Communication

A Candidate U1 Small Nuclear RNA for Trypanosomatid Protozoa*

(Received for publication, May 18, 1999)

Murray N. Schnare and Michael W. Gray‡
From the Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada

In trypanosomatid protozoa, all mRNAs obtain identical 5′-ends by trans-splicing of the 5′-terminal 39 nucleotides of a small spliced leader RNA to appropriate acceptor sites in pre-mRNA. Although this process involves spliced small nuclear (sn) RNAs, it is thought that trypanosomatids do not contain a homolog of the cis-spliceosomal U1 snRNA. We show here that a trypanosomatid protozoon, Crithidia fasciculata, contains a novel small RNA that displays several characteristic features of a U1 snRNA, including (i) a methylguanosine cap and additional 5′-terminal modifications, (ii) a potential binding site for common core proteins that are present in other trans-spliceosomal ribonucleoproteins, (iii) a U1-like 5′-terminal sequence, and (iv) a U1-like stem/loop I structure. Because trypanosomatid pre-mRNAs do not appear to contain cis-spliced introns, we argue that this previously unrecognized RNA species is a good candidate to be a trans-spliceosomal U1 snRNA.

In trypanosomatid protozoa, all mRNAs have identical 5′-terminal sequences. A 39-nucleotide (nt)1 spliced leader (SL) sequence is transferred from the 5′-end of a small SL RNA to the pre-mRNA in a process known as trans-splicing, which is very similar to the spliceosomal cis-splicing found in other eukaryotes (1). However, it is thought that trypanosomes and their relatives contain neither cis-spliced introns (1, 2) nor an equivalent of the U1 small nuclear (sn) RNA required for spliceosomal cis-splicing (3). The SL RNA is considered to be a trans-spliceosome-specific snRNA because it is found in a ribonucleoprotein particle that contains the same core proteins that are associated with other trans-spliceosomal snRNAs (4, 5). It has been proposed that during trans-splicing, sequences within the SL RNA are able to substitute for the function that U1 snRNA normally supplies in cis-splicing (3, 6, 7). Our discovery of a U1 snRNA homolog in Euglena gracilis (8), an organism that is specifically related to trypanosomatid protozoa (9), prompted us to re-evaluate a possible role for U1 snRNA in trypanosomatids.

In cis-splicing, during early stages of spliceosome assembly, the 5′-terminal region of U1 snRNA base pairs across the 5′-splice site (10). In the case of trans-splicing, it is commonly held that base pairing between U1 snRNA and the 5′-splice site may not be required for splicing of SL RNA sequences (1, 3, 6, 7). This view is supported by a study in which a trypanosomatid (Leptomonas collosoma) SL RNA sequence was placed upstream of a 5′-splice site, with the resulting chimeric substrate being efficiently spliced in a HeLa cell nuclear extract even after the 5′-end of >99% of the endogenous U1 snRNA had been removed by oligonucleotide-directed RNase H cleavage (7). This result provided support for an earlier proposal that U1 snRNA-like base pairing could be supplied by a region of the SL RNA upstream of the 5′-splice site (6). However, it was subsequently shown that U1 snRNP is, in fact, required for cis-splicing of the chimeric substrate, and that base pairing between the 5′-end of U1 snRNA and the SL RNA 5′-splice site does occur in these extracts when the 5′-end of U1 snRNA is intact (11). It has also been demonstrated recently that the proposed internal SL RNA base pairing across the 5′-splice site is not essential for trans-splicing in Leishmania tarentolae (12).

In sum, the proposal that trypanosomatid SL RNA substitutes for U1 snRNA in trans-splicing has not gained experimental support, and the data do not definitively rule out the possibility that a U1 snRNA homolog is present in trypanosomatid protozoa but has not yet been identified. Here we show that a representative trypanosomatid, Crithidia fasciculata, contains a novel small RNA that displays several characteristic features expected of a U1 snRNA homolog.

