ELSEVIER

Molecular & Biochemical Parasitology 112 (2001) 39–49

MOLECULAR & BIOCHEMICAL PARASITOLOGY

www.parasitology-online.com.

Identification of a spliced leader RNA binding protein from *Trypanosoma cruzi*☆

Ping Xu, Limin Wen, Gauri Benegal, Xu Wang, Gregory A. Buck *

Department of Microbiology and Immunology, Medical College of Virginia Campus, Box 980678, Virginia Commonwealth University, 1101 East Marshall, Rm. 5036 Sanger Hall, Richmond, VA 23298-0678, USA

Received 25 May 2000; accepted 13 September 2000

Abstract

Nuclear mRNAs in trypanosomatids are generated by *trans*-splicing. Although *trans*-splicing resembles *cis*-splicing in many ways and most of the U RNA participants have been characterized, relatively few involved proteins have been identified. Herein, we employed a yeast three-hybrid system to identify a protein, XB1, which binds to the *Trypanosoma cruzi* SL RNA. XB1 is a ~45 kDa protein which is homologous to the essential pre-mRNA-splicing factor PRP31p from *Saccharomyces cerevisiae*. Gel shift assays and UV cross-linking experiments with recombinant XB1 confirmed that this *T. cruzi* protein binds the SL RNA in vitro. The binding site of XB1 on the SL RNA was mapped to stem-loop II by deletion of the SL RNA 'bait' in the three-hybrid system. Finally, UV cross-linking SL RNA with S100 extract indicated native XB1 protein and SL RNA interaction in *T. cruzi* extract.

Keywords: RNA-protein interaction; SL RNA; SL RNA binding proteins; Three-hybrid screen; Trans-splicing factor

1. Introduction

Kinetoplastid protozoa process their pre-mRNA transcripts into mature mRNA via a bi-molecular *trans*-splicing mechanism. In this mechanism, a 39 base capped leader is spliced from the 5′ end of a small abundant nuclear transcript to the 5′ termini of the mature mRNAs [1,2]. This process occurs at 3′ splice sites (3′ SS) located upstream from the translation start codon of each protein-coding gene and defines the 5′ termini of mRNAs that were initially synthesized as polycistronic transcripts. The *trans*-splicing process also seems to be coupled with the 3′ cleavage and polyadenylation events that define the 3′ termini of mature mRNAs [3]. As a result of these two processes, intergenic regions of polycistronic pre-mRNAs are excised and turned over, and the mRNAs are processed into mature competent monocistronic units. An 'AG' dinucleotide and polypyrimidine (pPy) tract upstream from the 3′ splice site in the pre-mRNA, and a 'GU' dinucleotide downstream from the 39 nt SL in the SL RNA resemble the 3′ and 5′ splice sites that have been well-characterized in other eukaryotes [4,5].

Despite our detailed knowledge of the *cis*-splicing mechanism of other eukaryotes (for review, see Ref. [4]), there are several major gaps in our understanding of pre-mRNA processing in trypanosomatids. Thus, the mechanisms of splice site selection and activation remain only partially defined. It remains largely unknown how the 5′ and 3′ splice sites from two discrete RNAs, i.e. the SL RNA and the pre-mRNA, become juxtaposed prior to splicing. A complete or nearly complete complement of small nuclear RNAs (snRNAs), including U2 [6–8], U4 [7,8], U5 [9,10], U6 [11], and possibly others [12–14], and most recently a possible U1 homolog [15,16], have been identified in the trypanosomes.

Abbreviations: 3-AT, 3-aminotriazole; BLAST, basic local alignment search tool; bp, base pair(s); EMSA, electrophoretic mobility shift assays; GST, glutathione S-transferase; kb, kilobase pair(s); kDa, kilodalton; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; SL, spliced leader; snRNP, small nuclear ribonucleoprotein particle; X-gal, 5-bromo-4-chloro-3-indolyl-β-D-galactoside.

☆ *Note:* Nucleotide sequence data reported in this paper is available in GenBank under the accession number: AF140548.

* Corresponding author. Tel.: +1-804-828-2318; fax: +1-804-828-1397.

E-mail address: buck@hsc.vcu.edu (G.A. Buck).

0166-6851/01/$ - see front matter
PII: S0166-6851(00)00341-8

In contrast to the relatively complete panel of snRNAs known in these protozoans, the protein components of the *trans*-splicing process remain largely obscure. Thus, whereas several dozen protein factors are known to participate in splicing in *Saccharomyces cerevisiae* and mammalian cells [4], only a handful of splicing-related proteins have been identified in the trypanosomes. In an early report [17], several proteins ranging in size from ~40 to 90 kDa or greater were shown to cross-link to the SL RNA of *Trypanosoma brucei*. Another report [18] showed four proteins ranging in size from 27 to 36 kDa that co-purified with the SL RNP, and one of ~31 kDa that co-purified with the U2 RNP. A panel of low molecular mass (8.5–15 kDa) proteins were also identified as 'core' components of the SL, U2 and U4/U6 snRNPs of *T. brucei* [19]. These studies also identified a ~40 kDa protein specific for the U2 RNP and a possible ~25 kDa protein specific for the U4/U6 RNP. The ~40 kDa protein was cloned and shown to be partially homologous to human U2A′, and to participate in the assembly of the U2 snRNP [20,21]. In *T. brucei*, a 30–43 kDa SR protein homologue that binds the SL RNA (TSR1) was identified [22], and ~43 kDa protein displaying homology to the U1 70 kDa protein was demonstrated to interact with TSR1 [23], and a ~50 kDa protein was identified in the U6 snRNP. The latter was shown to interact with the 5′ terminus of the U RNA, and to play a role in γ-monomethylphosphate capping of the RNA [11]. An apparent homologue of the U5-specific protein PRP8 from *S. cerevisiae* and p220 from humans was recently cloned from *T. brucei* [24]. This protein apparently participates in the early stages of spliceosome assembly [25]. Despite these advances, it is clear that only a small fraction of the protein factors involved in *trans*-splicing in the trypanosomes have been identified and even fewer have been functionally characterized.

Central to the *trans*-splicing event is the SL RNA itself. This small abundant nuclear RNA is 5′-G capped, contains several associated novel 5′ modifications [26], and bears the 39 nt SL on its 5′ end. The SL RNA has been predicted to fold into a characteristic and highly conserved secondary structure [27–29]. Minimally, this secondary structure contains three conserved stem-loops [27,30], although alternative secondary and higher level structures have also been proposed [28,31,32]. Whereas the roles of these structures remain obscure, all three stem-loops, I–III, are probably required [33] for the in vivo splicing process [34,35].

Herein, we have begun to identify and dissect proteins that could be involved in *trans*-splicing and nuclear mRNA maturation in *T. cruzi* using a yeast three-hybrid system [36,37]. A cDNA library generated in a yeast three-hybrid vector was used to select clones expressing fusion proteins that bind to the 111 nt *T. cruzi* SL RNA. One clone selected encodes a ~45 kDa protein, XB1, that shows homology with yeast splicing factor PRP31p, which plays a role in spliceosome assembly in that organism. We show that XB1 interacts with the SL RNA when it is expressed in yeast. The XB1 protein also interacts with the *T. cruzi* SL RNA in vitro as demonstrated by gel mobility shift assays and UV cross-linking. Finally, we show that XB1 binds to stem-loop II region of the SL RNA.

2. Materials and methods

2.1. Cells, cultures, plasmids and other materials

T. cruzi CL Brener strain epimastigotes were grown in LIT medium supplemented to 10% with fetal calf serum under gentle agitation at 28°C essentially as previously described [30]. Dr M. Wickens (University of Wisconsin-Madison) generously provided the yeast three-hybrid system. Proteinase K, RNasin, rNTP, and dNTP were from Promega. T4 DNA ligase, T4 polynucleotide kinase and all restriction enzymes were from New England Biolabs. RNase A was from Sigma. [α-^{32}P]GTP, [α-^{32}P]dATP and [γ-^{32}P]ATP were from DuPont NEN.

