

U1 small nuclear RNP from *Trypanosoma brucei*: a minimal U1 snRNA with unusual protein components

Zsafia Palfi, Bernd Schimanski^{1,2}, Arthur Günzl^{1,2}, Stephan Lücke and Albrecht Bindereif*

Institut für Biochemie, Justus-Liebig-Universität Giessen, Heinrich-Buff-Ring 58, D-35392 Giessen, Germany,
¹Department of Genetics and Developmental Biology and ²Department of Molecular, Microbial and Structural Biology, University of Connecticut Health Center, Farmington, CT, USA

Received March 15, 2005; Revised and Accepted April 14, 2005

ABSTRACT

Processing of primary transcripts in trypanosomes requires *trans* splicing and polyadenylation, and at least for the poly(A) polymerase gene, also internal *cis* splicing. The trypanosome U1 snRNA, which is most likely a *cis*-splicing specific component, is unusually short and has a relatively simple secondary structure. Here, we report the identification of three specific protein components of the *Trypanosoma brucei* U1 snRNP, based on mass spectrometry and confirmed by *in vivo* epitope tagging and *in vitro* RNA binding. Both *T. brucei* U1-70K and U1C are only distantly related to known counterparts from other eukaryotes. The *T. brucei* U1-70K protein represents a minimal version of 70K, recognizing the first loop sequence of U1 snRNA with the same specificity as the mammalian protein. The trypanosome U1C-like protein interacts with 70K directly and binds the 5' terminal sequence of U1 snRNA. Surprisingly, instead of U1A we have identified a novel U1 snRNP-specific protein, TbU1-24K. U1-24K lacks a known RNA-binding motif and integrates in the U1 snRNP via interaction with U1-70K. These data result in a model of the trypanosome U1 snRNP, which deviates substantially from our classical view of the U1 particle and may reflect the special requirements for splicing of a small set of *cis*-introns in trypanosomes.

INTRODUCTION

Messenger RNA processing of primary polycistronic transcripts in trypanosomes involves both *trans*-splicing and polyadenylation (1,2). The resulting mRNAs always carry a 39 nt

non-coding spliced leader (SL) sequence at their 5' ends, which is derived from the SL RNA. The discovery of a single, *cis*-spliced intron in the poly(A) polymerase (PAP) genes from *Trypanosoma brucei* and *Trypanosoma cruzi* changed this view of gene expression in trypanosomes, suggesting there is a small subset of genes that require both *trans* and *cis* splicing (3). The interesting question about differential factor requirements leads to the U1 snRNP, which was identified in the trypanosome system only recently and probably represents a component specific to the *cis*-splicing machinery. As shown in the nematode system, U1 snRNA is not essential for *trans* splicing (4).

Among the spliceosomal snRNPs, U1 represents the best-characterized particle, both structurally and with regard to its role in spliceosome assembly (5,6). It consists of the U1 snRNA and 10 protein components highly conserved from yeast to man: the heteroheptameric Sm protein complex bound at the Sm site, and three U1-specific polypeptides: U1-70K, U1A and U1C. Both U1-70K and U1A associate with the U1 snRNA directly through their RNA recognition motifs, recognizing conserved loop positions in stem-loops I and II, respectively. On the other hand, U1C integrates via protein-protein interactions in the U1 snRNP, requiring 70K and the Sm complex (7).

In the classical role of the U1 snRNP, early 5' splice site recognition by RNA-RNA base pairing, the U1C protein appears to play a particularly important role (8,9). There is evidence that U1C alone is able to recognize the 5' splice site sequence in the absence of an intact U1 snRNP, at least in yeast (10–12). In contrast to U1C, U1A is not important for the assembly of early spliceosomal complexes, consistent with the finding that U1 snRNAs lacking binding sites for either 70K (stem-loop I) or U1A (stem-loop II) were active in *in vitro* splicing (13). Interestingly, the U1A protein also plays other roles: 5' and 3' splice site communication (14), and regulation of polyadenylation (15–17).

*To whom correspondence should be addressed. Tel: +49 641 99 35 420; Fax: +49 641 99 35 419; Email: albrecht.bindereif@chemie.bio.uni-giessen.de
Present address:

Stephan Lücke, Howard Hughes Medical Institute, Program in Gene Function and Expression, University of Massachusetts Medical School, LRB-670A, 364 Plantation Street, Worcester, MA 01605, USA

The trypanosomatid U1 snRNA is unusual in that it is much smaller than all other known homologs (~75 versus 164 nt for the mammalian and 568 for the yeast U1), suggesting that a minimal or specialized U1 snRNP operates in trypanosomes (18–20). Specifically, the trypanosomatid U1 snRNA lacks stem-loops II and III, but contains an Sm site for binding of a set of common proteins, which are present in all spliceosomal snRNPs (21). Moreover, the U1 snRNA can form an extensive base-pairing interaction with the PAP 5' splice site (19), and point mutations in the PAP 5' splice site that are predicted to disrupt base pairing, abolish splicing *in vivo* (3).

The *cis*-spliceosomal U1 snRNP of *T.brucei* has been affinity-purified in a functional form; an initial analysis of the protein components of this ~10S RNP yielded, in addition to the Sm protein set, preliminary evidence for a 70K-homologous protein (20). This provided the starting point for our current study: using the putative 70K protein from *T.brucei* (TbU1-70K) and tandem affinity purification (TAP), we established that TbU1-70K is U1 snRNP-specific. Moreover, TbU1-70K purification resulted in mass-spectrometric identification of two additional U1 snRNP components: a U1C homolog of 22 kDa (TbU1C) and a novel 24 kDa protein (TbU1-24K). For both of them, we demonstrated U1 specificity, and we investigated their interactions within the U1 snRNP. In conclusion, a model is presented for the trypanosomal U1 snRNP, which deviates substantially from our classical view of the U1 snRNP.

MATERIALS AND METHODS

T.brucei cell culture and extract preparation

Cells of wild-type and stably transfected procyclic form *T.brucei brucei* strain 427 were cultured as described previously (22). Extract was prepared in PA-150 buffer (20 mM Tris-HCl, pH 8.0, 150 mM KCl, 3 mM MgCl₂ and 0.5 mM DTT) using a Polytron PT 3100 cell homogenizer (Kinematica AG, Switzerland).

Antibodies and immunoprecipitation analysis

Polyclonal antibodies were raised in rabbits against recombinant glutathione *S*-transferase (GST) TbU1-70K protein. Anti common-protein antibodies against the *T.brucei* Sm proteins are described previously (23), as well as the U1 snRNP affinity selection (20). For immunoprecipitations, 50 µl of rabbit sera or the corresponding non-immune sera were reacted with 100 µl total cell extract (23). Coimmunoprecipitated RNAs were analyzed by primer extension or by 3' end labeling with [³²P]pCp.

Database analysis

The accession numbers of the trypanosomatid genes are annotations of GeneDB (<http://www.genedb.org/>).

Plasmid constructs

Construct pTb70K-PTP-NEO is a derivative of pBluescript SK+ (Stratagene) and contains two separate gene cassettes arranged in tandem. The first cassette comprises 745 bp of the TbU1-70K C-terminal coding region (nt 88–831), the PTP tag coding sequence (Bernd Schimanski, Tu N. Nguyen and

Arthur Günzl, manuscript submitted) and 470 bp of 3' flank from *TbRPA1*, the gene encoding the largest subunit of RNA polymerase I. The second cassette contains the neomycin phosphotransferase gene as a selectable marker (24).

Constructs where the original TAP tag (25) had been replaced by the PTP-tag are based on the same vector. Instead of the TbU1-70K coding region, they contain the TbU1C open reading frame (nt 13–582) or the TbU1-24K open reading frame (nt 112–636).

