Cloning and molecular characterization of *Trypanosoma cruzi* U2, U4, U5, and U6 small nuclear RNAs

DL Ambrósio, MTA Silva, RMB Cicarelli+

Departamento de Ciências Biológicas, Faculdade de Ciências Farmacêuticas, Universidade Estadual Paulista, Rod. Araraquara-Jaú km 1, 14801-902 Araraquara, SP, Brasil

Small nuclear RNAs (snRNAs) are important factors in the functioning of eukaryotic cells that form several small complexes with proteins; these ribonucleoprotein particles (U snRNPs) have an essential role in the pre-mRNA processing, particularly in splicing, catalyzed by spliceosomes, large RNA-protein complexes composed of various snRNPs. Even though they are well defined in mammals, snRNPs are still not totally characterized in certain trypanosomatids as Trypanosoma cruzi*. For this reason we subjected snRNAs (U2, U4, U5, and U6) from* T. cruzi *epimastigotes to molecular characterization by polymerase chain reaction (PCR) and reverse transcription-PCR. These amplified sequences were cloned, sequenced, and compared with those other of trypanosomatids. Among these snRNAs, U5 was less conserved and U6 the most conserved. Their respective secondary structures were predicted and compared with known* T. brucei *structures. In addition, the copy number of each snRNA in the* T. cruzi *genome was characterized by Southern blotting.*

Key words: *Trypanosoma cruzi* - ribonucleoprotein - small nuclear RNP - small nuclear RNA - spliceosome - trans-splicing

In trypanosomes, all mRNAs are processed by trans-splicing, in which a common spliced leader sequence (SL) is acquired at the 5'-end to yield mature transcripts (Agabian 1990). SL trans-splicing has been characterized mainly in the trypanosomes and nematodes and requires, in addition to the SL RNP, small nuclear ribonucleoproteins (U snRNPs) U2, U4/U6, and U5 (Tschudi & Ullu 1990). Recently, intervening sequences were described in the *PAP* gene of *Trypanosoma brucei* and *T. cruzi* and a U1 snRNA sequence was described in *T. brucei*, demonstrating that both cis and trans-splicing can occur in these organisms, with a prevalence of trans (Schnare & Gray 1999, Mair et al. 2000). Hannon et al. (1992) showed that the U1 snRNP is not essential for trans-splicing, but only cis-splicing, as in nematodes.

The ribonucleoproteins are complexes that consist of small uridine-rich RNAs (UsnRNAs), interacting with common and specific proteins for each snRNP. In addition these small RNAs possess a uridine-rich Sm site for protein binding, composed of two conserved regions named Sm1 and Sm2, separated by a non-conserved region (Hermann et al. 1995, Séraphin 1995). Seven common proteins of *T. brucei* snRNP (Sm proteins) have been identified. These include: Sm B, -D1, -D2, -D3, -E, -F, and -G (Palfi et al. 2000).

In most eukaryotic organisms, all snRNAs genes are transcribed by RNA polymerase II to produce primary transcripts with a 7-methylguanosine (m^7G) cap at their 5' end and short extensions at the 3' end. The precursors are exported from the nucleus to the cytoplasm where they assemble into a stable core ribonucleoprotein particle (RNP) through binding to common proteins and the m^7G cap is hypermethylated to 2,2,7-trimethylguanosine (m_3G or TMG), as essential maturation step. The complex, consisting of the m_3G cap and the common proteins, signals reimportation (nuclear reimport) of uridine-rich snRNPs (U snRNPs) into the nucleus (Günzl et al. 2000). The exception is U6 snRNA, which is synthesized by RNA polymerase III and acquires a γ-Me cap (Baserga & Steitz 1993). In trypanosomatids, the U snRNAs are transcribed by RNA polymerase II and they receive an m_3G cap at the 5' end, except U6 snRNA. The 5' cap protects mRNAs from 5' exoribonucleases and also has an important role in mRNA translation and processing and therefore in the splicing process (O'Mullane & Eperon 1998, Marchetti et al. 1998, Lodish et al. 2000). All U snRNA gene transcription depends on the class 2 promoter (box A and box B elements) of an upstream tRNA gene that is oriented in opposite direction, as observed for *T. brucei* U2 and U6 snRNAs (Nakaar et al. 1994,1997, Gunzl et al. 1995, Tschudi & Ullu 2002).

In this context we proposed the molecular characterization of the *T. cruzi* snRNAs (U2, U4, U5, and U6) by polymerase chain reaction (PCR) and reverse transcription-PCR. The corresponding amplified sequences were cloned, sequenced, and compared with those other from trypanosomatids, using bioinformatic tools. Parasite sequences were also compared with those from *Homo sapiens*, to analyse conserved and non-conserved regions. Additionally, Southern blotting studies revealed the copy number of each snRNA in the *T. cruzi* (Y strain) genome.

Financial support: Fundunesp (850/03), Fapesp (99/11393-4 and 04/01630-9), PADC/FCF-Unesp (2004/01-I), Capes
+Corresponding author: cicarell@fcfar.unesp.br
Received 18 May 2006
Accepted 23 January 2007

MATERIALS AND METHODS

Culture of T. cruzi, extraction of genomic DNA and total RNA - *T. cruzi* epimastigotes (Y strain - Silva & Nussenzweig 1953) were grown at 28ºC in LIT (liver infusion tryptose) medium (Fernandes & Castellani 1966), supplemented with 10% fetal bovine serum. The genomic DNA and the total RNA were extracted with *DNAzol* and *Trizol* reagents (Invitrogen Corporation, Carlsbad, CA, US), respectively, following the manufacturer's instructions.

Immunoprecipitation with anti-m_3G antibody - Immunoprecipitation assay was performed with *T. cruzi* total RNA (Y strain epimastigotes) and rabbit polyclonal anti-m_3G serum, following the method proposed by Steitz (1989).

cDNA and genomic cloning of T. cruzi *snRNAs* - The following primers were used in PCR and RT-PCR experiments: TbU2 dir (5'- ATATCTTCTCGGCT –3'), complementary to nucleotides 1 to 14 of *T. brucei*; TcU4 dir (5'- AAGCCTTGCGCAGNGAG -3'), complementary to nts 1 to 15 of *T. cruzi*; TcU4 inv (5'- GCGGTA GGTGTCAAATATTC -3'), complementary to nts 103 to 84 of *T. cruzi*; TcU5 dir (5'- GCATCATCATTTCGACTT -3'), complementary to nts 1 to 18 of *T. cruzi*; TcU5 inv (5'- GTTTAAATTGTTTGGGC -3'), complementary to nts 50 to 34 of *T. cruzi*; TcU6 dir (5'- GTCAAGCGAAGG ACATC -3'), complementary to nts 3 to 19 of *T. cruzi*.

