 μg) was used for transfection; phleomycin and neomycin resistant parasites were selected with 1.5 μg/ml phleomycin and 15 μg/ml G418. These parasites are referred to as TbRPB4-tagged cells.

2.2. Extract preparation and protein purification

Nuclear extracts were prepared from TbRPB4-tagged cells. Briefly, 2×10^{10} cells (from a 2 l culture) were harvested and washed once with wash buffer (20 mM Tris–HCl, 7.4, 100 mM NaCl, 3 mM $MgCl_2$) and once with TX buffer (10 mM HEPES, pH 7.9, 20 mM potassium glutamate (Kglu), 150 mM sucrose, 2.5 mM $MgCl_2$, 1 mM EDTA, 2.5 mM DTT, 1 μM each of pepstatin, leupeptin and phenylmethylsulfonyl fluoride). Cells were lysed in TX buffer in a 7 ml glass homogenizer with type A pestle by vigorously douncing for 10–15 min. Nuclei were collected by centrifugation (4000 × *g*, 40 min), resuspended in TX buffer and extracted by adding saturated (4.1 M at 4 °C) ammonium sulfate ($(NH_4)_2SO_4$) drop-wise to final concentration of 10%, while stirring for 20 min. The resulting extract was cleared by ultracentrifugation (100,000 × *g*, 45 min). Proteins from the clarified extract were precipitated using 0.33 g/ml $(NH_4)_2SO_4$; NaOH (0.01 mmol/10 g salt) was added to maintain pH. Precipitated protein was collected by centrifugation (15,000 × *g*, 20 min), resuspended in 1.2 ml of TX buffer and dialyzed against TX buffer containing 150 mM Kglu (TX-150). Dialyzate was cleared by centrifugation (10,000 × *g*, 10 min) and frozen on dry ice-ethanol.

The TAP-tagged protein purification protocol was adapted from [16]. Extract was bound to a 1 ml IgG-agarose column and washed extensively with TX-150 buffer. Bound proteins were released from the column by tobacco etch virus (TEV) protease digestion and subsequently loaded onto a 1 ml Calmodulin-agarose column. Bound proteins were eluted with 2 mM EGTA in three 0.5 ml fractions. Purified protein was used in transcription reactions, either directly or after concentration using a Centricon™ 10 (Millipore) filtration unit.

2.3. In-gel digestion and mass spectrometry analysis

The protein mixture was separated by SDS-PAGE and stained by Sypro Ruby™ dye (Molecular probes). The bands were excised and washed with 30% acetonitrile (ACN) in 50 mM ammonium bicarbonate. The proteins were digested using 25 ng/μl of trypsin for 2 h on a robotic platform (TECAN). The resulting peptides were extracted with 30 μl of 1% trifluoroacetic acid (TFA), desalted using a C_{18} Ziptip and injected onto a 75 μm × 150 mm capillary PepMap C_{18} column (Dionex), gradient eluted into a Q-TOF™ API tandem mass spectrometer system (Waters Corp) through a nano electrospray interface for

data-dependant MS/MS analysis. Protein identification was performed by searching the MS/MS spectra against the translated *T. brucei* protein sequence database using a local MASCOT search engine (V.1.9). Proteins resolved into peptides recognizable with confidence interval (CI) values no less than 95% were considered to be identified.

2.4. Promoter-independent and promoter-dependent in vitro transcription assays

To prepare $(dC)_n$-tailed templates, pUC19 DNA was restricted with *Nde*I and the resulting 3′ ends were extended with Terminal Transferase (TdT; NEB) in the presence of 50 μM dCTP. This $(dC)_n$-tailed DNA was further digested with *Pst*I which resulted in the generation of two DNA templates of 2433 bp and 253 bp. Promoter independent transcription assays [17] were performed on these $(dC)_n$-tailed templates using TAP-purified RNA pol II in 25 μl reaction volume containing 10 mM HEPES, pH 7.9, 3% PEG, 50 mM Kglu, 2.5 mM $MgCl_2$, 1 mM DTT, 1 μM leupeptin, 20 mM creatine phosphate, 0.48 mg/ml creatine kinase, 2 units porcine RNase inhibitor, 2.5 mM ATP, 0.8 mM GTP, 0.13 μM [α-^{32}P] CTP and 0.13 μM [α-^{32}P] UTP. Reactions were performed at 28 °C for 10 min and chased with a mix of 1 mM CTP and UTP for 30 min. The resulting RNAs were separated on a denaturing 6% polyacrylamide gel and detected by PhosphorImager™ analysis.

SL RNA transcription was performed as previously described and was programmed with pJP10 DNA [4]. Transcripts were detected by primer extension using a ^{32}P (5′ end)-labeled oligonucleotide primer that hybridized to the 20 nt RNA tag.

2.5. Immunodepletions and antibodies

Protein A-sepharose was used for mock-depletion of RNA pol II from nuclear extract prepared from TbRPB4-tagged cells. IgG-agarose was used to deplete RNA pol II enzyme. A 200 μl of nuclear extract (1 mg protein) was incubated with either 100 μl of protein A-sepharose or IgG-agarose in TX buffer containing 300 mM Kglu for two incubations (gentle mixing was at 4 °C for 1 h total). Protein from the depleted extracts (50 μg) was used for each transcription assay.

TbRPB1 antibody was generated from recombinant protein that contained the carboxyl terminal domain (amino acids 1480–1765) produced as a maltose binding protein (MBP) fusion (using the pMAL-c2X plasmid, New England Biolabs, Beverly, MA). Recombinant protein was expressed and purified per the manufacturer's instructions. Antibodies were raised in rabbits at Lampire Biological Laboratories (Pipersville, PA). Anti-TBP antibody is a rabbit polyclonal antibody [4].

2.6. RPB12 cDNA analysis

The amino terminus of TbRPB12 protein was determined experimentally using random hexameric oligonucleotides and reverse transcriptase to generate cDNA from total RNA prepared from *T. brucei*. Specific cDNAs were amplified by PCR using a primer set contain the sense strand of the SL (5′-CTATTATTAGAACAGTTTCTGTAC-3′) and the antisense strand of TbRPB12 open reading frame (5′-GTTCCTCATCCTCCAATGATCC-3′). The resulting ∼140 bp amplified DNA was directly sequenced using the antisense strand primer.