EXPERIMENTAL PROCEDURES

Immunoprecipitation (3)—100 μl of 2 × buffer (1 × buffer = 50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1 mM EDTA, 0.05% (v/v) Nonidet P-40) and 50 μl (5 μg) of anti-N7,N7,7-trimethylguanosine (m2,7G) monoclonal antibody (Oncogene Research Products, Ref. 13) were added to 50 μl (250 μg) of total RNA (prepared as described, see Ref. 14). After 1 h on ice, protein A-Sepharose (Amersham Pharmacia Biotec, 3 mg in 100 μl of 1 × buffer) was added, and the sample was incubated on ice (30 min) with occasional mixing. The immunoprecipitate was recovered by centrifugation (5 min) and washed (five times) by resuspension in 1 ml of 1 × buffer followed by centrifugation. The final pellet was resuspended in 200 μl of H2O. 20 μl of 3 × NaOAc were added, and RNA was recovered by extraction with phenol-cisul (twice) followed by ethanol precipitation. Contaminating RNA was removed by repeating the above immunoprecipitation procedure.

Amplification of DNA—The first primers were used in polymerase chain reaction (PCR) experiments (8, 15): CIU1 (5′-CCCT-C3AAATGAGTTGCGGAC-3′), CIU1′ (5′-ATCCAAGAGACCCAAGTCTG-3′), 3′ ARG (5′-ATCCGTCAGAGATGTGCAGA-3′), CIU5 (5′-AGATG-TCGGAAGGTCTAAG-3′). A reverse transcriptase (RT)-PCR product was generated (16, 17) by first polyadenylating total RNA using yeast run7P (5′-AATAAGAGCAGCAGCAGCAGCAGCAG-3′) and primer P-45 (5′-ATATGAGGTGGTCTAAG-3′). A specific product was obtained after two rounds of PCR; primary PCR employed P-55 (5′-GGAGCTCAATAAAGCGGCCGC-3′) and primer CIU1, while secondary PCR employed primer P-4 (5′-AATT-AAAGGCCGCCAGTGCA-3′) and primer CIU1.

Sequence Analysis—5′-End-labeled PCR and RT-PCR products, as well as a RT product generated using 5′-end-labeled primer CIU1A,
Fig. 1. Anti-m$_{2,2,7}$G immunoprecipitation of capped RNAs from C. fasciculata. RNAs were 3’-end-labeled and resolved in a 10% polyacrylamide, 7 M urea sequencing gel. Individual RNAs were identified by partial sequence analysis and comparison to known sequences. sno = small nucleolar. Notes: (a) the three U6 snRNA bands have identical 3’ termini (19) and are immunoprecipitated due to base pairing with mG-capped U4 snRNA (3); denaturation of the sample (90 °C, 2 min) prior to immunoprecipitation greatly reduces the yield of U6 snRNA. (b) The yield of tRNAs is greatly increased by prior heat treatment (65 °C, 5 min). (c) The yield of tRNA fragments varies for different RNA preparations. (d) The 5’-terminal tRNA$^{5\text{GM}}$ fragment ends at the 5-methyluridine at position 54. (f) The 3’-terminal tRNA$^{5\text{GR}}$ fragment starts at position A9. (f) The four SL RNAs (20–22) end at positions G83, U90, G91, and C92.

were sequenced by a modified chemical method. RNA was labeled with $^{33}$P at either the 5’ or 3’ end, sequenced by chemical and enzymatic methods, and subjected to terminal nucleotide analysis as described (8, 15, 18).

RESULTS AND DISCUSSION

Antibodies specific for m$_{2,2,7}$G have been used to enrich for capped snRNAs in Trypanosoma brucei (3). Although no U1 snRNA homolog was found among the four largest m$_{2,2,7}$G-capped RNAs, the experiments did detect other, smaller capped RNAs that were not further characterized (3). Additional snRNAs have also been detected by immunoprecipitation using antibodies directed against core proteins common to T. brucei spliceosomal snRNPs (5). Fig. 1 shows an electrophoretic profile of C. fasciculata RNAs that were immunoprecipitated using a monoclonal anti-m$_{2,2,7}$G antibody. The antibody reacted efficiently with homologs of the m$_{2,2,7}$G-capped RNAs identified previously in T. brucei. It also reacted efficiently with the 7-monomethylguanosine (m$^7$G)-capped SL RNA (20) and with an intermediate of the trans-splicing reaction, the 39-nr free SL RNA exon. In addition, the antibody enriched for a subset of tRNAs that appear to have internal m$^7$G in their variable loops, as judged by chemical reactivity during sequencing.

