

# Characterization of a Multisubunit Transcription Factor Complex Essential for Spliced-Leader RNA Gene Transcription in *Trypanosoma brucei*†

Bernd Schimanski, Tu N. Nguyen, and Arthur Günzl\*

Department of Genetics and Developmental Biology and Department of Molecular, Microbial, and Structural Biology, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, Connecticut 06030-3710

Received 15 March 2005/Returned for modification 28 April 2005/Accepted 20 May 2005

**In the unicellular human parasites *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania* spp., the spliced-leader (SL) RNA is a key molecule in gene expression donating its 5'-terminal region in SL addition *trans* splicing of nuclear pre-mRNA. While there is no evidence that this process exists in mammals, it is obligatory in mRNA maturation of trypanosomatid parasites. Hence, throughout their life cycle, these organisms crucially depend on high levels of SL RNA synthesis. As putative SL RNA gene transcription factors, a partially characterized small nuclear RNA-activating protein complex (SNAP<sub>c</sub>) and the TATA-binding protein related factor 4 (TRF4) have been identified thus far. Here, by tagging TRF4 with a novel epitope combination termed PTP, we tandem affinity purified from crude *T. brucei* extracts a stable and transcriptionally active complex of six proteins. Besides TRF4 these were identified as extremely divergent subunits of SNAP<sub>c</sub> and of transcription factor IIA (TFIIA). The latter finding was unexpected since genome databases of trypanosomatid parasites appeared to lack general class II transcription factors. As we demonstrate, the TRF4/SNAP<sub>c</sub>/TFIIA complex binds specifically to the SL RNA gene promoter upstream sequence element and is absolutely essential for SL RNA gene transcription *in vitro*.**

Trypanosomatid parasites of the genera *Trypanosoma* and *Leishmania* cause devastating diseases in humans and livestock. During infection, these organisms essentially depend on gene expression mechanisms not found in their hosts. In trypanosomatids, nuclear protein-coding genes are transcribed polycistronically and individual mRNAs are excised from large precursors by *trans* splicing and polyadenylation (reviewed in reference 22). *trans* splicing in which the capped 5'-terminal spliced leader (SL), also known as the mini-exon, is cleaved off the SL RNA and fused to the 5' end of a pre-mRNA is an essential maturation step for all trypanosomatid mRNAs. Since the SL RNA is consumed in the process, trypanosomatid organisms crucially depend on a high level of SL RNA synthesis, making it a promising target for parasite-specific inhibition. To accommodate the high synthesis rate, a trypanosome cell harbors approximately 200 SL RNA gene copies, which are transcribed by RNA polymerase II in a monocistronic fashion (3). The anatomy of the SL RNA gene promoter has been meticulously investigated in the three trypanosomatid species *Trypanosoma brucei* (7), *Leptomonas seymouri* (8, 13), and *Leishmania tarentolae* (38); it is conserved and consists of a bipartite upstream sequence element (USE) and a putative initiator element (23) at the transcription initiation site.

Trypanosomatids diverged from the main eukaryotic lineage very early in evolution (11, 34). As a consequence, protein identification in trypanosomatid genome databases is only suc-

cessful for the most conserved proteins, such as the TATA-binding protein (TBP)-related factor 4 (TRF4). Chromatin immunoprecipitation revealed that TRF4 is associated with SL RNA gene sequences, and an RNA interference analysis indicated that TRF4 functions in SL RNA gene transcription (30). A second factor associated with SL RNA gene transcription was biochemically purified in *L. seymouri* (2). The factor was termed PBP-1; it consists of three subunits with apparent molecular masses of 57, 46, and 36 kDa, and it binds specifically to the SL RNA gene USE (25). While p36 was not characterized, p46 appeared to be a parasite-specific protein (2). Conversely, p57 was identified as a divergent orthologue of SNAP50, a subunit of the human small nuclear RNA (snRNA)-activating protein complex (SNAP<sub>c</sub> [reference 9]) also described as a proximal sequence element-binding transcription factor (37), suggesting that PBP-1 is a SNAP<sub>c</sub>-like factor (2).

Human SNAP<sub>c</sub> is an essential factor for RNA polymerase II- or III-mediated transcription of snRNA genes (reviewed in reference 10). The protein complex binds to the proximal sequence element of snRNA gene promoters and consists of five subunits, three of which are essential for transcriptional activation, namely SNAP190, SNAP50, and SNAP43. The recent characterization of *Drosophila melanogaster* SNAP<sub>c</sub> revealed orthologues to these three proteins but no additional subunits (21). The sequence conservation between human and insect SNAPS is limited, and only functionally important domains exhibit substantial similarity (21). In addition to SNAP<sub>c</sub>, human snRNA gene transcription essentially depends on the basal transcription factors TBP, transcription factor IIA (TFIIA), TFIIB, TFIIF, and TFIIE (15). Orthologues of these factors may be involved in trypanosome SL RNA gene transcription as well because the 139-nucleotide-long SL RNA belongs to the class of snRNAs. However, except for the TBP

\* Corresponding author. Mailing address: Department of Genetics and Developmental Biology, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, CT 06030-3710. Phone: (860) 679-8878. Fax: (860) 679-8130. E-mail: gunzl@uchc.edu.

† Supplemental material for this article may be found at <http://mcb.asm.org/>.

homologue TRF4, they have not been identified in trypanosomatid genome databases which are virtually complete for *T. brucei* and *Leishmania major*.

We have recently cloned and epitope-tagged TbSNAP50, the *T. brucei* orthologue of human SNAP50 and *L. seymourei* p57, and confirmed that this protein binds to the SL RNA gene USE of *T. brucei* (32). Here, we have created a novel tag combination for tandem affinity purification (TAP) termed the PTP tag. By employing the PTP tagging and purification method initially with TRF4 and subsequently with other subunits, we isolated a stable multisubunit complex which directs SL RNA gene transcription in a cell extract. The complex consists of TRF4, three SNAP<sub>c</sub> subunits, the *T. brucei* orthologue of the small TFIIA subunit, and a sixth protein which appears to be an extremely divergent orthologue of the large TFIIA subunit.

## MATERIALS AND METHODS

**Plasmid construction.** The *T. brucei* genome integration construct pC-PTP-NEO is derived from pBluescript SK+ (Stratagene, La Jolla, CA) and contains two cassettes arranged in tandem. The first cassette is cloned into the ApaI and ClaI sites of the plasmid and, starting from the ApaI site, contains 745 bp of a C-terminal coding region of a trypanosomal protein, a NotI site, the coding sequence of the PTP tag, the translation stop codon TGA, and 470 bp of *TbRPA1* 3' flank. The second cassette (H23-NEO-T) was integrated into the HindIII and SacI restriction sites of the vector and contains the neomycin phosphotransferase gene (*NEO-R*) flanked 5' and 3' by the intergenic regions of heat shock protein 70 (*HSP70*) genes 2 and 3 and of the  $\beta$ - and  $\alpha$ -tubulin genes, respectively. In a further development of the original resistance marker cassette (20), we separated the *HSP70* intergenic region from *NEO-R* by an NdeI restriction site and *NEO-R* from the tubulin flank by BamHI, HpaI, and BstBI restriction sites. This allows precise replacement of the selectable marker gene within the context of pC-PTP-NEO. For PTP tagging of TbTRF4, TbSNAP50, TbSNAP2, and TbTFIIA-2 derivatives of pC-PTP-NEO were generated. In each case, the C-terminal protein-coding region preceding the PTP sequence was exchanged. In pTbTRF4-PTP-NEO, pTbSNAP50-PTP-NEO, pTbSNAP2-PTP-NEO, and pTFIIA-2-PTP-NEO the corresponding gene sequence comprised 783, 462, 692, and 400 bp, respectively. For genomic integration, both SNAP plasmids were linearized with restriction enzyme BstBI, whereas pTbTRF4-PTP-NEO and pTFIIA-2-PTP-NEO were linearized with SnaBI and StuI, respectively.

In the construct pPURO-HA-TbTFIIA-1, which was used for N-terminal hemagglutinin (HA) tagging of TFIIA-1, an H23-PURO-T selectable marker cassette precedes a second cassette comprising 677 bp of *TbRPA2* 5' flank (31), the translation initiation codon, the HA tag, and 478 bp of the TFIIA-1 N-terminal coding region. The template constructs for in vitro transcription assays, SLins 19 and GPEET-trm, have been described previously (7, 19).

**Cell culture and generation of cell lines.** Cultivation of procyclic forms of *T. brucei* strain 427 was carried out as described previously (19). For the generation of cell lines expressing epitope-tagged proteins, 10  $\mu$ g of linearized plasmid was transfected into procyclic 427 cells by electroporation as detailed elsewhere (5). In the cell line TbA3, the second *TbSNAP2* allele was knocked out by a PCR product in which the coding region of the hygromycin phosphotransferase was fused to 101 bp of *TbSNAP2* 5' flank and 102 bp of *TbSNAP2* 3' flank. Transfected cells were cloned by limiting dilution and selected with 40  $\mu$ g/ml G418, 20  $\mu$ g/ml hygromycin, and/or 4  $\mu$ g/ml puromycin (Sigma, St. Louis, MO). Correct integration of constructs was verified by PCR and Southern analysis in each case, and expression of the epitope-tagged proteins was analyzed by immunoblotting with monoclonal anti-ProtC and anti-HA antibodies (Roche, Indianapolis, IN).

