mRNA cap formation in trypanosomatid protozoa is mediated through trans-splicing of the capped spliced leader (SL) sequence of the SL RNA onto the 5′ end of all mRNAs. The SL RNA cap structure in Trypanosoma brucei is unique among eukaryotes and consists of 7-methylguanosine (m7G) followed by four methylated nucleotides (cap 4): m7Gpppm2Gppp4Mpm3Um. Using transcriptional arrest in permeable T. brucei cells, we have analyzed the temporal progression of cap 4 formation on the 140-nucleotide-long SL RNA. m7G capping of the SL RNA could be detected on prematurely terminated SL RNA transcripts of 56 nucleotides in length and longer. Subsequent modifications characteristic of the SL RNA cap 4 were added successively in a 5′ to 3′ direction and appeared to be independent of core ribonucleoprotein formation. Transcripts between 56 and 67 nucleotides in length were partially modified and carried methyl groups on the first two adenosine residues, whereas a fully modified cap 4 structure was present on transcripts arrested at position 117 and beyond. Taken together, our results are consistent with a cotranscriptional mechanism for generating the cap 4 structure on the SL RNA.

The m7G cap at the 5′ end of eukaryotic mRNAs is an essential modification that directs mRNAs to the processing and transport pathways in the cell nucleus and regulates mRNA turnover and translation initiation (1). In all eukaryotes examined, capping is the first detectable RNA-processing event, occurring cotranscriptionally by the time the transcript is 25 to 30 nucleotides long (2–5). Capping takes place by a series of three consecutive enzymatic reactions: 5′ RNA triphosphatase generates a diphosphate terminus, which becomes a substrate for the addition of a guanosine residue through a 5′-5′ triphosphate linkage by RNA guanylyltransferase, and finally, the cap is methylated at the N7 position of guanine by RNA (guanine-7) methyltransferase (6, 7). This results in the structure m7GpppN, or cap 0, which often is further modified by the addition of 2′-O-methyl groups to the first and second transcribed nucleotide, leading to cap 1 and cap 2 structures. Whereas the mechanism of cap 0 formation has been studied in great detail, the biogenesis of the more complex cap structures remains largely unexplored.

In contrast to other eukaryotic organisms, mRNA cap formation in trypanosomatid protozoa occurs by a post-transcriptional RNA processing event. During RNA maturation, trans-splicing transfers the spliced leader (SL) sequence and its cap from the SL RNA to the 5′ end of all mRNAs. Direct analysis of the SL RNA 5′ end in Trypanosoma brucei and Crithidia fasciculata revealed that the 5′-terminal m7G residue is followed by four methylated nucleotides, forming an unusual cap 4 structure: 7-methylguanosine-ppp-N6,N6,2′,2′-O-trimethylenosine-p-2′-O-methyladenosine-p-2′-O-methylcytosine-p-N3,2′-O-dimethyluridine (8). We have previously shown that proper modification of the SL cap 4 structure is essential for utilization of the SL RNA in transcription and, therefore, essential for the synthesis of mature mRNAs (9). However, these modifications are not required for the assembly of the SL RNA into a core ribonucleoprotein particle (RNP) nor for the stability of the SL RNA or for the proper folding of the SL RNA in vivo (9).

To further our understanding of cap 4 formation on the T. brucei SL RNA, we have investigated the temporal progression of cap synthesis in vivo. Using transcriptional arrest induced by 3′-O-methyl-GTP, we analyzed the 5′ ends of prematurely terminated SL RNA transcripts in permeable cells. Our data are consistent with a cotranscriptional event for both initial m7G cap addition and subsequent methylation of base and sugar moieties of the first four transcribed nucleotides resulting in a cap 4 structure.

EXPERIMENTAL PROCEDURES

Transcription in T. brucei Permeable Cells—Procyclic T. brucei YTat7.1 cells were permeabilized with lysolecithin as described previously (9, 10). Prematurely terminated SL RNA species were generated by adding 3′-O-methyl-GTP instead of GTP to the transcription reaction at a final concentration of 200 μM. After incubation at 26 °C for 10 min, total RNA was extracted with Trizol reagent (Life Technologies, Inc.) and stored as an ethanol precipitate for further analysis. Affinity Selection of SL RNA Transcripts—Total RNA was subjected to hybrid selection with the RNA oligonucleotide GM01 (5′-GGAGCUUCUCAUAC5555A-3′). The abbreviations used are: m7G, 7-methylguanosine; SL, spliced leader; RNP, ribonucleoprotein particle; PIPES, 1,4-piperazinediethanesulfonic acid; nt, nucleotide(s).

