A 5' exo-ribonuclease and RNA ligase of *T.brucei*

Jin Huang and Lex H.T. Van der Ploeg

Department of Genetics and Development, College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA

Received June 2, 1988; Revised and Accepted September 2, 1988

ABSTRACT

We identified a Mg^{2+} dependent 5' exo-ribonuclease and an RNA ligase in cell-free extracts of *Trypanosome brucei*. The exo-ribonuclease in S100 or nuclear extracts, removes about 20 nts from the 5' end of SP6 derived capped as well as uncapped RNA and then stops. In contrast to the activity of the exo-ribonuclease on capped SP6 mini-exon transcripts, the exonuclease cannot degrade trypanosome-derived mini-exon transcripts or the mini-exon located at hsp 70 mRNAs. We therefore assume that the four secondary base modifications adjacent to the mini-exon cap, generated *in vivo*, confer resistance to the exo-ribonuclease. After exonuclease shortening of SP6 transcripts, an RNA ligase catalyzes intramolecular ligation, generating a 3'-5' phosphodiester bond in a Mg^{2+} and ATP dependent reaction.

INTRODUCTION

Transcriptional control and RNA processing in the protozoan *Trypanosoma brucei* and related kinetoplastidae occur in a manner that drastically differs from those in other eukaryotes. In trypanosomes all protein coding genes are transcribed discontinuously and consist of two exons that are derived from separate genes (1-4). The most 5' exon is a capped 39 nt mini-exon, which is common to every mRNA (5-9, 54). The analyses of *in vivo* intermediates of the exon joining reaction have shown that the separately transcribed RNA precursors, encoding the 5' mini-exon and 3' protein coding exon, are joined as a result of trans-splicing in a process that is strikingly similar to cis-splicing (10-12; for review see, 13, 14). The mini-exon precursor (mini-exon donor, medRNA, or spliced leader RNA) in *T.brucei* is a capped, 140 nt RNA molecule (15-19). The exon-intron boundary of the medRNA has a GT dinucleotide at its 5' splice site, which serves as a signal for conventional cis-

splicing (20). Correspondingly, all the major coding exons are preceded by an intron containing the common AG splice acceptor site (21). While the cellular components involved in trans-splicing in trypanosomes remain obscure, an unusual feature of the capped mini-exon in trypanosomes is that it contains a heavily modified 5' end with four methylated nucleotides adjacent to the 7mG cap (7-9). Trans-splicing may therefore depend on the role played by this unusual cap structure in addition to that of the consensus signal at the 5' splice site. At present, there has been no success in developing a trypanosome in vitro trans-splicing system.

In analyzing the processing products of SP6 RNAs in cell free trypanosome extracts we have identified two enzymatic activities, which indicate the presence of an enzyme that can discriminate between mini-exon containing RNAs and putative pre-mRNAs which do not yet have a mini-exon cap. We characterized this enzyme, and show that it has 5'-3' exo-ribonuclease activity. An RNA ligase of the trypanosome extracts circularizes the accumulated exo-ribonuclease cleavage products. The finding of a 5' exo-ribonuclease that is capable of shortening any RNA except those that contain a common 5' mini-exon cap trans-spliced onto their 5' ends suggests a possible function in RNA processing.

MATERIALS AND METHODS

Trypanosome strains. *T.brucei* variant 1.208 (22) was used for extract preparation and variant 118a clone 1 (23) for RNA isolation. SP6 RNA polymerase was purchased from New England Biolabs and Gene Screen plus membrane from NEN products.

Construction of SP6 plasmids. A genomic clone of the rabbit B-globin gene was a gift from A.Efstratiadis (24). A Pvu II-Bam HI restriction fragment of this clone was inserted into the Hind III (made blunt with S1) -Bam HI restriction enzyme sites of pGEM-1 (Promega Biotec). Linearization of this plasmid at the Bam HI restriction enzyme site generated the templates used for SP6 transcription (25).