2.7. Protein alignments

Multiple alignments of the RPB4 and RPB12 proteins from *T. brucei*, *T. cruzi*, *L. major*, *S. cerevisiae* and *H. sapiens* were performed using ClustalW. Alignments were refined manually and visualized using GeneDoc (http://www.psc.edu/biomed/genedoc).

3. Results

3.1. Purification of T. brucei RNA polymerase II

To characterize *T. brucei* RNA pol II, we sought to purify the enzyme by affinity tagging a small subunit of the protein. Eukaryotic RNA pol II consists of 12 subunits, designated RPB1-12. Among the 12 subunits, RPB1, RPB2, RPB3 and RPB11 are unique to RNA pol II but have counterparts in both RNA pol I and III. Five subunits are shared among all three polymerases. Three subunits, RPB4, RPB7 and RPB9, are specific for RNA pol II. The highly conserved RPB4 subunit of eukaryotic RNA pol II enzymes has been affinity tagged in several systems without adversely effecting cell growth [18,19]. Therefore, we used homology searches to identify the *T. brucei* homolog of RPB4 and deduced that TbRPB4 is present in a single locus on chromosome 3. This homology was confirmed by performing a BLAST search of the non-redundant Swiss–Prot database with the *T. brucei* protein. A multiple sequence alignment of TbRPB4 and its Tri-Tryp homologs as well as human and *S. cerevisiae* RPB4 homologs are shown in Fig. 1. Although the primary sequence similarity of the Tri-Tryp RPB4 homologs and their human and yeast counterparts differ, a secondary structure prediction shows the presence of six helices (helices 1–6) in TbRPB4, which align structurally with that of human, and yeast RPB4.

We generated a stably transfected cell line in which one allelic copy of TbRPB4 was deleted and the second allele was tagged to encode a TAP-tagged fusion protein. Generation of this transgenic cell line was verified by several analytical PCR and Western blot analyses of total protein (data not shown). These transfected parasites, referred to as TbRPB4-tagged *T. brucei* procyclic cells, were used for biochemical purification of *T. brucei* RNA pol II. A two-step affinity purification scheme, which has been used with previous success in our laboratory, was used to purify the enzyme. Nuclei were enriched by centrifugation, proteins were extracted from the DNA using high salt, and two sequential affinity matrices were used to enrich for polypeptides associated with TAP-tagged TbRPB4. The protein purification scheme and aliquots from each purification step are shown in Fig. 2A and B. An initial indication that our purification scheme yielded RNA pol II was provided when the largest subunit of the enzyme, TbRPB1, co-fractionated with TbRPB4, as detected by western blot analysis (Fig. 2C).

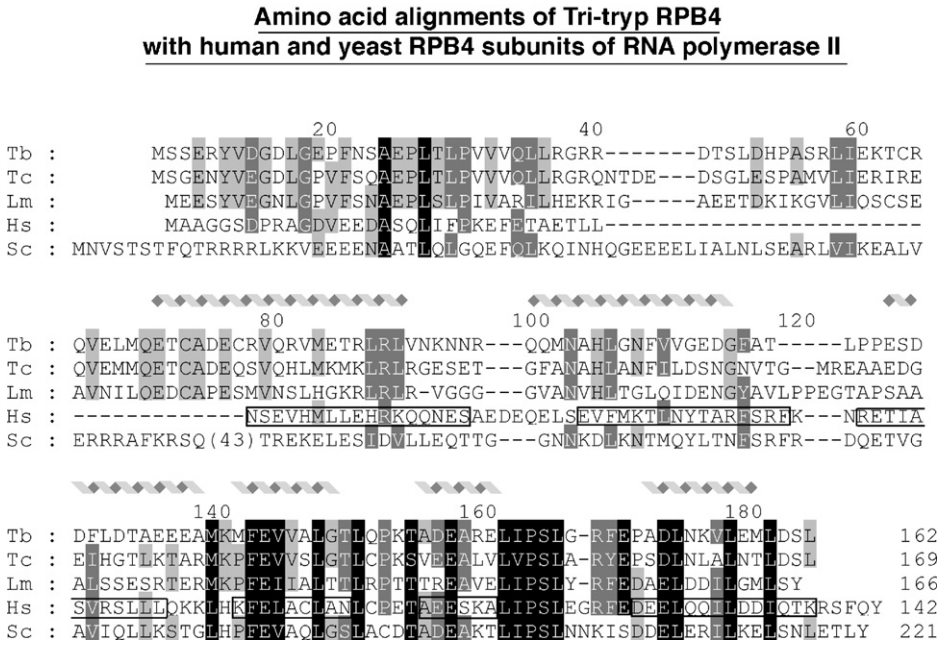

Fig. 1. Multiple sequence alignment of trypanosome RPB4 and homologs in other eukarya. A protein sequence alignment shows the similarity of *Trypanosoma brucei* RPB4 (Tb, Tb927.3.5270) with homologs from *T. cruzi* (Tc, Tc00.1047053508409.280), *Leishmania major* (Lm, LmjF29.0400), *Homo sapiens* (Hs, NP_002687) and *Saccharomyces cerevisiae* (Sc, CAA41112). Black shading signifies high conservation, gray shading with white lettering signifies 80% conservation, and gray shading with black letters signifies 60% conservation. Alignment was done using ClustalW and manually adjusted for optimal comparison. The boxed amino acids are helices H1–H6, as observed in the crystal structure of the human enzyme [34]. The corresponding H1–H6 helices in the trypanosome proteins are shown above the sequences with a drawn helix.