Received for publication, May 16, 2000, and in revised form, June 29, 2000
Published, JBC Papers in Press, July 3, 2000, DOI 10.1074/jbc.M004193200

Gunnar Maier, Elisabetta Ullu,§§, and Christian Tschudi¶¶
From the ‡Departments of Internal Medicine and §Cell Biology, Yale University School of Medicine, New Haven, Connecticut 06520-8022

The abbreviations used are: m7G, 7-methylguanosine; SL, spliced leader; RNP, ribonucleoprotein particle; PIPES, 1,4-piperazinediethanesulfonic acid; nt, nucleotide(s).
Trypanosome Cap 4 Biosynthesis

28995

500 μl of WB400 each, bound RNAs were eluted with 200 μl of 2 mM EDTA by incubation at 85 °C for 10 min with occasional mixing. If residual counts were present in the bead pellet, the elution procedure was repeated with 100 μl of 2 mM EDTA. Eluted RNA was precipitated by adding 10 μl of 20× SET (3 M NaCl, 20 mM EDTA, 0.6 M Tris-HCl), 10 μg of glycogen, and 2.5 volumes of ethanol.

Tobacco Acid Pyrophosphatase Cleavage—Oligonucleotide-selected RNAs were incubated in the presence of 2 units of tobacco acid pyrophosphatase (Sigma) for 60 min in 1× TAPP-buffer (50 mM sodium acetate (pH 5), 1 mM EDTA, 10 mM β-mercaptoethanol). Digestion products were displayed on denaturing 8% polyacrylamide gels.

Immunoprecipitations—Cell extracts for immunoprecipitations were prepared from permeable cells by lysis in 100 μl of 100 mM EDTA by incubation at 85 °C for 10 min, and the supernatant was then incubated with antibodies against the T. brucei common proteins (11) bound to protein A-Sepharose. Total RNA was subjected to immunoprecipitations with rabbit polyclonal antibodies against the mG cap (a kind gift of Dr. F. Richards). After incubation at 4 °C for 2 h, the beads were washed 10 times with NET-2, and RNA was prepared with Trizol reagent (Life Technologies, Inc.).

T2 Assay and Thin Layer Chromatography Analysis of the cap 4 Structure—Full-length SL RNA as well as prematurely terminated SL RNA species were excised from denaturing polyacrylamide gels and eluted at 120 °C overnight with 400 μl of water in the presence of 20 μg of tRNA. Eluted RNAs were purified through SPINX centrifuge filters (Costar), precipitated with ethanol, and digested with 4 μl each of RNase A (800 μg/ml) and RNase T2 (80 units/ml) overnight at 37 °C in a total volume of 50 μl of 50 mM ammonium acetate (pH 4.5) and 2 mM EDTA. Reactions were lyophilized and resuspended in 50 μl of water, and 500 μl of 10X ammonium formate (pH 8) was added. Digestion products were applied to a 100-μl DEAE-Sepharose CL-6B column equilibrated with 10 mM ammonium formate. RNA fragments were step-eluted with 3 times 250 μl each of 0.2, 0.3, 0.35, 0.4, and 0.5 mM ammonium formate. An aliquot of the eluate was fractionated on 8% polyacrylamide gel. Fragments containing T2-resistant fragments were pooled and lyophilized.