Stably transfected *T.brucei* cell lines were obtained by electroporation of 10 µg linearized plasmid, as described previously (24). Transfected cells were selected in the presence of 40 µg/ml of G418 (Geneticin, Gibco BRL). Expression of tagged proteins was analyzed by western blotting using the PAP reagent (Sigma) for detection.

PTP tandem affinity purification of the *T.brucei* U1 snRNP

For U1 snRNP TAP by means of the PTP-tagged TbU1-70K protein, a 2.5 l culture of stably transfected procyclics was grown to a density of 2×10^7 cells per ml, harvested and processed to a crude extract (26). The extract had a volume of 6 ml and contained 150 mM sucrose, 150 mM potassium chloride, 20 mM potassium L-glutamate, 3 mM MgCl₂, 20 mM Tris-HCl, pH 7.7, 1 mM DTT, 10 µg/ml leupeptin and 0.1% Tween-20. In addition, the extract was mixed with 0.5 ml of a 1 ml PA-150 buffer aliquot (20 mM Tris-HCl, pH 8.0, 150 mM KCl, 3 mM MgCl₂ and 0.5 mM DTT) containing 0.1% Tween-20, and a Complete Mini, EDTA-free protease inhibitor cocktail tablet (Roche). PTP purification was carried out as recently described in detail (Bernd Schimanski, Tu N. Nguyen, Arthur Günzl, manuscript submitted). For the first purification step, the extract was added to 200 µl settled bead volume of IgG-Sepharose 6 Fast Flow beads (Invitrogen) equilibrated with PA-150 buffer in a 0.8×4 cm Poly-Prep chromatography column (Bio-Rad) and rotated for 2 h at 4°C. Subsequently, the flow-through was collected by gravity flow and the column washed with 25 ml of PA-150 buffer. After equilibration in 15 ml of TEV protease buffer (PA-150 with 0.5 mM EDTA), the beads were resuspended in 2 ml of TEV protease buffer containing 300 U of AcTEV protease (Invitrogen) and rotated in the closed column at 4°C 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 200 µl settled bead volume of anti-protein C affinity matrix (Roche) equilibrated in PC-150 buffer. After rotation for 2 h, the flow-through was collected and the matrix washed with 60 ml of PC-150 buffer. Finally, the U1 snRNP was eluted by four consecutive steps in which the beads were resuspended in 1 ml of EGTA elution buffer (5 mM Tris-HCl, pH 7.7, 10 mM EGTA, 5 mM EDTA, 10 µg/ml leupeptin and 0.1% Tween-20) and rotated for 15 min at room temperature.

For protein analysis, the volume of the final eluate was reduced to ~800 µl by evaporation in a vacuum concentrator. Subsequently, the proteins were bound to 10 µl of the

hydrophobic StrataClean resin (Stratagene), released into SDS loading buffer at 80°C, separated on a 16% SDS/polyacrylamide gel and Coomassie-stained with the GelCode blue stain reagent (Pierce). Two bands (labeled X1 and X2 in Figure 2C) were excised from the gel and identified by LC/MS/MS mass spectrometry. The following peptide sequences were determined: (X2) TbU1C, VDDQEIVGGLPKPSR and VAGVLVASAAPPVVK; (X1) TbU1-24K, AAMEWESWKK, CVIPSSSTR and GSSVPIPAVVAGR.

TAP pull-down assays

For TAP-tag pull-down assays, extract from stably transfected cells was prepared and incubated with IgG–Sepharose beads. After washing, co-selected RNAs were released and analyzed by denaturing PAGE, followed by silver staining, or by northern blotting, using DIG-labeled probes (27). For subsequent immunoprecipitations, bound material was released by TEV protease (16 h at 4°C).

Protein–RNA or protein–protein interaction assays by GST pull-down

The open reading frames of TbU1-70K, TbU1C and TbU1-24K were amplified by PCR from genomic DNA, cloned into pGEX.2TK vector (Pharmacia) and expressed in *Escherichia coli* BL-21 (DE3) pLysS. For studying protein–RNA or protein–protein interactions, 5 µg of a GST-fusion protein or GST alone (as negative control) was immobilized on 25 µl glutathione–Sepharose 4B beads (Amersham) and incubated in 200 µl binding buffer (20 mM HEPES-KOH, pH 8.0, 150 mM KCl, 5 mM MgCl₂, 0.01% NP-40, 1 mM DTT and 1 µl RNasin), containing either *in vitro* translated ³⁵S-labeled proteins (Flexi Rabbit Reticulocyte Lysate System, Promega), or 50 ng *in vitro* transcribed ³²P-labeled U1 snRNA or mutant derivatives (see Supplementary Materials). The 17 nt control RNA was T7-transcribed from pCR2.1 TOPO vector linearized by XbaI. Total RNA (10 µg) purified from HeLa nuclear extract or *T. brucei* cells was used for GST pull-down assays in the same way, but including 10 µg of a GST protein. After a 1 h incubation at 25°C, the beads were washed and the selected proteins were analyzed on a 12.5% SDS–PAGE by fluorography (Amersham). RNAs were released, ethanol-precipitated, fractionated on a denaturing 10% polyacrylamide gel and detected by autoradiography.

DNA oligonucleotides

DNA oligonucleotides are listed in the Supplementary Materials.

RESULTS AND DISCUSSION

The *T. brucei* U1-70K protein (TbU1-70K): identification, domain structure and U1 snRNP specificity

We had previously affinity-purified the U1 snRNP from *T. brucei*, based on a biotinylated antisense 2'-*O*-methyl RNA oligonucleotide (20); this resulted in the identification of a putative *T. brucei* homolog of the 70K protein, which is a classical U1 snRNP-specific protein well characterized in the yeast and mammalian systems. The *T. brucei* 70K homolog

(Tb08.4A8.530) represents a minimal version of 70K, with only 278 amino acids (32 kDa), and—compared with the human protein (437 amino acids; 50 kDa)—missing the entire C-terminal half with mixed-charge (R/D; R/DE), Gly-rich and SR (serine-arginine rich) domains (see Figure 1A for the domain structure). To demonstrate definitively the U1 snRNP specificity of TbU1-70K, we raised polyclonal antibodies (Figure 1B) and used them in immunoprecipitation from *T. brucei* extract, followed by 3' end labeling of the immunoprecipitated RNA by [³²P]pCp (Figure 1C). The anti TbU1-70K antiserum specifically precipitated an RNA of 73 nt (compare lanes NIS and U1-70K); this RNA was also precipitated by antibodies specific for the common Sm proteins, together with the U2, SL, U4 and U5 snRNAs (lane Sm). The identity of this band as U1 snRNA was confirmed by primer-extension assays with the same samples (Figure 1D). We conclude that the trypanosome U1 snRNP contains, in addition to an Sm core, the 70K protein as a U1-specific component.

Mutational analysis of U1 snRNA-binding specificity of TbU1-70K

Next, we determined the RNA-binding specificity of the TbU1-70K protein. Although the 73 nt trypanosome U1 snRNA appears to be a minimal version of the U1 snRNA (Figure 1F for the *T. brucei* U1 snRNA), the loop sequence (nt 19–28) is highly conserved; for example, it deviates from the corresponding human sequence only in the two terminal positions, leaving eight continuous nucleotides identical (positions 20–27). Previous work in the mammalian system had established that this first loop is critical for U1-70K recognition of the U1 snRNA. Therefore, we studied by mutational analysis whether this 70K–U1 snRNA interaction is conserved in the trypanosome system (for the mutant derivatives, see Figure 1F). Besides the full-length U1 snRNA (Tb-U1 WT), we analyzed a truncated RNA, in which the 5' stem-loop was precisely deleted (Tb-U1 Δstem-loop), and a series of short stem-loop RNA derivatives. Those comprise the wild-type sequence (Tb-U1 stem-loop WT), a stem mutation (nt 15–18 and 24–27) with similar stability as the wild-type stem, an inverted loop sequence (nt 19–28), a chimeric RNA consisting of the human loop and the trypanosome stem sequence (human loop), and finally, a derivative thereof carrying a single point mutation in the human loop, predicted to disrupt the recognition of the human loop by the human 70K protein (U21A; mutant human loop).