To characterize the subunit composition of *T. brucei* RNA pol II, each of the 15 polypeptides observed in a silver-stained SDS-PAGE gel (Fig. 2B, second of the two 'EGTA elutions' lanes) were visualized using Sypro Ruby™ staining and analyzed by mass spectrometry. As expected, the 30 kDa- and 220 kDa-migrating bands are the TbRPB4 and TbRPB1 polypeptides. Additional polypeptides that are easily recognized homologs of known eukaryotic RNA pol II subunits were also identified; thus TbRPB2, 3, 5, 7 and 9 are associated with TbRPB4. Each of the biochemically identified RNA pol II subunits observed

Table 1
RNA polymerase II in *T. brucei*

Subunits of RNA polymerase II	Subunit molecular mass (kDa)	GeneDB systematic name	Amino acid length	Percent identity/similarity to			
				Sc		Hs	
				Identity	Similarity	Identity	Similarity
RPB1	**197.0**	**Tb927.4.5020**	**1765**	**34.7**	**34.2**	**37.5**	**33.1**
RPB2	**134.5**	**Tb927.4.3810**	**1190**	**44.5**	**30.5**	**46.7**	**29.0**
RPB3	**37.7**	**Tb927.3.5500**	**337**	**17.5**	**33.5**	**23.1**	**30.0**
RPB4	**18.2**	**Tb927.3.5270**	**162**	**22.2**	**40.0**	**11.7**	**28.4**
RPB5	25.2	Tb10.389.0110	219	33.8	32.0	31.1	34.2
RPB6	16.1	Tb927.4.3510	144	50.0	22.2	43.0	28.4
	14.4	Tb11.03.0935	131	22.9	33.5	19.0	32.8
RPB7	**21.2**	**Tb11.01.6090**	**190**	**25.3**	**37.9**	**22.6**	**43.2**
RPB8	16.3	Tb11.02.5790	149	26.1	44.9	30.2	44.2
RPB9	**15.4**	**Tb11.02.5180**	**133**	**21.8**	**38.3**	**27.8**	**34.6**
RPB10[a]	9.5	Tb927.3.1250	84	45.2	23.8	44.0	25.0
RPB11	14.7	Tb11.57.0004	132	28.0	36.3	31.0	35.6
	14.8	Tb11.01.0625	134	21.6	39.5	26.8	37.3
RBP12 (see Fig. 3)	**9.1**	**Tb927.1.1170**[b]	**82**	**10.0**	**16.8**	**6.8**	**13.0**

Proteins in bold were identified directly by protein purification followed by mass spectrometry in this paper. The two highly related *T. brucei* RPB11 proteins are listed in the order that reflects greater to lesser percent identity with human and yeast RPB11 orthologs. *S. cerevisiae* (Sc) RPB1–12 are: NP_010141; CAA99357; CAA86971; CAA41112; CAA85113; P20435; P34087; P20436; CAA96774122; CAA99425; CAA99004; YHR143W-A. *H. sapiens* (Hs) RPB1–12 are: POLRA2; P30876; NP_002685; NP_002687; BAA07406; NP_068809; P62487; CAB92189; NP_006224; NP_066951; NP_006225; NP_005025.

[a] This protein associates with RNA pol I; another RPB10 paralog may associate with RNA pol II [22].

[b] Current GeneDB annotation for **Tb927.1.1170** is an 18.4 kDa protein.

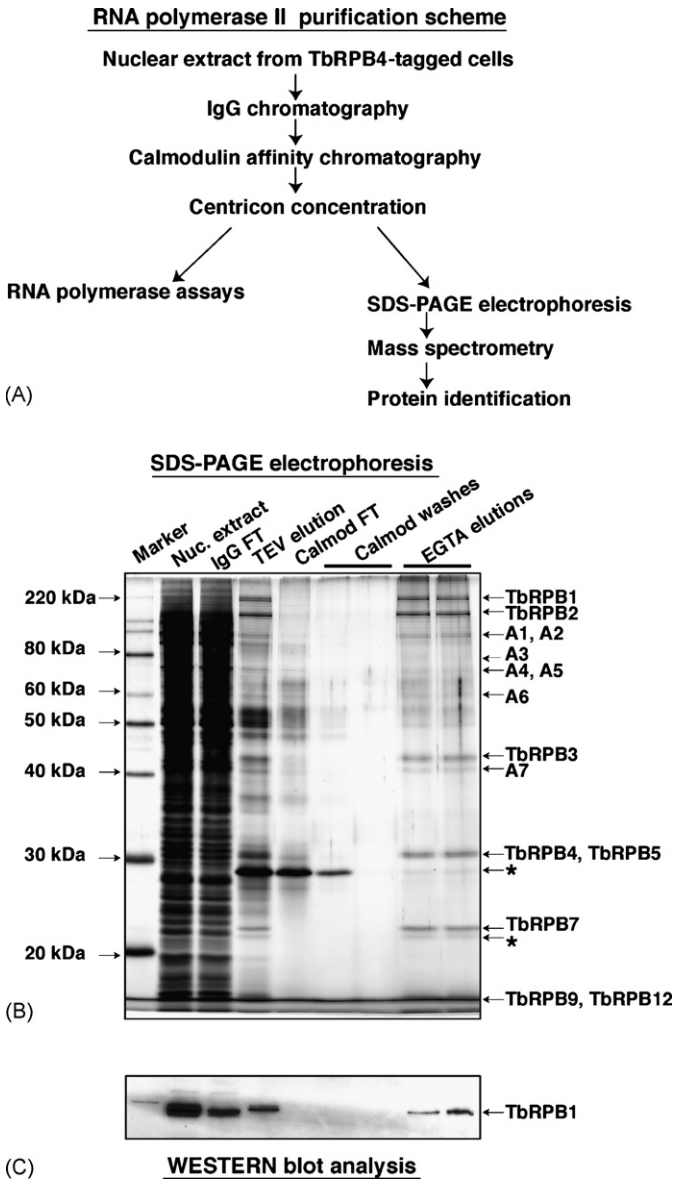

Fig. 2. Composition of *T. brucei* RNA polymerase II and associated proteins. (A) Purification scheme of tandem affinity tagged (TAP)-RPB4 and associated proteins. (B) Protein fractions from nuclear extract (lane 'Nuc. extract'), IgG chromatography flow through (lane 'IgG FT'), TEV protease-dependent elution (lane 'TEV elution'), Calmodulin-affinity chromatography flow through (lane 'Calmod FT'), Ca^{2+}-dependent washes (lanes 'Calmod washes') and two sequential fractions from EGTA elution (lane 'EGTA elutions'). The second fraction, shown in the right most lane is referred to as the RNA pol II-fraction in Figs. 4 and 5. Polypeptide markers are MagicMark™ (Invitrogen) and are labeled on the left side of the silver-stained SDS-10% polyacrylamide gel. RNA pol II subunits and RNA pol II-associated proteins identified by LC-ESI–MS/MS (Q-TOF™) (see Tables 1 and 2) are labeled. RNA pol II-associated proteins are numbered A1–A7; from slowest to fastest electrophoretic migration on SDS-PAGE. Two asterisks indicate unidentified proteins. (C) Western blot analysis shows the co-purification of TbRPB 1 with TbRPB4 through all chromatographic steps. The lanes in this panel correspond directly to those in (B).

in the purification using affinity tagged TbRPB4 is described in Table 1.