The snRNAs are noncoding molecules and thus have no poly-A tail; for this reason, it was necessary to add such a tail, by incubation for for 30 min at 37ºC in buffer (50 mM Tris-HCl pH 7.5; 250 mM NaCl; 10 mM $MgCl_2$; 2.5 mM $MnCl_2$; 0.25 mM ATP; 0.5 µg/µl BSA) containing 6.9 U Poly-A polymerase (Invitrogen) plus 4 ng/µl RNA before using the *3'RACE Kit for Rapid Amplification of cDNA Ends* (Invitrogen) to perform the RT-PCR. *T. cruzi* U2 and U6 snRNAs were amplified by RT-PCR from immunoprecipitated RNA template, using the forward primers, TbU2 and TcU6, respectively.

U4 and U5 snRNAs were also amplified by PCR, from *T. cruzi* genomic DNA, using TcU4 and TcU5 primers, respectively.

The RT-PCR and PCR products were cloned using the *TOPO TA Cloning for Sequencing kit* (pCR®4-TOPO® vector - Invitrogen), transformed in $CaCl_2$-competent *E. coli* DH5α and the plasmid DNA from positive colonies extracted and sequenced in *ABI PRISM 377 DNA Sequencer* (Perkin Elmer, Wellesley, MA, US), using *Big Dye Terminator* (Applied Biosystem, Foster City, CA, US).

Sequence analysis - The alignments were carried out with *GeneDoc* (Nicholas & Nicholas 1997) and RNA secondary structures at 37ºC were predicted with the *mfold* program version 3.1 (http://www.bioinfo.rpi.edu/applications/mfold/old/rna/). Sequences for various trypanosomatids and human snRNAs and their accession numbers in the NCBI-GenBank database are listed in Table I. The extra adenosine residues of the poly-(A) tail added to U2 sn RNA sequence were withdrawn after detection by comparison with the sequences present in the database of the *T. cruzi* and *T. brucei* genome projects.

Southern blotting - *T. cruzi* (Y strain) genomic DNA (25 µg) was digested separately with the restriction enzymes *Sau3A* I and *Hae* III, separately. These enzymes do not cut the sequences of any of the snRNAs analyzed, except that *Sau3A* I digests the U2 sequence at one site, resulting in two bands for one copy gene. After 0.7% agarose gel, the DNA was transferred to a nylon membrane that it was blocked with ULTRAHyb reagent (Ambion, Applied Biosystem, Austin, TX, US), hybridized with a probe containing the U2 and U4 snRNA sequences labeled by PCR with α-P^{32}-dATP (Amersham, GE Healthcare, Buckinghamshire, Norway). For U5 and U6 snRNA probes, the DNA was transferred onto Nylon-N+ membrane (Amersham), hybridized with a labeled PCR product containing these sequences and detected by chemiluminescence, using the *ECL Direct™ Nucleic Acid Labeling and Detection System* (Amersham).

RESULTS

PCR amplification, cloning and sequencing of T. cruzi U2, U4, U5, and U6 snRNAs - U2 and U6 snRNAs were amplified by RT-PCR as described in Materials and Methods. It was possible to amplify the U6 sequence, which lacks the m_3G 5'cap, probably because this snRNA is associated with U4 snRNA. The U5 and U4 snRNAs sequences were not obtained by this approach, so they were amplified by PCR from *T. cruzi* genomic DNA, using gene-specific primers as described in Materials

TABLE I

Accession numbers of snRNA sequences of various organisms in the NCBI-Genbank database or manuscript used for alignments

Organisms	U2 snRNA	U4 snRNA	U5 snRNA	U6 snRNA
Trypanosoma cruzi	AY205287	AY894522	AY894523	AY894521
Trypanosoma brucei	X04678	Mottram et al. 1989	AJ243568	X13017
Trypanosoma congolense	M58666	-	-	-
Leishmania amazonensis	M58665	-	-	-
Leishmania tarentolae	AY007788	AY007789	-	AF305715
Leishmania mexicana	-	-	-	X82228
Leptomonas seymouri	U23406	AJ245951	AJ243569	X78552
Leptomonas collosoma	X56455	AF204671	AF006632	X79014
Crithidia fasciculata	AF326335	AF326336	AF182356	X78551
Phytomonas sp.	-	-	-	X82229
Homo sapiens	X59360	X59361	X04215	X59362