The trypanosome mini-exon-hsp 70 chimeric molecule was constructed by fusing the 321 bp Hinf I restriction fragment containing the mini-exon (26) to the 879 bp Pst I-Eco RI restriction fragment of the hsp 70 gene (27). The Pst I and Hinf I sites were made blunt with Klenow polymerase and an 8 bp linker was added at the site of ligation. The fused fragment was then removed from its cloning vector (pUC 19) by digestion with Hinc II and Eco RI and cloned into the SP64 vector. Transcripts

of this construct were generated from the BstE II restriction enzyme-digested and linearized plasmid.

To generate the mini-exon construct, we digested the vector SP64 with Hind III, after which the sticky ends were removed by S1 digestion. The 951 bp mini-exon Hpa II fragment was treated with the exonuclease Bal 31 to release about 350 nt from either end. The fragments that were generated were cloned into the S1 treated Hind III site and Sma I site of the SP64 vector and their nucleotide sequence was determined by the Sanger di-deoxy chain termination sequencing method. One clone contained 34 bp of the 39 nt medRNA mini-exon and 96 bp of the medRNA intron. The SP6 derived medRNA has however, 5 additional nucleotides derived from the SP6 template, which shared 2 out of 5 nt with the mini-exon sequence. The SP6 mini-exon thus measures 40 instead of 39 nt.

Similarly the exonuclease Bal 31 was used to remove 59 bp (as determined by Sanger dideoxy sequencing of the final SP64 products) from the 5' end of the Pst-I restriction enzyme site of the intergenic regions of the *T.brucei* hsp 70 genes. The resulting hsp 70 fragment was released from its plasmid by Eco RI digestion and subcloned into SP64 (again the Hind III restriction enzyme site of SP64 was removed, while SP64 was subsequently cut with Eco RI).

All the SP64-derived transcripts were made under conditions in which over 95% of the RNA molecules initiate with the cap analogue (25). To prepare uncapped SP64 RNA the cap analogue was omitted from the reaction and the GTP concentration adjusted appropriately. The presence of cap analogues on the RNA was verified on capped alpha ³²P ATP labeled and T2 digested RNA. The nucleotides were separated on TLC plates after which the presence of the labeled cap was verified (data not shown). Decapping of SP64 RNA was performed by Tobacco acid pyrophosphatase (Promega) treatment of the RNA for 60 minutes at 37°C in 50 mM NaOAc (pH 5.0) 1 mM EDTA 10 mM 2-mercaptoethanol (28). Further purification and band isolation using polyacrylamide gel-electrophoresis were carried out as described (29).

Preparation of RNA from trypanosome extracts was done by proteinase K treatment (100 ug per ml) of the extracts in the presence of 0.1% SDS, followed by two phenol-chloroform extractions.

Preparation of nuclear extracts and S100 post-nuclear extracts from *T.brucei*. Trypanosome nuclear and S100 extracts were prepared following a modified protocol of Hernandez and Keller (30) and Dignam et al. (31). *T.brucei* variant 1.208 was grown in the bloodstream of male Wistar rats until a high density (10⁹ cells per ml) was reached. Experiments have been repeated with procyclic trypanosome derived extracts giving identical results. Trypanosomes were collected and separated from red blood cells by centrifugation at 5000 G. The cells were washed once with cold (0°C) PSG solution (phosphate buffered glucose) before being suspended in 5 parts per volume of cold (0°C) hypotonic buffer (10 mM Hepes pH 7.9, 10 mM KCl, 1.5 mM MgCl₂, 0.5 mM DTT and 0.1 mM PMSF). After a 10 minute incubation on ice, trypanosomes were lysed in a Stansted Cell disrupter (80 psi-55 psi). All subsequent steps were carried out between 0°C and 4°C. The cell lysate was centrifuged at 1000 G to separate crude

nuclei and supernatant, from which the NE and S100 extracts were respectively prepared.