These *in vitro* transcribed ³²P-labeled RNAs were incubated with recombinant GST TbU1-70K protein, and coprecipitated RNAs were analyzed by gel electrophoresis (Figure 1E, compare lanes I and P, representing 10% of the input and the total pellet, respectively). We conclude that the wild-type U1 snRNA was recognized efficiently by TbU1-70K (lanes 1/2), in contrast to the derivative without the stem-loop, which was not detectably precipitated (lanes 3/4); both the short stem-loop and its mutant stem derivative interacted with TbU1-70K at an efficiency similar to the full-length U1 snRNA (lanes 5/6 and 9/10), whereas no interaction was detected, when the loop was substituted (lanes 7/8). This clearly identified the loop of the *T. brucei* U1 snRNA as the critical determinant of U1-70K interaction. Finally, TbU1-70K recognized very well the short stem-loop RNA

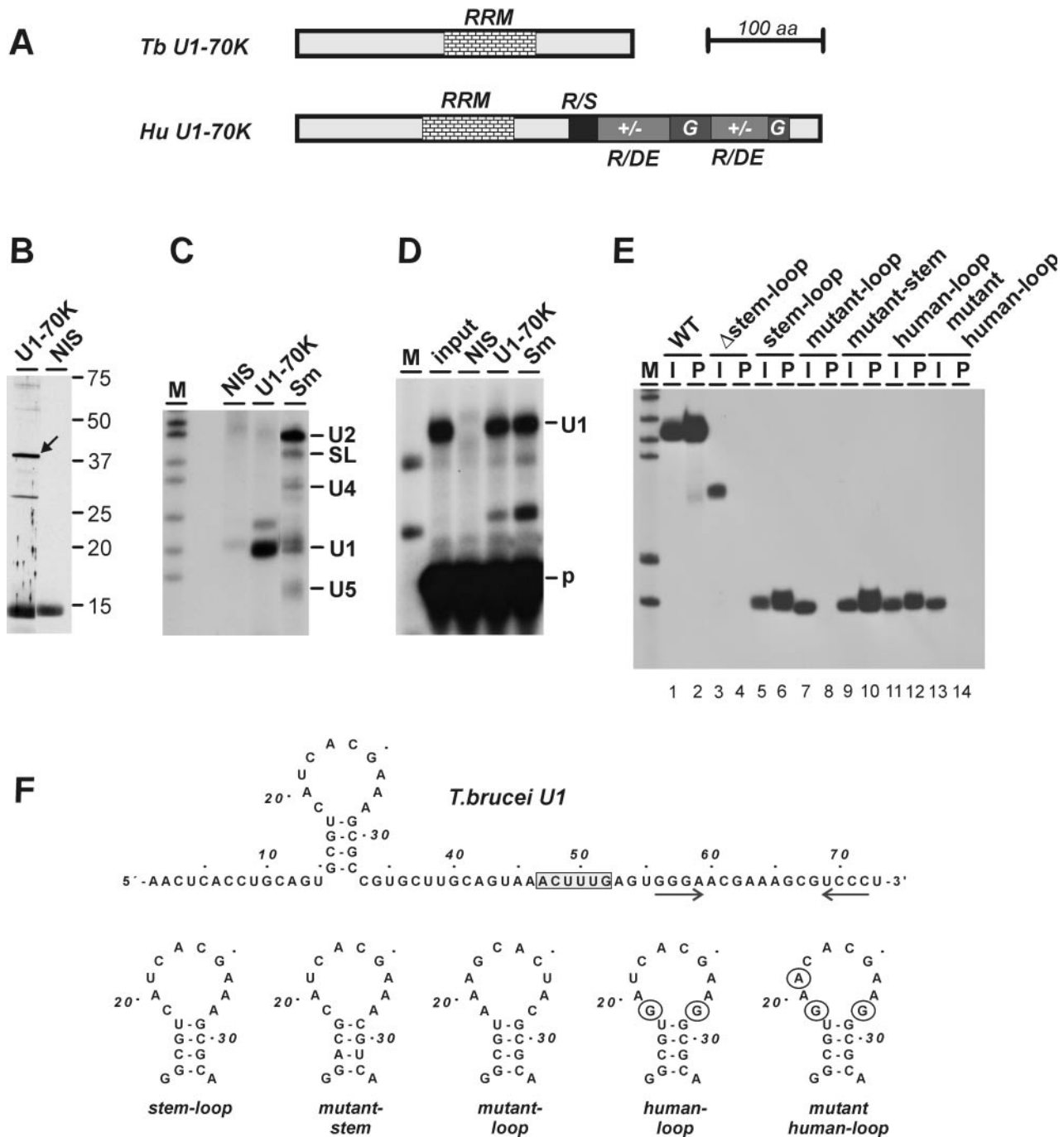


Figure 1. TbU1-70K is a U1 snRNP-specific protein and binds *in vitro* specifically to the 5' loop sequence of U1 snRNA. (A) Comparison of the domain structures of *T. brucei* (Tb08.4A8.530) and the human U1-70K (A25707) proteins. (B) Western blot analysis of *T. brucei* U1 snRNP proteins. U1 snRNPs were affinity-purified from *T. brucei* extract by a 2'-*O*-methyl RNA antisense oligonucleotide, protein was prepared and analyzed by SDS-PAGE and western blotting, using polyclonal rabbit antibodies against TbU1-70K (U1-70K) or non-immune serum (NIS). The arrow points to the immunostained TbU1-70K band of apparent molecular weight 40 kDa. Protein markers are on the right (in kDa). (C) U1 snRNA is specifically coprecipitated from *T. brucei* extract by anti-Tb U1-70 antibodies. Immunoprecipitations were carried out from *T. brucei* extract, using NIS, or with antibodies against the TbU1-70K protein (U1-70K) or against the trypanosome Sm proteins (Sm). RNA was purified from the immunoprecipitates and analyzed by 3' end labeling with [³²P]pCp. The positions of the SL RNA and snRNAs are marked on the right. M, ³²P-labeled pBR322/HpaII markers. (D) RNA from the same immunoprecipitates was also analyzed by primer extension with a U1-specific oligonucleotide. In addition, RNA from a 10% aliquot of the input was included; the positions of the primer (p) and the U1-specific primer-extension product (U1) are marked on the right. M, ³²P-labeled pBR322/HpaII markers. (E) ³²P-labeled *T. brucei* U1 snRNA and mutant derivatives [as indicated above the lanes; see (F)] were *in vitro* transcribed and incubated with GST-TbU1-70K, followed by GST pull-down. For each reaction, 10% of the input (I) and the total precipitated material (P) were analyzed. M, ³²P-labeled pBR322/HpaII markers. (F) Sequences and proposed secondary structures of the *T. brucei* U1 snRNA and its mutant derivatives. The boxed sequence in the *T. brucei* U1 snRNA indicates the Sm site; the two arrows indicate a potential second stem-loop. Below, the sequences of the stem-loop derivatives are given; the circled nucleotides mark the two positions in the human loop that differ from the *T. brucei* sequence, and the single-nucleotide mutation (A21) in the mutant human loop.

with the human loop sequence (lanes 11/12), but not at all the mutant version with a single point mutation in the loop (lanes 13/14).

The RNA-binding specificity of TbU1-70K was further subjected to a very stringent test, using GST TbU1-70K and total RNA from *T.brucei* or, in a heterologous assay, using total RNA from HeLa nuclear extract (Supplementary Figure S1). As a result, TbU1-70K was able to specifically recognize intact U1 snRNA from both total trypanosome and HeLa RNA, demonstrating the high specificity of the recombinant TbU1-70K protein.

