In vitro capping in *Trypanosoma cruzi* identifies and shows specificity for the spliced leader RNA and U-RNAs

Tadeusz A. Zwierzynski and Gregory A. Buck*

Department of Microbiology and Immunology, Medical College of Virginia Campus, Virginia Commonwealth University, Richmond, VA 23298, USA

Received March 21, 1990; Revised and Accepted June 8, 1990

ABSTRACT

Messenger RNA maturation in trypanosomes requires a *trans*-splicing event in which a capped 39 nucleotide leader sequence, the spliced leader (SL), from the 5' terminus of a small RNA (SL-RNA) is joined to the 5' termini of protein coding gene transcripts. We have developed nuclear extracts from *Trypanosoma cruzi* that label three small endogenous RNAs in the presence of [α - 32 P]GTP. Herein, we have characterized this labelling as 5' capping and shown that the capping activity exhibits an unusual ATP dependence. Moreover, partial sequence analysis identified the three cap-labelled RNAs as the *T. cruzi* SL-RNA, and two U-RNAs previously uncharacterized in *T. cruzi*, U2 and Ux. Finally, the capping reaction in the *T. cruzi* extracts showed apparent specificity for these RNAs—other endogenous or exogenous transcripts were not capped. The apparent specificity of this *in vitro* capping activity closely reflects the *in vivo* requirements; i.e., only the SL- and U-RNAs need to be capped since mature mRNAs are capped via *trans*-splicing. These observations are consistent with the hypothesis that one of the functions of *trans*-splicing is to supply 5' caps to mature trypanosome mRNAs.

INTRODUCTION

Many protein-coding RNAs in trypanosomes are transcribed as long 'multicistronic' transcripts that are processed into several distinct mature RNAs. This processing requires a *trans*-splicing event which places a common 39 nucleotide (nt) spliced leader (SL) at the 5' terminus of each mature mRNA (1–3). The mRNAs are also polyadenylated although sequences analogous to the hexanucleotide AAUAAA and downstream cleavage and polyadenylation signals of other eukaryotes have not been identified (4). Finally, the mechanism by which the trypanosome mRNAs obtain a 5' cap also differs from that of other eukaryotes; i.e., the caps are supplied from the SL via *trans*-splicing. The SL primary transcript (SL-RNA) is probably capped prior to *trans*-splicing of the SL to the mature mRNA (5–8).

In most eukaryotes, only RNA polymerase (Pol) II transcripts—mRNAs and U-RNAs—are capped; transcripts

synthesized by Pools I and III lack caps (9–11). The factors responsible for the specificity of guanylyl transferase (GTase) have not been identified, but are thought to reside within the Pol II transcription initiation complex. Although only Pol II transcripts are capped by cellular capping mechanisms, purified GTase non-specifically caps any RNA containing 5' di- or tri-phosphate. Thus, cellular GTases are bifunctional enzymes that carry out the following two step reaction:

where N is any ribonucleotide. Reaction 1 is mediated by a 5' triphosphatase (TPase) domain of the enzyme and reaction 2 is mediated by a GTase domain. Neither of these reactions requires ATP. In higher eukaryotes, methylation of the cap G and the 5'-most two or three nucleotides (nt) of the transcript completes the capping process.

The mechanisms by which the trypanosome capping machinery selects the SL-RNA for substrate are unknown. It is possible that all Pol II transcripts in trypanosomes are actively capped. However, we demonstrate herein that nuclear extracts from *Trypanosoma cruzi*, causative agent of Chagas' Disease, specifically caps three small endogenous RNAs. We have partially sequenced these RNAs directly and identified them as the *T. cruzi* SL-RNA and two previously uncharacterized *T. cruzi* U-RNAs, U2 and Ux [the latter was tentatively identified as U3 in *T. brucei* (12)]. Moreover, we have shown that this *in vitro* capping reaction shows unusual ATP dependence. Together, our data suggest that the capping activity in these nuclear extracts shows specificity for the SL- and U-RNAs. These observations are consistent with the hypothesis that one of the functions of the SL-RNA in these organisms is to provide a mechanism for capping mature mRNA via *trans*-splicing.

MATERIALS AND METHODS

Trypanosome culture, preparation of nuclear extracts and nuclear RNA

T. cruzi strain CL was grown to mid exponential phase (1×10^7 cells/ml) in LIT medium (13) at 28°C with moderate agitation

* To whom correspondence should be addressed

and the cultures were split every 24 hrs. Nuclear extracts were prepared following modifications of the procedure (4) described for mammalian cells by Dignam et al. (14). Briefly, washed trypanosomes were swollen for 10 min at 0°C in buffer A (14) and homogenized twice for 20 sec at full power in an Omnimixer (Omni Corporation). Lysis was monitored microscopically and the nuclei were pelleted and subsequently suspended and extracted in buffer C (14). Debris was removed by centrifugation and the extract was dialyzed, aliquoted and stored at -70°C as described (4). The final concentration of protein in the extracts was approximately 18 mg/ml as determined with the Bio-Rad Protein Assay using bovine plasma γ -globulin as standard protein. Total nuclear RNA was purified from the nuclear extracts by extraction once with phenol and once with chloroform followed by ethanol precipitation.

Capping reactions in the nuclear extracts and electrophoresis of the products

Capping reactions were performed for 30 min at 30°C in 25 μ l reactions containing 10 μ l nuclear extract, 2.5% (vol/vol) polyvinyl alcohol, 3.2 mM $MgCl_2$, 0.08 mM EDTA, 0.2 mM DTT, 0.5–1 mM ATP, 30 μ Ci [α - ^{32}P]GTP (3000 Ci/mmol, ICN), 0–20 mM creatine phosphate, 8% glycerol, 40 mM KCl, 12 mM Hepes (pH 7.9) and 0.2 mM PMSF. Reaction conditions varied as described in the figure legends. After incubation, the reactions were diluted to 300 μ l with 0.3 M sodium acetate (pH 5.3), extracted once with phenol and once with chloroform, precipitated with 95% ethanol, and analyzed by electrophoresis in 0.5 $\times$ TBE (15) at 15–20 watts constant power in 8% polyacrylamide/7 M urea gels (20 $\times$ 40 $\times$ 0.04 cm) run until the bromophenol blue dye reached the bottom of the gel (2–2.5 hrs). The gels were exposed to Fuji RX film at -70°C for 2–3 hrs or at -20°C for 12 hrs. For quantitative analysis the [^{32}P]labelled bands were excised and counted in a scintillation counter as Cerenkov cpm. Silver staining was carried out as suggested by the manufacturer (BioRad). The stained gels were photographed, dried at 50°C and subjected to autoradiography as described above.

