Spliced leader-associated RNA from *Crithidia fasciculata* contains a structure resembling stem/loop II of U1 snRNA

Murray N. Schnare, Michael W. Gray*

Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, N.S. B3H 4H7, Canada

Received 16 August 1999; received in revised form 3 September 1999

Abstract In contrast to earlier proposals, recent evidence suggests that *trans*-spliceosomes in trypanosomatid protozoa may contain a homolog of U1 small nuclear (sn) RNA (Schnare, M.N. and Gray, M.W. (1999) J. Biol. Chem. 274, 23 691–23 694). However, the candidate trypanosomatid U1 snRNA is unconventional because it lacks the highly conserved stem/loop II present in all other U1 snRNAs. Trypanosomatids also possess a unique spliced leader-associated (SLA) RNA of unknown function. We present the complete sequence of the SLA RNA from *Crithidia fasciculata* and propose that it may contribute a U1 snRNA-like stem/loop II to the *trans*-spliceosome.

Key words: Trypanosomatid; Spliced leader-associated RNA; Spliced leader; *Trans*-splicing

1. Introduction

A common 39 nucleotide spliced leader (SL) sequence is located at the 5′-terminus of all mRNAs in trypanosomatid protozoa [1]. The SL sequence is transferred by *trans*-splicing from the 5′-end of a small SL RNA to the appropriate site in pre-mRNA, via a mechanism that involves small nuclear (sn) RNAs operating in a *trans*-spliceosome [1]. It was initially shown that *Trypanosoma brucei* contains homologs of the U2, U4 and U6 snRNAs that are components of the *cis*-spliceosome in other eukaryotes [2–4]. Later, it was established that these snRNAs are required for SL RNA *trans*-splicing [5]. Early studies did not detect homologs of the *cis*-spliceosomal U1 and U5 snRNAs, leading to speculation that *trans*-splicing may not require these snRNAs and that their functions may be replaced by sequences within the *trans*-spliceosomal SL RNA itself [4,6–8]. These two *cis*-spliceosome-specific snRNAs were considered dispensable because trypanosomatid protozoa do not appear to have *cis*-spliced introns [1,9].

Analysis of RNAs that were cross-linked with psoralen in vivo to *T. brucei* SL RNA led to the discovery of a SL-associated (SLA) RNA that was initially postulated to be an analog of U5 snRNA, based on the presence of a short sequence that is universally conserved among U5 snRNAs [10]. However, this particular sequence is not conserved among trypanosomatid SLA RNAs [11] and a highly divergent *trans*-spliceosomal U5 snRNA has recently been identified in both *T. brucei* [12,13] and *Leptomonas collosoma* [14]. This finding left open the possibility that SLA RNA may be a highly divergent analog of typical U1 snRNA [10,12]. Although there is no obvious sequence similarity between SLA RNA and U1 snRNAs from other eukaryotes [10], SLA RNA does contain a binding site for the common core-protein complex that is associated with other trypanosomatid snRNAs [11]. It has also been suggested that SLA RNA may be a *trans*-spliceosome-specific RNA having a function in post-transcriptional modification of trypanosomatid SL RNA [15].

We have recently identified a candidate U1 snRNA in a representative trypanosomatid, *Crithidia fasciculata* [16]. This RNA contains a cap and a core-protein binding site as well as a 5′-terminal sequence and stem/loop I structure that are conserved relative to other U1 snRNAs. However, this candidate U1 snRNA lacks a stem/loop II structure, which in typical U1 snRNAs is the binding site for the U1-A protein [17]. We present here the sequence of *C. fasciculata* SLA RNA and propose that it may contribute a U1-like stem/loop II structure to the *trans*-spliceosome.

2. Materials and methods

The following oligonucleotide primers were used in this study (lower case residues do not match the final sequence of the SLA RNA): SLA1 = 5′-CgCTCCAGTTTCgTGCACcG-3′ (complementary to positions 43–62), SLA2 = 5′-gCaCGCACTACAGTAAGGGC-3′ (complementary to positions 4–23), SLA3 = 5′-CAAACCACCTAGGCA-CgGTG-3′ (corresponding to positions 28–47), SLA4 = 5′-gtGTC-TCTCGCCCTCCAGTTTC-3′ (complementary to positions 51–70).

