Purification of the Spliced Leader Ribonucleoprotein Particle from *Leptomonas collosoma* Revealed the Existence of an Sm Protein in Trypanosomes

**CLONING THE SmE HOMOLOGUE**

(Received for publication, December 29, 1998, and in revised form, February 11, 1999)

Igor Goncharov, Zsófia Palfi‡, Albrecht Bindereif‡, and Shulamit Michaeli§

*From the Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot, Israel 76100 and the ‡Institut für Biochemie, Humboldt-Universität/Charité Monbijou-Strasse 2a, D-10117 Berlin, Germany*

Trans-splicing in trypanosomes involves the addition of a common spliced leader (SL) sequence, which is derived from a small RNA, the SL RNA, to all mRNA precursors. The SL RNA is present in the cell in the form of a ribonucleoprotein, the SL RNP. Using conventional chromatography and affinity selection with 2′-O-methylated RNA oligonucleotides at high ionic strength, five proteins of 70, 16, 13, 12, and 8 kDa were co-selected with the SL RNA from *Leptomonas collosoma*, representing the SL RNP core particle. Under conditions of lower ionic strength, additional proteins of 28 and 20 kDa were revealed. On the basis of peptide sequences, the gene coding for a protein with a predicted molecular weight of 11.9 kDa was cloned and identified as homologue of the cis-spliceosomal SmE. The protein carries the Sm motifs 1 and 2 characteristic of Sm antigens that bind to all known cis-spliceosomal uridylic acid-rich small nuclear RNAs (U snRNAs), suggesting the existence of Sm proteins in trypanosomes. This finding is of special interest because trypanosome snRNPs are the only snRNPs examined to date that are not recognized by anti-Sm antibodies. Because of the early divergence of trypanosomes from the eukaryotic lineage, the trypanosome SmE protein represents one of the primordial Sm proteins in nature.

In trypanosomes all mRNAs are processed by trans-splicing, which joins a common spliced leader (SL)1 derived from the SL RNA to polycistronic pre-mRNAs (1, 2). The SL RNA fulfills a dual function, both recruiting splicing co-factors and serving as a substrate for splicing. In analogy to cis-splicing, trans-splicing requires the participation of U snRNPs. Initial studies identified U2, U4, and U6 snRNAs in trypanosomes (3), and ionic strength, additional proteins of 28 and 20 kDa were revealed. On the basis of peptide sequences, the gene coding for a protein with a predicted molecular weight of 11.9 kDa was cloned and identified as homologue of the *cis*-spliceosomal SmE. The protein carries the Sm motifs 1 and 2 characteristic of Sm antigens that bind to all known *cis*-spliceosomal uridylic acid-rich small nuclear RNAs (U snRNAs), suggesting the existence of Sm proteins in trypanosomes. This finding is of special interest because trypanosome snRNPs are the only snRNPs examined to date that are not recognized by anti-Sm antibodies. Because of the early divergence of trypanosomes from the eukaryotic lineage, the trypanosome SmE protein represents one of the primordial Sm proteins in nature.

In trypanosomes all mRNAs are processed by trans-splicing, which joins a common spliced leader (SL)1 derived from the SL RNA to polycistronic pre-mRNAs (1, 2). The SL RNA fulfills a dual function, both recruiting splicing co-factors and serving as a substrate for splicing. In analogy to cis-splicing, trans-splicing requires the participation of U snRNPs. Initial studies identified U2, U4, and U6 snRNAs in trypanosomes (3), and their essential role in trans-splicing was demonstrated (4). Only recently, the trypanosome U5 homologue was identified (5, 6), but no U1 homologue has been revealed so far (3).