In sum, these binding experiments demonstrate that the loop sequence of the *T.brucei* U1 snRNA is essential and sufficient for recognition by the trypanosome 70K protein. Remarkably, the detailed snRNA sequence requirements are highly conserved between the trypanosome and human systems, although the sequence and secondary structure drastically differ from each other.

TAP-tag affinity purification through TbU1-70K: mass-spectrometric identification of two new U1 snRNP protein components

After the identity of the TbU1-70K as a U1-specific protein was established, we tagged this protein at its C-terminus in a procyclic cell line with the PTP tag, which is a modified TAP tag consisting of two protein A domains, a TEV protease cleavage site and the protein C epitope (Bernd Schimanski, Tu N. Nguyen and Arthur Günzl, manuscript submitted). The PTP-tagged protein (TbU1-70K/PTP) was purified from crude extracts by IgG affinity chromatography, elution from the IgG column by TEV protease cleavage, followed by anti-protein C epitope immunoaffinity chromatography. Since the selection of TbU1-70K/PTP and its

complexes should result in the purification of the entire U1 snRNP, we followed the two stages of the affinity purification first by silver staining (Figure 2A), as well as by northern blotting, using a mixed snRNA probe (Figure 2B). Silver staining of RNA prepared from aliquots of the input, TEV-eluted and protein C-eluted material (lanes 1–3) clearly showed that an RNA of the expected size (73 nt) was selected and enriched by the second step (protein C elution). In fact, it was the only strong band detected. Since the input U1 snRNA was obscured by large quantities of tRNA, the purification was also followed by northern blotting, demonstrating both the highly efficient enrichment of U1 snRNA and the specificity of its selection through both steps (Figure 2B, compare lanes 1, 2 and 3).

Next, we determined the protein composition of the U1 snRNPs selected through two affinity steps, again following the enrichment from the extract to the TEV- and protein C-eluted material (Figure 2C, lanes 1–3). Protein analysis by Coomassie staining demonstrated that only after the second step a number of protein bands became predominant, the strongest of them representing by size the TEV-released TbU1-70K protein; in addition, a cluster of bands in the 10–15 kDa range most likely represents the Sm set of proteins. There were two further bands (labeled X1 and X2) of apparent molecular masses 24 and 22 kDa, respectively, that were strongly enriched after the second, protein-C elution step and that were subjected to mass-spectrometric analysis. Based on several peptides, each of these bands could be assigned to a hypothetical *T.brucei* protein (see Materials and Methods). Finally, there are three bands above U1-70K. The two lower ones in the 50 kDa range appear to be nonspecific, since they are not enriched during the second step of purification (compare lanes 2 and 3). The top band on the other hand is enriched and most likely represents the 65 kDa protein, which

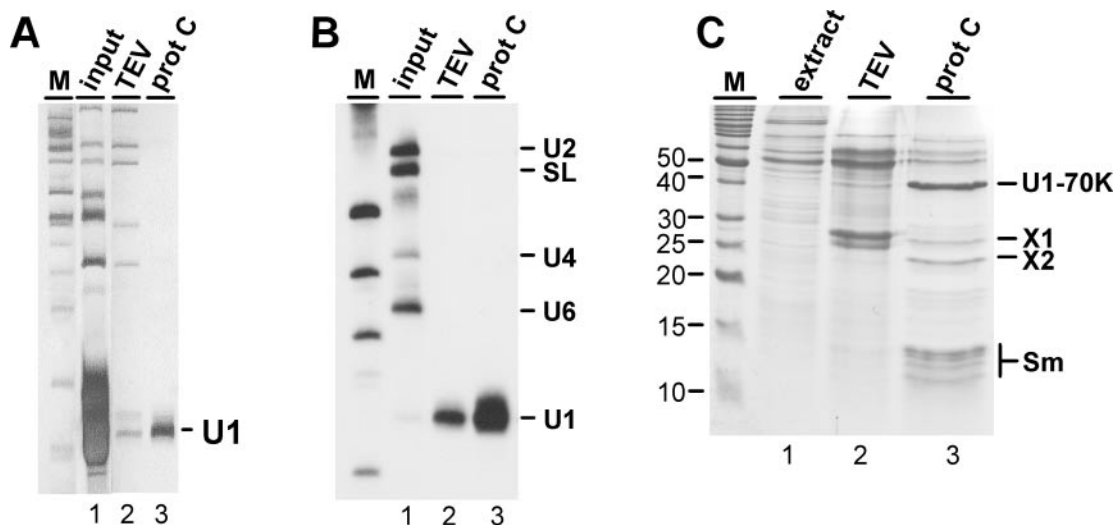


Figure 2. Identification of two new U1 snRNP-specific protein components. (A) Extract was prepared from a *T.brucei* cell line, which stably expresses PTP-tagged TbU1-70K protein, and used to affinity-purify PTP-tagged complexes. The three stages of purification were analyzed for copurifying RNAs by silver staining: 1% of the input (lane 1), 5% of the TEV-eluted material (lane 2) and 50% of the protein C-eluted material (lane 3). The position of the U1 snRNA is indicated on the right. Note that in the input the U1 snRNA band is completely obscured by large quantities of tRNA. *M*, pBR322/HpaII markers. (B) The same fractions as described in (A) were analyzed for copurifying RNAs by northern blotting, using a mixed snRNA probe (snRNA positions indicated on the right). *M*, DIG marker V (Roche). (C) Protein analysis of the same fractions as described in (A), using SDS-PAGE and Coomassie staining. On the right, positions are marked of the TbU1-70K protein, the cluster of Sm proteins and the two new polypeptides (X1 and X2), which were identified by mass spectrometry. *M*, protein marker (in kDa).

we previously identified as *T.brucei* poly(A) binding protein I (PABP I) (20).

***T.brucei* U1C (TbU1C): a second U1 snRNP-specific protein, which binds the 5' terminal sequence of U1 snRNA**

Sequence analysis and alignments revealed that one of the two hypothetical proteins, derived from the X2 protein band, a basic protein (pI 9.8) with a molecular mass of 21.7 kDa, represents a putative *T.brucei* homolog of U1C, which has been characterized well in the human and yeast systems (Tb10.70.5640). In addition, homologs exist as hypothetical proteins in *T.cruzi* and *Leishmania major* (Figure 3A).

A characteristic feature of U1C sequences from other known species is an N-terminal C₂H₂-type Zinc finger motif within the highly conserved first 40 amino acids, whereas the remaining C-terminal sequences are very diverged [e.g. for a sequence alignment, see (9)]. All three trypanosomatid U1C sequences possess a conserved C₂H₂ Zinc finger, but they carry an additional, highly basic N-terminal extension before the C₂H₂ Zinc finger: 44, 46 and 73 additional amino acids in *T.brucei*, *T.cruzi* and *L.major*, respectively. Because except for the Zn finger motif the sequences of the putative trypanosome U1C proteins diverge very much from other known U1C homologs, we cannot rule out that they are non-related proteins that have acquired a U1C-like Zn finger.

Since this putative U1C protein from *T.brucei* was identified by mass spectrometry, we had to demonstrate its U1 snRNP specificity *in vivo*. A C-terminal TAP-tag sequence was added to the TbU1C open reading frame, the construct was integrated into the *T.brucei* genome, and lysate was prepared from this cell line stably expressing TAP-tagged TbU1C. TbU1C-containing complexes were affinity-purified on IgG–Sepharose, followed by release through TEV protease cleavage. As the northern blot analysis shows, this single affinity step resulted in a fraction highly enriched in U1, and free of detectable levels of the other spliceosomal snRNAs (Figure 3B, compare lanes 1 and 2). Released U1 snRNP was further analyzed by immunoprecipitations, using non-immune serum (NIS) as a control, anti-70K and anti-Sm antibodies (lanes 3–5). These assays were positive for both 70K and Sm proteins, and no background was detectable in the non-immune control. Taken together, this establishes TbU1C as a second U1 snRNP-specific protein component; in addition, we have demonstrated that the affinity-purified *T.brucei* U1 snRNP contains, within the same RNP complex, TbU1C, TbU1-70K and the Sm core proteins.