Nuclease digestions and thin layer chromatography (TLC)

For analytical TLC of the nucleotides in the labelled RNAs the capping reactions were scaled up 2–3 fold and the RNA was fractionated in denaturing gels as above. The labelled RNAs were excised, eluted overnight at 25°C in 0.5 M sodium acetate, 2 mM EDTA, 2 mM $MgCl_2$ and 0.1% SDS, and ethanol precipitated. The purified RNAs were digested for 12 hr at 37°C with 1 μ g nuclease P1 (BRL) in 5 μ l 25 mM ammonium acetate (pH 5.3) and lyophilized for TLC. Enzymatic decapping of the P1 products was carried out with 2 units of tobacco acid pyrophosphatase (TAP, Promega) for 1 hr at 37°C in 10 μ l of 50 mM sodium acetate (pH 5.0), 10 mM 2-mercaptoethanol, 1 mM EDTA. Digestion products were analyzed by 2D TLC on 10 $\times$ 10 or 20 $\times$ 20 cm glass cellulose plates without indicator (EM Merck) using solvent systems 'a' [vertical] and 'c' [horizontal] as described in Silberklang et al. (16). The pm⁷G and pG standards were prepared by digesting ~2 A₂₆₀ units of m⁷GpppG cap structure (New England Biolabs) with 2 μ g snake venom phosphodiesterase (SVP, Boehringer) for 12 hrs at 37°C in 20 μ l 50 mM ammonium acetate buffer adjusted to pH ~8.0 with ammonium hydroxide. Cap dinucleotides GpppA, m⁷GpppA, m⁷GpppAm, GpppG and m⁷GpppG (Pharmacia) were also used as standards for TLC.

RNA sequencing reactions

The capping reactions were scaled up 25 fold and labelled G1A, G1B and G2 RNAs were gel purified as described above and sequenced by RNase digestion (17–19). In brief, approximately 500 Cerenkov cpm of each purified RNA was digested for 15 min at 55°C with RNases T1 (Boehringer), U2 or PhyM (BRL) in 5 μ l of 20 mM sodium acetate (pH 5.0), 7 M urea, 1 mM EDTA, 0.05% xylene cyanole, 0.05% bromophenol blue. Similar reactions with RNase CL3 (BRL) were carried out for 15 min at 37°C in 10 mM sodium phosphate (pH 6.0), 20 mM KCl, 10 μ M EDTA and 2 μ l formamide dye (15) was added prior to loading the sequencing gel. Alkaline hydrolysis was carried out by incubation of the RNAs in 5 μ l of formamide at 100°C for 30 min (20). No exogenous carrier was added to the reactions as other unlabelled RNA that co-purified with the labelled species was present. Sequencing gels were run as described in Maniatis et al. (15) and autoradiography was performed at -70°C for 7–10 days with intensifying screens.

RNA synthesis, in vitro capping with vaccinia virus GTase, chemical decapping and 3' labelling with [^{32}P]pCp

A 59 nt transcript was synthesized from EcoRI cleaved pGEM2 (Promega) using T7 RNA polymerase according to the manufacturers instructions. This synthetic RNA and chemically decapped (see below) and untreated nuclear RNA were capped with vaccinia virus GTase (BRL) in 15 μ l reactions containing approximately 4 μ g of RNA, 2 units of v-GTase, 50 mM Tris-HCl (pH 7.9), 1.25 mM $MgCl_2$, 6 mM KCl, 2.5 mM DTT, and 100 μ Ci [α - ^{32}P]GTP (3000 Ci/mmol, ICN). The reactions were incubated for 45 min at 37°C. Nuclear RNA was chemically decapped in 40 μ g batches by NaIO₄ oxidation and β -elimination essentially as described in Rose and Lodish (21) using reaction volumes proportional to the amount of RNA substrate. *T. cruzi* nuclear RNA was 3' end-labelled with 200 μ Ci 5'-[^{32}P]pCp and T4 RNA ligase (New England Biolabs) as described in England et al. (22). 5'-[^{32}P]pCp was prepared from Cp (Sigma) with [γ - ^{32}P]ATP (6000–7000 Ci/mmol, NEN) and T4 polynucleotide kinase (Promega) as described by Walseth and Johnson (23).

RESULTS

Three small endogenous RNAs are capped in *T. cruzi* nuclear extracts

Nuclear extracts were prepared from culture-grown *T. cruzi* following modifications of protocols developed for mammalian cells [(14); see Materials and Methods]. Incubation of these extracts in the presence of ATP and [α - ^{32}P]GTP resulted in labelling of three discrete endogenous RNA molecules; G1A, G1B and G2 (Fig. 1A). G1A and G1B migrated with apparent sizes of 145–150 nt and G2 migrated with an apparent size of ~110 nt. In addition, several small RNAs 30–40 nt in length (GF) were labelled in these reactions.