C. fasciculata RNA [18] and *T. brucei* RNA (a gift from M. McManus and S.L. Hajduk) were reverse-transcribed using AMV reverse transcriptase (RT) in the presence of 5′-end-labelled primer (either SLA1 or SLA4). Total RNA that was poly(A)-tailed with yeast poly(A)-polymerase was subjected to RT-polymerase chain reaction (RT-PCR) [16] using 5′-end-labelled SLA3 as the specific primer. The SLA repeat was amplified [19] from *C. fasciculata* DNA (a gift from D.F. Spencer) using primers SLA2 and SLA3 in the presence of 5–10% dimethylsulfoxide (DMSO) [20]. Usually, either primer SLA2 or SLA3 was 5′-end-labelled prior to use in PCR experiments.

End-labelled DNA, prepared as described above, was gel-purified and subjected to a modified chemical sequencing procedure (D.F. Spencer, unpublished). All products were sequenced twice. Some sequencing gels contained 40% formamide (deionized) as well as 7 M urea [19].

3. Results and discussion

The availability of a partial sequence of *C. fasciculata* SLA RNA [21] (see Fig. 1) as well as complete SLA RNA sequences from several other trypanosomatid protozoa [10,11, 15,21] allowed us to design primer SLA1, which was used to revise and extend the available partial sequence. This new sequence information (positions 3–42, containing a two nu-

*Corresponding author. Fax: (1) (902) 494-1355.
E-mail: mwgray@is.dal.ca

Abbreviations: SL, spliced leader; sn, small nuclear; SLA, spliced leader-associated; RT, reverse transcriptase; DMSO, dimethylsulfoxide

PII: S0014-5793(99)01235-1

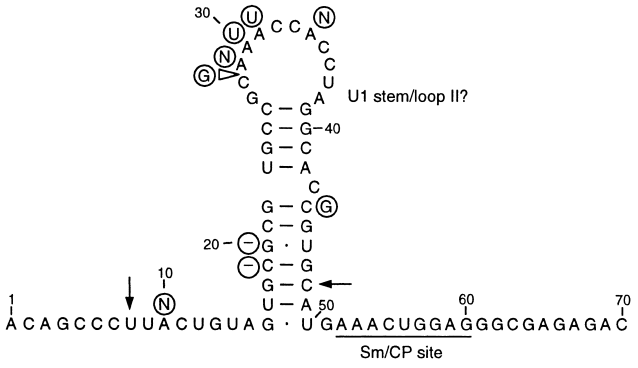

Fig. 1. Sequence and secondary structure of *C. fasciculata* SLA RNA. The location of a possible spliceosomal core-protein binding site analogous to the mammalian Sm site [11,12] is indicated (Sm/CP site). Arrows mark the ends of a previously published partial sequence [21]. Differences in the published sequence (circled residues) relative to our sequence are indicated. The *C. fasciculata* SLA RNA sequence has been submitted to GenBank (accession number AF177033).

cleotide deletion at positions 19 and 20 relative to the final sequence, Fig. 1, see below) was then used to design primers SLA2 and SLA3. These primers were used in PCR experiments to amplify between copies of the SLA RNA gene, assuming that the gene is tandemly repeated in *C. fasciculata*, as it is in other trypanosomatids [15,21]. Initial PCR experiments failed to yield a product (Fig. 2, lane 1), but the SLA repeat was efficiently amplified when 5–10% DMSO was added to the reaction mixture (Fig. 2, lane 2). These data establish that *C. fasciculata* SLA RNA genes are located within a tandem repeat that is approximately the same size as the 1.45 kb repeat described for *Leishmania tarentolae* [15] and substantially larger than the SLA RNA gene repeats of 0.83 kb in *T. brucei* [21] and 0.76 kb in *Trypanosoma cruzi* [15]. Preliminary sequence analysis of the *C. fasciculata* PCR product established that the SLA RNA gene is located downstream of a sequence homologous to the 250/270 nucleotide small nucleolar RNA gene that is found in this region of the SLA RNA gene repeat in other trypanosomatid protozoa [15,21].