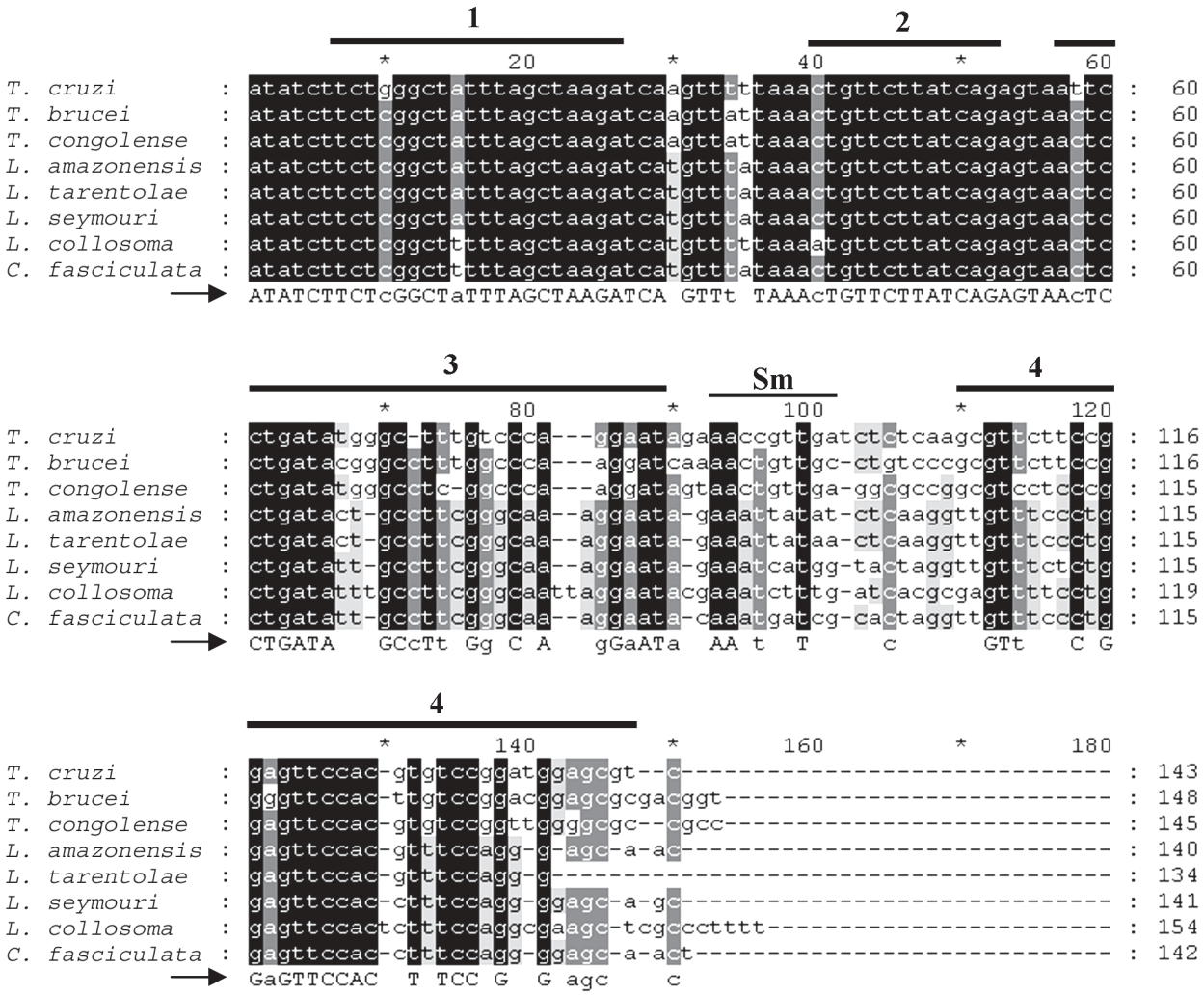

Fig. 1a: *Trypanosoma cruzi* U2 snRNA multiple alignment with other trypanosomatid U2 snRNA sequences. The arrows indicate the consensus sequence; the capital letters indicate identity between all the sequences (black boxes) and the small letters identity between some sequences (gray boxes). The Sm site is indicated by a thin bar, the loops by thick bars and numbers.

and Methods. The corresponding PCR products, 143 bp - U2, 102 bp - U4, 50 bp - U5, and 101 bp - U6, were cloned and sequenced.

Sequence analysis-alignment and secondary structure - The sequences of *T. cruzi* U2, U4, U6, and U5 snRNAs were aligned with those other from trypanosomatids, to identify the percentage of identity among these sequences, as shown in Table II and Figs. 1a, 2a, 2b, and 3a. On the whole, the U6 snRNA sequence was highly conserved across all the trypanosomatids, unlike U2, U4, and U5 snRNAs, the U5 snRNA being the least conserved among the four snRNAs. The secondary structures of U2, U4/U6, and U5 snRNAs (Figs 1b, 2c, and 3b, respectively) exhibited 4, 3/2 and 1 loops, respectively. The sequence of U4 and U6 snRNAs can be fold into a canonical U4/U6 base-paired secondary structure (Fig. 2c).

To reveal any differences between the snRNA sequences of cis-splicing *H. sapiens* and predominantly trans-splicing *T. cruzi*, these sequences were aligned. The degrees of identity in the U2, U4, U5, and U6 snRNAs were 50, 48, 19, and 60%, respectively. Again, it was observed that the U5 snRNA was the least conserved and the U6 snRNA the most (data not shown).

The *T. cruzi* U2 snRNA sequence was highly conserved in the first 65 nts from the 5' end in all organisms analyzed (Fig. 1a), probably because this region of loops 1 to 2 contains the branch point that is important to the trans-splicing mechanism. The *T. cruzi* U2 secondary structure (Fig. 1b) has four loops at the same sites as in *T. brucei*, with a few modifications in the primary structure (Fig. 1a) that are reflected only in the lenght and position of the loops, as observed in the Fig. 1 B, C, and D. These modifications were discussed by Hartshorne and Agabian (1990) and do not alter the general structure of *T. cruzi* U2 snRNA greatly compared to the *T. brucei* structure.

Multiple alignment of *T. cruzi* U4 snRNA with other trypanosomatids (Fig. 2a) showed a high degree of sequence conservation at several points. The secondary structure for *T. cruzi* U4 snRNA (Fig. 2c) revealed three

TABLE II
Percentage of identity of *Trypanosoma cruzi* U2, U4, U5. and U6 snRNAs with other organisms

	T. cruzi snRNAs			
Organisms	U2 snRNA	U4 snRNA	U5 snRNA	U6 snRNA
Trypanosoma brucei	83	87	58	90
Trypanosoma congolense	82	-	-	-
Leishmania amazonensis	72	-	-	-
Leishmania tarentolae	71	77	-	91
Leishmania mexicana	-	-	-	91
Leptomonas seymouri	61	78	50	92
Leptomonas collosoma	69	77	48	90
Crithidia fasciculata	72	77	50	89
Phytomonas sp.	-	-	-	87