To determine the nature of the labelling observed in these extracts, G1A, G1B, G2 and GF were purified from the gel and analyzed by P1 nuclease digestion and two dimensional thin layer chromatography (TLC). The labelled digestion products (Fig. 1B) did not migrate with any of the four mono-nucleotides (pNs). Further digestion of the P1 products with calf intestine phosphatase did not alter their mobilities (data not shown), demonstrating that the [α - ^{32}P] from the labelled GTP is not

Figure 1. In vitro capping in *T. cruzi* nuclear extracts. **A.** Labelling of endogenous RNAs with [α - 32 P]GTP. Nuclear extracts were incubated for 30 min at 30° (lane 1) or 37°C (lane 2) in the presence of 30 μ Ci [α - 32 P]GTP, 0.5 mM ATP and 20 mM creatine phosphate and the labelled products were analyzed as described in Materials and Methods. G1A, G1B, G2 are RNAs labelled in the reaction. GF shows the location of several RNAs 35–40 nt in length that are also labelled in these reactions. XC shows the position of the xylene cyanole dye after electrophoresis. **B.** TLC analysis of P1 nuclease digested G1A, G1B, G2 and GF. The RNAs were extracted from gels, digested with P1 nuclease, and the labelled products were analyzed by TLC as described in the Materials and Methods. (o) denotes origin of TLC. First (1) and second (2) dimensions are indicated. The positions of the mononucleotides pA, pG, pU and pC are labelled on the panel of P1 products from G1A and indicated on the remaining panels. Arrows show positions of faint spots on these exposures. **C.** Analysis of the putative cap di-nucleotides after digestion with tobacco acid pyrophosphatase (TAP). G1A, G1B, G2 and GF were isolated and treated with nuclease P1 as for **B**, and the products treated with TAP as described in the Materials and Methods and analyzed by TLC. The positions of the mono-nucleotides are indicated as in **B**. Arrows show the position of pm⁷G. **D.** G1A, G1B, G2, and GF were treated with nuclease P1 as in **B** and analyzed by one dimensional TLC on 20×20 cm plates in solvent 'a' as described in Materials and Methods. The positions of the standards are shown. GpppG and m⁷GpppG migrated near the origin and are therefore not indicated. **E.** Labelled G1A (lanes 1, 4, 7), G1B (lanes 2, 5, 8), G2 (lanes 3, 6, 9) were isolated after labelling in another nuclear extract preparation digested with nuclease P1, P1 plus SVP, or TAP and analyzed by TLC as in **D**. Asterisk denotes position of the pm⁷G marker.

present as a free 5' phosphate. Moreover, when the P1 products were digested with tobacco acid pyrophosphatase (TAP), which degrades GpppN cap into $\dot{p}G + pN + P_i$ (24), and analyzed by two dimensional TLC, the labelled species co-migrated with [32 P]pG or [32 P]pm⁷G (Fig. 1C, arrows). G2 and GF consistently showed a higher proportion of m⁷G than G1A or G1B, although the degree of methylation varied with each extract (see Fig. 1E, lanes 6,9). These observations verify that G1A, G1B, G2 and GF are capped and partially methylated in the *T. cruzi* nuclear extracts.

The P1 products from digestions of G1A and G1B were identical and migrated as expected for m⁷GpppA (Fig. 1D and E, lanes 1,2). However, TAP treatment of these RNAs yielded

labelled pG but little labelled pm⁷G (Fig. 1C), and a similar result was obtained after digestion of these RNAs with both P1 nuclease and snake venom phosphodiesterase (SVP, Fig. 1E, lanes 4,5). Together, these observations argue that the cap structure on G1A and G1B is GpppA*; i.e., a modification (methylation?) of adenosine nucleoside causes the structure to co-migrate with m⁷GpppA. The labelled G2 and GF digestion products were also similar to each other but differed from those of G1A and G1B. The major P1 products of G2 and GF co-migrated with modified cap structures; i.e., m⁷GpppAm and other unidentified highly modified cap structures (Fig. 1D). However, TAP or P1 plus SVP digestions of G2 and GF indicated that only a minority of these RNAs contain m⁷G (Fig. 1C and

1E, lanes 6 and 9), suggesting that the adenosine nucleoside is indeed highly modified.

These observations demonstrate that G1A, G1B, G2 and GF are capped *in vitro* in the *T. cruzi* nuclear extracts and suggest that the resulting cap structures of G2 and GF are more highly modified than those of G1A and G1B. We have not yet fully characterized the nature and extent of the modifications observed on these cap structures.

G2 is the *T. cruzi* SL-RNA

The SL-RNA in *T. cruzi* is ~110 nt in length (2,3)—the same length as G2—suggesting the intriguing possibility that G2 may be the SL-RNA. To test this possibility, *in vitro* [α - 32 P]GTP labelled (capped) G1A, G2 and GF were extracted from a polyacrylamide gel, hybridized with a 35 residue synthetic oligonucleotide complementary to the *T. cruzi* SL and digested with RNase H (Fig. 2). G2 and GF were selectively degraded by RNase H in the presence of the anti-SL oligonucleotide (lanes 8 and 12). In contrast, control reactions in the absence of RNase H and the oligonucleotide (lanes 6 and 10) or with RNase H but without the oligonucleotide (lanes 7 and 11) did not degrade G2 or GF. When purified G1A was subjected to similar RNase H analysis (lane 4), no degradation was detected. These results show that G2 and GF contain sequences homologous to the *T. cruzi* SL, and suggest that G2 is either the *T. cruzi* SL-RNA or an unusually small RNA species that obtains the SL via *trans*-splicing.

To confirm the relationship between G2 and the *T. cruzi* SL-RNA, [32 P]labelled capped G2 was isolated from polyacrylamide gels and subjected to sequence analysis using base specific nucleases (Fig. 3A). The sequence derived from the 5' capped G2 coincides to the sequence of the SL-RNA (Fig. 3B). Nucleotides +1–4 and +6 were not readable on this sequence analysis, probably due to modifications of SL-RNA cap structures [above, and see (6,7,25,26)]. Sequence data downstream from the 39 nt SL was required to discriminate the SL-RNA from other very small RNA molecules that obtained the SL by *trans*-splicing. The sequence of G2 was readable to position +49; i.e., 10 nt downstream from the 3' end of the 39 nt SL. These downstream sequences are identical to those of the SL-RNA and we therefore conclude that the ~110 nt G2 species that is capped in the *T. cruzi* nuclear extracts is the ~110 nt SL-RNA and not a product of *trans*-splicing.

We were unable to obtain yields of GF adequate for sequence analysis. However, the size of these RNAs (30–40 nt), their sensitivity to RNase H in the presence of the anti-SL oligonucleotides, and the fact that their cap structures resemble that of G2 suggest that these RNAs may be 5' fragments of the SL-RNA. We do not know if these fragments are artifacts or bona-fide products or intermediates in the *trans*-splicing reaction.