When the amount of antibody relative to RNA was reduced in immunoprecipitation experiments, the RNA yield decreased but the intensity of individual bands relative to each other remained unchanged (data not shown). Although the monoclonal antibody did not distinguish between mG and m$_{2,2,7}$G caps, this procedure did allow us to identify a previously unrecognized methylguanosine (mG)-capped RNA that was present at approximately the same concentration as the known trans-spliceosomal snRNAs (U1 snRNA?) in Fig. 1.

When the preliminary sequence of this candidate U1 snRNA was used as a query in a GenBank search, we detected a similar sequence in a tRNA gene cluster from L. tarentolae (23), located on the opposite strand between upstream tRNA$^{5\text{GM}}$ and downstream tRNA$^{5\text{GR}}$ genes. PCR experiments (primer combination CfU1A and 3’ ARG, see “Experimental Procedures”) confirmed that in C. fasciculata there is also a tRNA$^{5\text{GR}}$ gene on the opposite strand ~100 base pairs upstream of the candidate U1 snRNA sequence. Attempts to amplify the region in C. fasciculata DNA between this new sequence and a possible downstream tRNA$^{5\text{GM}}$ gene (primer combination CfU1 and 5’ LEU, see “Experimental Procedures”) were unsuccessful, indicating that this particular gene linkage may not be conserved between L. tarentolae and C. fasciculata.

The first two residues of the novel RNA were identified as A by DNA sequencing (Fig. 2A) but were resistant to cleavage by RNases and alkali during enzymatic sequencing (not shown), indicating that they are O$^2$-methylated. The first residue was also resistant to cleavage in the A reaction during chemical sequencing (Fig. 2B), suggesting that it contains additional modifications. No other post-transcriptional modifications were encountered during sequencing, but we did detect a single site of heterogeneity, A/C at position 60 (Fig. 2, C and D).

In order to further characterize the modified A at the 5’-end of the molecule, the RNA was 5’-end-labeled after removal of the mG cap structure by tobacco acid pyrophosphatase treat-
ment. Gel-purified RNA was digested with snake venom phosphodiesterase, and the resulting radioactive mononucleotide was analyzed by thin layer chromatography. This experiment demonstrates that the candidate U1 snRNA has the same hypermodified 5'-terminal nucleoside, (N6,N6,O2- trimethyladenosine (m2Am)) that is known to be present at the 5'-end of SL RNAs from C. fasciculata and T. brucei (20); in contrast, the U2 (Fig. 3A) and U4 (not shown) snRNAs have 5'-terminal O2'-methyladenosine (Am) residues. The similarity in methylation patterns observed between this new RNA and the SL RNA may be explained by the fact that six of the seven 5'-terminal nucleotides are identical in the two RNA species (Fig. 3C).

Trypanosomatid snRNAs contain sequences that bind spliceosomal core proteins (4–6, 24, 25). The C. fasciculata SL RNA core-protein binding site, which is capable of interacting with mammalian Sm proteins (6) as well as with T. brucei core proteins (4), appears to consist of an AAAAAUGA sequence followed by a short G + C-rich hairpin. The novel RNA reported here has an almost identical sequence, AUAUUUGA (residues 44–52), followed by a 3'-terminal hairpin structure that is supported by compensating base changes in the L. tarentolae sequence (Fig. 4B).