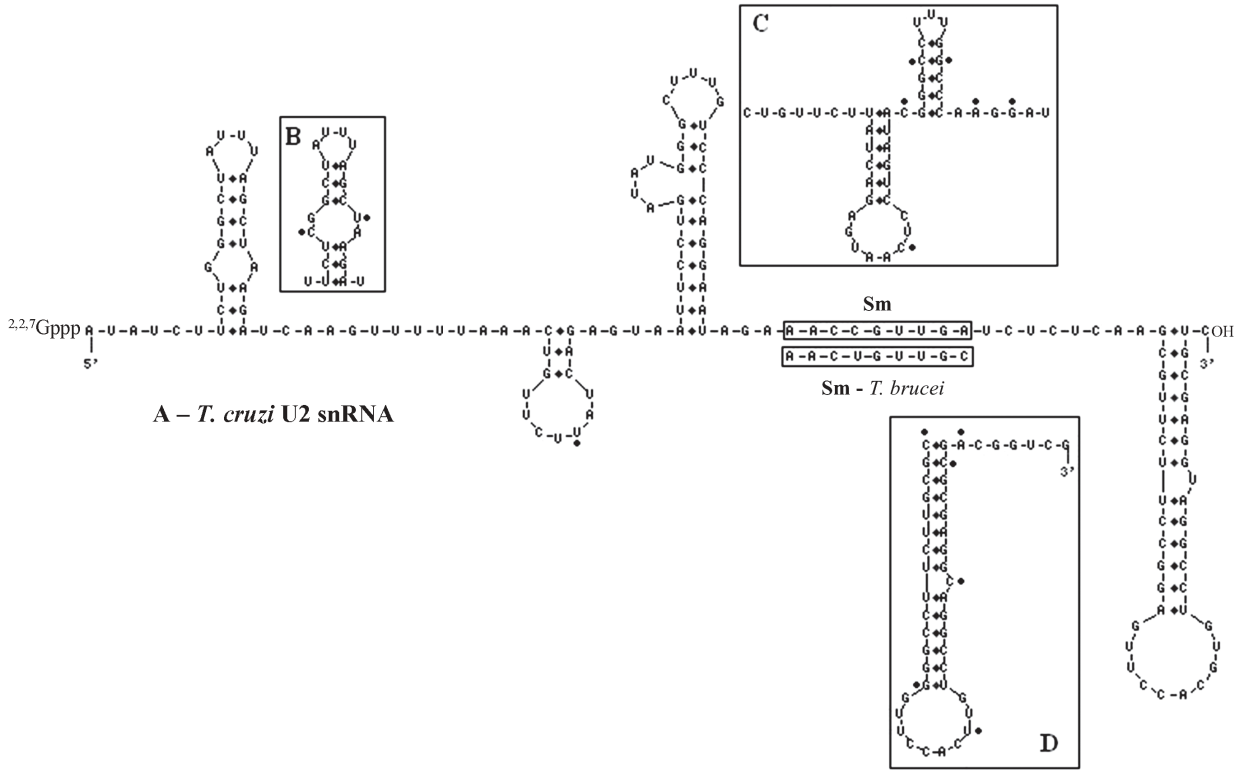

Fig. 1b: predicted secondary structure (panel A). Beside each of the loops (1, 2, 3,and 4), the panels B, C and D show the corresponding loops of *Trypanosoma brucei*. The dots show base changes between *T. cruzi* and *T. brucei*. Boxes indicate the Sm site in *T. cruzi* and *T. brucei*.

loops (1, 2, and 3) as observed in the *T. brucei* structure (Mottram et al. 1989) and *H. sapiens* (Baserga & Steitz 1993). Differences were observed between the sequences of the three U4 loops of *T. cruzi* and *T. brucei*. Loop 1 of U4/U6 snRNA has only one nucleotide change compared to *T. brucei* (Fig. 2a, indicated by arrow), but this change does not modify the loop structure (G-U) in Fig. 2c (base boxed). Fig. 2c (panel B) shows that the replacement of adenine in *T. brucei* by uridine in *T. cruzi* (indicated by square dots) modifies loop 2, stabilizing four new nucleotides in this structure (U-A and U-A). In loop 3, the change of uridine in *T. brucei* to citosine in *T. cruzi* (Fig. 2c, panel C, indicated by square dots) should destabilize the loop at its beginning. The other nucleotides that differ between these organisms do not change the loop structure (Fig. 2c, indicated by round dots). The *T. cruzi* Sm site is also conserved (Fig. 2c), with differences only in the first and last nucleotide of this site, compared to *T. brucei*, and in the last nucleotide, versus *L. collosoma*.

T. cruzi U6 snRNA sequence was highly conserved among all the trypanosomatids analyzed (Fig. 2b) and had a conserved central domain (ACAGAG hexanucleotide), indicated in Fig. 2b, which seems to be essential during both the first and the second step of splicing catalysis, and a conserved U4/U6 interaction region (stems 1 and 2) indicating that this trans-spliceosomal U6 snRNA conforms to the cis-spliceosomal U6 consensus se-

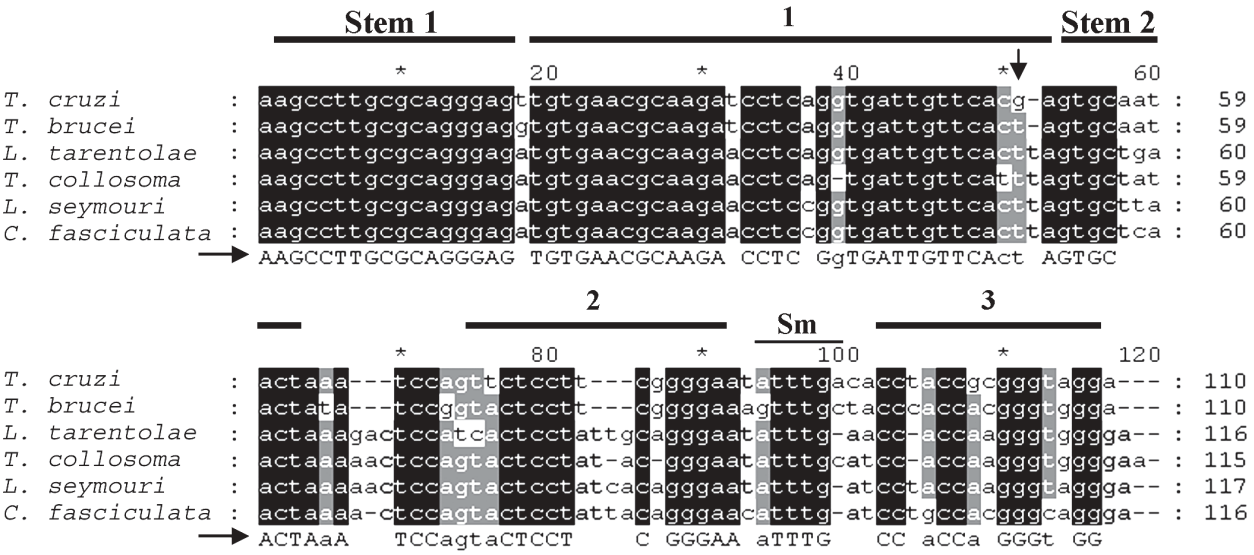

Fig. 2a: *Trypanosoma cruzi* U4 snRNA multiple alignment with other trypanosomatid U4 snRNA sequences. The arrows indicate the consensus sequence; the capital letters indicate identity between all the sequences (black boxes) and the small letters identity between some sequences (gray boxes). The vertical arrow shows a base change. The Sm site is indicated by a thin bar and the loops by thick bars and numbers. Stems 1 and 2 are the regions of interaction with U6 snRNA.