G1A and G1B are *T. cruzi* U-RNAs

Other than the SL-RNA and the mature mRNAs, the only trypanosome RNAs known to be capped are the U-RNAs (2,12,27). The 145–150 nt lengths of G1A and G1B are very similar to the 148 nt reported for *T. brucei* U2 (12,27) and to the 144 nt reported for an additional RNA proposed as *T. brucei* U3 (12), suggesting that these RNAs may be *T. cruzi* U-RNAs. To investigate this possibility, labelled G1A and G1B were extracted from polyacrylamide gels and subjected to RNA sequence analysis as described for G2 (Fig. 3A). This analysis resolved 40 nt of G1A that match very closely the sequence of

Figure 2. G2 and GF RNAs are homologous to the spliced leader. Labelled G1A, G2 and GF were purified and analyzed for homology with the *T. cruzi* SL by hybridization with a synthetic 35 residue oligonucleotide complementary to the SL (5'-CAATATAGTACAGAACTGTATCAATAATAGCGTT-3') followed by digestion with RNase H essentially as described in Donis-Keller (50). Briefly, approximately 500 Cerenkov cpm of each of the purified RNAs was hybridized for 5 min at 65°C with 1 μ g of the oligonucleotide in 10 μ l of 40 mM Tris (pH 7.9), 4 mM MgCl₂, 1 mM dithiothreitol, and 0.03 μ g/ μ l bovine serum albumin, and the temperature was decreased slowly to 30°C. RNase H (~1 unit) was added, the reactions were incubated 30 min at 32°C, the products were purified by phenol and chloroform extraction and ethanol precipitation and examined by electrophoresis in an 8% (G1A and G2) or 10% (GF) polyacrylamide/7 M urea gel and autoradiography. Lanes 1, 5 and 9 contain intact labelled RNA; lanes 2, 6 and 10 contain RNA incubated in buffer without oligonucleotide or RNase H; lanes 3, 7 and 11 contain RNA incubated with RNase H in the absence of oligonucleotide; and lanes 4, 8 and 12 contain RNA incubated with RNase H in the presence of the oligonucleotide. The gel shown in lanes 1–8 was run for a longer period than the gel shown in lanes 9–12. The small RNase H product in lane 12 is probably the 5'-most few nucleotides of the SL-RNA that do not hybridize to the antisense oligonucleotide and are therefore not degraded. XC indicates the position of the xylene cyanole dye after electrophoresis, B indicates bromophenol blue.

T. brucei U2 (Fig. 3B) and we therefore conclude that G1A is *T. cruzi* U2. Only 34 nt of sequence were obtained for G1B. However, this sequence matched very closely that of the putative 144 nt *T. brucei* U3 (Fig. 3B). Since this U-RNA has not been unequivocally assigned, we refer to it herein as Ux.

Conditions for *in vitro* capping in the *T. cruzi* extracts

To examine the requirements for *in vitro* RNA capping, the *T. cruzi* nuclear extracts were incubated with [α - 32 P]GTP in the presence or absence of exogenous Mg²⁺ and ATP (Fig. 4). Without exogenous ATP or Mg²⁺, labelling of the SL-RNA and GF fragments was abolished, but U2 and Ux were only slightly less intensely labelled (lane 1). Addition of 5 mM EDTA to these reactions eliminated all labelling (data not shown) suggesting that some divalent cation is required for capping. Addition of exogenous Mg²⁺ in the absence of ATP had little effect on the reaction (lane 2). In contrast, addition of ATP alone significantly enhanced SL-RNA and GF labelling (lane 3), and ATP plus Mg²⁺ (lane 4) had a synergistic effect on labelling of these RNAs and a smaller effect on labelling of U2 and Ux. These

Figure 3. Sequence analysis of purified cap-labelled *T. cruzi* RNAs. **A.** Sequence analysis of G2, G1A, G1B, respectively. Lanes: L, RNA ladder; G, T1 digest; A, U2 digest; A+U, Phy M digest; C, CL3 digest. Although all bases are not readable on reproductions of these gels (e.g., the A at position 11 of G2), the original autoradiograms provided unequivocal signals at these sites. **B.** Comparison of sequences of G2 and G1A and G1B to sequences of the *T. cruzi* SL-RNA and *T. brucei* U2 and putative U3 RNAs (12). The position of the 39 nt *trans*-spliced SL is shown, (N) indicates nucleotides that were unreadable probably due to modification (methylation), (+) indicates homology, (–) at position +7 in *T. brucei* U3 indicates a gap.

results imply that capping of the SL-RNA and U-RNAs proceeds via an ATP and Mg^{2+} dependent mechanism.

The ATP requirement of the *in vitro* capping reaction was examined over a range of concentrations (Fig. 5A). The optimal ATP concentration for capping was 1 mM (lane 5); greater than 3 mM seemed to inhibit activity (lanes 8–10). As expected from the experiment shown in Figure 4, excision and counting of these RNAs showed that capping of the SL-RNA was enhanced >25 fold at 1 mM ATP, whereas capping of the U-RNAs was enhanced only slightly (<5 fold). This difference is consistent with observation that, in the *T. cruzi* extracts, a high proportion of the U-RNAs is already capped whereas most of the SL-RNAs lack caps (see below).

The standard capping reactions contain very low ($\sim 0.4 \mu\text{M}$) concentrations of GTP. To assess the effect of the GTP

concentration, capping reactions were carried out over a range of GTP concentrations, while the specific radioactivity of the GTP was maintained constant (Fig. 5B). Over the 0.132–3.2 μM range of GTP concentrations examined, labelling of the SL-RNA and U2 and Ux increased approximately linearly with the GTP concentration. However, even at the highest GTP concentration, only the SL-RNA, U2, Ux and GF were capped; i.e., the apparent specificity of the reaction for these RNAs was maintained.

Capping of the SL-RNA, U2 and Ux progresses rapidly at 30°C (data not shown). Surprisingly, the reaction also progresses efficiently at very low temperatures; e.g., 5°C (Fig. 5C). The SL-RNA, U2 and Ux from the 45 min-5°C incubation (lane 6) were excised, counted and compared to the RNAs from the control reaction at 30°C (lane 1). The amount of labelling at 5°C

Figure 4. Requirements of the *in vitro* capping reaction. Capping reactions were carried out as described in Materials and Methods with 20 mM creatine phosphate in the presence or absence of 3.2 mM MgCl₂ or 500 μ M ATP and analyzed by polyacrylamide gel electrophoresis and autoradiography.

was approximately 50% for the SL-RNA and Ux, and 15% for U2, of the control reaction, suggesting that even at very low temperatures capping of these RNAs progresses with favorable kinetics.