Based on the presence of the Zn finger motif, we tested TbU1C for direct binding to the U1 snRNA (Figure 3C). ³²P-labeled full-length U1 snRNA and various derivatives were transcribed *in vitro* and incubated with the GST version of TbU1C. Bound RNA was recovered by GST pull-down and analyzed by denaturing gel electrophoresis. Full-length U1 snRNA bound TbU1C (with an efficiency of ~5%; lanes 1 and 2), whereas U1 snRNAs lacking either the 5' terminal stem-loop (U1 Δstem-loop), the first 14 nt [U1 Δ5'(1–14)], or the first 30 nt [U1 Δ5'(1–30)] bound undetectably or at very low efficiency (lanes 3–8). Note that the negative result for binding of the U1 Δstem-loop RNA is likely owing to a folding problem of this mutant derivative,

since after heat denaturation some binding could be recovered (data not shown). The stem-loop of U1 snRNA by itself (which is sufficient for U1-70K interaction; see Figure 2) was also negative in TbU1C binding (U1 5' stem-loop; lanes 9–10); in contrast, the first 14 nt of U1 were sufficient for TbU1C binding [U1 5'(1–14); lanes 11–12], but not a non-specific control RNA of similar length (control; lanes 13–14).

In sum, these results provide evidence for a direct RNA interaction of TbU1C, with specificity for the 5' terminal sequence of U1 snRNA (nt 1–14), the same region, where the U1 snRNP interacts with the 5' splice site (see below).

***T.brucei* U1-24K (TbU1-24K): a third, novel U1 snRNP-specific protein**

The other putative U1 snRNP protein (see protein band X1 in Figure 2C) was identified in the trypanosome database also as a hypothetical protein, based on three peptide sequences (Tb03.27F10.160). Unexpectedly and in contrast to TbU1C, it shows no significant homology or relationship to any other known snRNP protein and contains neither obvious protein motifs nor domain organization. In particular, it carries no identifiable RNA-recognition motifs and, therefore, cannot correspond to the missing U1A protein. In sum, this acidic *T.brucei* protein of 24.1 kDa (pI 4.5) represents a novel, putative U1 snRNP component. Among the trypanosomatids, two *T.cruzi* genes coding for almost identical hypothetical proteins are clearly homologous to the *T.brucei* U1-24K protein, indicating that it is conserved at least in trypanosomatid species and demonstrating two highly conserved regions in this novel protein, i.e. in TbU1-24K the N-terminal 37 amino acids and a region near the C-terminus with amino acids 120–200 (Figure 4A).

As performed for TbU1-70K and TbU1C, we next studied the U1 snRNP specificity of the novel TbU1-24K protein *in vivo*, using a stable *T.brucei* cell line, which expresses TbU1-24K protein with a C-terminal TAP tag. Cell lysate was prepared, and TbU1-24K protein-containing complexes were affinity-purified on IgG–Sepharose. As the northern blot analysis of co-selected snRNAs shows, this single affinity step resulted in a fraction highly enriched in U1, and almost free of detectable levels of the other spliceosomal snRNAs (Figure 4B, compare lanes 1 and 2). In another experiment shown in Figure 4C, bound material was released by TEV protease cleavage (compare lanes 1 and 2), and the resulting released fraction was further analyzed by immunoprecipitations, using NIS as a control, anti-70K and anti-Sm antibodies. These tests were positive for both 70K and Sm proteins (lanes 4 and 5), and no background was detectable in the non-immune control (lane 3). Taken together, this clearly establishes TbU1-24K as a third, novel U1 snRNP-specific protein component; in addition, we have demonstrated that the affinity-purified U1 snRNP contains, within the same RNP complex, U1-24K, 70K and the Sm core proteins.

Finally, TbU1-24K was tested *in vitro* for direct binding to U1 snRNA, using GST pull-down assays (Figure 4D). No significant coprecipitation was observed with TbU1-24K (compare lanes 1 and 3). In contrast, both GST TbU1-70K and TbU1C were able to associate individually with U1 snRNA (lanes 4 and 5), and there was no detectable background with GST protein (lane 2).

A

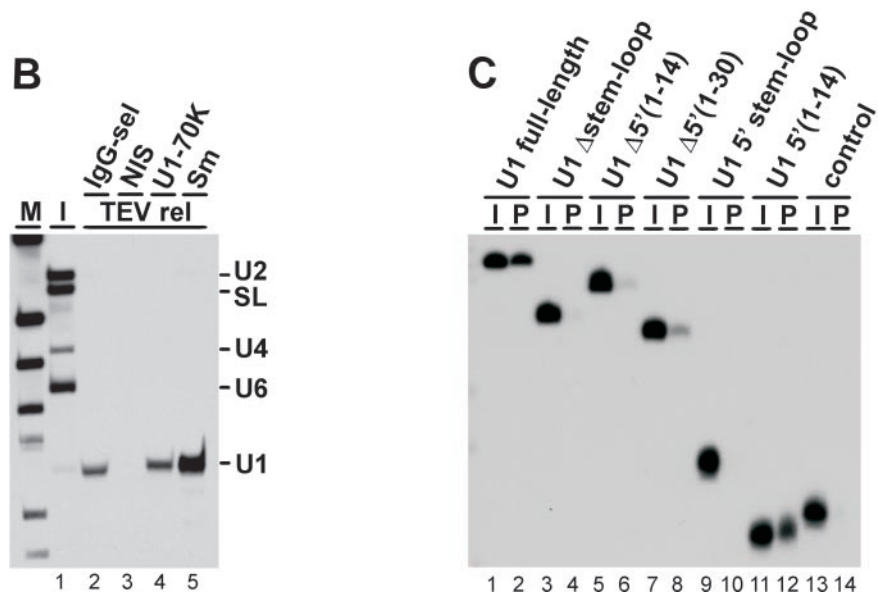


Figure 3. *T. brucei* U1C (TbU1C): a U1 snRNP-specific component binding specifically to the 5' terminal sequence of U1 snRNA. (A) ClustalW alignment of the protein sequences for the newly identified U1C homologs from *T. brucei*, *T. cruzi* and *L. major*, in comparison with the human U1C sequence. The conserved C₂H₂-type Zn finger within the boxed sequence is highlighted by large-size letters; asterisks indicate absolutely conserved amino acid positions. Accession numbers (GeneDB): *T. brucei* (Tb10.70.5640), *T. cruzi* (Tc00.1047053511367.354) and *L. major* (LmjF21.0320); human U1C (P09234). (B) Extract was prepared from a *T. brucei* cell line, which stably expresses TAP-tagged TbU1C protein, and used to affinity-purify TAP-tagged complexes. Purification was followed by analyzing copurifying RNAs by northern blotting, using a mixed snRNA probe (snRNA positions indicated on the right). M, DIG marker V (Roche). Lane 1, 1% of input; lane 2, 10% of IgG-selected and TEV-released material. Affinity-purified complexes were then immunoprecipitated with NIS (lane 3), anti TbU1-70K (lane 4) or anti-Sm antibodies (lane 5), using 30% for each immunoprecipitation. (C) TbU1C protein binds specifically to the 5' terminal sequence of U1 snRNA. GST TbU1C protein was incubated with ³²P-labeled full-length U1 snRNA (lanes 1 and 2) and various U1 snRNA derivatives: U1 Δstem-loop (lanes 3 and 4), U1 Δ5'(1-14) (lanes 5, 6), U1 Δ5'(1-30) (lanes 7 and 8), U1 5' stem-loop (lanes 9 and 10), U1 5'(1-14) (lanes 11 and 12) or a 17mer control RNA (lanes 13 and 14). In each case, 10% of the input (I) and the total GST pull-down material (P) were analyzed.