The PCR product was used to obtain the sequence of SLA RNA positions 1–3 and positions 48–70. These new data allowed for the design of primer SLA4, which was used to complete the sequence (positions 2–50). Sequencing gels that contained 7 M urea plus 40% formamide were used to resolve band compressions, attributable to a stable secondary structure (Fig. 1). This structure had initially caused us (see above) and others [21] to overlook two residues (positions 19 and 20). The 3′-terminus of *C. fasciculata* SLA RNA was mapped by sequencing a RT-PCR product. The A residues immediately following position 70 on these sequencing gels (not shown) must have been added to the 3′-hydroxyl of the RNA by yeast poly(A)-polymerase prior to RT-PCR because the genomic PCR product has a G residue after position 70 (not shown). It is possible that some A residues had already been added to the SLA RNA transcript in vivo, but this does not appear to occur in *T. brucei* [10]. The 5′-terminus of *C. fasciculata* SLA RNA was mapped by primer extension. In this case, we cannot rule out the possibility that the RT stop may be the result of a hypermodified nucleoside. However, sequence similarity among trypanosomatid SLA RNA genes ceases abruptly at this position. *C. fasciculata* [16] and *T. brucei* [21] SLA RNAs do not have the methylguanosine cap structure that is located at the 5′-terminus of several other trypanosomatid snRNAs [2–4,16].

Trypanosomatid SLA RNAs have a phylogenetically established secondary structure [11,21] that consists of a single hair-

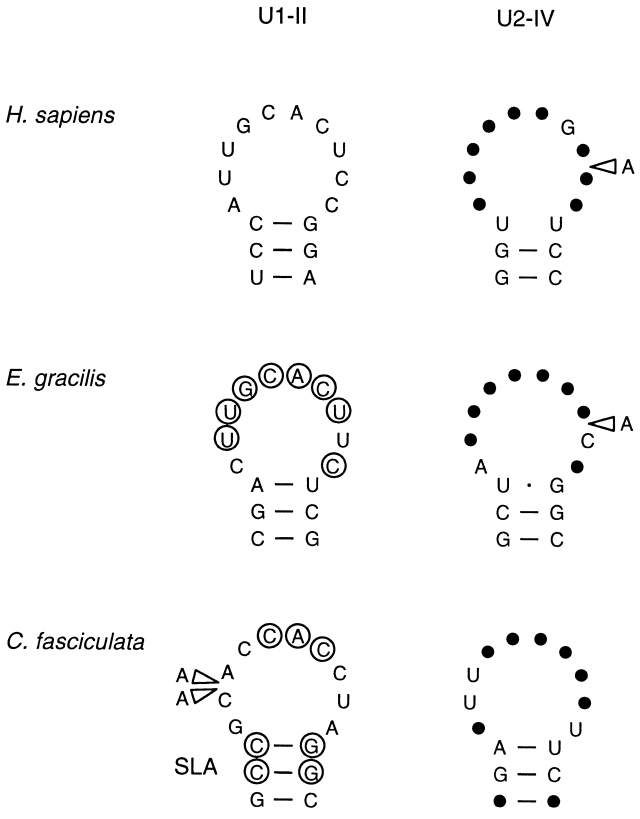

Fig. 3. Comparison of the stem/loop of *C. fasciculata* SLA RNA with U1 snRNA stem/loop II sequences (U1-II). Residues in the *Euglena gracilis* [19] and *C. fasciculata* sequences that are identical to the same positions in the *Homo sapiens* sequence [28] are circled. These sequences are also compared to the U2 snRNA stem/loop IV sequences (U2-IV) from the same organisms, *H. sapiens* [29], *C. fasciculata* [16] and *E. gracilis* [19]. Residues in the U2 sequences that are identical in the corresponding U1/SLA sequence are represented by closed circles.