Crude nuclei were suspended in 20 mM Hepes pH 7.9, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 25% glycerol, 0.5 mM DTT and 0.2 mM PMSF, followed by 10 strokes with a Dounce homogenizer (piston type A). One tenth volume of saturated $(NH_4)_2SO_4$ was added to the nuclear lysate which was subsequently stirred gently for 30 to 40 minutes to salt out chromatin. Centrifugation at 25,000 G separated soluble nuclear factors from chromatin pellets. The nuclear supernatant was further fractionated by adding solid $(NH_4)_2SO_4$ to 0.38 g per ml, and subsequently stirred gently for 4 hours. By centrifugation at 100,000 G for 30 minutes, most protein factors were precipitated, which were next dissolved in 10 ml of dialysis solution (20 mM Hepes pH 7.9, 100 mM KCl, 0.2 mM EDTA, 20% glycerol, 0.5 mM DTT and 0.2 mM PMSF). Dialysis was performed for 5 hours with three changes of dialysis solution.

To make S100 extracts, the post nuclear supernatant was supplemented with one tenth volume of buffer containing 0.3 mM Hepes pH 7.9, 1.4 mM KCl and 30 mM $MgCl_2$, and centrifuged at 100,000 G for one hour. Solid $(NH_4)_2SO_4$ was added to the supernatant to 0.33 g per ml which was subsequently gently stirred for 4 hours. Centrifugation to collect precipitated proteins was again performed at 100,000 G for 30 minutes. Pellets were suspended in 10 ml of dialysis solution and dialyzed against three changes of 100 volumes of dialysis solution. The protein content in NE and S100 was measured by Buriet assay (32) and adjusted to 14 mg/ml for NE and 17 mg/ml for S100. Extracts were aliquoted and stored at $-140^\circ C$ in dialysis buffer.

Assay conditions for cleavage and ligation reactions. The standard assay condition for RNA processing in trypanosomes extracts was adapted from that used in Hela cell splicing reactions (33). The assay included 60% extract, 12.8 mM Hepes pH 7.9, 60 mM KCl, 3.2 mM $MgCl_2$, 0.2 units per lambda RNasin, 20 mM phosphocreatine, 1 mM ATP, 14% glycerol, 0.5 mM DTT, 0.2 mM PMSF and 1-2 nM input SP6 transcript. Reactions were incubated at $30^\circ C$ and terminated by the addition of 100 ug per ml proteinase K and an equal volume of 2 x TES (20 mM Tris-HCl pH 7.5, 10 mM EDTA, 0.2% SDS). After 15 minutes for proteinase K treatment, the mixture was extracted once with phenol-chloroform. RNA was precipitated by addition of 2.5 volumes of ethanol. Titration experiments of Mg^{2+} , ATP and phosphocreatine were performed by changing one parameter at a time while leaving all other conditions constant.

RNA analysis by S1 nuclease mapping and primer extension. Protection against digestion by nuclease S1 was performed with uniformly labeled RNA (Figure 5A) and end-labeled DNA (Figure 5C) as probes. Conditions for DNA-RNA annealing and S1 nuclease digestion were as described by Dudler and Travers (34). Primer extension analysis of the 5' end of the hsp 70 mRNA was performed as described by Shea et al. (35). Direct di-deoxy nucleotide sequencing of the hsp 70 mRNA involved an appropriate mixing of dideoxy and deoxy oligonucleotides in the standard Sanger sequencing primer extension assay.

Northern blot analysis. Trypanosome total RNA was prepared using LiCl-Urea extraction of RNA as described by Auffray and Rougeon (36). RNA was size separated in 5% polyacrylamide 8M Urea gels, followed by electroblotting onto Gene Screen plus membrane as

Figure 1.