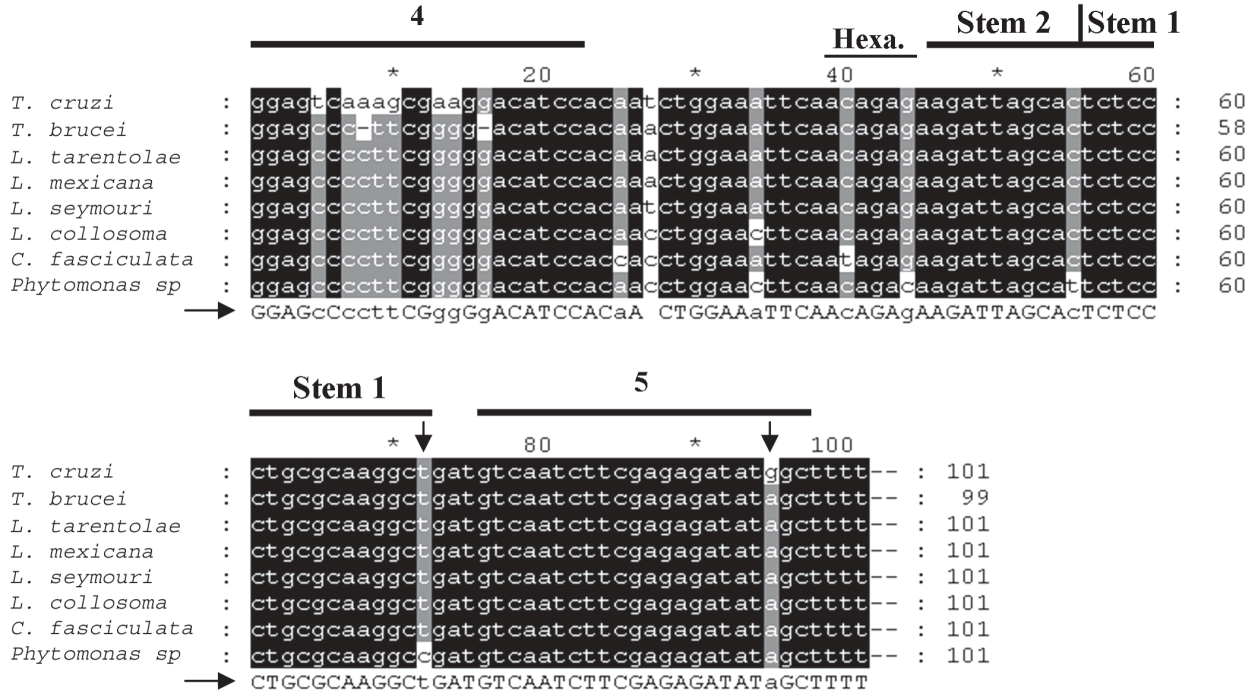

Fig. 2b: *Trypanosoma cruzi* U6 snRNA multiple sequence alignment with other trypanosomatid U6 snRNA sequences. The arrows indicate the consensus sequence; the capital letters indicate identity between all the sequences (black boxes) and the small letters the identity between some sequences (gray boxes). The vertical arrows show the base changes. Thick bars and numbers indicate the loops. Stems 1 and 2 are the regions of interaction with U4 snRNA. Hexa is the consensus hexanucleotide sequence.

quence (Guthrie & Patterson 1988). Genetic studies in yeast reveal that the first three nucleotides of the ACAGAG element contact the 5' splice site region (Kandels-Lewis & Séraphin 1993), while the last nucleotide is engaged in a tertiary interaction with U2 (Madhani & Guthrie 1994). The secondary structure has two loops (Fig. 2c - one at the 3' end and another at the 5' end), as observed in *T. brucei* (Mottram et al. 1989). Although highly conserved among the trypanosomatids, the U6 sequence was less conserved at the 5' end (Fig. 2b), with a 5'-AAAG-3' sequence that did not appear in the other trypanosomatids. This variation results in dif-

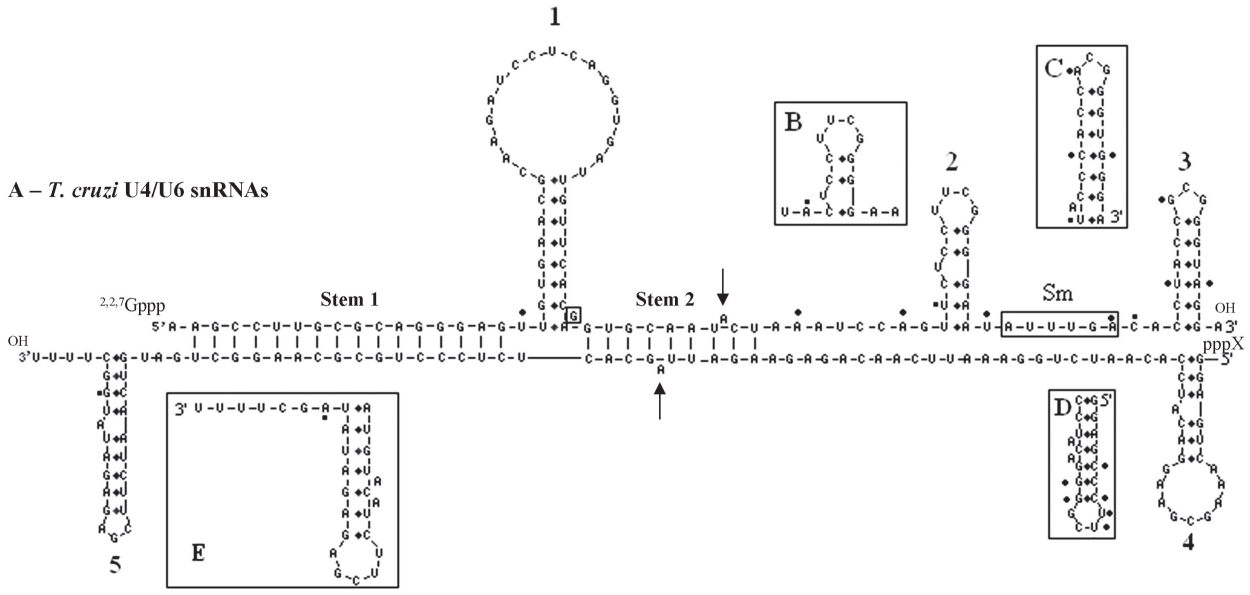

Fig. 2c: *Trypanosoma cruzi* U4/U6 snRNAs predicted secondary structure (panel A). The panels B, C, D and E show the loops of *T. brucei* that correspond to loops 2, 3, 4, and 5, respectively. The round dots show base changes between *T. cruzi* and *T. brucei*. The square dots indicate base changes between *T. cruzi* and *T. brucei* that result in structural modifications. Boxes indicate the Sm site and base changes. Stems 1 and 2 are the regions of interaction between U4 and U6 snRNAs.