2.2. Yeast three-hybrid screening

2.2.1. T. cruzi cDNA-Gal4 activation domain fusion

To prepare the Gal4AD fusion cDNA library, total RNA was isolated from the cultured *T. cruzi* CL Brener strain using guanidinium isothiocyanate extraction [38]. Poly(A) RNA was purified using an mRNA Separator kit (Clontech). cDNA was synthesized using a HybriZAP™ two-hybrid cDNA synthesis kit essentially as described by the manufacturer (Stratagene). Large scale double-stranded phagemid DNA for yeast transformation was prepared by CsCl centrifugation [39].

2.2.2. The RNA 'bait' construct

The RNA 'bait' fusion was prepared using the yeast RNA expression plasmid pIIIAMS2-1. pIIIAMS2-1 was digested with *Sma*I and dephosphorylated with calf intestine alkaline phosphatase (CIAP). Four oligonucleotides, SL5F, SL5R, SL3F and SL3R, were annealed to form two double-stranded DNA fragments with 'sticky' ends. The two double-stranded DNAs were ligated, phosphorylated and inserted in pIIIAMS2-1 to form the pIIIAMS2/SL RNA 'bait' plasmid.

2.2.3. Selection

The pIIIAMS2/SL RNA plasmid was transformed into *S. cerevisiae* L40-coat strain. Yeast cells containing the plasmid were transformed in large scale with the cDNA library using the lithium acetate method [39].

The transformants were spread on synthetic media lacking histidine and leucine but containing 2.5 mM 3-AT. The plates were incubated at 28°C for 3–6 days. Growing colonies were analyzed further for *lac*Z expression by streaking on synthetic medium lacking histidine and leucine in the presence of 3-AT and X-gal [39]. Plasmids were isolated from 2 ml cultures of positive clones using the glass beads method [39], and transformed into *E. coli* XL-1 Blue (Stratagene). Candidate plasmids were transformed into *S. cerevisiae* L-40-coat yeast cells containing pIIIAMS2/SL RNA plasmid to confirm the interaction or used for further study.

2.3. Plasmid construction and oligonucleotides

The GST fusion protein plasmid was generated by inserting an *Eco*RI-*Sal*I fragment from the candidate clone pXB1 into *Eco*RI and *Sal*I cleaved pGEX 4T-1 (Pharmacia) to generate pGST-XB1. SL RNA deletion plasmids were generated using inverse PCR [40]. The *Eco*RI fragment from pIIIAMS2-1/SL RNA was first inserted into pUC18 (Pharmacia) to decrease the PCR template size. SL RNA deletions were inverse amplified using Pfu Turbo™ DNA polymerase (Stratagene) with the intact plasmid. Different pairs of oligonucleotides, DR1/DF1, DR2/DF2, DR3/DF3 and DF1/DR2 were used to generate the constructs. Amplified DNA fragments were phosphorylated, self-ligated, and transformed into *E. coli*. The *Eco*RI fragments containing the altered SL RNA sequences were excised from the resulting plasmids and inserted into pIIIAMS2-1.

2.4. Genome walking, RT-PCR, and sequencing

DNA fragments flanking the XB I coding sequence were amplified by PCR using the Universal GenomeWalker™ Kit (Clontech) as described as manufacturer with four gene-specific primers, two nested primers g19-3 and g22-3 for upstream of the gene, and two nested primers g19-5 and g22-5 for downstream of the gene. The amplified fragments were separated on low melting temperature agarose gels and sequenced. To clone the genomic DNA fragment containing the full-length coding region, two oligonucleotides, gXBI-5 and gXBI-3, were designed from the genomic sequence data. A $\sim$1.7 kb genomic DNA fragment containing the putative open read frame encoding the full-length XB1 protein was obtained using Pfu Turbo™ DNA polymerase (Stratagene).

Spliced leader-containing XB1 gene-specific cDNA was obtained by RT-PCR using the Advantage™ cDNA Polymerase Mix (Clontech), and the SL-12 and nested gene specific primers, g19-3 and g22-3. Complementary DNA was synthesized from poly(A) mRNA using random hexamer primers. The primary PCR was performed using the SL-12 and g22 primers, and 20 ng of cDNA in a 50 μl reaction volume under the following cycling conditions: 94°C 1 min, five cycles with 94°C 10 s, 72°C 2 min, 25 cycles with 94°C 10 s, 50°C 30 s, 72°C 2 min, extension for 10 min at 72°C. The secondary PCR was performed using the SL-12 and g19-3 primers, 1 μl 1000 diluted primary PCR reaction in a 50 μl reaction volume under conditions: 94°C 1 min, 30 cycles of 94°C 10 s, 56°C 30 s, 72°C 1 min, extension for 10 min at 72°C. The amplified band from the second PCR reaction was excised from gel and directly sequenced. All oligonucleotide primers used in this research are listed in Table 1. All DNA sequencing reactions were performed on a Perkin-Elmer Applied Biosystems 377 Prism Automated DNA Sequencer using BigDye™ Sequencing Reagents (PE/ABI).

2.5. GST fusion protein purification

The GST-XB1 and GST recombinant fusion proteins were obtained by over-expression in *E. coli* BL21 cells containing pGST-XB1 or pGST alone, respectively. The recombinant GST-XB1 and GST proteins were affinity purified separately on glutathione-agarose beads as previously described [41]. Purified GST-XB1 and GST were judged to be more than 90% pure by Coomassie blue staining of SDS-polyacrylamide gels. The glutathione-agarose bead purified protein was further purified by SDS-PAGE for production of a rabbit polyclonal antibody essentially as described [39]. The

Table 1
Oligonucleotides used in this project[a]

DF1	GTCAATTTCTTTTGACCGGGGTCCAC
DF2	CCAACCCGCCTCTGGCGGCTATGT
DF3	GGGGGATCCACTAGTTCTAGCC
DR1	AAGCTTCGCGTACCAATATAGTACAGAA
DR2	TATCAATAATAGCGTTAGTTGGGCGATC
DR3	GGTCAAAAGAAATTGACCAAACATAGC
g19-3	ACACCACCGGCGATGGCAAACAGTTGTGA
g19-5	CGTCAACGGGAGATGGTGCGTTTAAAG
g22-3	TGACTAGAAGCTCTACGGTTTGCGTCAATG
g22-5	ACGCAAACCGTAGAGCTTCTAGTCAACTTG
gXB1-3	CGCAGATCTCCAGTCCCAGGTCTGTCAGGGATA TG
gXB1-5	ATAAGATCTGGAGCAGCCATGATGATGGCTACT G
SL-12	TATTGATACAGTTTCTGTACTA
SL3F	TCCAACCCGCCTCTGGCGGCTATGTTTGGTCAAT TTCTTTTGACCGGGGTCCACGAACCC
SL3R	GGGTTCGTGGACCCCGGTCAAAAGAAATTGACC AAACATAGCCGCCAGAGGCGGGTT
SL5F	AACTAACGCTATTATTGATACAGTTTCTGTACTA TATTGGTACGCGAAGC
SL5R	GGAAGCTTCGCGTACCAATATAGTACAGAAACT GTATCAATAATAGCGTTAGTT
T7SL	AAGGATCCCGAAATTAATACGACTCACTATAGG CTAACGCTATTATTGATACAG
T7SL-3	GGGTTCGTGGACCCCGGTC

[a] Oligonucleotides are shown 5′–3′.

rabbit antibody produced was purified using protein A and GST-linked resin chromatography as previously reported [39].

2.6. Electrophoretic mobility shift assay and UV-cross linking

A T7 promoter was incorporated upstream from the SL RNA by ligating PCR amplification products of the SL RNA gene generated with primers T7SL and T7SL-3 into the *Sma*I site of pUC18. To increase transcription efficiency, the two 5′ As of the SL RNA were replaced by Gs in the T7SL primer. The resulting pT7SL plasmid was purified and cleaved with *Bam*HI and *Sma*I, and the liberated fragment was recovered from a 5% polyacrylamide gel. Radioactively labeled SL RNA transcript was synthesized by incorporation of [α-^{32}P] GTP (800 Ci mmol^{-1}, NEN) using a Riboprobe kit (Promega). The transcription reaction was treated with RNase-free DNase I (Promega) for 15 min at 37°C to remove DNA template. Transcribed RNA was extracted once with phenol/chloroform and once with chloroform. Free ribonucleotides were removed by ethanol precipitation. The integration of radioactive labeled RNA was analyzed by denaturing polyacrylamide gel electrophoresis and quantified by liquid scintillation counting.