**PTP purification.** For purification of PTP-tagged proteins, a 2.5-liter culture of procyclic *T. brucei* cells was grown to a density of  $2 \times 10^7$  cells per ml, harvested, and extracted as described previously (19). All further steps were carried out at 4°C or on ice. The resulting 6 ml of unconcentrated extract was mixed with 0.5 ml of a 1-ml PA-150 buffer aliquot (150 mM potassium chloride, 20 mM Tris-HCl, pH 7.7, 3 mM MgCl<sub>2</sub>, 0.5 mM dithiothreitol, 0.1% Tween 20) in which a Complete Mini, EDTA-free protease inhibitor cocktail tablet (Roche, Indianapolis, IN) was dissolved. Subsequently, the extract was added to a 200- $\mu$ l settled bead volume of immunoglobulin G (IgG) Sepharose 6 Fast Flow beads (Amer-

sham Biosciences, Piscataway, NJ) and equilibrated with PA-150 buffer in a 0.8-by 4-cm Poly-Prep chromatography column (Bio-Rad, Hercules, CA). PTP-tagged proteins were bound to IgG Sepharose by rotating the closed column for 2 h. Subsequently, the flowthrough was collected by gravity flow and beads were washed with 25 ml of PA-150 buffer. After equilibrating the beads in 15 ml of tobacco etch virus (TEV) protease buffer (PA-150 with 0.5 mM EDTA), they were resuspended in 2 ml of TEV protease buffer containing 300 units of AcTEV protease (Invitrogen, Carlsbad, CA) and rotated overnight. The TEV protease eluate was collected by gravity flow, diluted to 6 ml by a wash of the IgG Sepharose beads with 4 ml of PC-150 buffer (PA-150 buffer containing 1 mM calcium chloride), and mixed with the remaining 0.5 ml of the protease inhibitor cocktail. For anti-ProtC affinity purification, calcium chloride was added to the TEV protease eluate to a final concentration of 2 mM, which was then combined in a new column with a 200- $\mu$ l settled bead volume of anti-protein C affinity matrix (Roche, Indianapolis, IN) equilibrated in PC-150 buffer. The column was rotated for 2 h, after which the flowthrough was collected and the matrix washed with 60 ml of PC-150. Finally, ProtC-tagged proteins were eluted either with EGTA elution buffer (5 mM Tris-HCl, pH 7.7, 10 mM EGTA, 5 mM EDTA, 10  $\mu$ g/ml leupeptin) or with peptide elution buffer (transcription buffer containing 0.5 mg/ml ProtC peptide and 0.1% Tween 20). In case of EGTA elution, five consecutive steps were carried out in which the beads were resuspended in 0.6 ml of EGTA elution buffer and rotated for 15 min at room temperature. For peptide elution, the beads were mixed with 300  $\mu$ l of peptide elution buffer and rotated at room temperature for 1 h.

The peptide eluate was used without further concentration in functional assays. The volume of the EGTA eluate was reduced from 3 ml to approximately 600  $\mu$ l by evaporation in a vacuum concentrator. Subsequently, the proteins were bound to 10  $\mu$ l of the hydrophobic StrataClean resin (Stratagene, La Jolla, CA), released into sodium dodecyl sulfate (SDS) loading buffer at 80°C, separated on SDS-polyacrylamide gels, and Coomassie stained with the GelCode blue stain reagent (Pierce, Rockford, IL). Proteins were identified by liquid chromatography-tandem mass spectrometry (see the list of identified peptides for each protein in the supplemental material).

**Western detection.** In immunoblot analyses, proteins were separated on SDS-polyacrylamide gels, electroblotted onto polyvinylidene difluoride membrane, and detected either by the protein A-specific PAP reagent (Sigma, St. Louis, MO), the anti-ProtC antibody HPC4, or a monoclonal anti-HA antibody in combination with the BM chemiluminescence blotting substrate (Roche, Indianapolis, IN). For detection of TbSNAP3, a polyclonal antibody against the peptide EMRRRINTESLLKRR was raised.

**Sedimentation on sucrose gradients.** Sedimentation analysis was carried out by ultracentrifugation of 3.8-ml 10 to 40% linear sucrose gradients containing 20 mM HEPES-KOH, pH 7.7, 20 mM potassium L-glutamate, 3 mM MgCl<sub>2</sub>, and 20 mM potassium chloride. Gradients were overlaid with 200  $\mu$ l of TbSNAP2-P eluate and centrifuged at 42,000 rpm in a Beckman SW55 rotor for 19 h at 4°C. Twenty fractions were collected from top to bottom, and protein was precipitated using StrataClean resin as described above. Proteins were separated on 12% SDS-polyacrylamide gels and visualized by silver staining using the Silver Stain Plus kit (Bio-Rad, Hercules, CA) according to the manufacturer's protocol.

**Promoter pull-down and in vitro transcription assays.** In vitro transcription reactions and promoter pull-down assays were carried out as described in detail elsewhere (18, 32). In brief, transcription reactions were carried out in a volume of 40  $\mu$ l for 60 min at 27°C and the reaction mixtures contained 8  $\mu$ l of extract, 20 mM potassium L-glutamate, 20 mM KCl, 3 mM MgCl<sub>2</sub>, 20 mM HEPES-KOH, pH 7.7, 0.5 mM of each nucleoside triphosphate (NTP), 20 mM creatine phosphate, 0.48 mg ml<sup>-1</sup> of creatine kinase, 2.5% polyethylene glycol, 0.2 mM EDTA, 0.5 mM EGTA, 4 mM dithiothreitol, 10 mg ml<sup>-1</sup> leupeptin, 10 mg ml<sup>-1</sup> aprotinin, and 40  $\mu$ g ml<sup>-1</sup> of exogenously added DNA. The latter comprised 7.5  $\mu$ g ml<sup>-1</sup> of SLins19 template, 20  $\mu$ g ml<sup>-1</sup> of GPEET-trm template, and 12.5  $\mu$ g ml<sup>-1</sup> of vector DNA. In addition, each reaction mixture contained 0.1 mg ml<sup>-1</sup> of ProtC peptide to exclude the possibility of a nonspecific peptide effect. To specifically detect GPEET-trm and SLins19 RNAs, total RNA was prepared from each reaction and analyzed by extension of <sup>32</sup>P-end-labeled primers Tag<sub>PE</sub> (19) and SLtag (7), which are complementary to unrelated oligonucleotide tags of GPEET-trm and SLins19, respectively. The primer extension products were separated on 6% polyacrylamide–50% urea gels and visualized by autoradiography.

For the pull-down assays, biotinylated promoter DNA fragments were generated by PCR using a 5'-biotinylated sense oligonucleotide. For each reaction, 500 ng of biotinylated DNA fragments was coupled to 10  $\mu$ l (100  $\mu$ g) of RNase-free, paramagnetic M-280 streptavidin Dynabeads (Dyna) according to the manufacturer's protocol. The beads were equilibrated and blocked for 30 min at room temperature in TK<sub>20</sub> buffer (150 mM sucrose, 20 mM HEPES-KOH, pH 7.7, 20

mM potassium L-glutamate, 20 mM KCl, 3 mM MgCl<sub>2</sub>, 2.5% [wt/vol] polyethylene glycol, 0.2 mM EDTA, 0.5 mM EGTA, 4 mM dithiothreitol, 10 μg ml<sup>-1</sup> leupeptin, 10 μg ml<sup>-1</sup> aprotinin) containing 5 mg ml<sup>-1</sup> bovine serum albumin and 5 mg ml<sup>-1</sup> polyvinylpyrrolidone. Subsequently, the beads were washed twice with 0.5 ml of TK<sub>20</sub> buffer and incubated in a 40-μl *in vitro* transcription reaction mixture for 15 min on ice and for 15 min at 27°C. Beads were washed three times with 0.5 ml TK<sub>20</sub> buffer and once with 0.5 ml of TN<sub>40</sub> buffer (150 mM sucrose, 20 mM Tris pH 8.0, 40 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.5 mM dithiothreitol, 10 μg ml<sup>-1</sup> leupeptin, 10 μg ml<sup>-1</sup> aprotinin) before proteins were eluted in standard SDS gel loading buffer by a 5-min-long incubation at 70°C.

## RESULTS

**PTP tagging and purification of TRF4.** The well-established TAP procedure allows for rapid purification of tagged proteins under native, nondenaturing conditions. The TAP tag consists of two protein A domains (ProtA) which bind to immobilized IgG, of a TEV protease cleavage site which facilitates release of the protein from the IgG column, and of the calmodulin binding peptide (CBP) (28, 29). We planned to use this technology to characterize *T. brucei* SNAP<sub>c</sub> and TRF4-associated proteins and analyze their potential function in SL RNA gene transcription. However, our attempts to purify TAP-tagged proteins from crude, transcriptionally competent extracts which consisted of a mix of cytoplasmic and extracted nuclear components failed due to the inefficiency of the second purification step, calmodulin affinity chromatography (data not shown). To overcome this problem, we modified the TAP tag by replacing the calmodulin binding peptide with the 12-amino-acid-short protein C epitope (ProtC), which is derived from human protein C, a vitamin K-dependent plasma zymogen specifically expressed in hepatocytes (Fig. 1A). As we have shown before, ProtC-tagged proteins can be efficiently purified from *T. brucei* extracts by anti-ProtC immunoaffinity chromatography using the monoclonal antibody HPC4 (31). HPC4 recognizes ProtC with very high affinity ( $K_d$ , ~1 nM) and, as a unique property, has a calcium binding site which needs to be occupied for epitope binding (35). To facilitate efficient PTP tagging of trypanosome proteins, we designed the genome integration vector pC-PTP-NEO, which allows fusion of a C-terminal protein-coding region to the PTP tag sequence in a single cloning step (Fig. 1B). Cell lines which stably express PTP-tagged proteins were established by targeted insertion of the vector to the gene of interest (Fig. 1C).