Protein-protein interactions in the *T. brucei* U1 snRNP U1-70K interacts both with U1C and U1-24K

After we had demonstrated two specific RNA-protein interactions in the *T. brucei* U1 snRNP, we next searched for protein-protein contacts, using *in vitro* GST pull-down assays

(Figure 5). Figure 5A shows the input of the three GST-fusion proteins and GST itself. Either of the three specific U1 snRNP proteins (TbU1-70K, TbU1C and TbU1-24K) was *in vitro* translated in the presence of ³⁵S-labeled methionine and incubated with the GST form of either of the two other proteins,

A

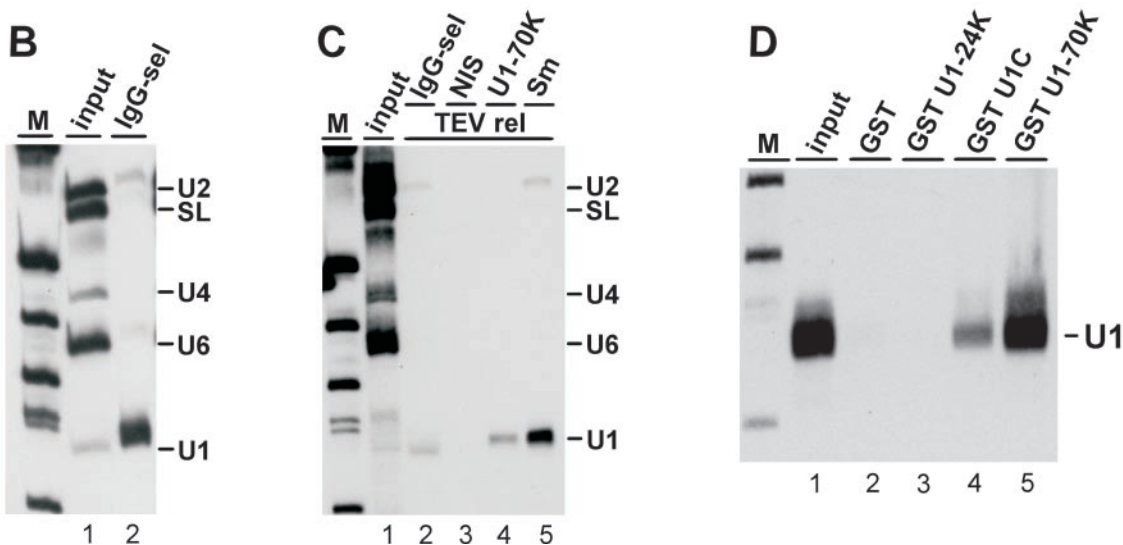
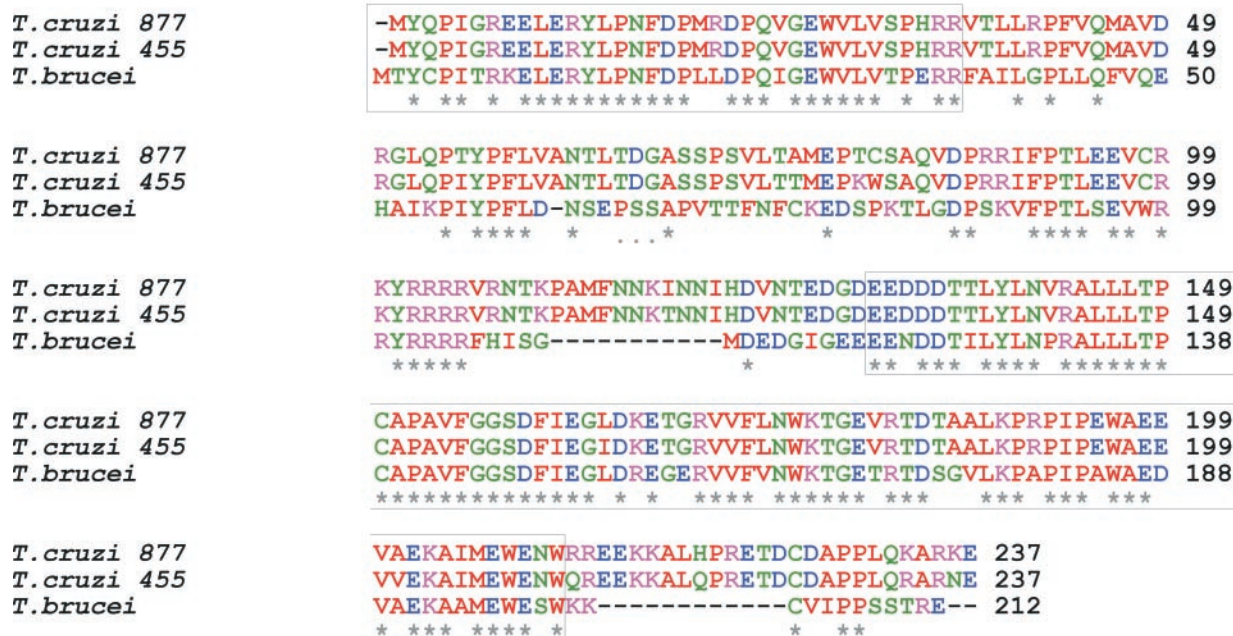


Figure 4. Identification and U1 snRNP association of a novel, U1-specific protein component: TbU1-24K. (A) ClustalW alignment of the novel protein component of the *T. brucei* U1 snRNP, TbU1-24K, with two putative homologs from *T. cruzi*. Two conserved regions are indicated by the boxed regions; the positions with asterisks are identical. Accession numbers (Gene DB): *T. cruzi* 877 (Tc00.1047053503877.10), *T. cruzi* 455 (Tc00.1047053509455.110) and *T. brucei* (Tb03.27F10.160). (B) TbU1-24K is a U1 snRNP-specific component. Extract was prepared from a *T. brucei* cell line, which stably expresses TAP-tagged TbU1-24K protein, and used to affinity-purify TAP-tagged complexes. Copurifying RNAs were analyzed by northern blotting, using a mixed probe (snRNA positions indicated on the right). Lane 1, 1% of input; lane 2, all of IgG-selected and TEV-released material; M, DIG marker V (Roche). (C) TbU1-24K coexists with TbU1-70K and Sm proteins in the same RNP complex. TAP-tag affinity purification of TbU1-24K complexes and RNA analysis was carried out as described in (A) (lane 1, 1% input; lane 2, 10% of IgG-selected and TEV-released material). Affinity-purified complexes were then immunoprecipitated with NIS (lane 3), anti TbU1-70K (lane 4) or anti-Sm antibodies (lane 5), using 30% for each immunoprecipitation. The snRNA positions are marked on the right. The slightly slower mobility of U1 snRNA in the immunoprecipitates (lanes 4 and 5) is most likely caused by comigrating tRNA released from the blocked protein A-Sepharose beads. M, DIG marker V (Roche). (D) *In vitro* U1 snRNA binding of U1 snRNP proteins. GST derivatives of the three U1-specific proteins TbU1-24K, TbU1C and TbU1-70K (lanes 3–5) were incubated with *in vitro* transcribed *T. brucei* U1 snRNA, followed by GST pull-down and analysis of coprecipitated RNA by northern hybridization with a U1-specific probe. A total of 10% of the input material was analyzed (lane 1), and a control precipitation was carried out with GST protein (lane 2). M, DIG marker V (Roche).

followed by GST pull-down. As a control, each ³⁵S-labeled protein was also incubated with GST protein. The data on Figure 5B clearly indicate that TbU1-70K interacts with TbU1C, an interaction that could be demonstrated in two

ways: GST U1-70K/³⁵S-labeled U1C (lane 4) and GST U1C/³⁵S-labeled U1-70K (lane 11). TbU1-70K also interacted with the novel protein, TbU1-24K; however, this contact could be demonstrated only in the combination GST