SDS-PAGE analysis of our RNA pol II purification, followed by mass spectrometric analysis of fast-migrating proteins, identified the *T. brucei* ortholog of RPB12 (Fig. 2). An ortholog of RPB12 (also known as ABC10α), a small polypeptide shared by all three RNA polymerases, was not detected in initial Tri-Tryp database and protein analyses [20,21], underscoring its high sequence divergence from other eukaryotic homologs. However, a subsequent *in silico* study, followed by cDNA analysis identified the *T. brucei* RPB12 polypeptide [14,22]. The protein shown in Fig. 3 shares an invariant $CX_2CX_{10-15}CX_2C$ zinc binding motif, followed by a positively charged C-terminus region conserved in all eukaryotic versions of ABC10α. An alignment of the Tri-Tryp orthologs of RPB12 with the human and yeast polypeptides were shown in Fig. 3A. Table 1 shows the identities and similarities among TbRPB12 and its *S. cerevisiae* and human orthologs. Although TbRPB12 migrates at about 10 kDa on SDS-PAGE gels (see Fig. 2B), the GeneDB annotation indicates an 18.4 kDa polypeptide. To resolve this apparent anomaly, we mapped the 5′ UTR and start codon of TbRPB12 by cDNA analysis. Fig. 3B shows the experimentally determined 5′ end of the TbRPB12 cDNA. We conclude that TbRPB12 is 9.1 kDa protein (82 amino acids), and not 18.4 kDa (161 amino acids) as annotated in GeneDB. Recently published cDNA analysis by another group also proposes that TbRPB12 is an 82 amino acid protein [22].

3.2. Transcriptional activity of purified T. brucei RNA polymerase II

To probe the function of *T. brucei* RNA pol II, we performed in vitro transcription studies. Specially, we determined if the proteins that co-fractionated with TAP-tagged TbRPB4 through the EGTA elution step (shown in Fig. 2) possessed RNA polymerase activity. Promoter-independent transcription was assayed on $(dC)_n$-tailed templates, on which enzyme loads on the single stranded poly $(dC)_n$ tail and then transcribes the double stranded region of the template until it 'runs off' the end of the molecule. The templates used in our assays are drawn in Fig. 4A. Transcription from the shorter template was expected to yield a short 253 nt RNA. Accumulation of full-length, 'run off' transcripts produced from the longer template would yield a 2433 nt RNA and measure the processivity of the input enzyme. Fig. 4B shows that the purified RNA pol II is enzymatically active and polymerizes rNTPs into RNA chains that are 253 nt and 2433 nt long (lanes 2 and 3). Increasing amount of enzyme yielded increased amounts of RNA products, as expected. We also observed an intermediate size transcript, denoted as 'pause', which is possibly due to premature pausing of the polymerase on the DNA. As expected for a functional RNA pol II enzyme, α-amanitin abrogated enzyme activity, as observed by a decrease in 'run-off' RNA products in the presence of titrated amounts of the inhibitor (compare lane 2 with lanes 4–6).

The next set of experiments tested the ability of the purified RNA pol II enzyme to reconstitute promoter-dependent SL RNA gene transcription in vitro. The SL RNA gene template used for transcription reactions is drawn in Fig. 5A. A 20 bp tag, present 63 nt downstream from the SL RNA transcription start (AACU, +1 to +4), enables us to detect correctly initiated, in vitro transcribed RNAs in the SL RNA promoter-dependent transcription assays. *T. brucei* nuclear extracts, competent for SL

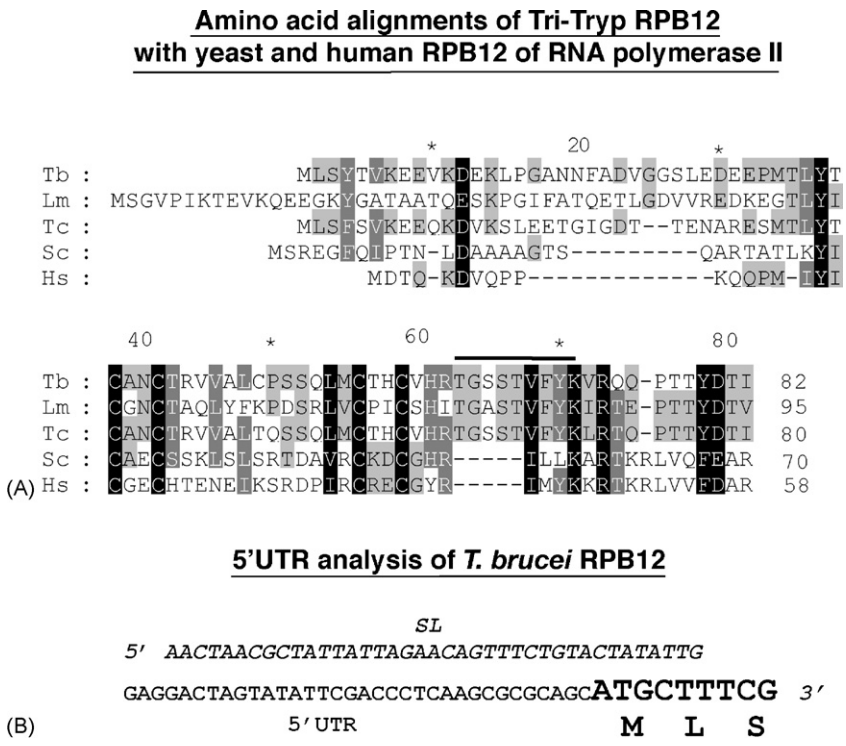

Fig. 3. Multiple sequence alignment of yeast and human RPB12 subunit of RNA polymerase II assist in the identification of TbRPB12. (A) A 9.1 kDa polypeptide that co-fractionated with TbRPB4 was identified by Q-TOF™ as Tb927.1.1170 (this GeneDB ORF is annotated as an 18.4 kDa RPB12 protein). This protein has homologs in all Tri-Tryp organisms (Lm, LmjF20.0490; Tc, Tc00.1047053507949.69) and is identified as the Tri-Tryp ortholog of eukaryotic RPB12 as shown in this alignment that contains *Saccharomyces cerevisiae* (Sc, AAB68994) and *Homo sapiens* (Hs, NP_005025) RPB12 polypeptides. Black shading signifies high conservation, gray shading with white lettering signifies 80% conservation, and gray shading with black letters signifies 40–60% conservation. Alignment was done using ClustalW and manually adjusted for optimal comparison. (B) The 5′ end of TbRBP12 is shown; the SL is in italics, the 34 nt long 5′ UTR is in plain text and the first three codons are in bold.