Electrophoretic mobility shift assays (EMSA) were performed essentially as previously described [42]. Briefly, protein eluted from GST beads was incubated on ice for 20 min with [^{32}P]SL RNA in buffer containing 0.1 mM ATP and fractionated on a native polyacrylamide gel. Gels were dried and visualized by conventional autoradiography or with a PhosphorImager (Molecular Dynamics).

For UV cross-linking, SL RNA was incubated with protein as described for EMSA. After incubation on ice for 20 min, the reactions were transferred to a flat sheet of parafilm on an ice-cold metal platform. Reaction mixtures were UV irradiated at 254 nm at 6 cm distance for 60 min and treated with RNase A (2 mg ml^{-1}) for 30 min at 37°C. RNA-protein complexes were boiled after addition of SDS-polyacrylamide gel sample buffer and resolved in 12% SDS-polyacrylamide gels (acrylamide/methylene:bisacrylamide, 29:1). Gels were dried and visualized by a conventional autoradiography for 3–6 days.

2.7. Northern and Western blot analyses

Total RNA was prepared from *T. cruzi* using Trizol reagent (Life Technologies). Altogether 15 μg of total RNA was analyzed by Northern hybridization [43] with the [α-^{32}P] labeled probe of *Eco*RI-*Sal*I fragment of pXB1 cDNA. The results were visualized by autoradiography.

S100 extract was separated on SDS-PAGE and blotted onto nitrocellulose membranes in a Mini Trans-blot Electrophoretic Transfer Cell (Bio-Rad) for Western analysis [39]. The membrane was blocked for 2 h with TTSB (10 mM Tris buffer, 0.05% Tween 20 and 150 mM NaCl) containing 2% BSA, and incubated for 2 h at room temperature with antibody against GST-XB1 antibody (1:5000 dilution), probed with alkaline-labeled goat anti-rabbit IgG (1:3000 dilution, Promega), and detected in NBT/BCIP system (Pierce). The molecular sizes of protein bands were determined from pre-stained bands on the protein ladder. As a control, a membrane containing separated protein bands from extracts were incubated with pre-immune rabbit serum (1:200). No significant labeling was observed (not shown).

2.8. Immunoprecipitation and competition interaction

A T7 promoter was incorporated upstream from SL RNA sequences by PCR amplification with T7SL and T7SL-3 primers, with the exception that M3 was generated using primers T7SL and DR3. Radioactively labeled 'wild-type' and mutated SL RNAs were synthesized by incorporation of [α-^{32}P] GTP (800 Ci mmol^{-1}, NEN) using a Riboprobe kit (Promega). The large scale cold RNA competitors were synthesized using the RiboMax kit (Promega). A 100-fold excess of cold RNA competitor was mixed with radiolabeled SL RNA. Immunoprecipitation was performed as described previously [44] except that the reactions were UV irradiated for 60 min and treated with RNaseA for 30 min at 37°C before antibody precipitation. The beads with precipitated proteins were boiled and proteins were separated by SDS-PAGE. The gel was dried and exposed on X-ray film for 3–6 days at −70°C.

3. Results

3.1. Yeast three-hybrid system selects a putative SL RNA-binding factor

The yeast three-hybrid system [36,37] was used to select *T. cruzi* genes encoding proteins that interact with the SL RNA. In this system (Fig. 1A), a hybrid RNA molecule bridges two fusion proteins, one containing the LexA DNA-binding domain (LexABD) fused with MS2 coat protein RNA-binding domain and the other containing the Gal4 transcriptional activation domain (Gal4AD) fused with the RNA-binding protein of interest. The LexABD-MS2 coat protein RNA-binding domain fusion is constitutively expressed from a construct integrated into the chromosome *S. cerevisiae* L40-coat [37], a yeast strain that also contains integrated reporter genes (*his*3 and *lac*Z) that require tran-

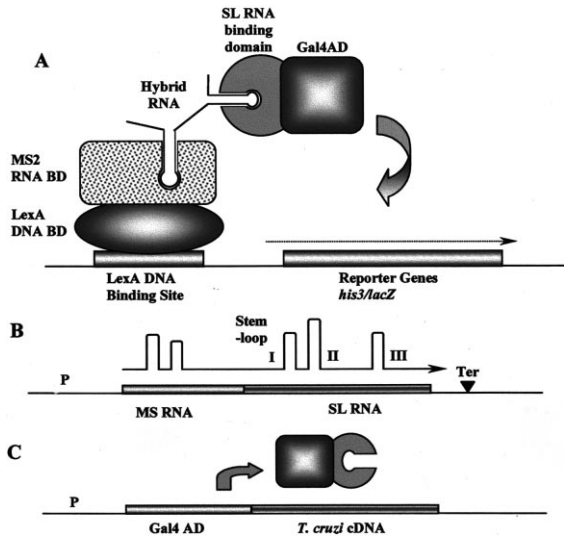

Fig. 1. Components of the yeast three-hybrid system. (A) Schematic illustration of the yeast three-hybrid system [36]. The yeast host constitutively expresses a fusion protein incorporating both the MS2 coat protein RNA binding domain (MS2 RNA BD) and the bacterial LexA DNA binding domain (LexA DNA BD). A second construct expresses a fusion RNA (Hybrid RNA) incorporating the MS2 coat protein binding site and the RNA target in question, in this case the *T. cruzi* SL RNA, flanked by small segments of RNAse P RNA [36]. The cDNA is fused with the yeast Gal4 activation domain (GalAD). LexA DNA binding sites are incorporated upstream from the selectable *his*3 gene, and screenable *lac*Z genes. When a cDNA encoding the desired RNA binding protein is fused with the Gal4AD, the Gal4AD is brought into proximity of the reporter genes (*his*3 and *lac*Z), activating their transcription to permit selection on histidine-medium and generation of blue colonies on *X*-gal. (B) Schematic representation of the 'bait' RNA construct; pIIIAMS2-1/SL RNA. The intact 111 nt *T. cruzi* SL RNA was fused downstream from the MS2 coat protein RNA binding site and upstream from a yeast RNA polymerase III terminator (Ter). (C) Schematic illustration of the *T. cruzi* cDNA library. The *T. cruzi* cDNA library was fused to the carboxyl terminal of the yeast Gal4AD.

scriptional activation via the LexA DNA binding site (Fig. 1A). A hybrid RNA containing the target or 'bait' RNA fused to the MS2 coat protein RNA-binding site is introduced to the yeast strain. A Gal4AD-cDNA library is introduced into LexA-MS2/fusion RNA-expressing yeast, and the yeast cells are plated on selective medium (histidine$^-$ with X-gal) to simultaneously poll expression of both the *his*3 and *lac*Z reporter genes. In this system, only yeast expressing a Gal4AD-cDNA fusion protein that binds to the *T. cruzi* SL RNA bait will survive on the histidine$^-$ selective medium and yield blue colonies on X-gal.

We prepared hybrid RNA in the yeast RNA expression plasmid pIIIAMS2-1, which contains an RNA polymerase III (Pol III) promoter transcribing MS2 RNA sequences that bind the MS2 coat protein upstream from the *S. cerevisiae* RNase P RNA gene Pol III terminator (RPR1) [37]. Our initial RNA fusion construct, pIIIAMS2-1/SL RNA (Fig. 1B), contains the entire 111 nt SL RNA, including stem-loops I, II and III, fused between the MS2 coat protein binding sequences and the Pol III terminator. The RNA fusion construct, pIIIAMS2-1/SL RNA, was transformed into *S. cerevisiae* L40-coat for use as 'bait' in the three-hybrid system to select *T. cruzi* cDNAs expressing proteins that bind to the SL RNA. A *T. cruzi* cDNA library fused to the Gal4AD (Fig. 1C) was generated in the HybriZAP™ two-hybrid cDNA Vector (Stratagene) as described in the Section 2. The titer of the primary library was $\sim 1.0 \times 10^7$, and PCR analysis indicated that $\sim 90\%$ contained inserts with sizes ranging from 0.5 to 2.5 kb, with an average of ~ 1.5 kb (data not shown). Yeast strain *S. cerevisiae* L40-coat, expressing the pIIIAMS2-1/SL RNA fusion transcript and the LexA-MS2 coat protein fusion, was transformed with the *T. cruzi* cDNA library. Transformants were selected on plates containing synthetic medium lacking histidine and leucine in the presence of 3-AT (Sigma) to eliminate yeast growth due to leaky expression of *his*3 [45,46]. Of $\sim 2 \times 10^6$ transformants screened, ~ 500 presumptive positive colonies grew on his$^-$ medium.