Specificity of the *in vitro* capping activity

The activity in the *T. cruzi* nuclear extracts capped only the SL-RNA, U2, Ux and GF, suggesting that it has specificity for these RNAs. To explore this phenomena, we assessed the ability of the extracts to cap synthetic transcripts synthesized from plasmid vector pGEM2 with T7 RNA polymerase (see Materials and Methods). The T7 transcript was not capped in the extracts (Fig. 6, lane 2) even when added at high concentrations (not shown). In contrast, the T7 transcript was efficiently capped by vaccinia virus guanylyl transferase (v-GTase, lanes 5,9), indicating that its 5' terminus is a competent cap acceptor. The T7 transcript and all of the expected endogenous transcripts are also efficiently capped by v-GTase after 30 min incubation in the extract (lanes 6,10), indicating that the extract does not rapidly degrade or alter these RNAs. Similar results were obtained with

Figure 5. Conditions for capping in the *T. cruzi* extracts. **A.** Effects of ATP concentration. Capping reactions were carried out in various ATP concentrations; i.e., 0, 0.1, 0.2, 0.5, 1.0, 2.0, 3.0, 4.0, 6.0, 10 mM (lanes 1–10, respectively), in the absence of creatine phosphate. The RNA bands were excised and Cerenkov counted for quantitative analysis (see text). **B.** Effects of GTP concentration on capping specificity. Capping reactions were performed in 1 mM ATP and 0.132, 0.4, 0.8, 1.6 and 3.2 μ M GTP (lanes 1–5, respectively). The specific radioactivity of the GTP was held constant at 3000 Ci/mmol. **C.** Capping activity at 5°C. Capping reactions were carried out at 30°C for 30 min (lane 1), or at 5°C for 0, 5, 15, 30 or 45 min (lanes 2–6, respectively). Positions of U2, Ux, SL-RNA (SL) and GF are indicated.

other synthetic transcripts (data not shown). Thus, synthetic transcripts efficiently capped by v-GTase are not capped in the *T. cruzi* nuclear extract.

To examine the possibility that the SL-RNA, U2, Ux and GF are the only endogenous RNAs with 5' termini that can be capped, we labelled purified nuclear RNA with v-GTase (Fig. 6, lane 3). Under these conditions, 5S rRNA, several other discrete RNAs, and a background smear of endogenous RNAs were labelled. Thus, the SL-RNA, U2, Ux and GF are not the only endogenous RNAs with competent 5' termini. Chemical decapping (see Materials and Methods) and recapping of the RNA with v-GTase showed enhanced labelling of only U2 and Ux (Fig. 6, lane 4), suggesting that these RNAs are partially capped prior to purification whereas other RNAs, including the SL-RNA, are largely uncapped. [Chemical decapping removes the cap from the SL-RNA in *T. brucei* (5,7), and we assume that the cap on the *T. cruzi* SL-RNA does not differ significantly from that of the *T. brucei* SL-RNA].

Figure 6. Specificity of the capping reaction for the SL-RNA, U2 and Ux. Lanes: 1, control capping reaction performed in 1 mM ATP and no creatine phosphate as described in Materials and Methods; 2, capping reaction as for lane 1 but in the presence of 0.2 μ g of the synthetic T7 transcript; 3, RNA purified from 10 μ l nuclear extract and labelled with v-GTase as described in Materials and Methods; 4, as for lane 3 except the nuclear RNA was chemically decapped prior to v-GTase labelling [the decrease in the intensity of labelling of several RNAs in this lane is probably due to differential sensitivity to degradation during the chemical decapping procedure]; 5, 0.2 μ g of the 59 nt synthetic T7 transcript labelled with v-GTase; 6, RNA labelled with v-GTase after isolation from 10 μ l of a capping reaction performed in the presence of 0.2 μ g T7 transcript but without GTP; 7, RNA purified from 10 μ l extract and 3' labelled with [32 P]pCp and T4 RNA ligase as described in Materials and Methods; 8, as for lane 7 but with 0.2 μ g T7 transcript; 9–12, 10 fold shorter exposure of lanes 5–8; 13, capping reaction performed as for lane 1 but with RNA purified from 10 μ l nuclear extract supplied exogenously; 14, as for lane 13 except the exogenous RNA was chemically decapped as described in Materials and Methods; 15, control capping reaction as for lane 1; 16, silver stained gel from lane 15; 17, silver stained gel of RNA from 10 μ l nuclear extract; 18, as for 17 except the RNAs were chemically decapped prior to electrophoresis. The positions of U2, Ux, SL-RNA (SL), GF, 5S rRNA, and the synthetic T7 transcript are indicated for each of the gels. [?] indicates the positions of unidentified RNAs that are labelled with v-GTase.

The specificity of the *T. cruzi* capping activity was also maintained on RNA purified from the extract, or purified and chemically decapped; i.e., only the SL-RNA, U2, Ux and GF were labelled after reintroduction of purified (Fig. 6, lane 13) or purified and decapped (lane 14) nuclear RNA into the extract. As expected from the experiment with v-GTase and nuclear RNA (lanes, 3 & 4), U2 and Ux labelling—but not SL-RNA labelling—was enhanced after chemical decapping, reinforcing the above conclusion that in the extracts, the U-RNAs but not the SL-RNA, are partially capped.