RNA gene transcription [4], were depleted for RNA pol II with antibodies that recognized the single TbRPB4 protein expressed in TAP-tagged TbRPB4 procyclic cells. Significant depletion of both TbRPB4 and TbRPB1 indicated that RNA pol II was efficiently removed from the extract (compare Mock and IgG-depleted Western blot analysis shown in Fig. 5C). As a control for non-specific protein depletion, we determined that TbTBP, a transcription factor not expected to be tightly associated with RNA pol II, was not immunodepleted (Fig. 5C, bottom panel). Fig. 5B shows that nuclear extracts depleted for RNA pol II had reduced transcriptional activity for SL RNA as compared to controls (compare lanes 3 and 13 with lanes 1, 2 and 12). The depletion was specific for RNA pol II as RNA pol III-transcribed U6 RNA transcription levels were unaltered (compare lanes 12 and 13). Although the extract was depleted of RNA pol II, addition of increasing amount of purified RNA pol II failed to restore transcription activity (lanes 4 and 5). Since transcription factor TbTFIIB remains associated with RNA pol II in *T. brucei* nuclear extract [1], we tested whether depletion of RNA pol II caused a consequent depletion of TbTFIIB. Indeed, addition of increasing amount of RNA pol II along with TbTFIIB significantly restored (lanes 8 and 9) the transcription activity of the depleted extract. As controls, we determined that addition of TbTFIIB alone (lanes 6 and 7) and BSA (lanes 10 and 11) did not have any effect on the transcription activity. These results demonstrate that the RNA pol II enzyme present in the EGTA-eluted fraction from the purification is competent for RNA pol II-dependent SL RNA transcription.

3.3. *T. brucei* RNA polymerase II-associated proteins

To initiate an investigation of RNA pol II partners, we have identified the small number of proteins that co-fractionated with RNA pol II using mass spectrometry. Seven proteins that co-purified with RNA pol II, designated A1–A7 (A is for 'associated' protein) are defined in Table 2. These proteins are specifically associated with RNA pol II, as they were not identified in several other TAP-tagged protein-based purifications [4] (Milone, J., Utter, C., Bellofatto, V., unpublished data). Two of the proteins, A1 and A2, clearly resemble known proteins. A1 is a metabolic enzyme, NADH-dependent fumarate reductase [23]. A2 protein is a member of the Retrotransposon Hot Spot Protein subfamily, RHS4 [24]. A3–A7 are undefined proteins, but A3–A5 contain known protein domains. A3 contains an amino-terminal ubiquitin-domain. Proteins A4 and A5 contain DEAD box helicase domains (Pfam domains PF00270 and PF00271), which are found in proteins of RNA metabolism, including transcription initiation, splicing, processing, and nucleoplasmic export of RNA. Two proteins, A6 and A7, are trypanosome-specific, as they are not orthologous to any known proteins. A6 is a glutamine-rich protein; glutamine-richness is a feature common to many RNA pol II-associated transcription factors in

Table 2
T. brucei RNA polymerase II-associated proteins

Associated proteins	Molecular mass (kDa)	GeneDB systematic name	Identification/salient features
A1	123	Tb05.28F8.460	NADH-dependent fumarate reductase
A2	98	Tb927.1.120	Retrotransposon hot spot protein 4
A3	68	Tb11.02.1600	Contains a ubiquitin domain
A4	72	Tb10.61.2130	RNA helicase-like
A5	69	Tb05.45E22.770	RNA helicase-like
A6	49	Tb11.02.4300	Glutamine-rich protein
A7	38	Tb06.26G9.840	Hypothetical protein

Seven proteins, A1–A7, co-fractionated with RNA pol II in TbRPB4-tagged procyclic *T. brucei*. Proteins were identified by Q-TOF™ tandem mass spectrometry, in which six peptides (A1 and A2), one peptide (A3), three peptides (A4 and A7), two peptides (A5), and four peptides (A6) matched to the corresponding protein sequences. Although only one peptide sequence was derived from A3, all the major peaks in the MS/MS spectrum derived either from y-ions or b-ions and produced a MASCOT score of 46 (CI value of 99.93%). Thus, A3 is identified with high confidence.