RESULTS

Transcriptional Arrest with 3′-O-Methyl-GTP Produces Distinct, Prematurely Terminated SL RNA Transcripts in Permeable Cells—Terminating ribonucleotide analogs have been used in several groups to examine RNA polymerase pausing and 5′ cap formation in vitro and in nuclear run-on assays (2, 3). We decided to apply this approach to permeable trypanosome cells and investigate the progression of cap 4 formation on SL RNA transcripts. To test whether we could prematurely terminate SL RNA transcription, we incubated procyclic trypanosomes permeabilized with lyssolecithin (10) in the presence of 3′-O-methyl-GTP. Fig. 1A shows a profile of total α-32P-labeled RNA synthesized in a 10-min incubation period (lane 1). The most abundant transcript is the SL RNA of 140 nucleotides, for which we have previously shown that 50 to 70% carries the modifications characteristic of the cap 4 structure. We found that the addition of 200 μM 3′-O-methyl-GTP to permeable cells significantly altered the profile of newly synthesized RNAs (Fig. 1A, lane 3). In particular, we observed a marked decrease in the accumulation of tRNA-size molecules and the appearance of a series of discrete transcripts in the SL range of 40 to 150 nucleotides. To analyze the origin of these RNA species, total RNA samples were subjected to gel electrophoresis alongside a dideoxy-GTP sequencing ladder of the cloned SL RNA gene revealed essentially the same pattern of bands and, thus, allowed us to assign each RNA species to a termination event at a particular G residue (Fig. 1B).

The difference in mobility between the RNA and DNA ladder most likely reflects the presence of modified nucleotides in the RNA fragments (see below). Taken together, our results are consistent with the hybrid-selected transcripts being prematurely terminated SL RNAs ending at G-5s. To further confirm the identity of the selected RNAs, we employed site-directed cleavage with RNase H and a DNA oligonucleotide complementary to the very 5′ end of the SL RNA (nucleotides 1–11). This resulted in the specific cleavage of all selected RNA species, except the one corresponding to the SL RNA intron of 100 nucleotides (data not shown).

Cotranscriptional m′G Capping of the SL RNA—The 5′ end of the SL RNA is capped with a 7-methyl GMP moiety, which is linked to the first transcribed nucleotide by a 5′-5′ pyrophosphate bond. To directly look at cap formation in 3′-O-methyl-GTP-treated trypanosome cells, oligonucleotide-selected transcripts were digested with tobacco acid pyrophosphatase, an enzyme capable of releasing GMP by cleaving the pyrophosphate bond (Fig. 2A). As a result, pyrophosphate-treated transcripts should move slightly faster than untreated control transcripts, if properly capped. All selected SL RNA transcripts,

FIG. 1. Prematurely terminated SL RNA transcripts in permeabilized T. brucei cells. A, RNA was synthesized in the absence (lane 1) or presence (lane 3) of the transcription terminator 3′-O-methyl-GTP and [α-32P]GTP and hybrid-selected with an oligonucleotide complementary to nt 40 to 54 of the SL RNA (lanes 2 and 4). RNA samples were separated on a 6% denaturing polyacrylamide gel. The position of the SL RNA and the SL intron is indicated. M, [α-32P]-labeled Msgl digest of pBR322. B, oligonucleotide-selected transcripts (lane 1) were electrophoresed alongside an 85S-labeled dideoxy-GTP DNA sequencing ladder of the SL leader gene generated with an oligonucleotide primer complementary to the very 5′ end (lane 2). Products were resolved on a 8% denaturing polyacrylamide gel. Asterisks indicate RNA bands not corresponding to transcript termination sites at G residues; they were not seen consistently.
The presence of 3'-O-methyl-GTP (lane 1) was immunoprecipitated with lane 1 on a 6% denaturing polyacrylamide gel. Anti-m7G cap specific antibodies (lane 2) and electrophoresed lane 2 -methyl-GTP-terminated RNA is provided for comparison (lane 9). Result suggested that the 111-nt-long transcript was partially synthesized in the presence of 3'-O-methyl-GTP, was gel-purified after digestion with RNase T2, the T2-resistant fragments (T2R) were analyzed by electrophoresis through a 20% denaturing polyacrylamide gel. Nucleotide 3'-monophosphates (Nps) are indicated.

Except the linear SL intron, were sensitive to this treatment, resulting in a slight but noticeable increased electrophoretic mobility (compare lanes 1 and 2). Next, we analyzed the methylation status of the cap nucleotide by performing immunoprecipitations with an antibody specific for the m7G residue. As shown in Fig. 2B, immunoprecipitations with anti-m7G antibodies (lane 2), but neither control precipitation (lanes 3 and 4), displayed the same transcripts that were shown by our oligonucleotide selection procedure to represent prematurely terminated SL RNA transcripts (lane 5). Thus, according to these two assays transcripts of 56 nucleotides in length or greater were capped at the 5' end with an m7G residue.