Previous investigations have shown that a single selection is not sufficient to confirm a putative positive clone in the one-, two- and three-hybrid systems [47,48], i.e. promoter up mutations, second site reversions, and other artifacts account for most of the primary positives. The three-hybrid system we used incorporates both a selection for *his*3 prototrophy and a screening for *lac*Z up regulation to separate the 'real' SL RNA binding activities from the artifacts. Thus, the ~ 500 presumptive positives from the histidine selection were directly screened on X-gal containing medium. Several clones giving strong blue colonies were picked and sequenced (data not shown). The sequences of two clones were identical, although one was longer than the other, indicating that these two clones probably arose independently from the same transcript. The shorter of these two clones, pXB1, encoding the putative XB1 SL RNA-binding factor, was selected for further characterization.

3.2. Verification of XB1-mediated transcriptional activation in yeast

Second site mutations or other host-related artifacts could be responsible for the activation of both *his* and *lac*Z in the initial experiments. We performed several experiments to verify that interaction between XB1 and the SL RNA was responsible for transcription of *his*3 and *lac*Z. First, pXB1 plasmid was recovered from the transformed *S. cerevisiae* L40-coat strain bearing the pIIIAMS2-1/SL RNA, amplified in *Escherichia coli*, and transformed anew into L40-coat yeast containing pIIIAMS2-1/SL RNA. Transformants were selected on medium lacking leucine and uracil to select for the

cDNA and RNA fusion constructs, respectively. Several independent colonies were streaked on synthetic media lacking leucine, uracil and histidine in the presence of 2.5 mM 3-AT to select for *his*3 expression driven by the interaction of the cDNA fusion with the SL RNA-MS2 RNA fusion. The medium also contained *X*-gal to screen for *lac*Z expression driven by the same interaction. All colonies grew well and were blue, confirming that expression of both the *his*3 and *lac*Z reporter genes is mediated by and dependent on expression of the XB1 fusion protein (data not shown).

Second, to confirm the specific interaction of the XB1 fusion protein with the SL RNA and not with other parts of the MS2/SL-RNA fusion transcript, several control plasmids were constructed and tested in yeast (Fig. 2). The pXB1 plasmid was introduced into yeast bearing plasmid vector pIIIAMS2-1, which expresses the MS2 RNA but lacks the SL RNA, to verify that the protein was not binding to the MS2 portions of the RNA. Also, pADGAL4, expressing the activation domain of Gal4 but lacking a *T. cruzi* cDNA, was introduced into yeast expressing the full-length MS2-SL RNA fusion RNA from pIIIAMS2-1/SL RNA to verify that the activation domain of Gal4 was not itself binding to the SL RNA. Finally, pRandom, expressing a randomly selected Gal4 fusion protein from the *T. cruzi* cDNA library, was introduced into yeast expressing the full-length fusion RNA from pIIIAMS2-1/SL RNA to verify that this construct would not artifactually activate expression. In each of these control experiments, no growth was observed in the absence of histidine (Fig. 2). Thus, *his*3 and *lac*Z expression occurred only in yeast that contain both the MS2-SL RNA fusion construct and the Gal4AD-cDNA protein fusion construct mediated *his*3 and *lac*Z expression. These results indicate a direct interaction between the *T. cruzi* XB1 protein and the SL RNA.

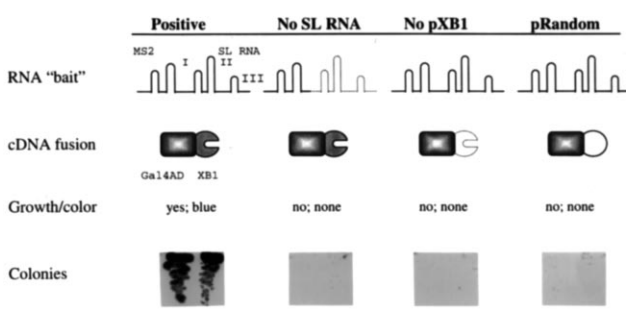

Fig. 2. Verification of transcriptional activation of *his*3 and *lac*Z by XB1. Constructs of the RNA 'bait' fusion, with or without the SL RNA, and cDNA fusions with or without the XB1 coding sequences, were tested for their ability to activate transcription of both *his*3 and *lac*Z in the three hybrid system as outlined in the text. Dotted lines indicate that a sequence or structure was deleted in that construct.

```
GGAGCAGCCATGATGATGGCTACTGACTCCCCCTTTTTTTTTTCTTTGTTTGTGATTGTG
TTGCCGTGGCTTTCATGCGCTCTCTCCCTTGTTTTTTTTTTTTCGCATATACAGTGGAAG  120
AAGACGACGGAGGAAGATGGCGAGCGGGCGAAGTGGGGACGAGAAAAGAACACGAGGCGG
CGTGACATACACTTGTTAATTTTTTTACATTTAAAAAATTATCTTGTTGTTATTTTAGT
TTTGGGTTCAGCACTGTCTTGTTCCCGTTTTCTTGCTCTTCTCCTCCTTGCGCACCTCAT
TTTATGCACGATTATTTTTTAAAAATTCTTTGCATTTTGTTAATGCAGCCGGATCAAG  360
ATGGATGAGAATGAGGGGGCTGGGAGAATTTGCCGAAGAGCCGCAACAATGGCGAA
M  D  E  N  E  G  G  W  E  E  F  A  E  E  P  Q  Q  Y  G  E
GCGGAGGACGCAGCGGATATATACGCCGAAGAGACGCTTGCAACAGCGCAGAAAATCGCA  480
A  E  D  A  A  D  I  Y  A  E  E  T  L  A  T  A  Q  K  I  A
TCTGATGACGACGCACTTCGCTTCGATAGCGTGAAGGAGGTGACCCTCCTTTTACGTTCA
S  D  D  D  A  L  R  F  D  S  V  K  E  V  T  L  L  L  R  S
GCATACATGAGTCGGATGTTGCAGAAGCTGGGCGACTACAGTGAACAAGAGGTGGTGAAG  600
A  Y  M  S  R  M  L  Q  K  L  G  D  Y  S  E  Q  E  V  V  K
AAGACCATCCTGCCGGAGGATCCTGAGTACCAATTTGTAATTGATTCTAGTACTCTTGTT
K  T  I  L  P  E  D  P  E  Y  Q  F  V  I  D  S  S  T  L  V
TTGCGTATTGAGGTGGAGAAGAGCAAAGCGGTCGTGTACCTGCGGGCACACTATGGACAA  720
L  R  I  E  V  E  K  S  K  A  V  V  Y  L  R  A  H  Y  G  Q
AGATTCCCGGAGTTGGCTATGTTCTTCTCCGACAGTGTGCTTTATGCCCGTATCGTGCGC
R  F  P  E  L  A  M  F  F  S  D  S  V  L  Y  A  R  I  V  R
TTGATTCAAAATAACATGGATCTCTCGTTGTCATTGACCAGCTGGACGAGTTGATACCT  840
L  I  Q  N  N  M  D  L  S  L  V  I  D  Q  L  D  E  L  I  P
TCGCAGCTCACTGCGGTTGTGATTGCGTGCGCCTCTACCACTGCGGGGAGGGAGCTGAGT
S  Q  L  T  A  V  V  I  A  C  A  S  T  T  A  G  R  E  L  S
GAGGAGGAGCTGCATCGTGTAGTCGAGGCTTGCCAGGAGATTGACATTCTCGAGGCTGCA  960
E  E  E  L  H  R  V  V  E  A  C  Q  E  I  D  I  L  E  A  A
AAACAGACATTTCTGGAGTATATTCAACGCTCCATGCCTCTGATCTGCCCAAATCTCTGC
K  Q  T  F  L  E  Y  I  Q  R  S  M  P  L  I  C  P  N  L  C
GCCTTTTGGGGACGGGTATTACGTCACAACTGTTTGCCATCGCCGGTGGTGTAAGTGCG 1080
A  F  L  G  T  G  I  T  S  Q  L  F  A  I  A  G  G  V  S  A
CTCTCCACGATGGACTCCACTGAGTTGGCACGGTTAGGTAGCAAACGCGCCGACTCCTCT
L  S  T  M  D  S  T  E  L  A  R  L  G  S  K  R  A  D  S  S
GGGGTTCTAATTAGAACAACGGGCTTTCTCAGTAACTCGGACCTTGTTGTCAACCACCCG 1200
G  V  L  I  R  T  T  G  F  L  S  N  S  D  L  V  V  N  H  P
CCGCAGATGCGACCCAAGGCTCTGCGTCTTGTGGCAAGTACCACTTCTATGCTTGCACGC
P  Q  M  R  P  K  A  L  R  L  V  A  S  T  T  S  M  L  A  R
ATTGACGCAAACCGTAGAGCTTCTAGTCAACATGAGGGCTATCGTCAACGGGAGATGGTG 1320
I  D  A  N  R  R  A  S  S  Q  Q  E  G  Y  R  Q  R  E  M  V
CGTTTAAAGATGTTATCATGGCTGGATCCGCCGGTTCTTCGTGGAGCTGCAAACAACACA
R  L  K  M  L  S  W  L  D  P  P  V  L  R  G  A  A  N  N  T
TATGCACGGCGCGGAAGAAAAAGGCCGAGGCGGCAAACACGGTAGCTACGCCATTACGTT 1440
Y  A  R  R  G  R  K  R  P  R  R  Q  T  R  *
AGCTATCATTATCCCTTTTGTATCAACCCGTTTTTTCTTGTTTGTTTTTTCCACTTCTAT
▼
ATGACGACCGTGTATGGGATGCTCCATTCATTGCACCCATTTGACCTTTACATTTTTATT 1560
TTTATTCCCTTTGCTGGAGGGGTGTGTATGTTTCTGTGTGCGTGCGTGTGTGGTTTCGTC
TTGATTTTTTTTTTTTTACAGTGTCTGAAAGAGGTTGGTCAAAGGCCCTAAAGGGTGCC
GAGTACAGGTGCGTGATGAGTGAGAACATATCCCTGACAGACCTGGGACTGG        1732
```