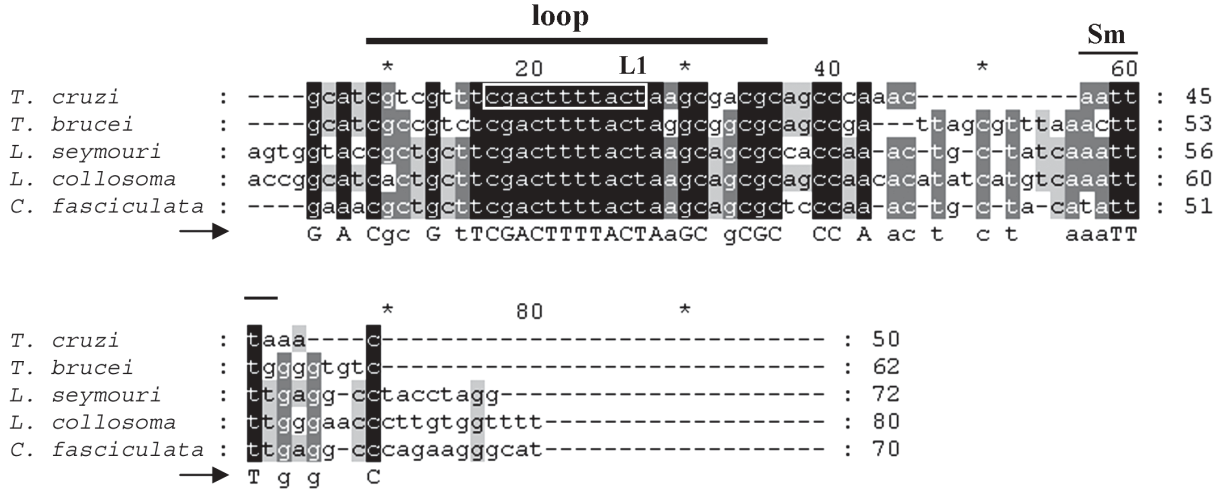

Fig. 3a: *Trypanosoma cruzi* U5 snRNA multiple alignment with other trypanosomatid U5 snRNA sequences. The arrows indicate the consensus sequence; the capital letters indicate identity between all the sequences (black boxes) and the small letters identity between some sequences (gray boxes). The thick bar indicates the single loop and the thin bar the Sm site. The box shows the 11-nt conserved sequence loop (L1).

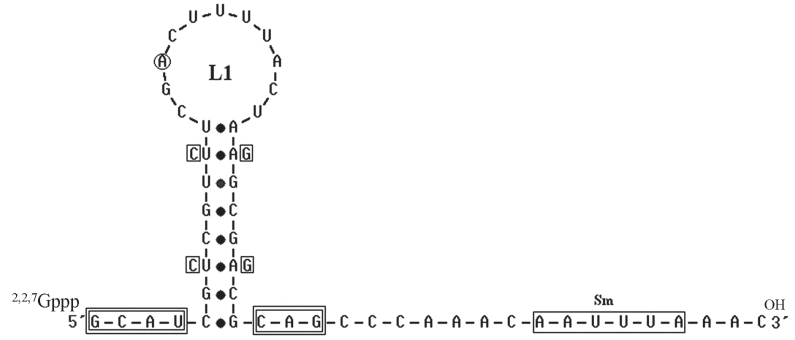

Fig. 3b: predicted secondary structure. Boxes indicate the Sm site and base changes. The L1 sequence is indicated with one base change (circle). Double-lined boxes show the probable sequences for binding U5-specific proteins.

ferences in the structure of the loop 4, compared to *T. brucei*, as demonstrated in Fig. 2c (panel D), which may be related to a larger open loop, formed out of *T. cruzi* nine nucleotides (only three in *T. brucei*). Another structural difference observed in loop 4 is in the region of paired bases, where *T. cruzi* has six nucleotides paired and *T. brucei* has seven (Fig. 2c, panel D).

The 3' end of *T. cruzi* U6 snRNA has highly conserved sequence conservation, with the fifth loop showing nine interacting bases as compared to seven in *T. brucei*. A guanosine is present only in *T. cruzi* (Fig. 2b, indicated by arrow and Fig. 2c, panel E, indicated by square dots) that stabilizes the additional base pairs.

The proposed structure has two regions of interaction between U4 and U6, separated by U4 loop 1 which can be folded into the canonical U4/U6 base-paired secondary structure (Fig. 2c). Stem 1 occurs on the intervals of nucleotides 2-17 of U4 and 57-70 of U6, highly conserved among the trypanosomatids, showing only one nucleotide change in U6, between *T. cruzi* (position 70) e *Phytomonas* sp. (Fig. 2b, indicated by arrow). Stem 2 is found between positions 53-62 in the U4 sequence and positions 46-55 in the U6; however the adenosines at positions 60 of U4 and 51 of U6 do not participate in the interaction (Fig. 2c, indicated by arrow), and this is also observed in *T. brucei*. The *T. cruzi* U4/U6 base-paired structure is conserved in other organisms as *T. brucei* (Mottram et al. 1989) and mammals (Baserga & Steitz 1993), confirming the presence of the U4 loop between two U4/U6 interaction sites, probably related to catalyzing the second transesterification step of the splicing process.

Among the spliceosomal snRNAs, U5 snRNA is the least conserved at the primary sequence level and the *T. cruzi* U5 is smaller than the other trypanosomatid U5 snRNAs, with only 50 nt, while *T. brucei*, *Leptomonas seymouri*, *L. collosoma,* and *Crithidia fasciculata* U5 snRNAs have 62, 72, 80, and 70 nt, respectively. The Sm site, however, was found to be highly conserved, with just a few modifications (Fig. 3a). It can be seen that *T. cruzi* lacks 11 nucleotides before the Sm site. The predicted *T. cruzi* U5 snRNA secondary structure (Fig. 3b) is similar with that predicted for *T. brucei* (Dungan et al. 1996). The sequence and structure of the 11-nt loop (L1) is conserved in *T. cruzi* (Fig. 3b), with the exception of the adenosine (circled in Fig. 3b), relative to the L1 consensus sequence observed in *T. brucei*, *L. collosoma,* and *L. seymouri*, which has been shown to interact directly with the 5' and 3' splice sites during both steps of splicing (Dungan *et al*. 1996, Bell & Bindereif 1999). Other features include nucleotides 1-4 (5'-GCAU-3') and 32-34 (5'-CAG-3') (double line boxes in Fig. 3b), which are similar in sequence and location to the conserved IL1 sequences implicated in binding U5-specific proteins. The loop is closed by a stem of eight bp, with two similar compensatory changes, on the second and fifth nucleotides (C-G → A-U), compared to the *T. brucei* sequence (Fig. 3b, indicated by boxes). The single loop of *T. cruzi* U5 snRNA is conserved among the other trypanosomatids, but *L. collosoma* (Xu et al. 1994), *C. fasciculata* (Schnare & Gray 2000) and *H. sapiens* (Baserga & Steitz 1993) show an additional loop at the 3' end.