**Cap 4 Formation Takes Place On Nascent Transcripts**—The trypanosomatid SL RNA 5' end is unusual in that, in addition to the m7G cap nucleotide, it contains four consecutive modified nucleotides, forming a cap 4 structure: m7G(5')ppp(5')NpNpNpN, 2'-O-trimethyladenosine-p-2'-O-methyladenosine-p-2'-O-methylcytosine-p-2'-O-dimethyluridine (8). Having established that formation of the m7G cap nucleotide occurs on transcripts as short as 56 nucleotides, we next asked whether hypermethylation of the SL RNA 5' end also takes place cotranscriptionally. The presence of 2'-O modifications at positions 1 to 4 of the SL RNA 5' end is revealed by digestion with ribonuclease T2, which cleaves RNA at every position, except pyrophosphate bonds or 5' bonds adjacent to 2'-O-modified nucleotides. Thus, digestion of fully modified SL RNA with ribonuclease T2 gives rise to a characteristic T2-resistant fragment with the sequence m7GpppAUCUA, which can be displayed on a denaturing 20% polyacrylamide gel (8, 12). Indeed, when full-length SL RNA, synthesized in the presence of 3'-O-methyl-GTP, was gel-purified and digested to completion with T2, a T2-resistant fragment characteristic of the cap 4 structure was obtained (Fig. 3, lane 2). Surprisingly, a similar analysis with SL transcripts terminated at position 111 also gave rise to a T2-resistant fragment, albeit with an increased electrophoretic mobility (Fig. 3, lane 3). This result suggested that the 111-nt-long transcript was partially modified at the 5' end and raised questions about the timing and progression of cap 4 formation. To address this issue, RNA synthesis in permeable cells was carried out in the presence of three α-32P-labeled nucleotide triphosphates ([α-32P]ATP, [α-32P]CTP, and [α-32P]UTP) and 3'-O-methyl-GTP, and the precise nucleotide composition at the 5' end of three different size classes (indicated in Fig. 1B) of prematurely terminated SL RNA transcripts was determined. The transcripts are named according to the last G residue they incorporated: G67 refers to a pool of transcripts terminating between positions 56 and 67, G111 includes transcripts terminating between positions 109 and 111, and G127 represents a pool of transcripts terminating between position 117 and 127. Following gel purification and T2 ribonuclease digestion, T2-resistant fragments were separated from mononucleotides by chromatography on DEAE-Sephacel columns (see “Experimental Procedures”).

Both G67 and G111 transcripts revealed a T2-resistant fragment that migrated faster in a 20% denaturing polyacrylamide gel than the one characteristic of the cap 4 structure (Fig. 3, lane 3, and data not shown). On the other hand, treatment of G127 RNA with RNase T2 resulted in two resistant fragments: one comigrated with the G67 and G111 fragment (G127-short), whereas the other (G127-long) had the same mobility as the one originating from a fully modified cap 4 structure (data not shown). Next, the various T2-resistant fragments were further digested to mononucleotides with nucleotide pyrophosphatase and nuclease P1 and separated by two-dimensional thin layer chromatography (Fig. 4 and Table I). As a control, the T2-resistant fragment characteristic of mature SL RNA was purified from total T. brucei RNA and processed similarly (panel E). Consistent with previously published TLC analysis (12, 13), the five spots of the cap 4 structure were identified as pm5Amp, pAmp, pCmp, pm7Um, and pA (note that pm7G is not visible here, since we did not include [α-32P]GTP in the transcription reaction). A sixth minor spot (pC), which accounts for less than 5% of the total C residues was also noted, but its origin is unclear. As shown previously (12) the intensity of the labeled nucleotide spots does not correlate with their representation in the SL cap 4 structure, which should theoretically be 1:1:1:1:1. This is caused by a difference in the pools of endogenous ribonucleotide triphosphates, which leads to different specific activities for each radiolabeled nucleotide. Indeed, in total RNA the ratio of [α-32P]CTP: [α-32P]UTP: [α-32P]ATP was found to be 9.5: 1.8:1. Correcting for the unequal specific activity of radiolabeled nucleotides, we deduced the following structures for the various T2-resistant fragments. As predicted from the size of the T2-resistant fragment, TLC analysis of G127-long gave rise to nucleotides with mobilities and intensities comparable with...
those of control RNA (compare panels D and E in Fig. 4). Thus, a proportion of transcripts arrested between positions 117 and 127 carried a fully modified cap 4 structure. In contrast, G127-short displayed seven nucleotide spots (one more than in control RNA), and the relative intensities of several spots were significantly different when compared with control RNA (compare panels C and E). In particular, 80% of cytidine was present as unmodified pC, and only 20% was modified to pCm. In contrast, in control RNA about 95% of cytidine was present as unmodified pC, and only 20% was modified to pCm.