Fig. 3. Sequence of XB1 binding protein. Sequence of XB1 (GenBank accession number: AF 140548). The sequence of the single genomic copy of the gene encoding XB1 is shown. The deduced amino acid sequence is indicated. The 'AG' dinucleotide splice site 10 nt upstream from the ATG start codon is underlined. The poly(A) cleavage and addition site is indicated with an arrowhead (▼).

3.3. XB1 shows homology with yeast splicing factor PRP31p

The 658 base pair cDNA insert in the pXB1 cDNA clone contained a 546 bp open reading frame (ORF) capable of encoding a 181 amino-acid polypeptide with a 78 bp 3′ untranslated region (UTR) and a 34 base poly(A) tail (not shown). As expected, the XB1 open reading frame was fused in frame with the Gal4 transcriptional activation domain of pADGAL4. The second clone encoding XB1 had an identical nucleotide sequence but lacked 15 bases at the 5′ end. A genomic fragment containing the full-length pXB1 gene was isolated as described in the Section 2. The 1732 bp genomic DNA fragment included a single open reading frame of 1062 bp (Fig. 3), encoding a predicted 354 amino acid protein with a molecular weight of ~45-kDa. The ATG start codon at position 361 (Fig. 3) is in a favorable context to be a eukaryotic translation initiation site [49], and a TAA translation stop codon 18 bp upstream from and in frame with the ATG confirms its role as the translation initiation site for this protein.

The splice site of the pXB1 gene was determined by nested RT PCR amplification using a 5′ primer homologous to the spliced leader (SL) and two 3′ primers complementary to the XB1 coding sequence as described in the Section 2. The RT-PCR fragment was directly sequenced revealing a 3′ splice site 10 bp upstream from the ATG translation start codon of XB1 (Fig. 3). The poly(A) addition site is located 78 bp

downstream from the translation stop codon, indicating a mRNA of ~1.2 kb plus the poly(A) tail. Northern analysis revealed a transcript of ~1.2 kb, supporting this conclusion (Fig. 4A). The sequences of the 3′ UTR of the cDNA and the 3′ flanking genomic DNA were identical upstream of the poly(A) tail on the cDNA. A polypyrimidine (pPy) tract and 'AG' dinucleotide located ~120 bp downstream of the poly(A) site is consistent with previous observations suggesting that accurate poly(A) site selection in trypanosomatids requires a downstream 3′ splice site [3,5].

BLASTP searches revealed homology (score 80, bits 194, E value 3e-14 with effective search space 36823744830) to splicing factor PRP31p, a splicing factor that plays a role in U4/U5/U6 snRNP function in *S. cerevisiae* [50]. An amino acid alignment between PRP31p and XB1 using GAP (GCG) showed 70 identities and 157 similarities with several small gaps (data not shown). Similarities spanned the XB1 protein, but PRP31p contains a ~120 amino acid carboxyl terminal extension not present in XB1.

3.4. The pXB1 gene is a ~1.2 kb transcript encoding a ~45 kDa protein

Expression of XB1 was examined by Northern analysis of total RNA from cultured *T. cruzi* epimastigotes or from heat-shocked epimastigotes (Fig. 4A and data not shown). This analysis revealed a single ~1.2-kb transcript in both normal and heat-shocked cells, and there was no detectable alteration of signal in response to heat shock. The molecular size of native XB1 was examined by Western analyses using antibody raised against a glutathione-S-transferase-XB1 (GST-XB1) fusion protein isolated from overexpressing *E. coli* as described in the Section 2. This analysis revealed a ~45 kDa protein in *T. cruzi* extracts (Fig. 4B).

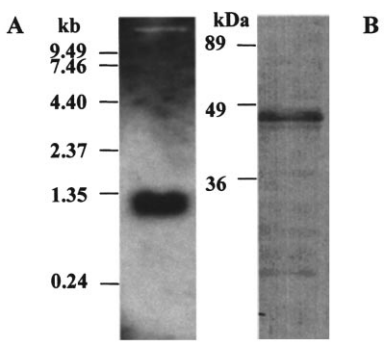

Fig. 4. The native XB1 mRNA and protein. (A) Northern analysis of XB1 transcript. Total RNA was isolated from *T. cruzi* epimastigotes and electrophoresed, blotted and hybridized to the XB1 DNA probe using the radiolabeled *Eco*RI/*Sal*I fragment of XB1 as described in the Section 2. (B) Western blot of native XB1 in *T. cruzi*. Proteins from *T. cruzi* epimastigotes were fractionated by SDS-PAGE, blotted and detected with polyclonal antibody against the XB1-GST fusion protein. A ~45 kDa putative native *T. cruzi* XB1 protein that reacts with the antibody is indicated.

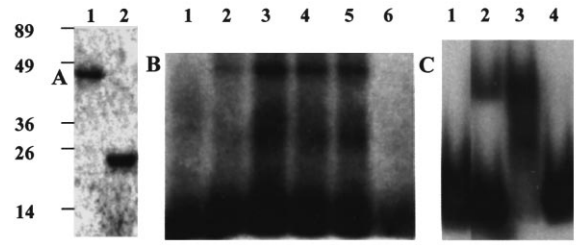

Fig. 5. The recombinant XB1-GST fusion induces a UV cross-link and gel shift with the SL RNA. (A) Recombinant XB1. Recombinant XB1 was isolated as a GST fusion protein from *E. coli*, fractionated on SDS PAGE, and stained with Coomassie Blue (lane 1) as described in the Section 2. GST alone was similarly purified and analyzed (lane 2). Molecular size markers are indicated. (B) UV cross-linking of the recombinant XB1-GST fusion protein and in vitro transcribed SL RNA. The SL RNA and control RNAs were synthesized and radiolabeled enzymatically in vitro as described in the Section 2. The radiolabeled RNA was incubated in the presence of recombinant XB1-GST. The mixture was UV irradiated, treated with RNase A, and analyzed by SDS-PAGE and autoradiography. Reactions in lanes 1–5 contained 20 ng radiolabeled SL RNA probe. Lanes: 1, no XB1-GST; 2, ~0.1 μg XB1-GST with ~20 fold excess cold SL RNA; 3, ~0.4 μg XB1-GST but no competitor RNA; 4, ~0.2 μg XB1-GST but no competitor RNA; and 5, ~0.1 μg XB1-GST but no competitor RNA. The reaction in lane 6 contained (20 ng) radiolabeled control RNA probe and ~0.1 μg XB1-GST. (C) Electrophoretic mobility shift assays with the recombinant XB1-GST and the radiolabeled synthetic SL RNA. SL RNA was synthesized and labeled as described above, electrophoresed without protein (lane 1), or incubated in the presence of XB1-GST (lane 3), with GST alone (lane 4), or in the presence of *T. cruzi* S100 extracts (lane 2). The complexes were electrophoresed and analyzed on native PAGE as described in the Section 2.