Finally, we examined the possibility that the specificity of the capping activity observed was due to relative abundance of the potential cap acceptor RNAs. The intensity of v-GTase labelling of the RNAs suggested that many of the potential cap acceptor RNAs are nearly as abundant as the SL-RNA, U2 and Ux (Fig. 6, lanes 3,4). The relative abundance of these RNAs was further

examined by 3' end labelling nuclear RNA with [32 P]pCp and T4 RNA ligase. These experiments confirmed that many v-GTase cappable RNAs, including 5S rRNA (lanes 7,11) and the synthetic T7 transcript (lanes 8, 12), were at least as abundant as U2, Ux and the SL-RNA. Examination of the RNAs by silver staining (Fig. 6, lane 16–18) showed that none of the extracts contained abundant RNAs at the positions of the SL and U-RNAs (compare lanes 15 & 16). In contrast, 5S rRNA and the synthetic T7 transcript (0.2 μ g, not shown) are readily visualized by the silver stain. Thus, the specificity of the capping reaction in the *T. cruzi* extracts cannot be explained by the relative abundance of the potential cap acceptor RNAs.

Taken together, our results suggest that the capping activity observed in the *T. cruzi* nuclear extracts shows unusual specificity for a small group of RNAs including the SL-RNA, U2, Ux and GF.

Figure 7. Trans-splicing as a required capping mechanism. SL-RNAs bear 5' tri-phosphate termini that are efficiently and specifically capped in *T. cruzi* nuclear extracts in the presence of ATP and GTP. In contrast, many trypanosome Pol II primary transcripts are 'multicistronic' RNAs that are processed into multiple mature protein coding mRNAs. Cleavage of the 'multicistronic' messages would yield RNAs with 5' mono-phosphates (or -OHs) which are incapable of being capped by cellular GTases that require 5' di- or tri-phosphate termini. These RNAs can only be capped by trans-splicing of the capped SL to their 5' termini. For ease of description, this figure shows an intermediate containing a 5' mono-phosphate. However, it is probably more likely that cleavage and trans-splicing occur as a concerted reaction to yield mature capped mRNA and this intermediate may never exist as a discrete RNA species.

DISCUSSION

In vitro capping in *T. cruzi* identifies the SL- and two U-RNAs

Several small RNA species were 5' capped in *T. cruzi* nuclear extracts incubated in the presence of ATP and GTP and we have identified the labelled RNAs as the *T. cruzi* SL-RNA and two U-RNAs, U2 and Ux, previously uncharacterized in *T. cruzi*. U2 and Ux were identified by the similarity of their sequences to *T. brucei* U2 and U3, respectively; i.e., 39 of the 40 nt of *T. cruzi* U2 sequenced were identical to corresponding sequences of *T. brucei* U2, and 33 of the 34 nt sequenced of *T. cruzi* Ux matched the putative *T. brucei* U3 sequence. Thus, at least the 5' ends of these RNAs are highly conserved in both size and sequence across trypanosome species. GF, a fourth labelled RNA species 30–40 nt in length, was degraded by RNase H in the presence of a synthetic oligonucleotide complementary to the *T. cruzi* SL indicating that it bears homology to the SL. Since GF functions as a cap acceptor, it must bear a 5' tri- or di-phosphate and therefore is probably derived from the 5' terminus of the SL-RNA. The role, if any, of this RNA species in the trans-splicing process remains to be understood.

Specificity of the *in vitro* capping activity

The only RNAs that are 5' capped in trypanosomes are the SL-RNA, the U-RNAs, and the mature mRNAs that obtain caps from the SL-RNA via trans-splicing. The capping activity reported herein in *T. cruzi* nuclear extracts reflects this requirement; i.e., only the SL-RNA and two U-RNAs are capped. The determinants of this apparent specificity are unknown. Other exogenous and endogenous RNAs were not capped in the *T. cruzi* extracts although these RNAs were readily capped by v-GTase,

Figure 8. Potential homologous interactions among *T. cruzi* U2, Ux and SL-RNA. The sequences of the 5' portions *T. cruzi* U2, Ux and the SL-RNA were aligned to maximize complementarity. A similar model for homologous interaction between *T. brucei* U2 and SL-RNA has been proposed by Tschudi et al. (27) and Mottram et al. (12). Small case nucleotides were extrapolated from the published sequences of *T. brucei* U2 and U3 (12) that have not yet been verified for *T. cruzi*.

demonstrating that inactivity of these RNAs is not due to lack of competent 5' termini. Moreover, since many RNAs in the extract were capped with v-GTase without prior decapping (Fig. 6, lane 3), it is unlikely that the observed specificity of labelling is due to decapping rather than the forward reaction; i.e., it is unlikely that the SL- and U-RNAs are specifically decapped and therefore only appeared to be specifically capped. Another possible explanation for the specificity is that TPase is lost from the extracts and only the SL- and U-RNAs bear cappable 5' diphosphates. However, purified chemically decapped (5' triphosphate) *T. cruzi* SL- and U-RNAs were efficiently recapped when re-supplied to the extract (Fig. 6, lane 14) demonstrating no deficiency in TPase function. Other explanations, such as relative abundance of the RNA substrates and specific exonuclease degradation have also been ruled out as explanations for the apparent specificity of the capping activity in these extracts (Fig. 6). Currently, we cannot eliminate the possibility that specificity is due to inhibitory factors present in the unfractionated nuclear extract which prevent capping of all RNAs except the SL- and U-RNAs. However, such a model invokes a requirement for inhibitory factors that specifically recognize and prevent capping of all RNAs other than the SL- and U-RNAs.

It has recently been shown that *T. brucei* U4-RNA is capped (12) and we assume that our *T. cruzi* extracts contain this and possibly other capped U-RNAs. However, *T. brucei* U4-RNA is ~111 nt, very similar in length to the ~110 nt *T. cruzi* SL-RNA and it is possible that labelling (capping) of a ~111 nt *T. cruzi* U4-RNA was obscured by labelling of the ~110 nt SL-RNA. Moreover, our results showed that, in the extract, the U2 and Ux RNAs are more highly capped than the SL-RNA. Other U-RNAs (e.g., U4) may also be largely capped and not observed because they are poorly labelled in the capping assay. U6 RNA has been identified in *T. brucei* but, as in other eukaryotes, lacks a 5' G cap (12) and would not be labelled by a GTase activity. Other capped U-RNAs have not been identified in trypanosomes.