The presence of a core-protein binding site and a modified 5'-terminus, both of which are specifically related to their trans-spliceosomal SL RNA counterparts, provides strong evidence that this new capped RNA is also a component of the trans-spliceosome. Given that U1 snRNA is the only typical spliceosomal snRNA that has not yet been discovered in trypanosomatid protozoa (3, 24, 25), it is not surprising that we also found primary sequence conservation between this C. fasciculata candidate U1 snRNA and known U1 snRNAs from other organisms (Fig. 4C; compare Fig. 4, A and B). Notably, the highly conserved 5'-terminal region contains the ACCU se-

---

**FIG. 3.** 5'-Terminal modified nucleotide analysis. Thin layer chromatography was carried out using unlabeled pN and pAm as markers. Co-migration of the SL and U1? products was verified by eluting them from the original plate (A) and re-chromatographing them (B) using the same solvent system (95% ethanol:H2O, 4:1; Ref. 18). Co-migration was also observed in a second solvent system (saturated ammonium sulfate:isopropanol, 40:1, Ref. 18; not shown). C, comparison of the 5'-terminal sequences of C. fasciculata U1 snRNA? and SL RNA (20–22). Note that the modifications present at positions three (O2'-methylcytosine, Cm) and four (3',O2'-dimethyluridine, m3Um) of the SL RNA are readily detected in RNA sequencing reactions (not shown) and on this basis are clearly not present in the candidate U1 snRNA. Terminal mG caps are not included.

**FIG. 4.** Comparison of the C. fasciculata candidate U1 snRNA with known U1 snRNAs. A, secondary structure (26) of Homo sapiens U1 snRNA (27, 28) showing the locations of modified nucleosides (Um = O2'-methyluridine, p = pseudouridine) and the Sm core-protein binding site. B, potential secondary structure of the C. fasciculata RNA showing differences in its homolog from L. tarentolae (circled residues next to the C. fasciculata sequence). A trypanosomatid-specific spliceosomal core-protein (CP) binding site, analogous to the Sm site in higher eukaryotes, is indicated (Sm/CP site). C, 5'-terminal sections of the C. fasciculata (C.f.) nucleotide sequence are compared with the corresponding sections of U1 snRNA sequences (8, 29) from Euglena gracilis (E.g.), Physarum polycephalum (P.p.), Saccharomyces cerevisiae (S.c.), and H. sapiens (H.s.). Dots represent residues that are identical to the C. fasciculata sequence. The loop region of stem/loop I is enclosed by parentheses. Secondary structure diagrams were generated using the program XRNA developed by B. Weiser and H. Noller (University of California, Santa Cruz, CA).
sequence (residues 6–9, overlined in Fig. 4) that has the potential to interact with conserved splice site sequences (10). The C. fasciculata candidate U1 snRNA also contains a stem/loop I sequence that, in mammalian systems, is the binding site for U1-70K, a U1 RNase-specific protein (30).

In view of the unusually small size of the C. fasciculata candidate U1 snRNA (69 nt), it is also not surprising that it lacks some of the characteristic features of a U1 snRNA (Fig. 4, A versus B). The highly conserved stem/loop II, which serves as the binding site for the U1-A protein in other eukaryotes (30), is the most notable structural element that is missing. One could argue that stem/loop II functions are either unnecessary for trans-splicing or are supplied by another component of the spliceosome (RNA and/or protein). Other trypanosomatid snRNAs show a similar pattern of reduced size and sequence divergence compared with their homologs from other systems (1, 3, 19, 24, 25).

The presence of structurally divergent snRNAs in trypanosomatid protozoa may indicate that the highly accurate splice site selection required for cis-splicing systems may not be as important in a “trans-splicing only” system (1). For example, the demands on 5′-splice site selection machinery (including U1 snRNA) are considerably reduced in trans-splicing, where the 5′-splice site is already present in the spliceosome as part of the spliceosomal SL RNA. Similarly, reduced accuracy of branch point and 3′-splice site selection (1) in SL RNA trans-splicing (which occurs upstream of coding sequences) is likely to be tolerated, because errors would not disrupt the reading frame in the spliced product (1). On the other hand, with the discovery of an apparent U1 snRNA homolog, trypanosomatid protozoa may possess a complete set of spliceosomal snRNAs, raising the possibility that the mechanisms of trans- and cis-splicing are not as different as was previously thought. In this regard, in the absence of a complete genomic sequence, we cannot rule out the possibility that a least a few cis-spliced introns may eventually be found in the trypanosomatid group of protozoa.