The trypanosome U snRNAs are generally smaller and show differences from their *cis*-spliceosomal counterparts (3, 7). Another unique property of the trypanosomatid snRNAs is that they possess a divergent Sm site. In other eukaryotes, the Sm site constitutes the binding site for the common Sm proteins (8) that are recognized by anti-Sm antibodies present in autoimmune patients (9). Surprisingly, so far, the trypanosome snRNPs represent the only exception in that they are not recognized by these antibodies, suggesting unique properties of the trypanosome core proteins (10, 11). Five core snRNP proteins were identified in *Trypanosoma brucei* ranging in size from 15 to 8.5 kDa, which are shared among all trans-spliceosomal snRNPs, including the SL, U2, U4/U6 (11), and U5 RNP (12).

Recent studies in both mammalian and yeast systems have demonstrated that all the known Sm proteins possess homology in two regions that were termed Sm motifs 1 and 2 (13–15). The Sm motifs are also conserved in all putative homologues including plant and yeast (13–15). This suggests that Sm proteins may have arisen from a single common ancestor. Recently, Sm-like proteins that do not belong to the canonical core proteins were also revealed, and one of them was identified as a U6-specific protein (14).

Very little is known about snRNP proteins in trypanosomes. The two trypanosome proteins that have been identified so far are the U2-specific 40 kDa, the human homologue of U2A (16) and the U5-specific PRP8/p220 homologue (12). Interestingly, not all the amino acid positions critical for PRP8 function are conserved in the trypanosome protein p277, suggesting that certain differences exist between PRP8 functions in *cis* and *trans*-splicing (12).

In this study, we have purified the SL RNP particle from the trypanosomatid *Leptomonas collosoma* using a combination of conventional chromatography and affinity selection with biotinylated 2′-O-methyl (2′-OMe) antisense RNA oligonucleotide, resulting in the identification of a subset of core proteins and loosely bound proteins that were selected with the SL RNA. The *L. collosoma* proteins show immunological relationship to the *T. brucei* core proteins. Based on internal peptide sequences, one of the core proteins was cloned and sequenced; the analysis identified it as a *bona fide* Sm protein harboring the conserved Sm motifs 1 and 2, mostly related to the SmE protein. In sum, although snRNP proteins are not recognized by anti-Sm sera in trypanosomes and the Sm site is degenerate, trypanosomes do possess Sm proteins.

**EXPERIMENTAL PROCEDURES**

*Extract Preparation and Antisense 2′-OMe RNA Affinity Purification of SL RNP—* *L. collosoma* cell culture and extract preparation (using 0.3 or 0.4 mM KCl) has been previously described (17), as well as postribosomal supernatant preparation and DEAE chromatography (10). Extracts were prepared from 10 liters of cell cultures (5 × 10⁶ cells/ml). The extract was fractionated on a DEAE-Sephael column, and the material eluted at 0.4 mM KCl was concentrated to 0.5–1 ml by chroma-
Fig. 1. A, RNA analysis of the SL RNP purification steps. Extracts were prepared from 5 × 10^11 cells and fractionated on a DEAE-Sepharose column. The RNA extracted from the following fractions (0.02 of each sample) was analyzed on a 6% denaturing gel and visualized by silver stain. Lane 1, whole cell extract; lane 2, post-ribosomal supernatant; lane 3, DEAE fraction eluted with 0.4 M KCl; lane 4, affinity selected fraction using U2-1 oligonucleotide; and lane 5, affinity selected fraction using SL-1 oligonucleotide. B, separation of particles from different purification steps on a native gel. Samples obtained as described in panel A (0.01 of each) were fractionated on a 5% native gel. The gel was subjected to Northern analysis with an anti-SL RNA probe (oligonucleotide 9825). Lane 1, total L. collosoma RNA (10 μg); lane 2, PRS; and lane 3, DEAE fraction eluted at 0.4 M KCl.