T. cruzi x H. sapiens snRNAs - A comparison between the sequences of human (cis-splicing) and *T. cruzi* (trans-splicing) snRNAs reveals an additional 38 nucelotides in the human U2 snRNA (data not shown). Although the secondary structure in *H. sapiens* seems to be longer (Lürhmann et al. 1990) and possesses an additional loop, the other loops of the secondary structure are in similar positions to those in *T. cruzi* U2 snRNA. *H. sapiens* U4 snRNA has 42 nucleotides more than the *T. cruzi* ho-

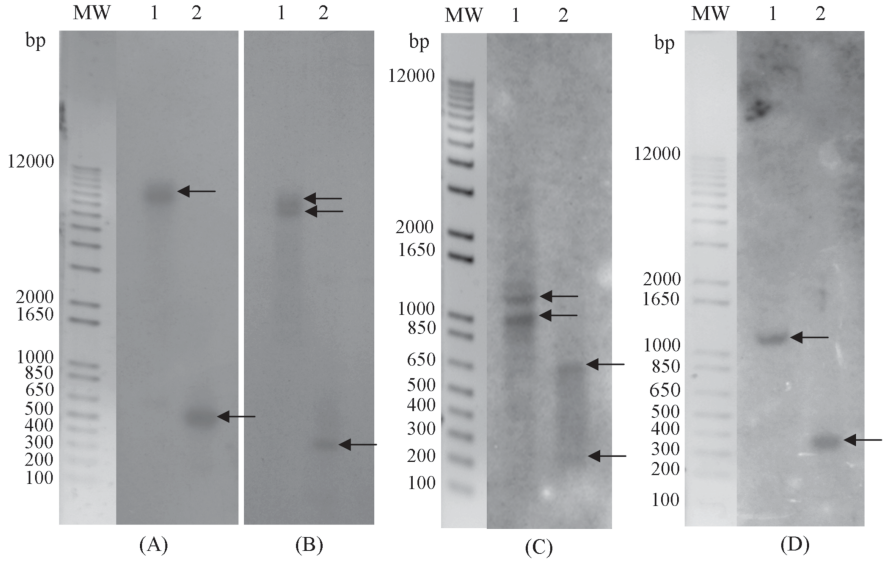

Fig. 4: *Trypanosoma cruzi* snRNAs Southern blot analysis. *T. cruzi* (Y strain) genomic DNA was digested with *Sau3A* I (1) and *Hae* III (2), separated by electrophoresis in 0.7% agarose gel and transferred to nylon membrane. The blot was probed with U2 (B), U4 (A), U5 (C) and U6 (D) *T. cruzi* snRNA sequences. Molecular weight marker (MW) 1 Kb plus DNA ladder (Invitrogen). The arrows indicate positive bands.

molog and its secondary structure is the most conserved, with three loops, complexed with U6 snRNA (Baserga & Steitz 1993), similar to the *T. cruzi* structure in this study. Interestingly, this snRNA was the only one that exhibited a conserved Sm site in the same position, in both *H. sapiens* (Lürhmann et al. 1990) and *T. cruzi*. The U6 snRNA seemed to be the most conserved spliceosomal U snRNA, both in length (101 nt - *T. cruzi* and 107 nt - *H. sapiens*) and sequence (data not shown), including the conserved hexanucleotide. The secondary structure of human U6 snRNA was similar to that of *T. cruzi* and it complexed in the same manner with U4, but possessed only one loop. The U5 snRNAs, besides being the least conserved snRNA between *H. sapiens* and *T. cruzi*, had an extended region of 63 nts, exclusive to the 5' end of the human sequence (data not shown); however, the L1 sequence was conserved in both organisms. As mentioned before, the human U5 snRNA has an additional loop at the 3' end (Baserga & Steitz 1993).

Southern blot analysis of T. cruzi snRNAs - To examine the genomic organization of the spliceosomal snRNAs, *T. cruzi* genomic DNA was subjected to Southern blotting (Fig. 4). *T. cruzi* U2, U4, and U6 snRNAs produced a single band (Fig. 4, panels B, A, and D, respectively), except for the U2 DNA digested with *Sau3A* I, which gave 2 bands (Fig. 4, panel B, lane 1), suggesting only one copy, as explained in Materials and Methods. U5 snRNA exhibited two bands in the DNA samples digested with either *Sau3A* I and *Hae* III (Fig. 4, panel C).

Analyzing the copy numbers of *T. cruzi* snRNAs, U2 snRNA yielded a single band (Fig. 4, panel B), suggesting a single-copy gene, in contrast to the results obtained by Hartshorne and Agabian (1990), which suggested additional copies. Their observation could result from using a different *T. cruzi* strain and the U2 sequence of *T. brucei* as a probe. Among other trypanosomatids, *T. brucei* and *L. seymouri* showed one copy of U2 snRNA in the genome and *L. collosoma* three copies (Hartshorne & Agabian 1990, He & Bellofatto 1995). Sequence analysis using the Vector NTI software (Invitrogen) revealed that the *T. cruzi* U2 sequence in this study has two *Sau3A* I restriction sites and the annotated sequence of the *T. cruzi* genome (CL Brener strain) contained the same restriction sites; the latter sequence for U2 has one *Hae* III-site and is one-copy gene (data not shown).

A single hybridization band was observed to U4 snRNA gene (Fig. 4, panel A), suggesting that it is a single copy, as for *L. collosoma* (Li et al. 2000). Amongst the analyzed sequences, only *T. cruzi* U5 snRNA showed two copies in the genome (Fig. 4, panel C). U6 snRNA gene is present in a single copy (Fig. 4, panel D), as similarly shown for *T. brucei* (Mottram *et al*. 1989), *L. seymouri* (Xu et al. 1994), *L. collosoma* (Goldring & Michaeli 1995), *Phytomonas* sp. and *Leishmania mexicana* (Wieland & Bindereif 1995).

DISCUSSION

The data presented in this paper result from an extensive study of *T. cruzi* U snRNA sequences, which are compared with those of other trypanosomatids and also *H. sapiens*. With the exception *T. cruzi* U5, all the snRNAs studied exhibited a higher degree of homology with *T. brucei* sequences than with those of other species. Since the *T. cruzi* genome project has yet to be completed, these data can contribute towards the annotation of these genes.

Analysis of the annotated sequences from the *T. cruzi* genome project (El-Sayed et al. 2005) reveals that the U4 and U6 snRNAs are present in two contigs, but they are the same genes, indicating a single-copy gene for these sequences. The U2 snRNA showed modifications in the primary structure of the nine contigs in which it appeared; however, it was not possible to decide the actual number of copies from our bioinformatics analyses. U5 snRNA was found in two contigs, which showed to be similar but not identical genes, indicating two copies for this sequence. Hence, the Southern blotting results provide validation of the data from the *T. cruzi* genome project.

Despite the fact that the U4 and U5 sequences in this study had to be amplified from the *T. cruzi* genomic DNA, the evidence of their percentage of identity above 55% and highly conserved secondary structures allowed their confirmation as these snRNAs. The partially annotated genome sequence of *T. cruzi* has recently published (El-Sayed et al. 2005). This study provides a more detailed analysis of U2, U4, U5, and U6 snRNA sequences, confirming the findings of the genome project, mainly for U5 snRNA, which was annotated as a putative sequence.