Capping in other eukaryotes is specific for Pol II transcripts—the U-RNAs and the pre-mRNAs. However, partially purified GTases lack specificity and indiscriminately cap any transcript with the appropriate 5' di- or tri-phosphate (9,10). It has been proposed that capping specificity in these organisms is mediated by the Pol II transcription initiation complex (9,10) based on observations that: 1) capping is limited to Pol II transcripts; 2) Pol II transcripts >50 nt in length are quantitatively capped (28,29); and, 3) *in vitro* transcription by RNA Pol II, but not by Pals I or III, is inhibited by the capping inhibitor adenosylhomocysteine (30). However, Pol II transcripts <20 nt

lack 5' caps (28), certain viral Pol II transcripts lack caps (31), and *in vitro* Pol II transcription from eukaryotic promoters proceeds in the absence of capping although the mature transcripts are properly capped (32), suggesting that capping is not coincident with initiation. Thus, the mechanisms by which Pol II transcripts are selected for capping have yet to be precisely defined.

Formation of a transcription initiation complex is not requisite for the capping observed in the *T. cruzi* extracts since the reaction proceeded in the absence of transcription; i.e., pre-existing endogenous SL- and U-RNAs were capped in the absence of UTP and CTP. A more likely explanation for the specificity of capping in the *T. cruzi* extracts is that it is mediated by a direct or indirect interaction between the SL and U-RNAs and the capping enzyme(s). Formation of the γ -monomethyl phosphate cap of mammalian U6 RNA requires a conserved stem-loop structure and six residue primary sequence in the RNA substrate (33), and trimethylation of U-RNA caps requires recognition of the Sm binding site on the RNAs by specific proteins (34). Moreover, U-RNAs in other eukaryotes exist as components of small nuclear ribonucleoprotein particles (snRNPs) and participate in many mRNA maturation events through direct and indirect interactions among primary sequences and secondary structures of relevant RNA components (35–41).

The SL-RNA in *T. brucei* exists in a $\sim 10S$ snRNP (2), and it is likely that the U-RNAs also form snRNPs. Thus, the capping specificity in the *T. cruzi* extracts could be dictated by the particle(s) with which the SL- and U-RNAs are associated. In fact, preliminary results suggest that capping activity from *T. cruzi* extracts co-sediments with the SL- and U-RNAs in a $\sim 20S$ complex (T.A.Z., unpublished data). SnRNP formation is mediated by recognition of RNA primary and secondary structure motifs by relevant protein factors. Thus, specificity of the *in vitro* *T. cruzi* capping reaction may ultimately be dictated by the primary and/or secondary structure of the SL- and U-RNAs. In fact, several sequence motifs in the SL-RNA are conserved and secondary structure models for this RNA have been proposed (2,3,42), and conserved secondary structures have also been proposed for U-RNAs (35,36).

The ATP requirement of the *T. cruzi* capping reaction was unexpected—other soluble capping systems require only the high energy phosphate bonds in GTP (9)—and it seems unlikely that the *T. cruzi* GTase itself is ATP dependent. Alternatively, ATP may enhance capping activity indirectly by stabilizing the RNA or maintaining GTP pools. However, RNA is quite stable in the *T. cruzi* extracts (4) and we have no evidence that added GTP is rapidly depleted (T.A.Z., unpublished data). ATP is required in the assembly of snRNPs and for interactions of some snRNPs during formation of spliceosomes (39,43–45). Thus, it is possible that the ATP requirement for capping in *T. cruzi* extracts derives from ATP dependent assembly of components that mediate capping activity or its specificity for the SL-RNA and the U-RNAs.

Trans-splicing may be required for capping many trypanosome mRNAs

The *in vitro* capping activity in the *T. cruzi* nuclear extracts accurately reflects the *in vivo* capping requirements for kinetoplastid protozoa; i.e., only the SL- and U-RNAs need to be capped since mature mRNAs obtain their caps from the SL-RNA via *trans*-splicing, and is consistent with the hypothesis that one function of *trans*-splicing in these organisms is to provide a mechanism for capping mRNA (Fig. 7). Many Pol II transcripts

in kinetoplastid protozoa are 'multicistronic' and consist of long precursors that are processed into several discrete protein coding mRNAs [(1), and Fig. 7]. Processing of these precursors into discrete RNAs creates transcripts lacking the 5' di- or tri-phosphates required for capping by known GTases. Moreover, some protein coding genes, including those encoding variable surface glycoproteins in African Trypanosomes, may be transcribed by Pol I (46,47). If capping were coupled to Pol II transcription initiation, these transcripts could not be properly capped. Therefore, the role of *trans*-splicing may be to provide a mechanism for capping all of the protein coding transcripts including those processed from 'multicistronic' pre-transcripts and those not synthesized by Pol II. A capping mechanism specific for the SL-RNA and the U-RNAs is consistent with this model.

Complementary interactions among U2, Ux and the SL-RNA

A potential interaction between the *T. brucei* U2 and SL-RNA was recently described (12,27). A similar interaction is possible between the corresponding regions of the *T. cruzi* U2 and SL-RNAs (Fig. 8). Moreover, the 5' terminal sequence of *T. cruzi* Ux will also potentially hybridize with an overlapping region of U2 (Fig. 8). U2 in other eukaryotes interacts with introns in pre-mRNA transcripts very early in the splicing process at the site of formation of the branched nucleotide (43,45,48,49). The U1, U2, and U4/U5/U6 snRNPs interact with the pre-mRNA to form a spliceosome (45) and these interactions are due in part to homologous recognition between the U-RNAs. Thus, it is possible that potential interactions between *T. cruzi* U2, Ux and the SL-RNA function in *trans*-splicing.

ACKNOWLEDGEMENTS

We thank Drs. Harvey Eisen, Eric H. Westin and Francis L. Macrina for critical evaluations of this manuscript. This work was supported by grants #J81 from the Jeffress Memorial Trust, #AI22946 from the National Institutes of Health and from the A.D. Williams Foundation. T.A.Z. was supported in part by a fellowship from the American Heart Association, VA Affiliate. Synthetic oligonucleotides were provided by the MCV/VCU Nucleic Acids Core Laboratory.