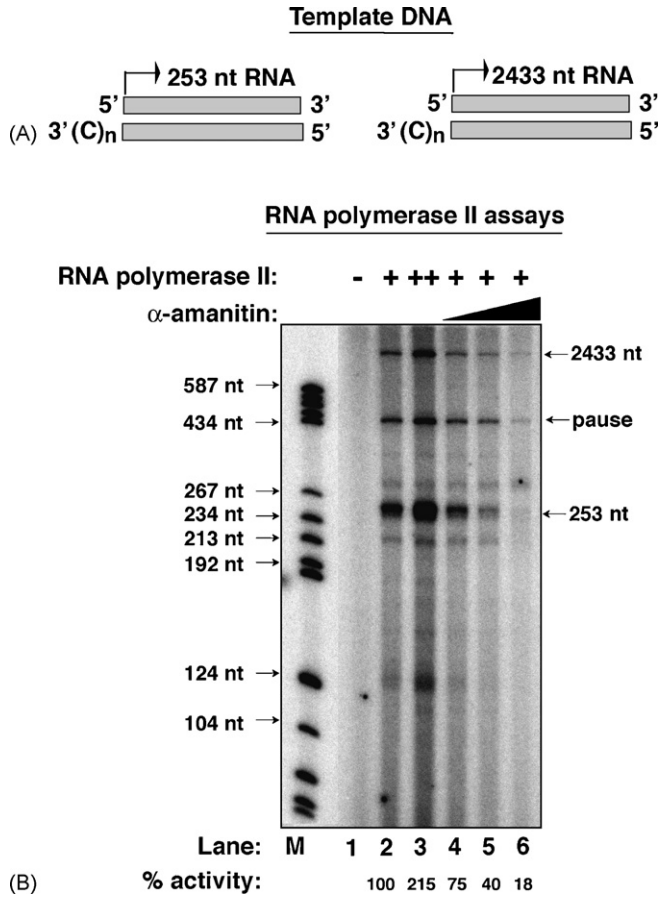

Fig. 4. Purified *T. brucei* RNA polymerase II is a functional enzyme that displays α-amanitin sensitivity. (A) A schematic showing the $(dC)_n$-tailed DNA template used in transcription run-off assays. 'Template DNA' contains a set of double stranded, $(dC)_n$-tailed molecules, one template is transcribed to yield a 253 nucleotide (nt) RNA, the other is transcribed to yield a 2433 nt RNA. The ~440 nt RNA is probably generated by premature pausing of RNA pol II. (B) In vitro transcription 'run-off' assays. Transcription activity was measured using 100 ng (+) and 200 ng (++) of the RNA pol II-fraction (see Fig. 2B). 'Run-off' transcripts are labeled with arrows on the right of the PhosphorImager™ analysis of the 7 M urea-6% polyacrylamide gel. Ablation of activity was observed with the addition of 1 μg/ml (lane 4), 10 μg/ml (lane 5) and 100 μg/ml of α-amanitin (compare lane 2 with lanes 4–6). Percent transcription activity, relative to lane 2, is shown below each lane. Nucleotide length markers are shown to the left of the gel.

eukaryotic cells [25]. These data provide an important initial cataloging of proteins that may function with RNA pol II to achieve both SL RNA and pre-mRNA transcription in trypanosomes.

4. Discussion

Over a decade ago, sequence comparisons identified an unusual RPB1 protein in trypanosomes that lacked a heptapeptide carboxyl terminal repeat and generated an interest in how RNA pol II-dependent transcription occurs in these parasites [12,13]. Since then, RNA pol II-dependent transcription studies have focused on the SL RNA gene promoter, as this is the only defined trypanosomal RNA pol II promoter. Recently, biochemical studies have begun to reveal the transcription factors that recruit the RNA pol II machinery to the SL RNA gene. These studies have established that SL RNA gene transcription in *T. brucei* requires a set of proteins that include tSNAPc, a TFIIAγ ortholog, a TFIIAα/β-like polypeptide, TbTBP (TRF4) and TbTFIIB [1–5]. The biochemical elucidation of these transcription factors has coincided with the completion of the Tri-Tryp genomes. The powerful tools of bioinformatics, proteomics and biochemistry now allow us to address RNA pol II function in trypanosomes.

Towards this aim, *T. brucei* RNA pol II activity has been isolated from procyclic trypanosomes. Eukaryotic RNA pol II enzymes contain 12 subunits, RBP1-12 [18,26]. In this study we have clearly identified trypanosomal orthologs for eight of these 12 subunits, RPB1-5, 7, 9, and 12. The biochemical identification of TbRPB12 was gratifying as this subunit was not recognized as such during initial sequence analysis. Our SELDI mass spectrometry analysis suggests that three additional small subunits, RPB 6, 8, and 11 (16 kDa, 16.3 kDa, 14.7 kDa, respectively), are present in our purified TAP-tagged protein preparation (data not shown). This supports a recent study by Devaux et al. [21] that detected RPB 6, 8 and 11 in *T. brucei* RNA polymerase II. A small RNA pol II subunit, RPB10, remains difficult to detect in RNA pol II; however, its counterpart in RNA pol I has been recently identified [22]

Structural analyses of yeast and human RNA pol II show that RPB4 and RPB7 are a subcomplex and are associated with a 10-subunit core enzyme [27]. The co-fractionation of TbRPB7 with TbRPB4 suggests a similar scenario in trypanosomes. Bioinfor-