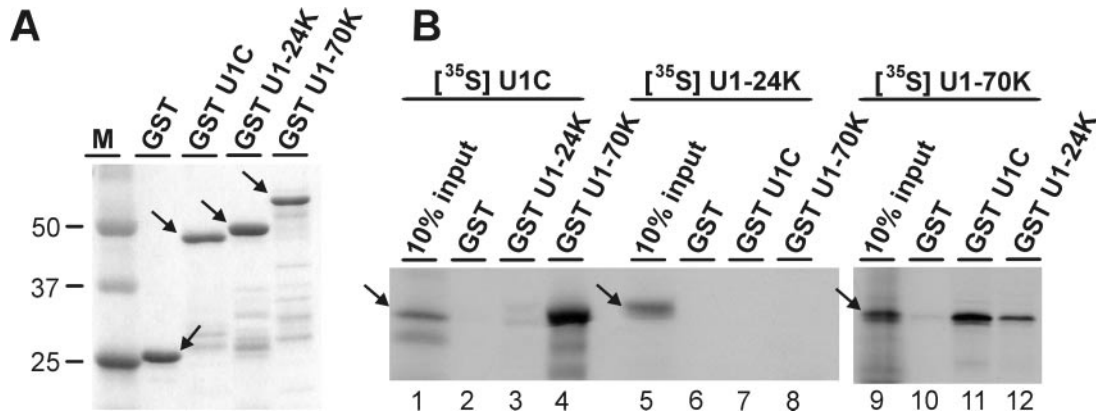


Figure 5. Protein-protein interactions in the trypanosome U1 snRNP: TbU1-70K interacts with both TbU1C and TbU1-24K. (A) GST-fusion proteins of TbU1C, TbU1-24K and TbU1-70K, as well as GST alone as a control were immobilized on glutathione-Sepharose. Corresponding aliquots of immobilized proteins were analyzed for their protein content by SDS-PAGE and Coomassie staining. The arrows point to the proteins listed above the lanes. *M*, protein marker (in kDa). (B) Immobilized GST proteins (as indicated above the lanes) were incubated with ^{35}S -labeled TbU1C (lanes 1–4), TbU1-24K (lanes 5–8) or TbU1-70K (lanes 9–12). After washing, bound proteins were recovered and analyzed by SDS-PAGE and fluorography. The arrows point to the respective ^{35}S -labeled proteins.

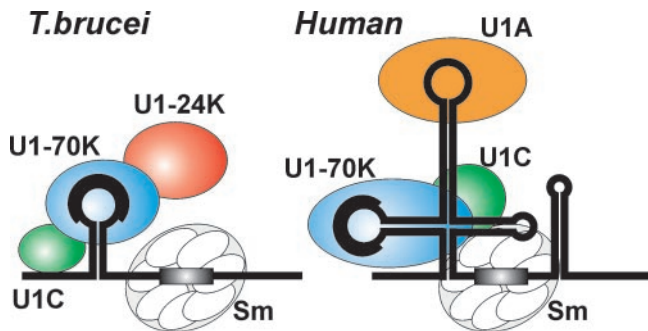


Figure 6. Model of the trypanosome U1 snRNP, in comparison with the human counterpart. Schematic model of RNA-protein and protein-protein interactions known for the human U1 snRNP (right) and, as analyzed in this study, for the trypanosome U1 snRNP (left). Lines indicate the RNA secondary structures; the boxed region indicates the Sm site; heavy lines indicate the conserved loop sequence. Protein homologies are represented by the same colors (U1-70K; U1C; Sm core). It is not known whether the Sm core proteins in the trypanosome U1 snRNP interact with U1-specific protein components. Note that in contrast to the human system, the trypanosome U1C contacts the 5' terminal sequence of U1 snRNA (see Figure 3C).

U1-24K/ ^{35}S -labeled U1-70K (lane 12), not the other way (lane 8), which is most likely owing to steric hindrance by the GST tag. Finally, for the combination U1C/U1-24K, these assays resulted in either orientation in very low or undetectable levels of interaction (lanes 3 and 7). In controls, none of the three ^{35}S -labeled proteins interacted significantly with GST protein (lanes 2, 6 and 10). Taken together, these data can be summarized in a model for the trypanosome U1 snRNP shown in Figure 6 and discussed below.

Model of the *T.brucei* U1 snRNP

Although the trypanosome U1 snRNA represents the shortest known U1 homolog, it contains a 5' terminal stem-loop with highly conserved loop positions as well as an Sm site (Figure 6). Therefore, it was not surprising to also find the protein components binding these elements: a U1-70K homolog, which recognizes the loop nucleotides (see Figure 1E),

and the Sm core polypeptides (20). The *T.brucei* U1-70K homolog represents a minimal version, yet we can be certain about its identity, based on its RNA-binding specificity. Since the RS domain in the human 70K protein engages in protein-protein interactions with ASF/SF2 (28), the absence of an RS domain in the trypanosomatids indicates differences in the early stages of *cis*-spliceosome assembly. Alternatively, in trypanosomes an yet undiscovered separate polypeptide may have taken over this function and may cooperate with TbU1-70K. An earlier report (29) had proposed an SR-domain containing protein, called TSR1 IP, as the *T.brucei* 70K homolog; based on three-hybrid assays in yeast, the protein had been suggested to bind specifically to SL RNA. No evidence for this RNA-binding property, however, had been obtained in the trypanosome system, and using epitope tagging in *T.brucei*, we were unable to confirm the SL RNA-specificity of TSR1 IP (Z. Palfi and A. Bindereif, unpublished data).

More unusual and surprising are the other two specific protein components of the trypanosome U1 snRNP, TbU1C and a novel protein, TbU1-24K. Both integrate into the U1 snRNP via protein-protein interaction with TbU1-70K. In addition, we have identified and mapped a new RNA-protein interaction between the TbU1C protein and the 5' terminal, single-stranded region of U1 snRNA, which is in contrast to the mammalian U1 snRNP, where human U1C has failed to bind U1 snRNA (30). Our result raises the interesting mechanistic question whether 5' splice site interaction of the U1 snRNP might be accompanied by a conformational switch, with the 5' terminal U1 sequence either in an inactive, U1C-bound state (RNA-protein interaction) or in the active state, available for RNA-RNA base pairing with the 5' splice site. Further mutational studies should address whether this intriguing RNA-binding specificity of the trypanosome U1C is related to certain differences within the Zn finger region, or to the highly basic N-terminal extension.

Strikingly, one of the classical U1 snRNP core components, U1A, appears to be missing in the trypanosome U1 snRNP. Although the absence of a U1A homolog in the trypanosome U1 snRNP is in fact very unusual and unprecedented, it is plausible, considering that the *T.brucei* U1 snRNA lacks a

canonical U1A binding site. Therefore, the classical direct RNA contact of U1A with the second stem-loop of U1 snRNA may be replaced in the trypanosome U1 snRNP by the novel TbU1-24K protein interacting with U1 snRNA-bound TbU1-70K.

In sum, compared with the mammalian and yeast complexes, our analysis of the trypanosome U1 snRNP revealed an RNP with at least 10 polypeptides, the same number as the mammalian U1 snRNP contains. On the other hand, in the yeast U1 snRNP at least 16 polypeptides are assembled on the U1 snRNA, which with 568 nt is much larger than the mammalian counterpart (31,32). The relatively simple composition of the trypanosome U1 snRNP may reflect not only species-specific differences, but also the restricted requirements of splicing a small set of *cis*-introns in these organisms.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Pingping Wang for a construct, Mary Ann Gawinowicz (Protein Core Facility, Columbia University) for good mass spectrometric analysis, and Andrey Damianov for critically reading the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (Sonderforschungsbereich 535, to A.B.), by the German-Israeli Foundation (GIF, to A.B.), and a grant of the National Institute of Health (AI059377, to A.G.). Funding to pay the Open Access publication charges for this article was provided by the German-Israeli Foundation.