For TRF4 purification, we established the TRF4-PTP-expressing procyclic cell line TbT1 (Fig. 2A). As starting material, extract was prepared from a 2.5-liter TbT1 culture which corresponded to  $5 \times 10^{10}$  cells or a packed cell volume of approximately 4 ml. Each purification step was monitored by immunoblot analysis (Fig. 2B). In both IgG and anti-ProtC affinity chromatography steps, tagged TRF4 bound very efficiently to the column, leaving only trace amounts of the protein in the flowthrough (compare lanes 2 and 4 with lanes 1 and 3, respectively). As a consequence of TEV protease cleavage, TRF4-PTP was shortened to TRF4-P, resulting in a faster-migrating protein band (compare lane 1 with lanes 3 to 5). Due to the calcium-dependent interaction of HPC4 with ProtC, TRF4-P was efficiently eluted in the presence of EGTA (lane 5). SDS gel electrophoresis and Coomassie staining revealed six major protein bands with apparent sizes of 75, 55, 46, 37, 26, and 25 kDa which copurified in approximately stoichiometric amounts. In addition, a minor protein band with an apparent

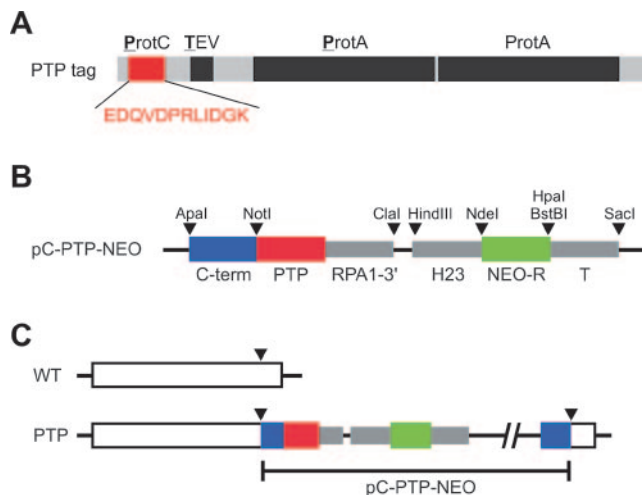


FIG. 1. PTP tagging. (A) Schematic outline to scale of the protein C epitope/TEV protease cleavage site/protein A (PTP) tag. ProtC and the ProtC sequence are red, the TEV site and the two protein A domains are black, and spacer sequences are gray. (B) Graphical depiction (not to scale) of the *T. brucei* genome integration vector pC-PTP-NEO. Large rectangles represent coding regions, and small rectangles depict trypanosome gene flanking regions which encode signals for *trans* splicing and polyadenylation. Vector sequences are indicated by the black line. The construct is a derivative of pBluescript SK+ and designed for in-allele integration to fuse the PTP tag C terminally to a target protein. It contains two tripartite cassettes. The first cassette consists of a C-terminal protein-coding region (C-term; blue), the PTP tag (red), and the *TbRPA1* 3' flank (RPA1-3'; gray). The second cassette comprises the neomycin resistance gene (NEO-R; green) flanked by the *HSP70* genes 2 and 3 (H23) intergenic region and the  $\beta/\alpha$ -tubulin (T) intergenic region. Unique restriction sites of this vector are indicated by arrowheads. For PTP tagging of a protein, the C-terminal coding region (blue) in pC-PTP-NEO has to be replaced by the corresponding sequence of the target protein via the *Apa*I and *Not*I restriction sites. (C) Illustration of pC-PTP-NEO integration into the genome. Shown are a wild-type allele (WT) and an allele with the integrated construct (PTP). For site-directed integration by homologous recombination, the vector is linearized (arrowhead) in the C-terminal coding region (blue). Colors correspond to the description above.

size of 70 kDa and several faint bands below the 55-kDa band were detectable. Major and minor bands appeared to be specific because they did not copurify with PTP-tagged RNA polymerase I (T. N. Nguyen, B. Schimanski, and A. Günzl, unpublished results). In contrast, mass spectrometric analysis of the faint bands identified these to be contaminations consisting mainly of  $\alpha$ - and  $\beta$ -tubulin and of IgG heavy chain (Fig. 2C) (data not shown).

**TRF4 is associated with SNAP<sub>c</sub>.** The six major bands and the minor band were excised from the gel and individually analyzed by mass spectrometry. In accordance with the immunoblot results, the protein with an apparent molecular mass of 37 kDa was identified as TRF4-P. The minor band of 70 kDa was identified as the *T. brucei* orthologue of the TFIIIB-related protein 1 (BRF1; GeneDB entry Tb11.03.0670), indicating that TRF4 is a component of trypanosome TFIIIB (reviewed in reference 16). Given the fact that TFIIIB is a conserved and essential factor for eukaryotic transcription of tRNA genes, 5S rRNA genes, and class III snRNA genes, it was surprising to



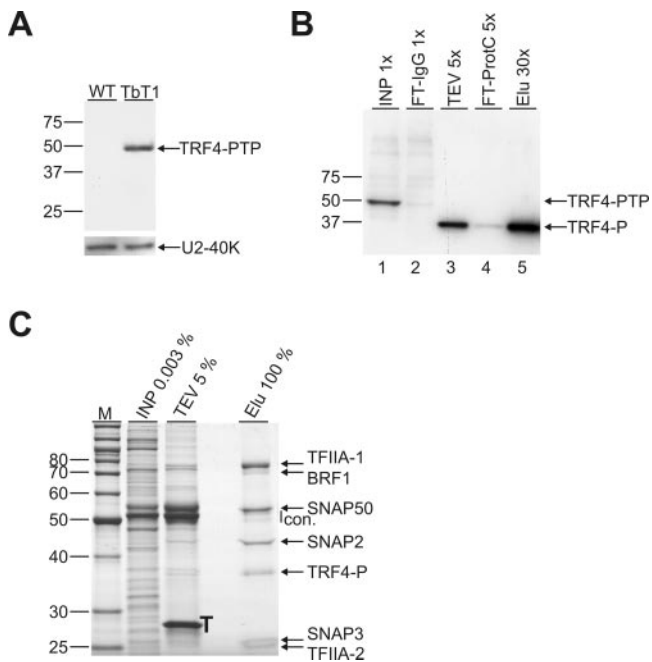


FIG. 2. Purification of TRF4-PTP. (A) Immunoblot analysis of wild-type (WT) and TbT1 whole-cell lysates. TRF4-PTP was detected with the anti-ProtC antibody HPC4. Protein loading was controlled by reprobing the same blot with an antibody against the U2 snRNP-specific protein 40K (U2-40K). Sizes of marker proteins are indicated on the left. (B) Immunoblot monitoring of TRF4-PTP purification. Aliquots of the input material (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 EGTA eluate (Elu) were separated on a 10% SDS-polyacrylamide gel, blotted, and probed with anti-ProtC antibody. The relative amount of each sample to the input material is specified. On the left are indicated marker protein sizes and on the right positions of TRF4-PTP and TRF4-P after TEV protease cleavage. (C) Coomassie staining of TRF4-PTP purification was separated on a 12.5% SDS-polyacrylamide gel and stained with Coomassie. For comparison, 0.003% of the input material (INP) and 5% of the TEV protease eluate (TEV) were loaded. Arrows on the right point to proteins which were identified by mass spectrometry (con., contamination). T marks the TEV protease in the corresponding eluate, and on the left, sizes of marker proteins are indicated.

discover that a putative TFIIIB component copurified with TRF4 in much smaller amounts than five other proteins.

The major 55-kDa protein was identified as TbSNAP50, indicating that TRF4 interacts with TbSNAP<sub>c</sub>. The other four major proteins were new discoveries and are described here for the first time. The protein with an apparent size of 26 kDa was unambiguously identified by sequence homology to be an orthologue of human and insect SNAP43 (Fig. 3A). The protein consists of 234 amino acids and has a predicted size of 26.1 kDa. A comparison between human and insect SNAP43 had shown that only the N-terminal region is conserved (21) due to its interacting function with SNAP50 (24). Correspondingly, sequence conservation in the 26-kDa protein is restricted to the N terminus. We named this protein TbSNAP3 because of the finding of another SNAP of larger size (see below) and because of its deviant size in comparison to its human orthologue. Interestingly, TbSNAP3 is 10 kDa smaller than the smallest yet uncharacterized SNAP<sub>c</sub> subunit of *L. seymouri*

(25). We have identified the gene sequences encoding the TbSNAP3 orthologues in the genome databases of *T. cruzi* (LM.36.1.Contig1; Sanger Center, Cambridge, United Kingdom) and *L. major* (t\_cruzi chr\_0 1047053510181 5857; TIGR, Rockville, MD). While the predicted sizes of the *T. cruzi* protein TcSNAP3 and TbSNAP3 are similar, the *L. major* protein LmSNAP3 has a predicted size of 38 kDa, in accordance with the apparent size of the *L. seymouri* protein. As revealed by a sequence alignment, the larger size of LmSNAP3 is due to an internal domain which is lacking in the trypanosome proteins (Fig. 3A) (data not shown).