Acknowledgments—We thank F. B. St. C. Palmer and R. Brecken for supplying cultures of C. fasciculata, D. F. Spencer for a gift of purified C. fasciculata DNA and the chemical sequencing protocol, and Y. Watanabe for the primers used for RT-PCR.

REFERENCES

1. Agabian, N. (1990) Cell 61, 1157–1160
2. Logsdon, J. M., Jr. (1998) Curr. Opin. Genet. Dev. 8, 637–648
3. Mottram, J., Perry, K. L., Lizardi, P. M., Luhrmann, R., Agabian, N., and Nelson, R. G. (1989) Mol. Cell. Biol. 9, 1212–1223
4. Cross, M., Gunzli, A., Palfi, Z., and Bindereif, A. (1991) Mol. Cell. Biol. 11, 5516–5526
5. Palfi, Z., and Bindereif, A. (1992) J. Biol. Chem. 267, 20159–20163
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. Breckenridge, D. G., Watanabe, Y., Greenwood, S. J., Gray, M. W., and Schnare, M. N. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 852–856
9. Sogin, M. L. (1991) Curr. Opin. Genet. Dev. 1, 457–463
10. Ares, M., Jr., and Weiser, B. (1995) Prog. Nucleic Acid Res. Mol. Biol. 50, 131–159
11. Seiwert, S. D., and Steitz, J. A. (1993) Mol. Cell. Biol. 13, 3135–3145
12. Sturm, N. R., Fleishmann, J., and Campbell, D. A. (1998) J. Biol. Chem. 273, 18689–18692
13. Krainer, A. R. (1988) Nucleic Acids Res. 16, 9415–9429
14. Gray, M. W. (1979) Can. J. Biochem. 57, 914–926
15. Greenwood, S. J., Schnare, M. N., and Gray, M. W. (1996) Curr. Genet. 30, 338–346
16. Zaug, A. J., Lingner, J., and Cech, T. R. (1996) Nucleic Acids Res. 24, 532–533
17. Thweatt, R., Goldstein, S., and Shmoakler Reis, R. J. (1990) Anal. Biochem. 190, 314–316
18. MacKay, R. M., Spencer, D. F., Doolittle, W. F., and Gray, M. W. (1980) Eur. J. Biochem. 112, 561–576
19. Xu, G.-L., Wieland, B., and Bindereif, A. (1994) Mol. Cell. Biol. 14, 4565–4570
20. Bangs, J. D., Crain, P. F., Hashizume, T., McCloskey, J. A., and Boothroyd, J. C. (1990) J. Biol. Chem. 265, 9805–9815
21. Muhlch, M. L., Hughes, D. E., Simpson, A. M., and Simpson, L. (1987) Nucleic Acids Res. 15, 3141–3153
22. Gabriel, A., Stosid, S. S., and Cleveland, D. W. (1987) J. Biol. Chem. 262, 16192–16199
23. Shi, X., Chen, D.-H. T., and Suyama, Y. (1994) Mol. Biochem. Parasitol. 65, 23–37
24. Dungan, J. M., Watkins, K. P., and Agabian, N. (1996) EMBO J. 15, 4016–4029
25. Xu, Y., Ben-Shlomo, H., and Michaeli, S. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 8473–8478
26. Guthrie, C., and Patterson, B. (1988) Annu. Rev. Genet. 22, 387–419
27. Brandlant, 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
28. Reddy, R., Hennings, D., and Busch, H. (1981) Biochem. Biophys. Res. Commun. 98, 1076–1083
29. Zwieb, C. (1997) Nucleic Acids Res. 25, 102–103
30. Luhrmann, R., Kastner, B., and Bach, M. (1990) Biochim. Biophys. Acta 1087, 265–292