Fig. 2. A, protein analysis of affinity selected L. collosoma SL RNP. After selection of SL RNPs from 10^11 cells with SL-1 RNA oligonucleotide, proteins were released, separated on 12% Tris-Tricine gels, and stained with Coomassie Blue. Markers were low molecular mass markers from Life Technologies, Inc., and their sizes are indicated on the right. Extract preparation and affinity selection were performed in the absence of 0.4 M KCl. Lane 1, whole cell extract; lane 2, post-ribosomal supernatant; lane 3, DEAE elution with 0.4 M KCl; lane 4, affinity selection with SL-1 oligonucleotide; and lane 5, control selection in the absence of SL-1 oligonucleotide. B, the same as in panel A, but extraction was performed at 0.2 M KCl. Designation of the lanes is as in panel A.

Fig. 3. Immunoprecipitation analysis of L. collosoma SL RNP. A, 50-μl DEAE fraction obtained from 5 × 10^11 cells was immunoprecipitated with 40 μl of anti-CP antibodies (21). RNA was deproteinized and analyzed by primer extension with anti-SL oligonucleotide 9825, and the products were separated on a 6% denaturing gel. Lane 1, pellet obtained with anti-CP antibodies; lane 2, pellet with nonimmune serum, and lane 3, supernatant of anti-CP precipitation. The marker was 32P-labeled pBR322 HpaII digest. B, immunoblot analysis of affinity selected proteins. SL RNP of L. collosoma and SL and U2 RNPs of T. brucei were affinity purified from a DEAE fraction (as described above and in Ref. 21), separated by SDS-polyacylamide gel electrophoresis, and probed with rabbit anti-CP antibodies as described previously (11). Immunocomplexes were visualized with goat anti-rabbit IgG conjugated to horseradish peroxidase and a fluorescent substrate (ECL). Molecular mass markers are indicated. Lane 1, L. collosoma SL RNP; lane 2, T. brucei U2 RNP; and lane 3, T. brucei SL RNP.

Primers—The primers used were: 1) 9825, 5'-GCGCGAAGAGCTC-GTC-3', complementary to nucleotides 80–95 of the L. collosoma SL RNA; 2) 24279, 5'-TTTCATATCTCCTATCC-3', antisense to the Sm1 peptide IEGNLLGYDEFMNVVL; 3) 24280, 5'-GGNTAXGAYGAX- TTYATGAA-3', sense to Sm1 motif; and 4) 24281, 5'-ACNACXCC- NACXTRTCT-3', antisense to Sm2 peptide sequence LIRRSXDVGVY- VVAI. X denotes inosine, N is a mixture of all four nucleotides, R denotes purines, and Y denotes pyrimidines.

Cloning and Sequencing—500 ng of degenerate oligonucleotides 24279, 24280, and 24281 were labeled each with 250 μCi of [γ-32P]ATP (7000 Ci/mmol) and polynucleotide kinase. Each labeled oligonucleotide (106 Cerenkov counts) was used to screen a bacterial genomic library (50,000 plaques) of L. collosoma (17, 19). Two plaques gave positive signals with all three probes. Pure DNA was digested with HpaII and AluI, and DNA fragments of 0.6 and 1.6 kilobases were cloned into pBluescript SK and sequenced, using T3 and T7 primers.

Native Gels—RNPs from the different purification steps were separated on a 5% nondenaturing gel buffered with 50 mm Tris-glycine, pH 8.8.