3.5. The purified recombinant XB1 protein binds to the T. cruzi SL RNA in vitro

The results of the yeast three-hybrid system experiments showed that XB1 interacts with the SL RNA in vivo in *S. cerevisiae*. To confirm this interaction in vitro, we purified recombinant XB1 protein from *E. coli* (Fig. 5A, lane 1) and used it in UV-cross linking and electrophoretic mobility shift assays (EMSA). The SL RNA sequence was transcribed with T7 RNA polymerase, radiolabeled, and purified as described in the Section 2. A radiolabeled RNA transcribed from the T7 promoter to the *Bam*HI site of pBluescript II was used as a control. Purified GST-XB1 protein was incubated with in vitro transcribed [^{32}P] labeled SL RNA and UV irradiated. The SL RNA was degraded with RNase A and the labeled protein was analyzed by SDS-PAGE and autoradiography. A [^{32}P]labeled band was observed when the SL RNA was cross-linked in the presence of GST-XB1 (Fig. 5B, lanes 2–5), whereas no labeled cross-linked band appeared in the absence of XB1 (lane 1), or when a control RNA was incubated with GST-

XB1 (lane 6). These data suggest that the recombinant XB1 protein binds to the SL RNA in vitro.

The in vitro association of XB1 with the SL RNA was further confirmed in electrophoretic mobility shift assays (EMSA). EMSA was performed by incubation of the in vitro transcribed and [^{32}P]labeled SL RNA with the recombinant GST-XB1 fusion protein as described in the Section 2. As shown in Fig. 5C, the GST-XB1 fusion protein (lane 3), but not the GST protein alone (lane 4), induces an electrophoretic mobility shift of the SL RNA, comparable to the shift induced by incubation of the SL RNA in the presence of *T. cruzi* S100 extract (lane 2). The UV-cross linking and RNA EMSA indicated the XB1 protein interacts with the SL RNA in vitro. We also attempted to immunoprecipitate the SL RNA directly from the *T. cruzi* S100 extract with XB1-specific polyclonal antibody. Synthetic radiolabeled SL RNA was incubated in *T. cruzi* S100, antibody was added, and the putative complexes were immunoprecipitated as described in the Section 2. These initial experiments were inconclusive in that little signal was observed after the immunoprecipitation (data not shown). Together, these results indicate a specific SL RNA-XB1 interaction that may be quite fragile and not stable during immunoprecipitation.

3.6. XB1 interacts with stem-loop II of the SL RNA

SL RNA secondary structure models predict three stem-loops (I, II and III). Since it is conserved across all trypanosomatid species [27,30], we assume that this structure has functional implications. To determine which segment of the SL RNA is recognized and bound by XB1, constructs bearing different portions of the SL RNA transcript (Fig. 6) were cloned into pIIIAMS2-1, transformed into yeast in the presence of pGal4AD-XB1, and assayed for their ability to activate *his*3 and *lac*Z expression as described in the Section 2. In these experiments, constructs bearing the intact SL RNA, or constructs lacking only stem-loop I or III, permitted growth of the yeast strain on histidine$^-$ medium (Fig. 6). Constructs lacking stem-loop II were unable to restore growth. These data suggest that stem-loop II plays important roles in the interaction between the SL RNA and the XB-1 protein and lead to the conclusion that the interaction of XB-1 and the SL RNA requires stem-loop II. These results are consistent with the previous observations that stem-loop II is required for in vivo *trans*-splicing [34,35].

To further confirm the interaction between XB1 and the stem-loop II domain of the SL RNA, the mutated SL RNA constructs were inserted downstream from the T7 promoter as described in the Section 2. Radiolabeled 'wild type' SL RNA was incubated in *T. cruzi* S100 extract as described above in the presence of cold mutated competitor SL RNAs. The extracts were subjected to UV cross-linking, immunoprecipitated as described in the Section 2, and analyzed by SDS-PAGE and autoradiography (Fig. 7). As described above, the intact 'wild type' SL RNA was specifically immunoprecipitated in these experiments (lane 1). More important, excess mutation M1 and M3 RNA, which both retain stem-loop II, efficiently competed immunoprecipitation of the radiolabeled SL RNA (lanes 3 and 5). In contrast, excess RNA from mutations M2 and M4, which both lack stem loop II, fail to compete (lanes 4 and 6).

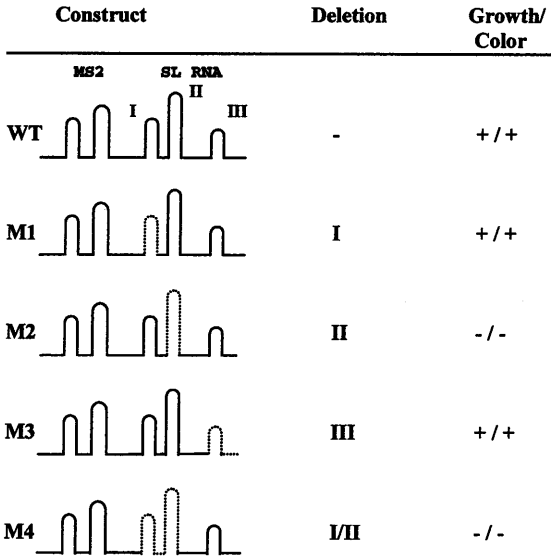

Fig. 6. *T. cruzi* XB1 binds to stem-loop II of the SL RNA. Fusion RNAs bearing different sequences and structures of the *T. cruzi* SL RNA were generated in yeast expression vector pIIIAMS2-1 and tested for their ability to complement a *his*3 mutation and induce *lac*Z expression in yeast expressing the XB1-Gal4AD fusion protein as described in the Section 2. Dotted lines indicate that a sequence or structure was deleted in that construct, and the structures deleted are indicated (I, stem-loop I; II, stem-loop II; and III, stem-loop III of the SL RNA). Growth and color of three independent colonies are reported to indicate activation or no activation of the *his*3 and *lac*Z expression.

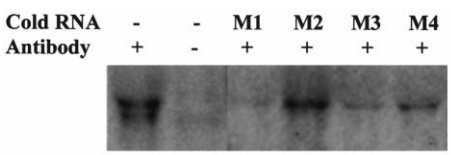

Fig. 7. Stem-loop II is required for binding by XB1 during UV cross-linking. *T. cruzi* S100 extract was incubated with [^{32}P]labeled SL RNA and competed by SL RNA deletion constructs. UV cross-linking experiments were performed as described in the Section 2. The UV cross-linked XB1 was immunoprecipated with XB1 antibody (lane 1) and not immunoprecipated by pre-immune antibody (lane 2). The interaction between XB1 and SL RNA was inhibited by (~ 100-fold excess) of M1 and M3, which both maintain stem-loop II (lanes 3 and 5), but not by M2 and M4, which both lack stem-loop II (lanes 4 and 6).

4. Discussion

Trans-splicing is related to yet distinct from the well-characterized canonical *cis*-splicing process. However, although most or all of the small RNA participants in *trans*-splicing have been identified, very few *trans*-splicing related proteins have been identified or even partially characterized [18,19]. Moreover, many if not all of the SL RNA-associated proteins that have been studied are members of a group of 'core' RNP proteins present on the U RNPs as well as the SL RNP [19]. A putative guanylyl transferase enzyme has been identified in *Trypanosoma brucei* [51] and several proteins that reportedly interact on protein interacted with SL RNA have been observed [17] or cloned [22,23] from *T. brucei*. A comprehensive understanding of the complex *trans*-splicing process will require the identification and characterization of the multiple protein factors that undoubtedly interact stably or transiently with the SL RNA, the U RNAs and pre-mRNA transcripts.