Schematic representation of gene constructs. All four constructs were made in vectors containing an SP6 promoter and only the linearized inserts which were transcribed into capped SP6 RNA are shown. The black boxes indicate vector derived sequences and the lines indicate intron sequences that are vector-derived. Exons and introns are schematically indicated with boxes and thin bars respectively. The trypanosome mini-exon sequence is highlighted by the dotted box. The arrow labeled SP6 indicates the transcription initiation site. The sizes of the introns and exons are given in nucleotides below each construct. The artificially generated 169 nt intron of the trypanosome cis-construct contains a 75 nt sequence derived from the medRNA intron and an additional 86 nt sequence derived from the intergenic region located upstream of the hsp 70 coding exon. The GT-AG splice junctions at the intron boundaries have thus been retained.

described (37). The membrane was subsequently dried at room temperature. Prehybridization and hybridization was carried out at 42°C in 25% formamide, 0.75 M NaCl, 0.15 M Tris-pH 7.5, 0.1 M Na-phosphate, 0.1 % Na-pyrophosphate, 1% SDS, 5% Dextran sulfate, 5 x Denhardtts and 50 ug per ml salmon sperm DNA. Post hybridizational washes were at 3 x SSC, 1% SDS and 40°C.

Debranching of structured RNAs. Hela cell nuclear extracts with de-branching activity were used to treat structured RNAs. RNA samples were incubated with 60% Hela cell nuclear extract, 0.8 units RNasin, and 10 mM EDTA for 45 minutes at 30°C. Reactions were stopped by proteinase K treatment and phenol-chloroform extraction. To test for de-branching activity rabbit B-globin derived lariats were tested in parallel as controls.

RESULTS

Processing of SP6 RNA in cell-free T.brucei extracts.

We compared the efficiency of processing of B-globin pre-mRNA, which contained introns that can be removed by splicing in cis,

with a trypanosome pre-mRNA that contained an artificial co-linear trypanosome intron. The SP6 constructs: rabbit B-globin and the trypanosome hsp 70 construct containing an intron flanked by the mini-exon and a 3' hsp 70 coding exon, are schematically indicated in Figure 1. In addition, three gene constructs were made in order to assay trans-splicing activity (Figure 1). The first is a 140 nt medRNA construct which differs from the endogenous medRNA in that it contains three base substitutions and one base addition (see Figure 5 for details).

Figure 2.

RNA processing in HeLa cell nuclear extracts. Capped SP6 RNA transcripts made from the B-globin (panel A) and trypanosome cis-construct (panel B) were incubated in HeLa cell extracts for various periods of time as indicated on top of each panel. The RNA was subsequently size separated in a 4% denaturing polyacrylamide gel. The size of the different RNAs and the structure of the splicing intermediates and products is shown in the left hand margin. Exons 1 and 2 are represented by boxes, and introns are represented by lines. Asterisks indicate lariat-shaped splicing intermediates.

The capped SP6 derived medRNA precursor also lacks the four secondary base modifications that immediately follow the cap of in vivo generated RNA. The other RNA substrates for trans-splicing, contained the 3' intron and adjacent hsp 70 coding exon, extending 37 and 86 nt (data not shown) respectively upstream of the hsp 70 coding exon.

We first tested whether or not the trypanosome consensus signals at splice junctions can be recognized in a Hela cell extract possessing splicing activity (Figure 2). From these experiments it became clear that the trypanosome SP6 cis-construct RNA could not be spliced in Hela cells even though control B-globin SP6 pre-RNA is efficiently processed (Figure 2; 30, 33, 38). Therefore, despite the presence of trans-splicing signals at the 5' and 3' junctions of the cis-construct, this artificial substrate containing a co-linear intron does not fulfill the required sequences for conventional RNA splicing.