The protein with an apparent size of 46 kDa was identified as an orthologue of the p46 subunit of the *L. seymouri* SNAP<sub>c</sub> (2) and therefore was termed TbSNAP2. TbSNAP2 consists of 372 amino acids with a predicted size of 42.1 kDa. Although it was possible to detect the homologous sequences in *T. cruzi* and *L. major* genome databases, the protein is not well conserved and again, only the N-terminal regions exhibit a significant similarity (Fig. 3B). Further support for the classification of p46 as a SNAP came from two observations. First, p46 cosedimented with TbSNAP50 in a sucrose gradient in the absence of other proteins (see below and see Fig. 5B). Second, in PTP tagging of TbSNAP50, we generated a cell line which for yet unknown reasons expressed TbSNAP50-PTP which did not bind TRF4. The only two proteins which copurified with this apparently defective protein were p46 and TbSNAP3 (data not shown).

Since TbSNAP50 and TbSNAP3 are orthologues of human SNAP50 and SNAP43, respectively, it is possible that TbSNAP2 is an extremely divergent orthologue of the third essential human subunit, SNAP190. A sequence comparison between human and insect SNAP190 revealed that functional conservation is restricted to four and a half unusual Myb repeats (21) which play a role in specific DNA binding (36). Two and a half of these Myb repeats appear to be present in TbSNAP2 aligning best with repeats a to c of the human and insect sequences (Fig. 3C). Surprisingly, however, the putative Myb repeats are not conserved among the trypanosomatid orthologues (data not shown). Hence, in the absence of conserved sequence domains, a functional analysis will be necessary to determine whether SNAP2 is a parasite-specific protein or functionally equivalent to SNAP190 of higher eukaryotes.

Finally, to prove correct identification of TbSNAP2 and TbSNAP3, we HA tagged either protein in cells expressing TbSNAP50-PTP and showed that both proteins copurified with TbSNAP50, in contrast to an unrelated HA-tagged protein (Fig. 4, compare lanes 3, 6, and 9).

**TRF4 is associated with TFIIA.** The smallest protein which copurified with TRF4-P has an apparent molecular size of 25 kDa and was unambiguously identified as the orthologue of the small TFIIA subunit, which is known as the  $\gamma$  subunit in higher eukaryotes and as TOA2 in the budding yeast *Saccharomyces cerevisiae*. The conservation of this protein is sufficient for an unambiguous assignment, and therefore it was termed TbTFIIA-2 (Fig. 5A). As has been analyzed in yeasts and mammals, TFIIA is a basal factor for RNA polymerase II-mediated transcription (26) and absolutely essential for snRNA gene transcription (15). While yeast TFIIA consists of the two subunits TOA1 and TOA2, the larger subunit of higher eukaryotes is proteolytically cleaved into  $\alpha$  and  $\beta$  subunits. The

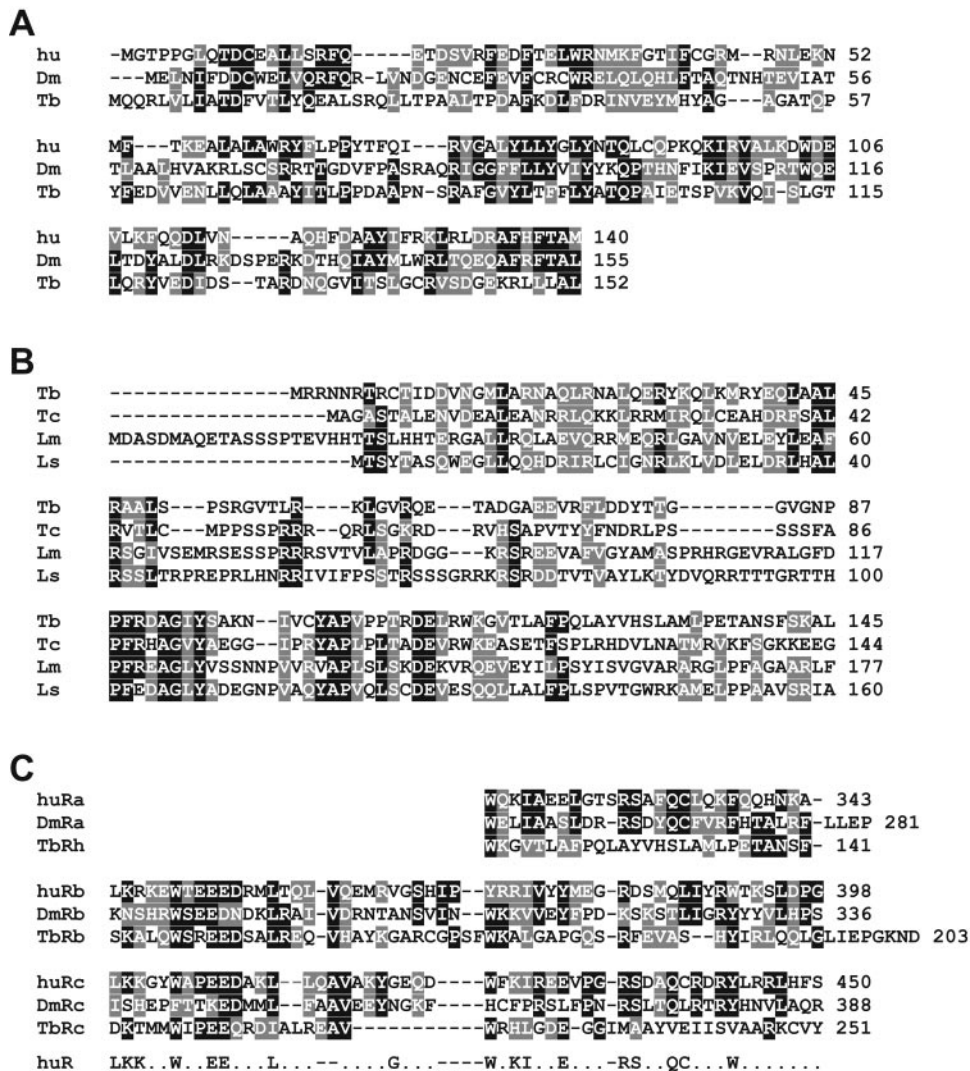


FIG. 3. SNAP sequence alignments. Black- and gray-shaded letters indicate identical and similar residues, respectively. (A) Alignment of N-terminal sequences of human SNAP43 (hu; accession no. CAA87590), *D. melanogaster* SNAP43 (Dm; NP\_650499), and *T. brucei* SNAP3 (Tb; AJ879576). (B) Alignment of SNAP2 N-terminal sequences of *T. brucei* (Tb; AJ879575), *T. cruzi* (Tc; t\_cruzi chr\_0 1047053506947 7150; TIGR, Rockville, MD), *L. major* (Lm; LM16.2.Contig21; Sanger Center, Cambridge, United Kingdom), and *L. seymouri* (Ls; AY160127). Only positions with a minimum of three identical or three conserved residues are shaded. (C) Alignment of the human SNAP190 Myb repeats huRa-huRc with the homologous sequences of *D. melanogaster* (DmRa-DmRc) and the putative Myb repeats in TbSNAP2 (TbRh; half-repeat). The alignment is according to Wong et al. (36) with minor modification. The human consensus sequence (huR) is based on the four and a half SNAP190 Myb repeats. Letters indicate positions with not more than one mismatch. The characteristic tryptophans of Myb domains which may be replaced by tyrosines or phenylalanines are underlined.

bi- and tripartite nature of TFIIA in other eukaryotes strongly argues that the remaining major protein which copurified with TRF4-P is the orthologue of TOA1/TFIIA- $\alpha$ B. However, the corresponding *T. brucei* protein which was identified from the 75-kDa protein band is the least conserved among the identified proteins. While a putative orthologue with 22.3%/34.2% sequence identity/similarity could be identified in the *T. cruzi* genome database (t\_cruzi chr\_0 1047053508897 7919; The Institute for Genomic Research [TIGR], Rockville, MD), no convincingly conserved sequence was obtained from the database of the more distantly related *L. major*. Among TOA1 and TFIIA- $\alpha$ B sequences, only the 50 N-terminal residues and 70 C-terminal residues are conserved. As has been shown for

yeast TOA1, the internal region serves mainly as a spacer region, contributing little to TOA1 function (14). Sequence alignments of the N- and C-terminal regions showed that some of the most conserved amino acid residues at both termini are present in the *T. brucei* sequence, supporting the notion that this protein is the orthologue of the large TFIIA subunit (Fig. 5B). Although it will require a functional analysis for a final assignment, we have tentatively named this protein TbTFIIA-1.