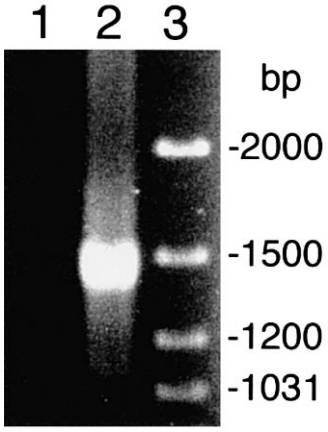

Fig. 2. Amplification of the *C. fasciculata* SLA RNA gene repeat. Lane 1, minus DMSO; lane 2, plus 5% DMSO; lane 3, DNA size markers (MBI Fermentas).

```
                      ---------->(              )<---  -------
C.f.  ACAGCCCUUAC-UGUAGUGCGCGUGCC GCAAACCACCUA GGCACCGUGCAUGAAACUGGAGGGCGAGAGAC
L.s.  AC-GCCCUCAC-UGUAGUGC-CGUGCC GCAAACCACGUA GGCACCGUGCAUGAAACUGGAGCGCAAGAUUC
L.t.  AC-GCGCUCAC-UGUAGUGCGCGUGCC GCAAGACCCACA GGCACCGUGCAUGAAACUGGAGCGCUAGAUUG
T.c.  AAAGCUCUUUUUUGUAGUGCGCUCGUC GCUAAACCAACA GACGCGGUGCACGAAACUGGAGAGCAAGAGAU
T.b.  AAAGCUCUUUUAUGUAGUGUGCGUACC ACGAAAGUAGCA GGUACUGCACACGAAACUGGAGAGCGAGACUC
                                                 * *
```

Fig. 4. Alignment of SLA RNA sequences from *C. fasciculata* (C.f.), *Leptomomas seymouri* (L.s.) [11], *L. tarentolae* (L.t.) [15], *T. cruzi* (T.c.) [15,21] and *T. brucei* (T.b.) [10,21]. The location of the potential U1 snRNA stem/loop II structure is indicated (dashed lines and arrowheads denote sequences that comprise the stem). Chemical sequencing of a primer SLA4-directed RT product verified a large portion of the *T. brucei* sequence (positions 9–50 in the *C. fasciculata* numbering system of Fig. 1). The published *T. brucei* sequence [10,21] contains an inserted G residue relative to all other sequences (between the two asterisks) that was not present in our sequencing gels (data not shown).

pin with a 12 nucleotide loop (see Fig. 1). Because the candidate *trans*-spliceosomal U1 snRNA that we previously identified does not contain a stem/loop II structure [16], we considered the possibility that the stem/loop in SLA RNA might be analogous to U1 stem/loop II. Fig. 3 shows that if *C. fasciculata* SLA RNA does contain a U1 stem/loop II, the loop sequence has diverged significantly from the sequence in typical U1 snRNAs. We further evaluated the hypothesis that SLA RNA contains a U1 stem/loop II by comparing the SLA stem/loop to stem/loop IV of *C. fasciculata* U2 snRNA. The loop sequences of U1 stem/loop II and U2 stem/loop IV are closely similar in other eukaryotes (Fig. 3), reflecting the fact that the two structures bind very similar proteins [17,22]. There is a high degree of sequence conservation between the *Crithidia* SLA RNA loop and the loop sequence of *Crithidia* U2 stem/loop IV (Fig. 3), supporting the idea that *C. fasciculata* SLA RNA supplies a U1 snRNA-like stem/loop II to the *trans*-spliceosome.

The loop region of U2 stem/loop IV sequences is well conserved among trypanosomatid protozoa [23,24] and this region has been implicated in RNA-protein interactions in *T. brucei* [25]. In contrast, the SLA RNA loop sequence is not very highly conserved (Fig. 4), which could indicate that this sequence does not participate in a conserved RNA-protein interaction. On the other hand, if the SLA RNA loop is involved in U1-A protein binding (as expected for a U1 stem/loop II [17,22]), this could be an interesting system in which to study co-evolution of the RNA and protein domains involved in the interaction. Note that different alignments of the SLA loop region are possible, but these variations would not improve the degree of conservation in the case of the very divergent *T. brucei* sequence (verified in this study, Fig. 4).

Whether or not the stem/loop in SLA RNA represents the binding site for a trypanosomatid analog of the mammalian U1-A protein remains to be experimentally tested. The U1-A protein has also been implicated in polyadenylation [26]. Thus, a *trans*-spliceosomal U1-A protein, bound to SLA RNA, would provide a conceptual basis for the functional coupling of polyadenylation of upstream mRNAs and *trans*-splicing of downstream mRNAs in trypanosomatid polycistronic transcripts [27].