We next replaced the Hela cell nuclear extract with a comparable trypanosome nuclear and S100 post-nuclear extract, generated from T.brucei bloodstream forms (see materials and methods for details), as well as from procyclic trypanosomes (data not shown). At first we used the standard assay conditions chosen for RNA splicing with Hela cell nuclear extracts (33). The time course analyses with the trypanosome bloodstream form and procyclic extracts gave very different processing patterns for globin as well as trypanosome cis constructs when compared to those of Hela cell extracts (Figure 3). First, neither the nuclear extract, the S100 extract, nor their combination (data not shown) were able to support RNA processing of capped B-globin or trypanosome cis-construct pre-RNAs. The degradation of RNA in figure 3 appears extensive but is over-represented because of the relatively long exposures (compare exposures in panels A and C). Second, a number of high molecular weight RNA species are generated with each construct, as a function of time, in the trypanosome S100 extract (panels labeled S100, bands labeled with asterisks) but not in the nuclear extract (panels labeled NE, see also following sections). The efficiency of formation of the high molecular weight bands varies depending on the pre-RNA sequence and the size of the input RNA. For

instance, at a 3 hour time point, about 50% of B-globin pre-RNA has been converted into high molecular weight fragments. With the much larger pre-RNA of the trypanosome cis-construct, the conversion only reaches about 5%. Third, in addition to the high molecular weight RNA species, two abundant cleavage products of the capped heat shock SP6 RNA are generated by both the S100 and the nuclear extracts. In the gel in Figure 3C the degradation is best visualized by the shortening of the 292 nt hsp 70 construct

Figure 3.

RNA processing in trypanosome extracts. Capped polymerase SP6-generated transcripts from the B-globin (panel A), trypanosome cis-construct (panel B), and hsp 70 3' exon construct (panel C) were incubated in nuclear extracts (NE) and S100 extracts (S100) for various periods of time as indicated on top of each of the panels. Sp6 RNA was purified and size separated in 4% denaturing polyacrylamide gels. The size of each input RNA molecule is given in the left hand margin. High molecular weight RNA generated from incubation with S100 extract is marked with asterisks and major cleavage products are marked with arrows. pBR322/Hinf I (panel A) and pBR322/Hpa II (panel B and C) are the DNA size standards.

by 14 nt and 20 nt respectively (bands indicated with arrows). As described in the following sections, similar major cleavage products of the B-globin and trypanosome cis-construct exist, but these large RNAs do not separate well from the input RNA. Cleavage is also a time-dependent reaction, while its progressive pattern suggests that the reaction begins from one end of the RNA molecule and terminates at either one of two strong termination sites.

Formation of high molecular weight RNA species involves a 5' exo-ribonuclease and an RNA ligase.

To characterize the nature of the enzymatic reaction by which high molecular weight RNA is generated in the trypanosome S100 extract, we gel-purified the high molecular weight RNA species derived from the B-globin pre-RNA, and analyzed it by S1 protection analysis and primer extension. The exact location of the 5' end of the abnormally migrating high molecular weight RNA species was determined using S1 protection with a Bst NI-Nae I restriction enzyme fragment which covers 159 nt from the Bst NI restriction enzyme site in exon 1 to the SP6 transcription initiation site of the B-globin gene (see Figure 4A). The 496 nt RNA band, which was purified at 2 hours of incubation and which co-migrates with the input B-globin pre-RNA, protected two major fragments of 159 and 149 nt from degradation by S1 nuclease (Figure 4A, lane 2). The 159 nt fragment corresponds to the RNA molecule with the intact 5' end of B-globin SP6 pre-RNA (data not shown and figure 5B). The 149 nt fragment results from a 10 nt shortening at the 5' terminus of SP6 RNA after incubation with S100 extract. The RNA band which comigrates with input B-globin RNA (Figure 4A, lane 3) is therefore a mixture of the input RNA molecule and a 10 nt shortened cleavage product. S1 mapping of the high molecular weight structured RNA species generated only a single S1-resistant fragment of 149 nt (Figure 4A, lane 5). Pre-treatment of the high molecular weight RNA species with a HeLa cell nuclear extract with debranching activity did not change the S1 digestion pattern, nor did it alter the migrational behavior of this RNA species (Figure 4A, lanes 7 and 6, respectively). The high molecular weight RNA could, therefore, be a circular RNA generated from a 10 nt

Figure 4.