**TRF4, SNAP<sub>2</sub>, and TFIIA form a stable protein complex.** The protein profile of the TRF4-PTP purification raised the possibility that the six major proteins form a single complex. Alternatively, TRF4 may form independent complexes with

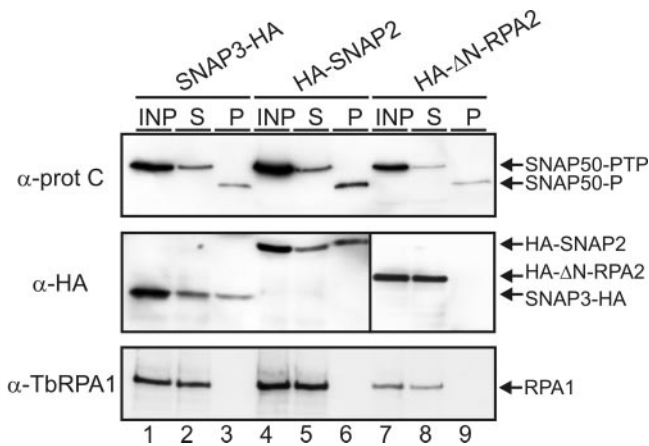


FIG. 4. TbSNAP2 and TbSNAP3 copurify with TbSNAP50-PTP. Three cell lines were generated, all of which expressed TbSNAP50-PTP together with HA-tagged versions of TbSNAP3 (C-terminal tag), TbSNAP2 (N-terminal tag), or a truncated version of TbRPA2, the second largest RNA polymerase I subunit (HA- $\Delta$ N-RPA2). In small-scale reactions (100  $\mu$ l), TbSNAP50-PTP was precipitated by IgG beads and released by TEV protease cleavage. Comparable amounts of input material (INP), of supernatant (S), and of released proteins from the precipitate (P) were analyzed by immunoblotting. On the same blot, TbSNAP50-PTP was detected with anti-ProtC antibody, HA-tagged proteins with anti-HA epitope antibody, and as a control, TbRPA1 with a specific polyclonal antiserum.  $\alpha$ -, anti-

TFIIA and SNAP<sub>c</sub>. To discriminate between these two possibilities, we C-terminally PTP-tagged TbSNAP2, TbSNAP50, and TbTFIIA-2 in separate cell lines and conducted PTP purifications. In all three cases, the same six major proteins were purified, strongly indicating that they form a single complex (Fig. 6A). Moreover, the purification pattern did not change when the potassium chloride concentration of the extract was increased to 1 M prior to purification, demonstrating that the tripartite complex is highly salt resistant (data not shown). Due to its larger size, the tagged protein in each purification shifted up in denaturing gel electrophoresis, again confirming that we identified and tagged the correct proteins (Fig. 6A, see arrows).

While we did not reproducibly detect a minor band copurifying with either TbSNAP2-P or TbSNAP50-P, TbTFIIA-2-PTP purification resulted in three minor bands and several faint bands (Fig. 6A, lane 6). These proteins have not been identified yet, but it is unlikely that they are contaminations because the absence of TEV protease in the final eluate suggested that the washing steps in this purification were as efficient as in the other purifications (Fig. 6A, compare lanes 2 and 6). As expected for the presence of an independent TFIIIB complex, BRF1 copurified only with TRF4 and not with TFIIA or SNAP<sub>c</sub> (Fig. 6A, compare lane 5 with lanes 3, 4, and 6).

To further establish the existence of a tripartite TRF4/SNAP<sub>c</sub>/TFIIA complex, we analyzed cosedimentation of the subunits in sucrose gradients. As expected, the six major proteins of the TbSNAP2-P eluate cosedimented in stoichiometric amounts in fractions 12 through 14 (Fig. 6B). A complex composed of six monomers has a predicted molecular mass of 230.6 kDa. In accordance with this size, the complex sedimented faster than the 150-kDa large IgG control. Furthermore, in-

creasing the potassium chloride concentration of the gradient to 400 mM did not change the sedimentation properties of the complex, verifying its salt stability (data not shown). Fraction 10 contained detectable amounts of only TbSNAP50 and TbSNAP2, indicating that the TbSNAP2-P final eluate contained both the tripartite complex and a separate SNAP<sub>c</sub>. Taken together, we concluded that in *T. brucei* TRF4, SNAP<sub>c</sub>, and TFIIA form a stable tripartite complex.

**TRF4/SNAP<sub>c</sub>/TFIIA subunits specifically bind to the SL RNA gene USE.** Previously we employed a promoter pull-down assay to show that TbSNAP50 binds specifically to the SL RNA gene USE (32, 33). We therefore explored the possibility that the other components of the TRF4/SNAP<sub>c</sub>/TFIIA complex bind the SL RNA gene promoter with the same specificity. Thus far, we had generated cell lines expressing PTP-tagged TRF4, SNAP2, and TFIIA-2. For the specific detection of the other two components, we generated a cell line which expressed TFIIA-1 with an HA tag fused to its N terminus and raised a polyclonal antiserum against SNAP3. The pull-down assay was carried out with transcription extract of the various cell lines and with linear, biotinylated DNAs which were immobilized on streptavidin beads. The SL RNA gene promoter fragment extended from position -126 to -18 relative to the transcription initiation site covering the USE but not the putative initiator element (Fig. 7A, SL RNA gene fragment -126/-18). As a negative control, we used a DNA fragment covering the GPEET promoter from position -246 to -3 because this DNA did not bind TbSNAP50 in our previous studies (32, 33). Correspondingly, the GPEET -246/-3 pull-down did not reveal binding of any of the TRF4/SNAP<sub>c</sub>/TFIIA subunits above background levels (Fig. 7B). In contrast, all subunits bound to the SL RNA gene fragment -126/-18. Subunit binding was independent of the PTP tag because a PTP-tagged version of the largest subunit of RNA polymerase I did not bind to SL RNA gene fragment -126/-18. Importantly, subunit binding to SL RNA gene fragment -126/-18 was sequence specific because mutation of USE1, the first sequence block of the bipartite USE, abolished subunit binding in all cases. Mutation of USE2 reduced subunit binding but did not abolish it, indicating that USE1 is the primary site for stable DNA-protein interaction. The exception was HA-tagged TFIIA-1, which in contrast to the other subunits, was unable to bind to the USE2 mutation, a phenotype which has likely been caused by the tag (Fig. 7B, panel HA-TFIIA-1). For comparison, extract corresponding to 20% of the input material was loaded. Although the binding efficiencies of the tagged subunits to SL RNA gene fragment -126/-18 were variable, binding of TFIIA-2-P was especially low, indicating that the tag partially impaired protein function here as well (panel TFIIA-2-P). Finally, we analyzed whether the purified protein complex was capable of specific interaction with the SL RNA gene USE. The pull-down was carried out with SNAP2-P eluate. While purified protein clearly bound to the wild-type USE, mutation of either USE1 or USE2 abolished the interaction, demonstrating the sequence specificity of protein binding. The latter finding was in contrast to the corresponding experiment carried out with extract, in which the same complex was able to bind to the DNA carrying the USE2 mutation (compare panel SNAP2-PTP with panel SNAP2-P). The discrepancy suggests that the extract harbored proteins which stabilized the DNA-