RESULTS AND DISCUSSION

Purification of the L. collosoma SL RNP Core Particle—Previously, using affinity selection with biotinylated RNA oligonucleotide, the SL RNP of T. brucei was purified and a subset of five low molecular mass proteins of 15 to 8.5 kDa was identified. Only a single protein of 14.4 kDa appeared to be specific to SL RNP, whereas other proteins are shared with U2 and U4/U6 RNPs (11). Because of the DEAE fraction (as described above) and the advantage of performing biochemical analysis with the monogenetic trypanosomatids, we have set up the purification of the SL RNP from L. collosoma extracts. Purification steps were scaled-up such that microgram amounts of proteins sufficient for microsequencing could be prepared. SL RNP was purified from high-salt extract because only under these conditions was...
it possible to recover about 80% of the particles. The SL RNP was highly enriched by two purification steps: depletion of the ribosomes and fractionation on a DEAE-Sephacel column. RNA was extracted from the different purification steps. Results presented in Fig. 1 demonstrate that this fractionation procedure efficiently enriches for the major snRNAs, U2, U4, SL, and U6, but that other small RNAs such as 7SL RNA and tRNAs were still present in these fractions. To investigate the status of the SL RNP during purification, we analyzed the RNPs by native gel electrophoresis and Northern hybridization. The results presented in Fig. 1 indicate that in the initial steps of purification, the SL RNP migrated as a slow form (lane 2) and was converted to a fast migrating form during fractionation on a DEAE column (lane 3). Fractionation of the DEAE eluate by CsCl gradient centrifugation suggests that these particles withstand fractionation in high salt gradients and band at a density of 1.4 g/ml, characteristic of core SL RNP particles (10, 20) (data not shown). It is currently unknown whether, during the conversion from the slowly to the fast migrating form of the SL RNP, only particle-specific proteins are dislodged or whether the SL RNP dissociates from a multisnRNP complex. We have previously demonstrated that, when RNPs extracted at low ionic strength are fractionated on sucrose gradients, at least 50% of the SL RNP migrates at S values higher than 11 S (core SL RNP), suggesting that the SL RNP associates with other spliceosomal components (6). Affinity selection of the SL RNP and U2 RNP from the DEAE eluate demonstrates its high specificity and efficiency because, in both selections, a single RNA band of the expected size was observed (Fig. 1A, lanes 4 and 5).

Identification of the SL RNP Proteins from L. collosoma—To analyze the protein composition of the SL RNP, a concentrated DEAE column fraction prepared from 10^11 cells was used in affinity selection with biotinylated antisense SL-1 RNA oligonucleotide. Proteins were analyzed on a 12% Tris-Tricine gel and stained with Coomassie Blue. Results presented in Fig. 2 show the enrichment of five proteins of 70, 15, 13, 12, and 8 kDa (lanes 1, 2, 3, and 4). The proteins of 55 kDa are the most abundant proteins in the DEAE fraction (lanes 4 and 5) and most likely represent α and β tubulins. These proteins were nonspecifically selected because they were present in the control selection without the SL oligonucleotide (lane 5). The pattern of the selected small proteins of the L. collosoma SL RNP seems to differ only slightly from the pattern of the proteins selected in T. brucei. In addition, the L. collosoma SL RNP contains a protein of 70 kDa. Because this SL RNP purification was done at 0.4 M KCl, we will refer to this RNP as the core SL RNP.

To investigate whether purification under less stringent conditions reveals additional SL RNP associated proteins, particles were extracted at 0.2 M KCl and were purified by DEAE chromatography. The selected proteins were stained with Coomassie Blue (Fig. 2B). The results indicate the presence of proteins of 28, 20, and 18 kDa in addition to the four low molecular mass proteins characteristic of the core SL RNP. However, the 18-kDa protein was bound nonspecifically to the
FIG. 5. Alignment of the SmE homologues. The alignment was performed using the ClustalW Multiple Sequence Alignment program and is illustrated using the program SeqVu 1.10. Identity is represented by gray background, and similarity is boxed. Abbreviations used are: H. sap., Homo sapiens; G. gal., Gallus gallus (chicken); M. mus., Mus musculus; A. thianiana, Arabidopsis thaliana; O. sativa, Oryza sativa (rice); S. cerevisiae, Saccharomyces cerevisiae; C. elegans, Caenorhabditis elegans; L. collosoma, Leptomonas collosoma. The sequences were obtained from the following references or data base accession numbers: human, (27); G. gallus, GenBank™ X65702; Mouse (28); A. thaliana, EST, GenBank™ Z29055; O. sativa, EST, GenBank™ D23404; C. elegans, GenBank™ Z81071; S. cerevisiae, (22); L. collosoma, GenBank™ AF126283. The Sm motifs 1 and 2 are underlined, and the consensus sequences are given (13).