Although a putative U1 RNA has now been characterized from *Crithidia fasciculata* [15,16], a *cis*-splicing reaction has also been observed [52]. Thus, the hypothesis that the SL RNA plays a dual role as splicing substrate and U1 homologue in *trans*-splicing [27] is in question. Nonetheless, the specific functions of the SL RNA sequences and structure are largely unknown. In addition to its role in *trans*-splicing, the SL RNA certainly participates in other functions, e.g. 5′-G capping, nuclear export, or translation. Each of these functions is likely to involve enzymes and other protein factors that have yet to have been identified or characterized.

This study employs a genetic approach for identification of factors involved in these important processes. Previous investigations of SL RNA-binding factors have used biochemical approaches. Such studies are complicated by the difficulty of culturing the massive quantities of potentially hazardous parasites and the relative lack of sophistication of known cell fractionation technology for these organisms. Our genetic approach, using the yeast three-hybrid system, largely circumvents these problems. Thus, generation of a single representative cDNA library fused to the activation domain of the yeast Gal4 transcription factor permits the sensitive selection of cDNAs expressing proteins that interact with the desired 'bait' RNA molecules. In this study, we elected to use the nearly intact *T. cruzi* SL RNA as 'bait' to begin to identify protein factors that play roles in the functions of this small conserved molecule, although other RNAs such as the U RNAs can also be studied.

In our initial selection using the yeast three-hybrid system, the full-length SL RNA was incorporated as 'bait', thus precluding any pre-selection of the function to be played by the selected proteins. This selection identified a *T. cruzi* protein, XB1, which is homologous to the PRP31p splicing factor of *S. cerevisiae*. Control experiments using variations of the three-hybrid screen used to select the pXB1 clone clearly demonstrate that XB1 binds specifically to the SL RNA, and in vitro EMSA and UV-cross-linking experiments using recombinant XB1 verify this interaction. Finally, co-precipitation of the SL RNA with XB1-specific specific antibody also strongly suggests that XB1 is a component of the SL RNP in *T. cruzi*.

Homology between XB1 and the yeast splicing factor PRP31, which plays a role in assembly of U4/U5/U6, is suggestive of a similar function for XB1 in *T. cruzi*. But, why should a participant in U4/U5/U6 assembly bind to the SL RNA? We speculate that XB1 could play the role of bringing the SL RNA into the *trans*-spliceosome, or in securing it in place during one or more steps of the *trans*-splicing process. It is possible that the protein could function through a simultaneous interaction between the SL RNA and the U4/U5/U6 RNP. The strongest homology between XB1 and its homologue from *S. cerevisiae* (PRP31p) is in the central domain. Less homology is observed in the amino terminal 50–100 amino acids, and the carboxyl terminal ~100 amino acids of these homologous proteins are not present in the *T. cruzi* protein. This fact may be suggestive that the amino and carboxyl terminal domains of these proteins perform divergent functions. We have not yet identified the SL RNA binding domain of the XB1 protein, and no RNA binding motifs [53] are evident. Thus, it is possible if not likely that XB1 has multiple RNA binding activities; i.e. for the SL RNA and one of the U RNAs; or that it binds the SL RNA and other proteins bound to other participants in *trans*-splicing; e.g. the U4/U5/U6 snRNP.

Our in vivo deletion studies suggest that XB1 requires stem-loop II for full binding activity. Previous work has identified stem-loop II as a requirement for in vivo *trans*-splicing activity [33–35], although other roles are also possible. It is unlikely that XB1 plays a role in post-splicing events because cleavage at the 5′ splice site dissociates stem-loop II from the mature mRNA. However, it remains possible that XB1 plays a role in 'intron' turnover; i.e. degradation of the branched intermediate or 3′ fragment of the SL RNA that remains after *trans*-splicing.

Since the SL RNA had been proposed as a U1 homologue in *trans*-splicing [27,54], we had anticipated possible selection of U1 70 k, A and C proteins in our screening. We also anticipated possible selection of clones expressing the previously characterized 'core' SL RNA binding proteins [19,55], SR proteins [22,23], or other small proteins previously shown to bind and co-purify with the SL RNA [18]. Other proteins that we might expect in this screen would be capping enzymes and factors, translation initiation factors, etc. We be-

lieve that the fact that we did not observe any of these factors in our initial screens reflects the fact that we used non-normalized cDNA libraries in these screens. Therefore, we probably selected only the most abundantly transcribed SL RNA binding factor and most of the interesting SL RNA-binding proteins remain to be discovered. Subsequent screenings will be performed with normalized T. cruzi cDNA libraries [56] to identify these proteins.

Acknowledgements

This work was supported by grants from the National Institute of Allergy and Infectious Disease at the National Institutes of Health, the National Chapter of the American Heart Association, the Virginia Chapter of the American Heart Association, the Virginia Thoracic Society, the American Lung Association, the A.D. Williams Fund, and the Jeffress Memorial Trust. We thank M. Ruth C. de Carvalho, Yingping Wang and Tadeusz Zwierzynski for assistance in various aspects of the performance of the studies described in this manuscript. All oligonucleotide synthesis and DNA sequencing were performed in the Nucleic Acids Research Facility at VCU. We thank Dr Wickens for the three-hybrid system.