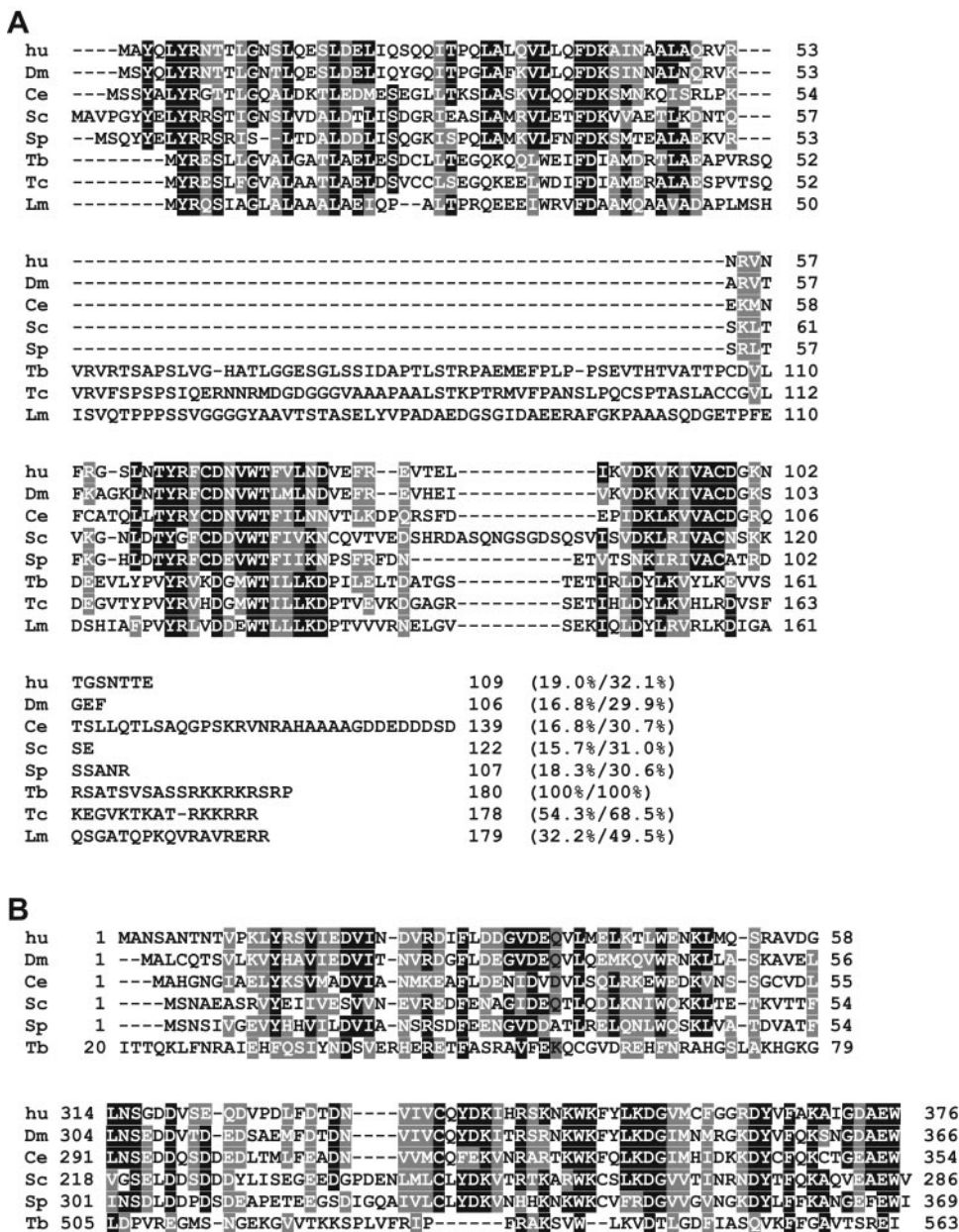


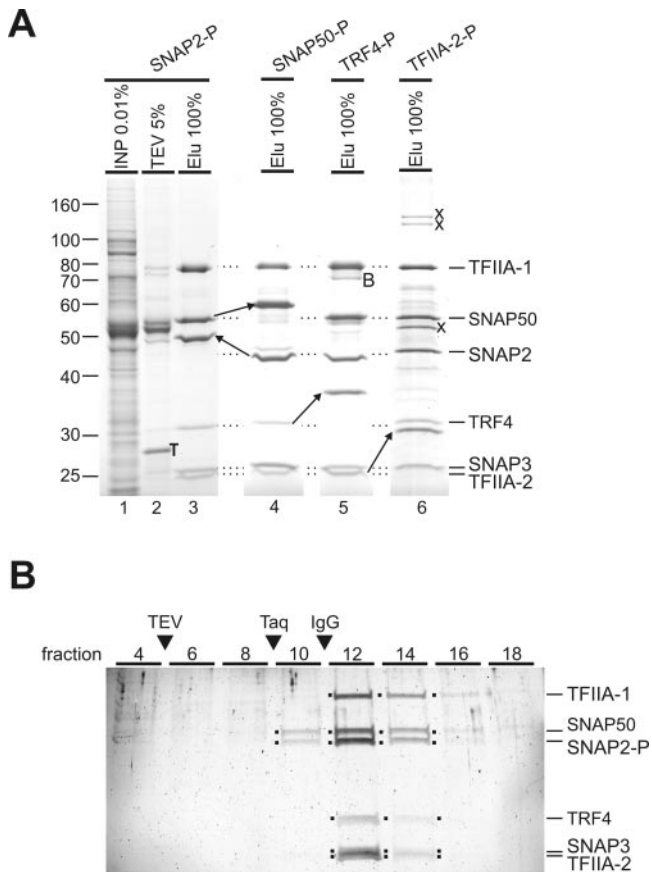
FIG. 5. TFIIA sequence alignments. (A) Sequence alignment of TFIIA- $\gamma$ /TOA2 orthologues of humans (hu; accession no. P52657), *D. melanogaster* (Dm; P52656), *Caenorhabditis elegans* (Ce; NP\_498226), *Saccharomyces cerevisiae* (Sc; NP\_012865), *Schizosaccharomyces pombe* (Sp; CAA19263), *T. brucei* (Tb; AJ887988), *T. cruzi* (Tc; t\_cruzi chr\_0 1047053510969 8557; TIGR, Rockville, MD), and *L. major* (Lm; LM36.1.Contigl; Sanger Center, Cambridge, United Kingdom). Identity/similarity values, determined in pairwise alignments between the *T. brucei* sequence and each of the other sequences, are indicated at the end of each sequence. At positions with a minimum of five identical or five conserved residues, letters are shaded black and gray, respectively. (B) Sequence alignment of N- and C-terminal regions of TFIIA- $\alpha$ /TOA1 orthologues. Shading is restricted to positions with four identical or similar residues. Accession numbers for the human, *D. melanogaster*, *C. elegans*, *S. cerevisiae*, *S. pombe*, and *T. brucei* sequences are P52655, A49076, T32660, NP\_014837, CAB57938, and AJ887987, respectively.

protein interaction but which did not copurify with TbSNAP2-P.

Taken together, these data demonstrate that all TRF4/SNAP<sub>2</sub>/TFIIA subunits specifically bound to the SL RNA gene USE and they suggest that the purified complex is capable of binding to this promoter element by itself.

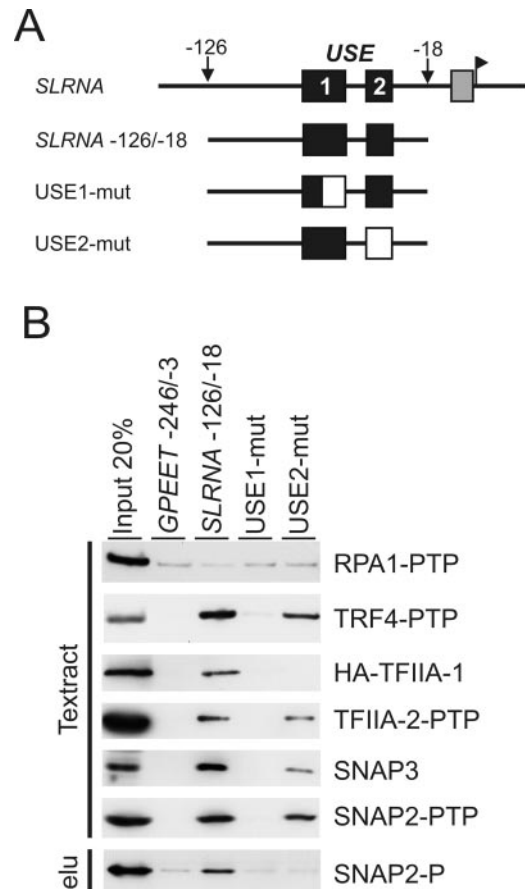
**Purified TRF4/SNAP<sub>2</sub>/TFIIA directs SL RNA gene transcription in vitro.** In a final step, we analyzed the function of

TRF4/SNAP<sub>2</sub>/TFIIA in SL RNA gene transcription. We first generated the clonal cell line TbA3, which exclusively expressed TbSNAP2-PTP (Fig. 8A). Importantly, the fact that this cell line grew normally proved that the PTP tag does not inherently interfere with protein function. Moreover, it allowed us to deplete TbSNAP2 from the extract through IgG affinity chromatography (Fig. 8B). Since the depletion step was carried out exactly in the same way as the first PTP purification



**FIG. 6.** *T. brucei* TRF4, SNAP<sub>o</sub>, and TFIIA form a complex. (A) Purification of SNAP2, SNAP50, and TFIIA-2 which were C-terminally PTP tagged in individual cell lines. Elu, eluate. Purified protein fractions were separated on a 12.5% SDS-polyacrylamide gel and stained with Coomassie. For comparison, the input material (INP) and the TEV protease eluate (TEV) of the SNAP2-PTP purification as well as a second TRF4-PTP purification are shown. Arrows indicate size shifts of tagged proteins, and dotted lines connect the corresponding untagged proteins across the purifications. T, B, and x mark the TEV protease, BRF1 in the TRF4-P eluate, and specific minor bands in the TFIIA-2-P eluate, respectively. Identified proteins are specified on the right, and protein marker sizes are indicated on the left. (B) Sedimentation of SNAP2-P eluate in a linear 10 to 40% sucrose gradient. Twenty fractions were collected from top to bottom. Proteins from the specified fractions were separated on a 12% SDS-polyacrylamide gel and stained with silver. Cosedimented proteins are indicated by dots. For comparison, sedimentations of TEV protease (29 kDa), Taq DNA polymerase (95 kDa), and IgG (150 kDa) were analyzed in parallel gradients (arrowheads).

step, it can be inferred from the protein profile of the Tb-SNAP2-PTP purification (Fig. 6A, lane 3) that depletion included the TRF4/SNAP<sub>o</sub>/TFIIA complex. To analyze the effect of complex depletion on SL RNA gene transcription, we carried out in vitro reactions in which we cotranscribed the template constructs GPEET-trm and SLins19. SLins19 contains a complete SL RNA gene tagged with an unrelated oligonucleotide sequence inside its coding region (7). The construct GPEET-trm is similarly tagged and served as a control template because it harbors the GPEET procyclin (*GPEET*) promoter, which recruits RNA polymerase I (6, 19), and because, as shown in Fig. 7, this promoter does not detectably bind any



**FIG. 7.** TRF4/SNAP<sub>o</sub>/TFIIA subunits specifically bind to the SL RNA gene (*SLRNA*) USE. (A) Schematic outline to scale of the SL RNA promoter. It consists of USE1 and USE2 which form the bipartite USE (black rectangles) and of a putative initiator element (gray rectangle). DNA fragments which were used in the promoter pull-down assays are depicted below. Mutations in USE1-mut and USE2-mut correspond to LS -71/-62 and LS -53/-42 described by Günzl et al. (7). (B) Promoter pull-down assays were carried out with the specified promoter fragments either in transcription extract (Texttract) containing epitope-tagged subunits of TRF4/SNAP<sub>o</sub>/TFIIA or in PTP-purified SNAP2-P eluate (elu). DNA-bound proteins were analyzed by immunoblotting with anti-epitope antibodies. Untagged SNAP3 was detected by antiserum directed against an internal peptide. For comparison, extract corresponding to 20% of the input material was co-analyzed.