The 12-kDa Protein Is an Sm Protein Homologous to SmE—To clone the genes coding for core SL RNP proteins, the low molecular mass proteins were eluted from the gel and digested with LysC protease. Several internal peptide sequences were obtained from the four core proteins. In particular, two long peptides (peptide 1, IEGNNLLGYDFMNVVL, and peptide 2, ILLRSDNVGVHAII) could be determined from the 12-kDa protein. Using degenerate oligonucleotides as probes, a genomic fragment of 1220 base pairs was cloned and sequenced (Fig. 4). An open reading frame was identified coding for a 11.9-kDa protein of 94 amino acids. The spliced leader addition site (AG3’ splice site) was mapped 16 nucleotides upstream of the first ATG (data not shown), and 22 consecutive pyrimidines were found upstream (Fig. 4, underlined in the sequence). These pyrimidines resemble the polypyrimidine tract found in intergenic regions of trypanosomatid genes that was shown to be essential for accurate trans-splicing (22). The polyadenylation site was mapped by reverse transcriptase polymerase chain reaction and is located 417 nucleotides downstream of the stop codon (results not shown). Southern analysis indicated that the gene is present in a single copy in the L. collosoma genome (data not shown). A multiple alignment of the L. collosoma protein with its homologues is presented in Fig. 5. The L. collosoma protein is 27% identical and 64% similar to the human SmE protein. Interestingly, the protein possesses the bipartite Sm motif that carries two conserved domains Sm1 and Sm2 underlined in Fig. 5 (boundaries according to the nomenclature of Hermann et al. (13); Sm motif 1, 32 amino acids; Sm motif 2, 14 amino acids). The Sm motifs are the most conserved domains in SmE protein (~45% identity). Among the few deviations in the Sm motifs of the L. collosoma protein from
its homologues is position 7 of the Sm2 motif (Arg instead of Lys in most of the other species). This is especially interesting because Arg is conserved in all Sm2 motifs, but in all E proteins this Arg was changed to Lys, except for the L. collosoma protein where the canonical Arg is kept. Recently, functional analysis was performed on the yeast SmE protein (23, 24). The results pointed out the importance of certain hydrophobic residues for cell viability; these positions are essential for interactions of the protein with SmF and SmG (24). Interestingly, of these seven important residues, all but one are either similar or identical in L. collosoma, suggesting that their function may be conserved in the trypanosome protein.

One of the most intriguing questions regarding the trypanosome core proteins is the lack of recognition by anti-Sm sera. This finding is even more surprising after revealing the presence of Sm motifs in the L. collosoma protein. It was previously suggested that important domains for Sm autoimmunization are the conserved Sm motifs (13). This current study suggests that not only are the Sm motifs involved in B-cells epitope recognition. Alternatively, the trypanosome Sm motifs may differ from the canonical motifs in those positions that are essential for recognition by the anti-Sm sera. The number of the core proteins described in this study is smaller than that observed in T. brucei (11). Either the L. collosoma SL RNP composition is less complex than that of T. brucei, or more than a single polypeptide migrates at the same position of the protein gel. The latter possibility is more likely, based on the differential staining intensity of the protein bands (Fig. 2, A and B). Cloning of the entire subset of the L. collosoma Sm proteins should clarify this issue.

Core proteins assembled on the Sm site of snRNAs were shown to serve as the recognition signal for the m3G cap methyltransferases and hence dictate the particular type of cap modification on the trypanosome snRNA.

Acknowledgments—We thank Ora Asher for excellent technical assistance and Dr. Christian Tschudi for advice in library screening with degenerate oligonucleotides.