Acknowledgements: We thank R. Breckon and F.B.St.C. Palmer for cultures of *C. fasciculata*, D.F. Spencer for *C. fasciculata* DNA, M. McManus and S.L. Hajduk for *T. brucei* RNA and Y. Watanabe and D.F. Spencer for technical advice. We also thank B. Weiser and H. Noller for the computer program XRNA, which was used to generate the secondary structure figure. A Grant (MT-11212) to M.W.G. from the Medical Research Council of Canada and salary support from the Canadian Institute for Advanced Research (Program in Evolutionary Biology) are gratefully acknowledged.

References

[1] Agabian, N. (1990) Cell 61, 1157–1160.
[2] Tschudi, C., Richards, F.F. and Ullu, E. (1986) Nucleic Acids Res. 14, 8893–8903.
[3] Tschudi, C., Krainer, A.R. and Ullu, E. (1988) Nucleic Acids Res. 16, 11375.
[4] Mottram, J., Perry, K.L., Lizardi, P.M., Lührmann, R., Agabian, N. and Nelson, R.G. (1989) Mol. Cell. Biol. 9, 1212–1223.
[5] Tschudi, C. and Ullu, E. (1990) Cell 61, 459–466.
[6] Bruzik, J.P., Van Doren, K., Hirsh, D. and Steitz, J.A. (1988) Nature 335, 559–562.
[7] Bruzik, J.P. and Steitz, J.A. (1990) Cell 62, 889–899.
[8] Steitz, J.A. (1992) Science 257, 888–889.
[9] Logsdon Jr., J.M. (1998) Curr. Opin. Genet. Dev. 8, 637–648.
[10] Watkins, K.P., Dungan, J.M. and Agabian, N. (1994) Cell 76, 171–182.
[11] Palfi, Z., Xu, G.-L. and Bindereif, A. (1994) J. Biol. Chem. 269, 30620–30625.
[12] Dungan, J.M., Watkins, K.P. and Agabian, N. (1996) EMBO J. 15, 4016–4029.
[13] Lücke, S., Klöckner, T., Palfi, Z., Boshart, M. and Bindereif, A. (1997) EMBO J. 16, 4433–4440.
[14] Xu, Y., Ben-Shlomo, H. and Michaeli, S. (1997) Proc. Natl. Acad. Sci. USA 94, 8473–8478.
[15] Roberts, T.G., Sturm, N.R., Yee, B.K., Yu, M.C., Hartshorne, T., Agabian, N. and Campbell, D.A. (1998) Mol. Cell. Biol. 18, 4409–4417.
[16] Schnare, M.N. and Gray, M.W. (1999) J. Biol. Chem. 274, 23691–23694.
[17] Lührmann, R., Kastner, B. and Bach, M. (1990) Biochim. Biophys. Acta 1087, 265–292.
[18] Gray, M.W. (1979) Can. J. Biochem. 57, 914–926.
[19] Breckenridge, D.G., Watanabe, Y., Greenwood, S.J., Gray, M.W. and Schnare, M.N. (1999) Proc. Natl. Acad. Sci. USA 96, 852–856.
[20] Shen, W.-H. and Hohn, B. (1992) Trends Genet. 8, 227.
[21] Roberts, T.G., Dungan, J.M., Watkins, K.P. and Agabian, N. (1996) Mol. Biochem. Parasitol. 83, 163–174.
[22] Rimmele, M.E. and Belasco, J.G. (1998) RNA 4, 1386–1396.
[23] Hartshorne, T. and Agabian, N. (1990) Genes Dev. 4, 2121–2131.
[24] Tschudi, C., Williams, S.P. and Ullu, E. (1990) Gene 91, 71–77.
[25] Günzl, A., Cross, M. and Bindereif, A. (1992) Mol. Cell. Biol. 12, 468–479.
[26] Proudfoot, N. (1996) Cell 87, 779–781.
[27] Nilsen, T.W. (1994) Science 264, 1868–1869.
[28] Branlant, C., Krol, A., Ebel, J.-P., Lazar, E., Gallinaro, H., Jacob, M., Sri-Widada, J. and Jeanteur, P. (1980) Nucleic Acids Res. 8, 4143–4154.
[29] Westin, G., Zabielski, J., Hammarström, K., Monstein, H.-J., Bark, C. and Pettersson, U. (1984) Proc. Natl. Acad. Sci. USA 81, 3811–3815.