**DISCUSSION**

Here we describe the temporal progression of cap 4 formation on the SL RNA in permeable *T. brucei* cells by taking advantage of the incorporation of the transcription terminator 3′-O-methyl-GTP into the growing RNA chain. This allowed us to arrest transcription at guanosine residues downstream of the transcription start site and then to determine the state of modification at the 5′ end of these transcripts.

Using oligonucleotide-selected RNAs, we found that SL RNA transcripts of 56 nucleotides or greater contain an m7G cap structure (Fig. 2, A and B), and immunoprecipitations with anti-m7G antibodies indicated that transcripts as short as 30 nucleotides are capped (Fig. 2 and data not shown). These results are consistent with previous reports on mRNA cap formation in other systems. In a *Drosophila* nuclear run-on analysis cap addition was observed on transcripts between 20 and 30 nucleotides in length (3), and similarly, in a vaccinia virus system Hagler and Shuman report capping on nascent RNA chains 31 nucleotides in length or longer (2). More recent studies have explored the implications of these findings and demonstrated that formation of the cap early in transcription is mediated by recruitment of the capping machinery to the phosphorylated carboxyl-terminal domain of the largest subunit of RNA polymerase II (15–18). At present the RNA polymerase II (15–18). At present the RNA polymerase II (15–18).
Trypanosome Cap 4 Biosynthesis

Quantitation of cap 4 constituents of 3'-O-methyl-GTP terminated SL RNA transcripts

Panels A–E shown in Fig. 4 were scanned with phosphorimaging, and the relative intensities of the various spots are given in arbitrary units (first number). The numbers in parentheses are corrected for the unequal specific activity of radiolabeled nucleotides (see "Results").

| Cap 4 Formation | G67 | G111 | G127-short | G127-long |
|-----------------|-----|------|------------|-----------|
| pA              | 9   | 14   | 45 (428)   | 56 (532)  |
| pAm             | 98  | 100  | 188 (1786)| 65 (618)  |
| pm6Am (?)       | 12  | 09   | 0 (0)      | 0 (0)     |
| pm3Am (?)       | 27  | 23   | 37 (352)   | 41 (390)  |
| pC              | 842 | 820  | 559 (539)  | 39 (38)   |
| pCm             | 121 | 34   | 122 (122)  | 724 (724) |
| pU              | 0   | 0    | 7 (37)     | 0 (0)     |
| pm6Um           | 0   | 0    | 21 (111)   | 75 (398)  |
| pm3Um           | 0   | 0    | 0 (0)      | 94 (498)  |