Taken together, these findings represent a step towards understanding how the *T. cruzi* snRNA/RNPs participating in the trans-splicing reaction. Further studies are currently in progress in this laboratory to characterize the involvement of these ribonucleoproteins in the mRNA processing of these parasites.

ACKNOWLEDGEMENTS

To Dr Montserrat Bach-Elias (Consejo Superior de Investigaciones Cientificas, Barcelona, Spain) for donating the anti-m3G polyclonal antibody. To Dr Benson Nyambega (International Center of Insect Physiology and Ecology, Nairobi, Kenya) for reading the manuscript.

REFERENCES

Agabian N 1990. Trans-splicing of nuclear pre-mRNAs. *Cell 61*: 1157-1160.

Baserga SJ, Steitz JA 1993. The diverse world of small ribonucleoproteins. In RF Gesteland, JF Atkins (eds), *The RNA World*, Cold Spring Laboratory Press, New York, 361 pp.

Bell M, 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. *Nucl Acid Res 27*: 3986-3994.

Dungan JA, Watkins KP, Agabian N 1996. Evidence for the presence of a small U5-like RNA in active trans-spliceosomes of *Trypanosoma brucei*. *EMBO J 15*: 4016-4029.

El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G, Westenberger SJ, Caler E, Cerqueira GC, Branche C, Haas B, Anupama A, Arner E, Aslund L, Attipoe P, Bontempi E,

Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A, Mizuno Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U, Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J, Simpson AJ, Sisk E, Tammi MT, Tarleton R, Teixeira S, Van Aken S, Vogt C, Ward PN, Wickstead B, Wortman J, White O, Fraser CM, Stuart KD, Andersson B 2005. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science 309*: 409-415.

Fernandes JF, Castellani O 1966. Growth characteristics and chemical composition of *Trypanosoma cruzi*. *Exp Parasitol 18*: 195-202.

Goldring A, Michaeli S 1995. The U6 snRNA-encoding gene of the monogenetic trypanosomatid *Leptomonas collosoma*. *Gene 156*: 139-144.

Günzl A, Bindereif A, Ullu E, Tschudi C 2000. Determinants for cap trimethylation of the U2 small nuclear RNA are not conserved between *Trypanosoma brucei* and higher eukaryotic organisms. *Nucl Acid Res 28*: 3702-3709.

Günzl A, Tschudi C, Nakaar V, Ullu E 1995. Accurate transcription of the *Trypanosoma brucei* U2 small nuclear RNA gene in homologous extract. *J Biol Chem 270*: 17287-17291.

Guthrie C, Patterson B 1988. Spliceosomal snRNAs. *Annu Rev Genet 22*: 387-419.

Hannon GJ, Maroney PA, Yu YT, Hannon GE, Nilsen TW 1992. Interaction of U6 snRNA with a sequence required for function of the nematode SL RNA in trans-splicing. *Science 258*: 1775-1780.

Hartshorne T, Agabian N 1990. A new U2 RNA secondary structure provided by phylogenetic analysis of trypanosomatid U2 RNAs. *Genes Dev 4*: 2121-2131.

He P, Bellofatto V 1995. Structure of the *Leptomonas seymouri* trans-spliceosomal U2 snRNA-encoding gene; potential U2-U6 snRNA interactions conform to the cis-splicing counterpart. *Gene 165*: 131-135.

Hermann H, Fabrizio P, Raker VA, Foulaki K, Hornig H, Brahms H, Lürhamann R 1995. snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein interactions. *EMBO J 14*: 2076-2088.

Kandels-Lewis A, Séraphin B 1993. Role of U6 RNA in 5' splice site selection. *Science 262*: 2035-2039.

Li L, Otake LR, Xu Y, Michaeli S 2000. The trans-spliceosomal U4 RNA from the monogenetic trypanosomatid *Leptomonas collosoma*. *J Biol Chem 275*: 2259-2264.

Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell JE 2000. *Molecular Cell Biology*, 4th ed., W. H. Freeman, New York, 115pp.

Lürhmann R, Kastner B, Bach M 1990. Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. *Bioch Bioph Acta 1087*: 165-292.

Madhani HD, Guthrie C 1994. Randomization-selection analysis of snRNAs in vivo: evidence for a tertiary interaction in the spliceosome. *Genes Dev 8*: 1071-1086.

Mair G, Shi H, Li H, Djikeng A, Aviles HO, Bishop JR, Falcone FH, Gavrilescu C, Montgomery JL, Santori MI, Stern LS, Wang Z, Ullu E, Tschudi C 2000. A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. *RNA 6*: 163-169.

Marchetti MA, Tschudi C, Silva E, Ullu E 1998. Physical and transcriptional analysis of the *Trypanosoma brucei* genome reveals a typical eukaryotic arrangement with close interspersion of RNA polymerase II and III-transcribed genes. *Nucl Acids Res 26*: 3591-3598.

Mottram J, Perry KL, Lizardi PM, Lürhmann R, Agabian N, Nelson RG 1989. 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 9*: 1212-1223.

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

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

Nicholas KB, Nicholas Jr HB 1997. GeneDoc: a tool for editing and annotating multiple sequence alignments. Available from URL: http://www.psc.edu.biomed/genedoc.

O'Mullane L, Eperon IC 1998. The pre-mRNA 5' cap determines whether U6 small nuclear RNA succeeds U1 small nuclear ribonucleoprotein particle at 5' splice site. *Mol Cell Biol 18*: 7510-7520.

Palfi Z, Lücke S, Lahm H, Lane WS, Kruft V, Bragado-Nilsson E, Séraphin B, 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.

Schnare MN, Gray MW 1999. A candidate U1 small nuclear RNA for Trypanosomatid Protozoa. *J Biol Chem 274*: 23691-23694.

Séraphin B 1995. Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J 14*: 2089-2098.

Silva LHP, Nussenzweig V 1953. Sobre uma cepa de *Trypanosoma cruzi* altamente virulenta para o camundongo branco. *Fol Clín Biol 20*: 191-208.

Steitz JA 1989. Immunoprecipitation of ribonucleoproteins using autoantibodies. *Methods Enzymol 180*: 468-481.

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

Tschudi C, Ullu E 2002. Unconventional rules of small nuclear RNA transcription and cap modification in trypanosomatids. *Gene Expression 10*: 3-16.

Wieland B, Bindereif A 1995. Unexpected diversity in U6 snRNA sequences from trypanosomatids. *Gene 161*: 129-133.

Xu GL, Wieland B, Bindereif A 1994. Trans-spliceosomal U6 snRNAs of *Crithidia fasciculata* and *Leptomonas seymouri*: deviation from the conserved ACAGAG sequence and potential base pairing with spliced leader RNA. *Mol Cell Biol 14*: 4565-4570.