REFERENCES

1. Borst, P. (1986) Annu. Rev. Biochem. 55, 701–732.
2. Agabian, N., Perry, K.L. and Murphy, W.M. (1987) in Molecular Biology of RNA: New Perspectives, Inouye, M. and Dudock, B.S. Eds. pp. 151–171, Academic Press, Inc., New York.
3. McCarthy Burke, C., Taylor, Z.A. and Buck, G.A. (1989) Gene. 82, 177–189.
4. Zwierzyński, T.A., Widmer, G. and Buck, G.A. (1989) Nucleic. Acids Res. 17, 4647–4660.
5. Laird, P.W., Kooter, J.M., Loosbroek, N. and Borst, P. (1985) Nucleic. Acids Res. 13, 4253–4266.
6. Perry, K.L., Watkins, K.P. and Agabian, N. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 8190–8194.
7. Lenardo, M.J., Dorfman, D.M. and Donelson, J.E. (1985) Mol. Cell Biol. 5, 2487–2490.
8. Freistadt, M.S., Cross, G.A. and Robertson, H.D. (1988) J. Biol. Chem. 263, 15071–15075.
9. Mizumoto, K. and Kaziro, Y. (1987) Prog. Nucl. Acid. Res. Mol. Biol. 34, 1–28.
10. Moss, B. (1984) in Processing of RNA, Apirion, D. Ed. pp. 119–131, CRC Press, Inc., Boca Raton.
11. Shatkin, A.J. (1976) Cell 9, 645–653.

12. Mottram, J., Perry, K.L., Lizardi, P.M., Luhrmann, R., Agabian, N. and Nelson, R.G. (1989) *Mol. Cell Biol.* 9, 1212–1223.
13. Morel, C., Chiari, E., Camargo, E.P., Mattei, D.M., Romanha, A.J. and Simpson, L. (1980) *Proc. Natl. Acad. Sci. U. S. A.* 77, 6810–6814.
14. Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) *Nucl. Acids Res.* 11, 1475–1489.
15. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
16. Silberklang, M., Gillum, A.M. and RajBhandary, U.L. (1979) *Meth. Enzymol.* 59, 58–109.
17. Donis Keller, H. (1980) *Nucl. Acids Res.* 8, 3133–3142.
18. Donis Keller, H., Maxam, A.M. and Gilbert, W. (1977) *Nucl. Acids Res.* 4, 2527–2538.
19. Boguski, M.S., Hieter, P.A. and Levy, C.C. (1980) *J. Biol. Chem.* 255, 2160–2163.
20. Simoncsits, A., Brownlee, G.G., Brown, R.S., Rubin, J.R. and Guilley, H. (1977) *Nature* 269, 833–836.
21. Rose, J.K. and Lodish, H.F. (1976) *Nature* 262, 32–37.
22. England, T.E., Bruce, A.G. and Uhlenbeck, O.C. (1980) *Meth. Enzymol.* 65, 65–74.
23. Walseth, T.F. and Johnson, R.A. (1979) *Biochim. Biophys. Acta* 562, 11–31.
24. Efstratiadis, A., Vournakis, J.N., Donis Keller, H., Chaconas, G., Dougall, D.K. and Kafatos, F.C. (1977) *Nucl. Acids Res.* 4, 4165–4174.
25. Sutton, R.E. and Boothroyd, J.C. (1988) *Mol. Cell Biol.* 8, 494–496.
26. Freistadt, M.S., Cross, G.A., Branch, A.D. and Robertson, H.D. (1987) *Nucl. Acids Res.* 15, 9861–9879.
27. Tschudi, C., Richards, F.F. and Ullu, E. (1986) *Nucl. Acids Res.* 14, 8893–8903.
28. Jove, R. and Manley, J.L. (1984) *J. Biol. Chem.* 259, 8513–8521.
29. Coppola, J.A., Field, A.S. and Luse, D.S. (1983) *Proc. Natl. Acad. Sci. U. S. A.* 80, 1251–1255.
30. Jove, R. and Manley, J.L. (1982) *Proc. Natl. Acad. Sci. U. S. A.* 79, 5842–5846.
31. Sonenberg, N. and Pelletier, J. (1989) *Bioessays* 11, 128–132.
32. Bunick, D., Zandomeni, R., Ackerman, S. and Weinmann, R. (1982) *Cell* 29, 877–886.
33. Singh, R., Gupta, S. and Reddy, R. (1990) *Mol. Cell Biol.* 10, 939–946.
34. Mattaj, I.W. (1986) *Cell* 46, 905–911.
35. Busch, H., Reddy, R., Rothblum, L. and Choi, Y.C. (1982) *Annu. Rev. Biochem.* 51, 617–654.
36. Reddy, R. and Busch, H. (1983) *Prog. Nucl. Acid. Res. Mol. Biol.* 30, 127–162.
37. Reed, R., Griffith, J. and Maniatis, T. (1988) *Cell* 53, 949–961.
38. Ruskin, B., Zamore, P.D. and Green, M.R. (1988) *Cell* 52, 207–219.
39. Konarska, M.M. and Sharp, P.A. (1987) *Cell* 49, 763–774.
40. Cheng, S.C. and Abelson, J. (1987) *Genes. Dev.* 1, 1014–1027.
41. Schaufele, F., Gilmartin, G.M., Bannwarth, W. and Birnstiel, M.L. (1986) *Nature* 323, 777–781.
42. Bruzik, J.P., Van Doren, K., Hirsh, D. and Steitz, J.A. (1988) *Nature* 335, 559–562.
43. Konarska, M.M. and Sharp, P.A. (1986) *Cell* 46, 845–855.
44. Konarska, M.M. and Sharp, P.A. (1988) *Proc. Natl. Acad. Sci. U. S. A.* 85, 5459–5462.
45. Bindereif, A. and Green, M.R. (1987) *EMBO. J.* 6, 2415–2424.
46. Shea, C., Lee, M.G. and Van der, Ploeg, L.H. (1987) *Cell* 50, 603–612.
47. Kooter, J.M. and Borst, P. (1984) *Nucleic. Acids Res.* 12, 9457–9472.
48. Black, D.L., Chabot, B. and Steitz, J.A. (1985) *Cell* 42, 737–750.
49. Parker, R., Siliciano, P.G. and Guthrie, C. (1987) *Cell* 49, 229–239.
50. Donis Keller, H. (1979) *Nucl. Acids Res.* 7, 179–192.