Fig. 5. Assembly of SL RNA transcripts into a core SL RNP. Total cell lysates of 3'-O-methyl-GTP-treated trypanosomes were immunoprecipitated with anti-common protein antibodies. This result helps to clarify an interesting aspect we observed previously in our in vitro cap 4 modification system (12). It appeared that cap 4 formation required the SL RNA substrate to be in an RNP, although at the time we did not exclude the possibility that a structural determinant or a specific sequence at the 5' end becomes masked in deproteinized SL RNA. The observation in the present study that the onset of cap 4 formation precedes binding of common proteins would argue for the latter scenario. Consistent with this conclusion is an in vivo mutational analysis of the Leptomonas SL RNA (21). Sequence elements necessary for cap 4 modification were found to reside exclusively within the modification site itself and stem loop I, whereas substitution or deletion of stem loops II and III had no effect on cap formation. The observation that common protein binding occurs when the SL RNA is at least 112 nt long suggests that the synthesis of the structural determinant for common protein binding is not yet completed in transcripts shorter than 112 nt. Interestingly, transcripts shorter than 112 nt do not contain the entire single-stranded region between nt 110 and 120 of the SL RNA, which has been proposed to be the region analogous to the Sm binding site of spliceosomal U small nuclear RNAs of other eukaryotic organisms. Last, our results have implications for the biogenesis of the SL RNA. Since cap 4 synthesis and assembly with the SL RNP proteins occurs on prematurely terminated transcripts, SL RNP biogenesis most likely occurs in the nucleus, as we have previously suggested.

Acknowledgment—We thank Albrecht Bindereif for providing anti-common protein antibodies.

REFERENCES

1. Varani, G. (1997) Structure (London) 5, 855–858
2. Hagler, J., and Shuman, S. (1992) Science 255, 983–986
3. Rasmussen, E. B., and Lis, J. T. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 7923–7927
4. Coppola, J. A., Field, A. S., and Luse, D. S. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 1251–1255
5. Jove, R., and Manley, J. L. (1984) J. Biol. Chem. 259, 8513–8521
6. Mizumoto, K., and Kaziro, Y. (1987) Prog. Nucleic Acid Res. Mol. Biol. 34, 1–28
7. Shuman, S. (1995) Prog. Nucleic Acid Res. Mol. Biol. 50, 101–129
8. Bangs, J. D., Crain, P. F., Hashizume, T., McCloskey, J. A., and Boothroyd, J. C. (1992) J. Biol. Chem. 267, 9805–9815
9. Ullu, E., and Tschudi, C. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 10774–10778
10. Ullu, E., and Tschudi, C. (1990) Nucleic Acids Res. 18, 3319–3326
11. Palfi, Z., and Bindereif, A. (1992) J. Biol. Chem. 267, 20159–20163
12. Ullu, E., and Tschudi, C. (1995) J. Biol. Chem. 270, 20365–20369
13. Perry, K. L., Watkins, K. P., and Agabian, N. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 8190–8194
14. Silberklang, M., Gillum, A. M., and RajBhandary, U. L. (1979) Methods Enzymol. 59, 58–109
15. Cho, E. J., Takagi, T., Moore, C. R., and Buratowski, S. (1997) Genes Dev. 11, 3319–3326
16. McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessell, A., Foster, S., Program, A. E., Shuman, S., and Bentley, D. L. (1997) Genes Dev. 11, 3306–3318
17. Yue, Z., Maldonado, E., Pillutla, R., Cho, H., Reinberg, D., and Shatkin, A. J. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 12898–12903
18. Ho, C. K., Sriskanda, V., McCracken, S., Bentley, D., Schwer, B., and Shuman, S. (1998) J. Biol. Chem. 273, 9577–9585
19. Gunzl, A., Ullu, E., Dorner, M., Fragoso, S. P., Hoffmann, K. F., Milner, J. D., Morita, Y., Nguu, E. K., Vanacova, S., Wünsch, S., Dare, A. O., Kwon, H., and Tschudi, C. (1997) Mol. Biochem. Parasitol. 85, 67–76
20. Yu, M. C., Sturm, N. R., Saito, R. M., Roberts, T. G., and Campbell, D. A. (1998) Mol. Biochem. Parasitol. 94, 265–281
21. Lücke, S., Xu, G. L., Palfi, Z., Cross, M., Bellofatto, V., and Bindereif, A. (1996) EMBO J. 15, 4380–4391
Cotranscriptional Cap 4 Formation on the Trypanosoma brucei Spliced Leader RNA
Gunnar Mair, Elisabetta Ullu and Christian Tschudi

J. Biol. Chem. 2000, 275:28994-28999.
doi: 10.1074/jbc.M004193200 originally published online July 3, 2000

Access the most updated version of this article at doi: 10.1074/jbc.M004193200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 21 references, 13 of which can be accessed free at
http://www.jbc.org/content/275/37/28994.full.html#ref-list-1