of the TRF4/SNAP<sub>o</sub>/TFIIA subunits. The sequence tags in both constructs allow specific detection of the corresponding transcripts by primer extension assays (Fig. 8C). IgG affinity chromatography was carried out by rotating the extract for 2 h at 4°C. When we mock treated transcription extract, we saw that the conditions of IgG affinity chromatography reduced the transcription signals only slightly and that the production of a transcriptionally competent extract of the IgG flowthrough was feasible (compare lanes 1 and 2). Indeed, *GPEET* transcription was as efficient in TRF4/SNAP<sub>o</sub>/TFIIA-depleted extract as in mock-treated extract, confirming that the chromatography step did not interfere with transcriptional activity. In contrast to *GPEET* transcription, SL RNA gene transcription was completely abolished in the depleted extract, indicating that TRF4/



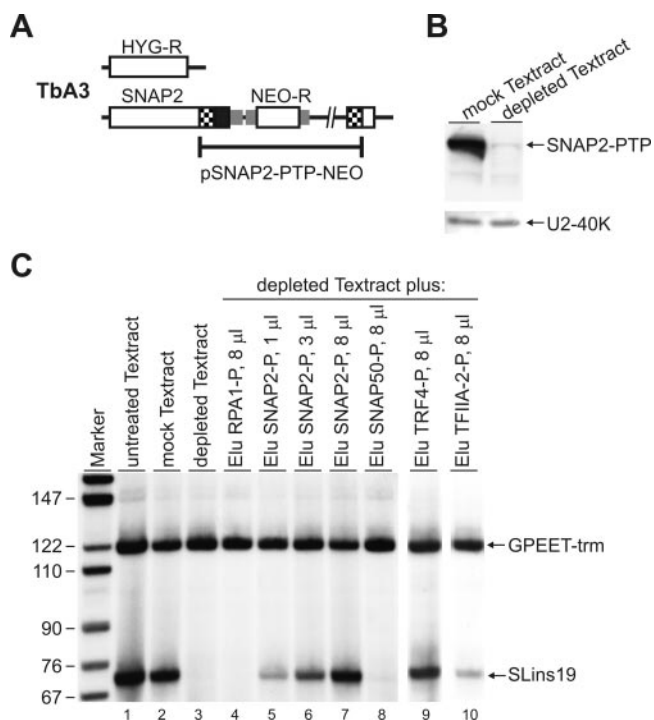


FIG. 8. TRF4/SNAP<sub>c</sub>/TFIIA is essential for SL RNA gene transcription in vitro. (A) Illustration of the two SNAP2 alleles in the clonal cell line TbA3. In one allele, the SNAP2 coding region was replaced by a hygromycin resistance gene (HYG-R) and in the second allele, the PTP coding region was fused to the 3' end by targeted insertion of pSNAP2-PTP-NEO. Coding regions are represented by open boxes, the PTP tag by a black box, homologous sequences in the construct by checkered boxes, and introduced gene flanks by small gray boxes. (B) Western detection of SNAP2-PTP with anti-ProtC antibody in mock-treated and depleted transcription extract (Textract). As a loading control, U2-40K protein was detected on the same blot. (C) Templates GPEET-trm and SLins19 were cotranscribed in untreated, mock-treated, or depleted transcription extract (Textract). Depleted extract was reconstituted with peptide eluates (Elu) from PTP purifications as specified. GPEET-trm and SLins19 transcripts were detected by primer extension of total RNA prepared from transcription reactions with <sup>32</sup>P (5' end)-labeled oligonucleotides TAG<sub>PE</sub> and SLtag, respectively. Reaction products were separated on a 6% polyacrylamide–50% urea gel and visualized by autoradiography. On the left, sizes of pBR322-MspI marker fragments are indicated, and on the right, GPEET-trm and SLins19 primer extension products are identified.

SNAP<sub>c</sub>/TFIIA depletion was causing this effect (compare lanes 2 and 3). To prove that this is the case, we reconstituted SL RNA gene transcription by adding final eluates of the PTP purifications. It is important to note here that we were unable to reconstitute SL RNA gene transcription with EGTA-derived eluates even after dialysis against transcription buffer, indicating that depletion of divalent cations irreversibly destroyed the TRF4/SNAP<sub>c</sub>/TFIIA transcription function (data not shown). However, ProtC-tagged protein can also be eluted in the presence of ProtC peptide. Therefore, we synthesized this peptide and eluted ProtC-tagged proteins from anti-ProtC matrix in the presence of 0.5 mg/ml peptide. Peptide elution was not as efficient as EGTA elution but resulted in protein concentrations comparable to those in the transcription extract (data not shown). Since transcription reaction mixtures con-

tained 8  $\mu$ l of extract, addition of 8  $\mu$ l of peptide eluate approximately restored the original SNAP2 concentration. Accordingly, the SL RNA gene transcription activity was completely reconstituted by the addition of 8  $\mu$ l of TbSNAP2-P final eluate (Fig. 8C, lanes 5 to 7). Similarly, the TRF4-P eluate was competent for full reconstitution (lane 9). It can be excluded that epitope tagging caused nonspecific activation of SL RNA gene transcription in these assays because, in a control reaction, addition of PTP-purified RNA polymerase I had no effect (lane 4). SL RNA gene transcription was also not reconstituted by the TbSNAP50-P eluate and incompletely restored by the TbTFIIA-2-P eluate, indicating that C-terminal tagging of these proteins interfered with transcriptional function (lanes 8 and 10). Our unsuccessful attempts to establish cell lines exclusively expressing TbSNAP50-PTP (data not shown) are in accordance with the former finding, whereas the latter result is in accordance with the inefficient binding phenotype of TFIIA-2-PTP in the promoter pull-down assays (Fig. 7).

In sum, we conclude that *T. brucei* TRF4/SNAP<sub>c</sub>/TFIIA is essential for SL RNA gene transcription in vitro. Since this complex was able to reconstitute transcription in an extract with no detectable SL RNA gene transcription activity, we infer that TRF4/SNAP<sub>c</sub>/TFIIA has a basal function in SL RNA gene transcription.

## DISCUSSION

In this study, we have purified and identified a *T. brucei* transcription factor complex consisting of TRF4 and TbSNAP50 and four newly discovered proteins. Two of these proteins were unambiguously identified as highly divergent orthologues of human SNAP43 and of the small TFIIA subunit and therefore were named TbSNAP3 and TbTFIIA-2, respectively. In contrast, the two remaining proteins have no substantial sequence homology to known transcription factors. Copurification of p46 with TbSNAP50 in the absence of TRF4 or TFIIA, though, has identified this protein as a trypanosomatid SNAP<sub>c</sub> subunit (Fig. 6B) (2; data not shown). Conversely, the assignment of p75 as TbTFIIA-1 is tentative because it is only based on several conserved residues in the N and C termini of the protein and on the fact that TFIIA in other eukaryotes consists of two subunits (neglecting proteolytic cleavage of the large subunit in higher eukaryotes). Thus far, we have been unable to disrupt the complex to verify that p75 is specifically associated with TFIIA-2. Nevertheless, our study clearly identified p75 as a subunit of the TRF4/SNAP<sub>c</sub>/TFIIA complex: the protein copurified specifically in PTP purifications of four other subunits, it remained bound to the complex at 1 M salt, it cosedimented with the other subunits, and it bound specifically to the SL RNA gene USE.

Thus far, our analysis has shown that SL RNA gene transcription depends on highly divergent orthologues of human factors which are essential for class II snRNA gene transcription. Since the basal transcription factors TFIIIB, TFIIIF, and TFIIIE are also essential for the latter (15), it is possible that trypanosomatids harbor orthologues of these factors for SL RNA gene transcription. In analyzing proteins which copurified with TRF4-P, we identified TbBRF1. BRF1 is a subunit of TFIIIB and belongs to the highly conserved TFIIIB family of proteins which extends to archaea (27). Interestingly, while

BRF1 orthologues can be found in the databases of other trypanosomatids, we were unable to identify an orthologue for TFIIB (data not shown). This is surprising because TFIIB is a key molecule in RNA polymerase II recruitment to a promoter and binds the enzyme directly (26). Similarly, TFIIF and TFIIE orthologues have not been identified in trypanosomatid genome databases. This apparent lack of basal transcription factors may be explained by the polycistronic mode which trypanosomatids use to transcribe their protein-coding genes (reviewed in references 1 and 4). It is not known how RNA polymerase II is recruited to these genes, and no class II promoter with a clearly defined transcription initiation site has been characterized for these genes thus far. Hence, it is possible that trypanosomatids have invented a mode of transcription initiation which is independent of most basal transcription factors. On the other hand, characterization of the TRF4/SNAP/TFIIA subunits in this study suggests that the SL RNA gene promoter forms a class II preinitiation complex comparable to that of other eukaryotes. It may very well be that it is the only such promoter in trypanosomatids. If this is true, the basal factors of trypanosomatids may have diverged much faster than in other organisms because the constraint on function on many different promoters did not exist. Sequence divergence may have progressed to an extent that hampers factor identification by database mining. Hence, biochemical purification will be required to identify other components of the SL RNA gene preinitiation complex. It is possible that some of these factors copurified with TFIIA-2-PTP as minor bands (Fig. 6A, lane 6). Support for this notion comes from the finding that TFIIE and TFIIF directly interact with TFIIA in the human system (17).