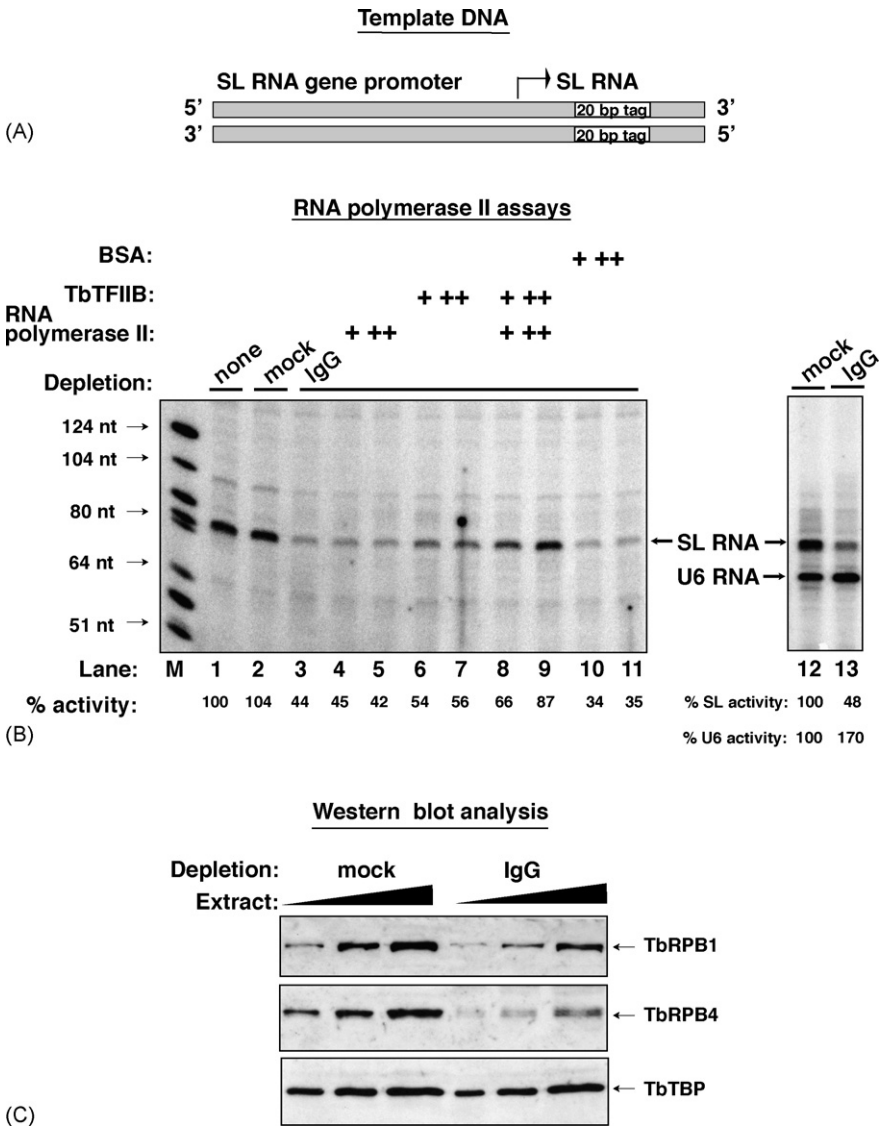

Fig. 5. SL RNA transcription in vitro can be restored with purified *T. brucei* RNA polymerase II along with TbTFIIB. (A) Schematic showing the wild-type SL RNA gene promoter and a portion of the transcribed tagged SL RNA used in promoter-dependent transcription assays. (B) In vitro transcription reactions. SL RNA transcription activity was lost in IgG-depleted extracts (compare lanes 3 and 13 with lanes 1, 2 and 12). Previous studies suggested that TbTFIIB coimmunoprecipitates with RNA pol II [1]. Restoration of transcription activity was observed when both RNA pol II and TbTFIIB were added together (lanes 8 and 9). Neither RNA pol II, 50 ng and 100 ng (+ and ++, lanes 4 and 5) nor TbTFIIB, 100 ng and 200 ng (+ and ++, lanes 6 and 7) alone restored activity. Comparable amounts of BSA (200 ng and 400 ng, lanes 10 and 11) were added as control reactions. The RNA pol III driven U6 RNA transcription activity remains unaffected (compare lane 12 and 13); indicating that only RNA pol II was depleted from the nuclear extracts. Percent SL RNA transcription activity, relative to lane 1, is shown below lanes 2–11. Percent SL RNA and U6 transcription activity, relative to lane 12, is shown below lane 13. Arrows that flank the PhosphorImager™ analysis of the 7 M urea-10% polyacrylamide gel indicate the SL RNA and U6 RNAs as well as nucleotide length markers. (C) Western blot analysis of protein A-sepharose-depleted (mock) and IgG-depleted (IgG) nuclear extracts. Antibodies specific for TbRPB1 and TbTBP identified their cognate proteins in 2 μl, 5 μl and 10 μl aliquots of nuclear extract. Specific depletion of TbRPB4 (detected by IgG antibodies that recognize TAP-tagged TbRPB4) and TbRPB1, and retention of TbTBP established that RNA pol II was removed from nuclear extracts.

matic analyses of TbRPB5 and TbRPB6 identified two isoforms of these proteins, annotated as RPB5 and 5z, and RPB6 and 6z [14]. RPB5z but not RPB5 co-purified with RNA pol I [28]. The presence of RPB5 in our RNA pol II purification and similar data from Devaux et al. [21] indicate that *T. brucei* RNA pol I and II rely on different RPB5 isoforms. We cannot predict which isoform of the RPB6 subunit functions as a part of *T. brucei* RNA pol II.

In this paper we describe the isolation of an RNA polymerase enzyme that contains the predicted subunits of RNA pol II and, more importantly, possesses RNA pol II activity. Namely, this enzyme exhibits α-amanitin sensitivity during transcription elongation on a promoterless template. In addition, this enzyme also restores SL RNA synthesis in extracts depleted of RNA pol II. Our work is significant as it provides a foundation for reconstituting the trypanosomal RNA pol II machinery in vitro, which may ultimately aid in the understanding RNA pol II-dependent gene expression.

While cataloging the factors involved in SL RNA gene transcription initiation has begun, the proteins involved in many

other trypanosome transcriptional processes remain elusive. In biochemical studies of human and yeast RNA pol II, a limited set of associated proteins involved in transcription initiation and elongation co-fractionate with the enzyme [19]. The co-purification of seven proteins with *T. brucei* RNA pol II is an important step toward understanding these processes in trypanosomes. RNA pol II-associated protein A1 is an NADH-dependent fumarate reductase. Increasing evidence indicates a direct role of metabolic enzymes in the regulation of gene expression. In mammals, the transcription factor Oct-1 coactivator, OCA-S regulates the expression of histone H2B. Interestingly, OCA-S is stimulated by NAD+, likely through its glyceraldehyde-3-phosphate dehydrogenase subunit [29]. We predict that the association of NADH-dependent fumarate reductase with *T. brucei* RNA pol II links transcriptional regulation and redox sensing of the parasite to its environment. RNA pol II-associated protein A3 contains an amino terminal ubiquitin domain. Ubiquitin as well as proteins that contain ubiquitin domains function in the regulation of eukaryotic mRNA synthesis [30]. Proteins A4 and A5 are DEAD-box helicase family members [31]. Helicases function in many aspects of RNA metabolism, including nuclear transcription, processing and transport of RNA. Interestingly, an RNA helicase that may be distinct from the proteins we identified associates with RNA pol II in the purification performed by Devaux et al. [21]. In addition, both our work and that of Devaux et al. identified RHS4 as a co-fractionating protein in RNA pol II preparations.

The A6 and A7 RNA pol II-associated proteins are unique to trypanosomes as they lack homologs in other eukaryotes. However, Pfam and ProDom database searches [32,33] indicate modest similarity to proteins involved in DNA binding and transcription regulation. As several RNA pol II-associated initiation and elongation factors appear to be absent in the Tri-Tryp databases, we speculate that A6 and A7 functionally replace one or more of these proteins.