REFERENCES

1. Agabian, N. (1990) Cell 61, 1157–1160
2. Ullu, E., Gunzl, A., and Tschudi, C. (1995) in Molecular Biology of Parasitic Protozoa, pp. 115–133, Tokyo: Oxford University Press, Oxford, NY
3. Mettman, J., Perry, K. L., Lizardi, P. M., Luhrmann, R., Agabian, N., and Nelson, R. G. (1989) Mol. Cell. Biol. 9, 1212–1223
4. Tschudi, C., and Ullu, E. (1990) Cell 61, 459–466
5. Dungan, J. D., Watkins, K. P., and Agabian, N. (1996) EMBO J. 15, 4016–4029
6. Xu, Y.-X., Ben-Shlomo, H., and Michaeli, S. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 8473–8478
7. Tschudi, C., Richards, F. F., and Ullu, E. (1986) Nucleic Acids Res. 14, 8893–8903
8. Branlant, C., Krol, A., Ebel, J., Lazar, E., Haendler, B., Jacob, M., Sri-Widada, J., and Jeanteur, P. (1988) EMBO J. 1, 1259–1263
9. Lerner, E. A., Lerner, M. R., Janeway, C. A. J., and Steitz, J. A. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 2737–2741
10. Michaeli, S., Roberts, T. G., Watkins, K. P., and Agabian, N. (1997) J. Biol. Chem. 265, 10582–10588
11. Palfi, Z., Gunzl, A., Cross, M., and Bindereif, A. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 9097–9101
12. Lücke, S., Klockner, T., Palfi, Z., Boshart, M., and Bindereif, A. (1997) EMBO J. 16, 4433–4440
13. Hermann, H., Fabrizio, P., Raker, V. A., Foulaik, K., Hornig, H., Brahms H., and Luhrmann, R. (1995) EMBO J. 14, 2076–2088
14. Séraphin, B. (1995) EMBO J. 14, 2089–2098
15. Cooper, M., Johnston, L. H., and Beggs, J. D. (1995) EMBO J. 14, 2066–2075
16. Cross, M., Wieland, B., Palfi, Z., Gunzl, A., Rothlisberger, U., Lahm, H.-W., and Bindereif, A. (1993) EMBO J. 12, 1239–1248
17. Goldring, A., Karchi, M., and Michaeli, S. (1995) Exp. Parasitol. 80, 333–338
18. Judd, R. C. (1994) in Methods in Molecular Biology (Walker, J. M., ed) pp. 49–58, Humana Press Inc., Totowa, NJ
19. Goldring, A., and Michaeli, S. (1995) Gene 156, 139–144
20. Cross, M., Gunzl, A., Palfi, Z., and Bindereif, A. (1991) Mol. Cell. Biol. 11, 5516–5526
21. Palfi, Z., and Bindereif, A. (1992) J. Biol. Chem. 267, 20159–20163
22. Matthews, K. R., Tschudi, C., and Ullu, E. (1994) Genes Dev. 8, 491–501
23. Bordonn, B., and Tarassov, I. (1996) Gene (Amst.) 176, 111–117
24. Camasses, A., Bragado-Nilsson, E., Martin, R., Séraphin, B., and Bordonné, R. (1996) Mol. Cell. Biol. 16, 4593–4603
25. Mattaj, I. W. (1986) Cell 46, 965–971
26. Ullu, E. and Tschudi, C. (1995) J. Biol. Chem. 270, 20365–20369
27. Stanford, D. R., Kehl, M., Perry, C. A., Holicky, E. L., Harven, S. E., Rohleder, A. M., Rehder, K. Jr., Luhrmann, R., and Wieben, E. D. (1988) Nucleic Acids Res. 16, 10595–10605
28. Fauthsch, M. P., Thompson, M. A., Holicky, E. L., Schultz, P. J., Hallett, J. B., and Wieben, E. D. (1992) Genomics 14, 883–890