In contrast to these basal transcription factors, SNAP<sub>c</sub> had been characterized only in higher eukaryotes and was not found in *S. cerevisiae* and *S. pombe* (12), indicating that it originated late in eukaryotic evolution. The identification of SNAPs in trypanosomatids, however, showed that this transcription factor has a very early evolutionary origin.

Finally, the finding that SL RNA gene transcription in trypanosomatids depends on proteins with extremely divergent sequences raises the possibility that this has structural consequences which can be exploited as antiparasitic targets. Our study represents a critical step in this direction.

#### ACKNOWLEDGMENTS

This work was supported by grants of the National Institute of Health (AI059377) and the R. and C. Patterson Trust to A.G. The *T. brucei* genome sequence was sequenced at TIGR and the Sanger Centre with support from the NIH and the Wellcome Trust.

We thank Jens Brandenburg for his advice on keeping purified proteins in solution, Laurie Lomask for excellent technical assistance, and Mary Ann Gawinowicz (Protein Core Facility, Columbia University) for superb mass spectrometric analysis.

#### REFERENCES

- Campbell, D. A., S. Thomas, and N. R. Sturm. 2003. Transcription in kinetoplastid protozoa: why be normal? *Microbes Infect.* **5**:1231–1240.
- Das, A., and V. Bellofatto. 2003. RNA polymerase II-dependent transcription in trypanosomes is associated with a SNAP complex-like transcription factor. *Proc. Natl. Acad. Sci. USA* **100**:80–85.
- Gilinger, G., and V. Bellofatto. 2001. Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. *Nucleic Acids Res.* **29**:1556–1564.
- Günzl, A. 2003. Transcription, p. 47–65. *In* J. J. Marr, T. W. Nielsen, and R. W. Komuniecki (ed.), *Molecular medical parasitology*. Academic Press, London, United Kingdom.
- Günzl, A., A. Bindereif, E. Ullu, and C. Tschudi. 2000. Determinants for cap trimethylation of the U2 small nuclear RNA are not conserved between *Trypanosoma brucei* and higher eukaryotic organisms. *Nucleic Acids Res.* **28**:3702–3709.
- Günzl, A., T. Bruderer, G. Laufer, B. Schimanski, L. C. Tu, H. M. Chung, P. T. Lee, and M. G. Lee. 2003. RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. *Eukaryot. Cell* **2**:542–551.
- Günzl, A., E. Ullu, M. Dörner, S. P. Fragoso, K. F. Hoffmann, J. D. Milner, Y. Morita, E. K. Nguu, S. Vanacova, S. Wünsch, A. O. Dare, H. Kwon, and C. Tschudi. 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.
- Hartree, D., and V. Bellofatto. 1995. Essential components of the mini-exon gene promoter in the trypanosomatid *Leptomonas seymouri*. *Mol. Biochem. Parasitol.* **71**:27–39.
- Henry, R. W., C. L. Sadowski, R. Kobayashi, and N. Hernandez. 1995. A TBP-TAF complex required for transcription of human snRNA genes by RNA polymerases II and III. *Nature* **374**:653–656.
- Hernandez, N. 2001. Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J. Biol. Chem.* **276**:26733–26736.
- Hirt, R. P., J. M. Logsdon, Jr., B. Healy, M. W. Dorey, W. F. Doolittle, and T. M. Embley. 1999. Microsporidia are related to fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc. Natl. Acad. Sci. USA* **96**:580–585.
- Huang, Y., and R. J. Maraia. 2001. Comparison of the RNA polymerase III transcription machinery in *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae* and human. *Nucleic Acids Res.* **29**:2675–2690.
- Huie, J. L., P. He, and V. Bellofatto. 1997. *In vitro* transcription of the *Leptomonas seymouri* SL RNA and U2 snRNA genes using homologous cell extracts. *Mol. Biochem. Parasitol.* **90**:183–192.
- Kang, J. J., D. T. Auble, J. A. Ranish, and S. Hahn. 1995. Analysis of the yeast transcription factor TFIIA: distinct functional regions and a polymerase II-specific role in basal and activated transcription. *Mol. Cell. Biol.* **15**:1234–1243.
- Kuhlman, T. C., H. Cho, D. Reinberg, and N. Hernandez. 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. *Mol. Cell. Biol.* **19**:2130–2141.
- Kumar, A., A. Grove, G. A. Kassavetis, and E. P. Geiduschek. 1998. Transcription factor IIB: the architecture of its DNA complex, and its roles in initiation of transcription by RNA polymerase III. *Cold Spring Harbor Symp. Quant. Biol.* **63**:121–129.
- Langelier, M. F., D. Forget, A. Rojas, Y. Porlier, Z. F. Burton, and B. Coulombe. 2001. Structural and functional interactions of transcription factor (TF) IIA with TFIIE and TFIIF in transcription initiation by RNA polymerase II. *J. Biol. Chem.* **276**:38652–38657.
- Laufer, G., and A. Günzl. 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., G. Schaaf, S. Bollgönn, and A. Günzl. 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.
- Lee, M. G.-S. 1996. An RNA polymerase II promoter in the *hsp70* locus of *Trypanosoma brucei*. *Mol. Cell. Biol.* **16**:1220–1230.
- Li, C., G. A. Harding, J. Parise, K. J. McNamara-Schroeder, and W. E. Stumph. 2004. Architectural arrangement of cloned proximal sequence element-binding protein subunits on *Drosophila* U1 and U6 snRNA gene promoters. *Mol. Cell. Biol.* **24**:1897–1906.
- Liang, X. H., A. Haritan, S. Uliel, and S. Michaeli. 2003. *trans* and *cis* splicing in trypanosomatids: mechanism, factors, and regulation. *Eukaryot. Cell* **2**:830–840.
- Luo, H., G. Gilinger, D. Mukherjee, and V. Bellofatto. 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.
- Ma, B., and N. Hernandez. 2001. A map of protein-protein contacts within the small nuclear RNA-activating protein complex SNAPc. *J. Biol. Chem.* **276**:5027–5035.
- Matkin, A., A. Das, and V. Bellofatto. 2001. The *Leptomonas seymouri* spliced leader RNA promoter requires a novel transcription factor. *Int. J. Parasitol.* **31**:544–548.
- Orphanides, G., T. Lagrange, and D. Reinberg. 1996. The general transcription factors of RNA polymerase II. *Genes Dev.* **10**:2657–2683.
- Ouzounis, C., and C. Sander. 1992. TFIIB, an evolutionary link between the transcription machineries of archaeobacteria and eukaryotes. *Cell* **71**:189–190.
- Puig, O., F. Caspary, G. Rigaut, B. Rutz, E. Bouveret, E. Bragado-Nilsson, M. Wilm, and B. Seraphin. 2001. The tandem affinity purification (TAP)

- method: a general procedure of protein complex purification. *Methods* **24**:218–229.
29. Rigaut, G., A. Shevchenko, B. Rutz, M. Wilm, M. Mann, and B. Seraphin. 1999. A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17**:1030–1032.
  30. Ruan, J.-P., G. K. Arhin, E. Ullu, and C. Tschudi. 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.
  31. Schimanski, B., B. Klumpp, G. Laufer, R. J. Marhöfer, P. M. Selzer, and A. Günzl. 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.
  32. Schimanski, B., G. Laufer, L. Gontcharova, and A. Günzl. 2004. The *Trypanosoma brucei* spliced leader RNA and rRNA gene promoters have interchangeable TbSNAP50-binding elements. *Nucleic Acids Res.* **32**:700–709.
  33. Schimanski, B., L. Ma, and A. Günzl. 2004. Failure to detect binding of *Trypanosoma brucei* SNAPc to U2 and U6 snRNA gene sequences by in vitro transcription competition and pull-down assays. *Mol. Biochem. Parasitol.* **137**:293–296.
  34. Sogin, M. L., J. H. Gunderson, H. J. Elwood, R. A. Alonso, and D. A. Peattie. 1989. Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. *Science* **243**:75–77.
  35. Stearns, D. J., S. Kurosawa, P. J. Sims, N. L. Esmon, and C. T. Esmon. 1988. The interaction of a Ca<sup>2+</sup>-dependent monoclonal antibody with the protein C activation peptide region. Evidence for obligatory Ca<sup>2+</sup> binding to both antigen and antibody. *J. Biol. Chem.* **263**:826–832.
  36. Wong, M. W., R. W. Henry, B. Ma, R. Kobayashi, N. Klages, P. Matthias, M. Strubin, and N. Hernandez. 1998. The large subunit of basal transcription factor SNAP<sub>c</sub> is a Myb domain protein that interacts with Oct-1. *Mol. Cell. Biol.* **18**:368–377.
  37. Yoon, J.-B., S. Murphy, L. Bai, Z. Wang, and R. G. Roeder. 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. *Mol. Cell. Biol.* **15**:2019–2027.
  38. Yu, M. C., N. R. Sturm, R. M. Saito, T. G. Roberts, and D. A. Campbell. 1998. Single nucleotide resolution of promoter activity and protein binding for the *Leishmania tarentolae* spliced leader RNA gene. *Mol. Biochem. Parasitol.* **94**:265–281.