Conflict of interest statement. None declared.

REFERENCES

- Ullu, E., Tschudi, C. and Günzl, A. (1996) *Trans*-splicing in trypanosomatid protozoa. In Smith, D.F. and Parsons, M. (eds), *Molecular Biology of Parasitic Protozoa*. IRL Press, Oxford, pp. 115–133.
- Liang, X.H., Haritan, A., Uliel, S. and Michaeli, S. (2003) *Trans* and *cis* splicing in trypanosomatids: mechanism, factors, and regulation. *Eukaryot. Cell*, **2**, 830–840.
- Mair, G., Shi, H., Li, H., Djikeng, A., Aviles, H.O., Bishop, J.R., Falcone, F.H., Gavrilescu, C., Montgomery, J.L., Santori, M.I. *et al.* (2000) A new twist in trypanosome RNA metabolism: *cis*-splicing of pre-mRNA. *RNA*, **6**, 163–169.
- Hannon, G.J., Maroney, P.A. and Nilsen, T.W. (1991) U small nuclear ribonucleoprotein requirements for nematode *cis*- and *trans*-splicing *in vitro*. *J. Biol. Chem.*, **266**, 22792–22795.
- Kambach, C., Walke, S. and Nagai, K. (1999) Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. *Curr. Opin. Struct. Biol.*, **9**, 222–230.
- Stark, H., Dube, P., Lührmann, R. and Kastner, B. (2001) Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature*, **409**, 539–542.
- Nelissen, R.L., Will, C.L., van Venrooij, W.J. and Lührmann, R. (1994) The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. *EMBO J.*, **13**, 4113–4125.
- Heinrichs, V., Bach, M., Winkelmann, G. and Lührmann, R. (1990) U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site. *Science*, **247**, 69–72.
- Muto, Y., Krummel, D.P., Oubridge, C., Hernandez, H., Robinson, C.V., Neuhaus, D. and Nagai, K. (2004) The structure and biochemical properties of the human spliceosomal protein U1C. *J. Mol. Biol.*, **341**, 185–198.
- Du, H. and Rosbash, M. (2002) The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature*, **419**, 86–90.
- Lund, M. and Kjems, J. (2002) Defining a 5' splice site by functional selection in the presence and absence of U1 snRNA 5' end. *RNA*, **8**, 166–179.
- Rossi, F., Forne, T., Antoine, E., Tazi, J., Brunel, C. and Cathala, G. (1996) Involvement of U1 small nuclear ribonucleoproteins (snRNP) in 5' splice site-U1 snRNP interaction. *J. Biol. Chem.*, **271**, 23985–23991.
- Will, C.L., Rumpel, S., Klein Gunnewiek, J., van Venrooij, W.J. and Lührmann, R. (1996) *In vitro* reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein enhances the formation of early (E) spliceosomal complexes. *Nucleic Acids Res.*, **24**, 4614–4623.
- Tarn, W.Y. and Steitz, J.A. (1995) Modulation of 5' splice site choice in pre-messenger RNA by two distinct steps. *Proc. Natl Acad. Sci. USA*, **92**, 2504–2508.
- Gunderson, S.I., Beyer, K., Martin, G., Keller, W., Boelens, W.C. and Mattaj, J.W. (1994) The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. *Cell*, **76**, 531–541.
- O'Connor, J.P., Alwine, J.C. and Lutz, C.S. (1997) Identification of a novel, non-snRNP protein complex containing U1A protein. *RNA*, **3**, 1444–1455.
- Phillips, C., Pachikara, N. and Gunderson, S.I. (2004) U1A inhibits cleavage at the immunoglobulin M heavy-chain secretory poly(A) site by binding between the two downstream GU-rich regions. *Mol. Cell Biol.*, **24**, 6162–6171.
- Schnare, M.N. and Gray, M.W. (1999) A candidate U1 small nuclear RNA for trypanosomatid protozoa. *J. Biol. Chem.*, **274**, 23691–23694.
- Djikeng, A., Ferreira, L., D'Angelo, M., Dolezal, P., Lamb, T., Murta, S., Triggs, V., Ulbert, S., Villarino, A., Renzi, S. *et al.* (2001) Characterization of a candidate *Trypanosoma brucei* U1 small nuclear RNA gene. *Mol. Biochem. Parasitol.*, **113**, 109–115.
- Palfi, Z., Lane, W.S. and Bindereif, A. (2002) Biochemical and functional characterization of the *cis*-spliceosomal U1 small nuclear RNP from *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **121**, 233–243.
- Palfi, Z., Lücke, S., Lahm, H.-W., Lane, W.S., Kruft, V., Bragado-Nilsson, E., Séraphin, B. and Bindereif, A. (2000) The spliceosomal snRNP core complex of *Trypanosoma brucei*: cloning and functional analysis reveals seven Sm protein constituents. *Proc. Natl Acad. Sci. USA*, **97**, 8967–8972.
- Cross, M., Günzl, A., Palfi, Z. and Bindereif, A. (1991) Analysis of small nuclear ribonucleoproteins (RNPs) in *Trypanosoma brucei*: structural organization and protein components of the spliced leader RNP. *Mol. Cell Biol.*, **11**, 5516–5526.
- Palfi, Z. and Bindereif, A. (1992) Immunological characterization and intracellular localization of *trans*-spliceosomal small nuclear ribonucleoproteins in *Trypanosoma brucei*. *J. Biol. Chem.*, **267**, 20159–20163.
- Schimanski, B., Laufer, G., Gontcharova, L. and Günzl, A. (2004) The *Trypanosoma brucei* spliced leader RNA and rRNA gene promoters have interchangeable TbSNAP50-binding elements. *Nucleic Acids Res.*, **32**, 700–709.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M. and Séraphin, B. (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.*, **17**, 1030–1032.
- Laufer, G., Schaaf, G., Bollgönn, S. and Günzl, A. (1999) *In vitro* analysis of alpha-amanitin-resistant transcription from the rRNA, procyclic acidic repetitive protein, and variant surface glycoprotein gene promoters in *Trypanosoma brucei*. *Mol. Cell Biol.*, **19**, 5466–5473.
- Bell, M. and Bindereif, A. (1999) Cloning and mutational analysis of the *Leptomonas seymouri* U5 snRNA gene: function of the Sm site in core RNP formation and nuclear localization. *Nucleic Acids Res.*, **27**, 3986–3994.
- Kohtz, J.D., Jamison, S.F., Will, C.L., Zuo, P., Lührmann, R., Garcia-Blanco, M.A. and Manley, J.L. (1994) Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature*, **368**, 119–124.

29. Ismaili,N., Perez-Morga,D., Walsh,P., Cadogan,M., Pays,A., Tebabi,P. and Pays,E. (2000) Characterization of a *Trypanosoma brucei* SR domain-containing protein bearing homology to *cis*-spliceosomal U1 70 kDa proteins. *Mol. Biochem. Parasitol.*, **106**, 109–120.
30. Nelissen,R.L., Heinrichs,V., Habets,W.J., Simons,F., Lührmann,R. and van Venrooij,W.J. (1991) Zinc finger-like structure in U1-specific protein C is essential for specific binding to U1 snRNP. *Nucleic Acids Res.*, **19**, 449–454.
31. Gottschalk,A., Tang,J., Puig,O., Salgado,J., Neubauer,G., Colot,H.V., Mann,M., Seraphin,B., Rosbash,M., Lührmann,R. and Fabrizio,P. (1998) A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. *RNA*, **4**, 374–393.
32. Fabrizio,P., Esser,S., Kastner,B. and Lührmann,R. (1994) Isolation of *S.cerevisiae* snRNPs: comparison of U1 and U4/U6.U5 to their human counterparts. *Science*, **264**, 261–265.