In summary, the purification of a trypanosomal RNA polymerase II activity in vitro and the identification of seven associating proteins, A1–A7, lays the groundwork for extending our knowledge of RNA polymerase II function in these parasites.

Acknowledgements

This work was supported by NIH-NIAID grant AI 29478 to VB and NS046593 to HL. We thank Gina Conforti for the anti-TbRPB1 antibody. We thank Purnima Bhanot, David Lukac and Jennifer Palenchar for critical reading of this manuscript.

References

[1] Palenchar JB, et al. A divergent transcription factor TFIIB in trypanosomes is required for RNA polymerase II-dependent spliced leader RNA transcription and cell viability. Eukaryot Cell 2006;5(2):293–300.
[2] Schimanski B, et al. A TFIIB-like protein is indispensable for spliced leader RNA gene transcription in *Trypanosoma brucei*. Nucl Acids Res 2006;34(6):1676–84.
[3] Ruan JP, et al. Functional characterization of a *Trypanosoma brucei* TATA-binding protein-related factor points to a universal regulator of transcription in trypanosomes. Mol Cell Biol 2004;24(21):9610–8.
[4] Das A, et al. Trypanosomal TBP functions with the multisubunit transcription factor tSNAP to direct Spliced leader RNA gene expression. Mol Cell Biol 2005;25(16):7314–22.
[5] Schimanski B, Nguyen TN, Gunzl A. Characterization of a multisubunit transcription factor complex essential for spliced-leader RNA gene transcription in *Trypanosoma brucei*. Mol Cell Biol 2005;25(16):7303–13.
[6] Martinez-Calvillo S, et al. Transcription initiation and termination on *Leishmania major* chromosome 3. Eukaryot Cell 2004;3(2):506–17.
[7] Ullu E, Matthews KR, Tschudi C. Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts. Mol Cell Biol 1993;13(1):720–5.
[8] Hahn S. Structure and mechanism of the RNA polymerase II transcription machinery. Nat Struct Mol Biol 2004;11(5):394–403.
[9] Proudfoot NJ, Furger A, Dye MJ. Integrating mRNA processing with transcription. Cell 2002;108(4):501–12.
[10] Meinhart A, et al. A structural perspective of CTD function. Genes Dev 2005;19(12):1401–15.
[11] Ahearn Jr JM, et al. Cloning and sequence analysis of the mouse genomic locus encoding the largest subunit of RNA polymerase II. J Biol Chem 1987;262(22):10695–705.
[12] Evers R, et al. *Trypanosoma brucei* contains two RNA polymerase II largest subunit genes with an altered C-terminal domain. Cell 1989;56(4):585–97.
[13] Smith JL, et al. In trypanosomes the homolog of the largest subunit of RNA Polymerase II Is encoded by two genes and has a highly unusual C-terminal domain structure. Cell 1989;56:815–27.
[14] Kelly S, Wickstead B, Gull K. An in silico analysis of trypanosomatid RNA polymerases: insights into their unusual transcription. Biochem Soc Trans 2005;33(Pt 6):1435–7.
[15] Wirtz E, et al. A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. Mol Biochem Parasitol 1999;99(1):89–101.
[16] Rigaut G, et al. A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol 1999;17(10):1030–2.
[17] Kadesch TR, Rosenberg S, Chamberlin MJ. Binding of Escherichia coli RNA polymerase holoenzyme to bacteriophage T7 DNA. Measurements of binding at bacteriophage T7 promoter A1 using a template competition assay. J Mol Biol 1982;155(1):1–29.
[18] Bushnell DA, Kornberg RD. Complete, 12-subunit RNA polymerase II at 4. 1 Å resolution: implications for the initiation of transcription. Proc Natl Acad Sci USA 2003;100(12):6969–73.
[19] Jeronimo C, et al. RPAP1, a novel human RNA polymerase II-associated protein affinity purified with recombinant wild-type and mutated polymerase subunits. Mol Cell Biol 2004;24(16):7043–58.
[20] Ivens AC, et al. The genome of the kinetoplastid parasite, *Leishmania major*. Science 2005;309(5733):436–42.
[21] Devaux S, et al. Characterization of RNA polymerase II subunits of *Trypanosoma brucei*. Mol Biochem Parasitol 2006;148(1):60–8.
[22] Nguyen TN, et al. Purification of an eight subunit RNA polymerase I complex in *Trypanosoma brucei*. Mol Biochem Parasitol 2006;149(1):27–37.
[23] Besteiro S, et al. Succinate secreted by *Trypanosoma brucei* is produced by a novel and unique glycosomal enzyme, NADH-dependent fumarate reductase. J Biol Chem 2002;277(41):38001–12.
[24] Bringaud F, et al. A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. Eukaryot Cell 2002;1(1):137–51.
[25] Titz B, et al. Transcriptional activators in yeast. Nucl Acids Res 2006;34(3):955–67.
[26] Armache KJ, Kettenberger H, Cramer P. Architecture of initiation-competent 12-subunit RNA polymerase II. Proc Natl Acad Sci USA 2003;100(12):6964–8.
[27] Choder M. Rpb4 and Rpb7: subunits of RNA polymerase II and beyond. Trends Biochem Sci 2004;29(12):674–81.
[28] Walgraffe D, et al. Characterization of subunits of the RNA polymerase I complex in *Trypanosoma brucei*. Mol Biochem Parasitol 2005;139(2):249–60.

[29] Zheng L, Roeder RG, Luo Y. S phase activation of the histone H2B promoter by OCA-S, a coactivator complex that contains GAPDH as a key component. Cell 2003;114(2):255–66.
[30] Conaway RC, Brower CS, Conaway JW. Emerging roles of ubiquitin in transcription regulation. Science 2002;296(5571):1254–8.
[31] Rocak S, Linder P. DEAD-box proteins: the driving forces behind RNA metabolism. Nat Rev Mol Cell Biol 2004;5(3):232–41.
[32] Bateman A, et al. The Pfam protein families database. Nucl Acids Res 2004;32(Database issue):D138–41.
[33] Servant F, et al. ProDom: automated clustering of homologous domains. Brief Bioinform 2002;3(3):246–51.
[34] Meka H, et al. Crystal structure and RNA binding of the Rpb4/Rpb7 subunits of human RNA polymerase II. Nucl Acids Res 2005;33(19): 6435–44.