References

[1] Agabian N. *Trans* splicing of nuclear pre-mRNAs. Cell 1990;61:1157–60.
[2] Bonen L. *Trans*-splicing of pre-mRNA in plants, animals, and protists. FASEB J 1993;7:40–6.
[3] LeBowitz JH, Smith HQ, Rusche L, Beverley SM. Coupling of poly(A) site selection and *trans*-splicing in *Leishmania*. Genes Dev 1993;7:996–1007.
[4] Kramer A. The structure and function of proteins involved in mammalian pre-mRNA splicing. Annu Rev Biochem 1996;65:367–409.
[5] Lopez-Estrano C, Tschudi C, Ullu E. Exonic sequences in the 5′ untranslated region of alpha-tubulin mRNA modulate trans splicing in *Trypanosoma brucei*. Mol Cell Biol 1998;18:4620–8.
[6] Hartshorne T, Agabian N. A new U2 RNA secondary structure provided by phylogenetic analysis of trypanosomatid U2 RNAs. Genes Dev 1990;4:2121–31.
[7] Mottram J, Perry KL, Lizardi PM, Luhrmann R, Agabian N, Nelson RG. Isolation and sequence of four small nuclear U RNA genes of *Trypanosoma brucei* subsp. brucei: identification of the U2, U4, and U6 RNA analogs. Mol Cell Biol 1989;9:1212–23.
[8] Tschudi C, Ullu E. Destruction of U2, U4, or U6 small nuclear RNA blocks *trans* splicing in trypanosome cells. Cell 1990;61:459–66.
[9] Dungan JM, Watkins KP, Agabian N. Evidence for the presence of a small U5-like RNA in active *trans*-spliceosomes of *Trypanosoma brucei*. EMBO J 1996;15:4016–29.
[10] Xu Y, Ben Shlomo H, Michaeli S. The U5 RNA of trypanosomes deviates from the canonical U5 RNA: the *Leptomonas collosoma* U5 RNA and its coding gene. Proc Natl Acad Sci USA 1997;94:8473–8.
[11] Groning K, Palfi Z, Gupta S, Cross M, Wolff T, Bindereif A. A new U6 small nuclear ribonucleoprotein-specific protein conserved between *cis*- and *trans*-splicing systems. Mol Cell Biol 1991;11:2026–34.
[12] Cross M, Günzl A, Palfi Z, Bindereif A. Analysis of small nuclear ribonucleoproteins (RNPs) in *Trypanosoma brucei*: structural organization and protein components of the spliced leader RNP. Mol Cell Biol 1991;11:5516–26.
[13] Roberts TG, Sturm NR, Yee BK, et al. Three small nucleolar RNAs identified from the spliced leader-associated RNA locus in kinetoplastid protozoans. Mol Cell Biol 1998;18:4409–17.
[14] Watkins KP, Dungan JM, Agabian N. Identification of a small RNA that interacts with the 5′ splice site of the *Trypanosoma brucei* spliced leader RNA in vivo. Cell 1994;76:171–82.
[15] Schnare MN, Gray MW. A candidate U1 small nuclear RNA for trypanosomatid protozoa. J Biol Chem 1999;274:23691–4.
[16] Schnare MN, Gray MW. Spliced leader-associated RNA from *Crithidia fasciculata* contains a structure resembling stem/loop II of U1 snRNA. FEBS Lett 1999;459:215–7.
[17] Pelle R, Murphy NB. In vivo UV-cross-linking hybridization: a powerful technique for isolating RNA binding proteins. Application to trypanosome mini-exon derived RNA. Nucleic Acids Res 1993;21:2453–8.
[18] Michaeli S, Roberts TG, Watkins KP, Agabian N. Isolation of distinct small ribonucleoprotein particles containing the spliced leader and U2 RNAs of *Trypanosoma brucei*. J Biol Chem 1990;265:10582–8.
[19] Palfi Z, Günzl A, Cross M, Bindereif A. Affinity purification of *Trypanosoma brucei* small nuclear ribonucleoproteins reveals common and specific protein components. Proc Natl Acad Sci USA 1991;88:9097–101.
[20] Cross M, Wieland B, Palfi Z, et al. The *trans*-spliceosomal U2 snRNP protein 40 K of *Trypanosoma brucei*: cloning and analysis of functional domains reveals homology to a mammalian snRNP protein. EMBO J 1993;12:1239–48.
[21] Gunzl A, Cross M, Palfi Z, Bindereif A. Assembly of the U2 small nuclear ribonucleoprotein from *Trypanosoma brucei*. A mutational analysis. J Biol Chem 1993;268:13336–43.
[22] Ismaili N, Perez-Morga D, Walsh P, et al. Characterization of a SR protein from *Trypanosoma brucei* with homology to RNA-binding *cis*-splicing proteins. Mol Biochem Parasitol 1999;102:103–15.
[23] Ismaili N, Perez-Morga D, Walsh P, et al. Characterization of a *Trypanosoma brucei* SR domain-containing protein bearing homology to *cis*-spliceosomal U1 70 kDa proteins. Mol Biochem Parasitol 2000;106:109–20.
[24] Lucke S, Klockner T, Palfi Z, Boshart M, Bindereif A. Trans mRNA splicing in trypanosomes: cloning and analysis of a PRP8-homologous gene from *Trypanosoma brucei* provides evidence for a U5-analogous RNP. EMBO J 1997;16:4433–40.
[25] Brown JD, Beggs JD. Roles of PRP8 protein in the assembly of splicing complexes. EMBO J 1992;11:3721–9.
[26] Bangs JD, Crain PF, Hashizume T, McCloskey JA, Boothroyd JC. Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. J Biol Chem 1992;267:9805–15.
[27] Bruzik JP, Van Doren K, Hirsh D, Steitz JA. Trans splicing involves a novel form of small nuclear ribonucleoprotein particles. Nature 1988;335:559–62.
[28] LeCuyer KA, Crothers DM. The *Leptomonas collosoma* spliced leader RNA can switch between two alternate structural forms. Biochemistry 1993;32:5301–11.
[29] Nilsen TW. *Trans*-splicing: an update. Mol Biochem Parasitol 1995;73:1–6.
[30] McCarthy-Burke C, Taylor ZA, Buck GA. Characterization of the spliced leader genes and transcripts in *Trypanosoma cruzi*. Gene 1989;82:177–89.

[31] Harris KA, Jr, Crothers DM, Ullu E. In vivo structural analysis of spliced leader RNAs in *Trypanosoma brucei* and *Leptomonas collosoma*: a flexible structure that is independent of cap4 methylations. RNA 1995;1:351–62.
[32] LeCuyer KA, Crothers DM. Kinetics of an RNA conformational switch. Proc Natl Acad Sci USA 1994;91:3373–7.
[33] Sturm NR, Campbell DA. The role of intron structures in *trans*-splicing and cap 4 formation for the *Leishmania* spliced leader RNA. J Biol Chem 1999;274:19361–7.
[34] Lucke S, Xu GL, Palfi Z, Cross M, Bellofatto V, Bindereif A. Spliced leader RNA of trypanosomes: in vivo mutational analysis reveals extensive and distinct requirements for trans splicing and cap4 formation. EMBO J 1996;15:4380–91.
[35] Goncharov I, Xu YX, Zimmer Y, Sherman K, Michaeli S. Structure-function analysis of the trypanosomatid spliced leader RNA. Nucleic Acids Res 1998;26:2200–7.
[36] SenGupta DJ, Zhang B, Kraemer B, Pochart P, Fields S, Wickens M. A three-hybrid system to detect RNA-protein interactions in vivo. Proc Natl Acad Sci USA 1996;93:8496–501.
[37] Zhang B, Gallegos M, Puoti A, et al. A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. Nature 1997;390:477–84.
[38] Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 1987;162:156–9.
[39] Ausuble FM, Brent R, Kingston RE, et al. Current Protocals in Molecular Biology. NY, USA: John Wiley and Sons, 1998.
[40] Ochman H, Gerber AS, Hartl DL. Genetic applications of an inverse polymerase chain reaction. Genetics 1988;120:621–3.
[41] Guan KL, Dixon JE. Eukaryotic proteins expressed in *Escherichia coli*: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase. Anal Biochem 1991;192:262–7.
[42] Sun XL, Antony AC. Evidence that a specific interaction between an 18-base *cis*-element in the 5′-untranslated region of human folate receptor-alpha mRNA and a 46-kDa cytosolic *trans*-factor is critical for translation. J Biol Chem 1996;271:25539–47.
[43] Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning, a Laboratory Manual. Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press, 1989.
[44] Bourgeois CF, Popielarz M, Hildwein G, Stevenin J. Identification of a bidirectional splicing enhancer: differential involvement of SR proteins in 5′ or 3′ splice site activation. Mol Cell Biol 1999;19:7347–56.
[45] Fields S. The two-hybrid system to detect protein-protein interactions. Methods 1993;5:116–24.
[46] Durfee T, Becherer K, Chen PL, et al. The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. Genes Dev 1993;7:555–69.
[47] Xu P, Narasimhan ML, Samson T, et al. A nitrilase-like protein interacts with GCC box DNA-binding proteins involved in ethylene and defense responses. Plant Physiol 1998;118:867–74.
[48] Frederickson RM. Macromolecular matchmaking: advances in two-hybrid and related technologies. Curr Opin Biotechnol 1998;9:90–6.
[49] Kozak M. The scanning model for translation: an update. J Cell Biol 1989;108:229–41.
[50] Weidenhammer EM, Ruiz-Noriega M, Woolford JL, Jr. Prp31p promotes the association of the U4/U6 × U5 tri-snRNP with prespliceosomes to form spliceosomes in *Saccharomyces cerevisiae*. Mol Cell Biol 1997;17:3580–8.
[51] Silva E, Ullu E, Kobayashi R, Tschudi C. Trypanosome capping enzymes display a novel two-domain structure. Mol Cell Biol 1998;18:4612–9.
[52] Mair G, Shi H, Li H, et al. A new twist in trypanosome RNA metabolism: *cis*-splicing of pre-mRNA. RNA 2000;6:163–9.
[53] Burd CG, Dreyfuss G. Conserved structures and diversity of functions of RNA-binding proteins. Science 1994;265:615–21.
[54] Bruzik JP, Steitz JA. Spliced leader RNA sequences can substitute for the essential 5′ end of U1 RNA during splicing in a mammalian in vitro system. Cell 1990;62:889–99.
[55] Palfi Z, Xu G-L, Bindereif A. Spliced leader-associated RNA of trypanosomes. Sequence conservation and association with protein components common to *trans*-spliceosomal ribonucleoproteins. J Biol Chem 1994;269:30620–5.
[56] Bonaldo MF, Lennon G, Soares MB. Normalization and subtraction: two approaches to facilitate gene discovery. Genome Res 1996;6:791–806.