Characterization of ligation and cleavage products. Capped SP6 RNA of the B-globin gene (panel A) was incubated with trypanosome S100 extracts for two hours after which RNA samples were resolved on 4% denaturing polyacrylamide gels. After a

short exposure to locate the different bands, the high molecular weight fragment and the RNA fragment which still co-migrated with the input SP6 RNA were purified from the gel. As indicated in the physical maps of the two different constructs, a Bst N1-Nae I restriction fragment was used to protect B-globin RNA from degradation by S1 in panel A.

Panel A contains: Lane 1, linear RNA fragment purified after two hours of incubation (this RNA still co-migrates with the input RNA); lane 2, S1 nuclease protection of the fragment shown in lane 1; lane 3 control S1 digestion with no added DNA; lane 4, band purified high molecular weight RNA containing circular RNA and nicked circle; lane 5, S1 protection of the high molecular weight fragment; lane 6, band purified high molecular weight RNA treated with Hela cell debranching extract; lane 7, S1 protection of the RNA from lane 6.

Panel B contains the cDNAs generated from primer extension with a 15 nt oligonucleotide located from 18 to 32 nt downstream of the B-globin splice acceptor site of exon 2 (see schematic below the panel for primer location). The lanes contain, lane M, pBR322/Hpa II DNA marker; lane 1, size separated primer extension product of input B-globin RNA; lane 2, size separated primer extension product of high molecular weight fragments.

Panel C. S1 mapping analysis of the cleavage products of the SP6 hsp 70 3' exon transcript. The lanes contain: Lane 1, control lane with band-purified high molecular weight RNA without S1 treatment (the low molecular weight band results from nicking in the circle); Lane 2, S1 protected end labeled DNA fragment of the zero time point input RNA; lane 3, S1 protected fragments of the incubation with circular RNA recovered after a two hour incubation with nuclear extract; lane 4, S1 protected fragment with circular RNA recovered after a two hour incubation in S100 extract. Bands that are not marked are derived from protection of RNA with SP6 DNA template left in the in vitro assay.

Panel D. 5' exonuclease digestion as revealed in a time course incubation of capped trypanosome hsp 70 3' exon construct. Arrowheads point to a series of cleavage products visible at the early time points only before the major cleavage products accumulate.

Panel E. Resistance of the exo-ribonuclease cleavage products to continued exo-ribonuclease digestion. SP6 heat shock 70 RNA (lane labeled input) was exo-ribonuclease shortened by incubation in S100 extracts for two hours (lane 0). The SP6 cleavage products were purified from gels, re-introduced into S100 extract aliquots and incubated for up to two hours (timepoints at which the samples were taken from the S100 incubation are indicated on top of each lane).

shortened and thus decapped linear precursor. This was confirmed by performing primer extension analysis, showing that the high molecular weight RNA species was indeed a 10 nt shortened circular molecule (Figure 4B). The primer used was an end-labeled 15 nt oligonucleotide complimentary to a portion of B-globin exon 2, located close to the 3' splice junction (see

Comparison of exonuclease sensitivity of capped SP6 medRNA, trypanosome derived medRNA and hsp 70 mRNA. Panel A shows the progressive degradation of capped SP6 medRNA by the 5' exonuclease in a S100 extract incubation. The RNAs were size separated in a 5% denaturing polyacrylamide gel and visualized by autoradiography. Panel B contains the size separated RNA from an incubation of S100 extracts under the standard condition without added RNA. The endogenous RNAs of the extract were size separated on the same 5% polyacrylamide gel as indicated in panel A, after which the RNA was blotted to gene screen plus membranes. The medRNA was detected by hybridization with a 24 mer oligonucleotide complementary to nucleotides 90-114 of the medRNA intron. The nucleotide sequence below the two panels indicates the medRNA and SP6 medRNA 5' end. Only trypanosome derived medRNA contains secondary base modifications as indicated by the nucleotides marked with asterisks. Panel C presents the 5' mini-exon sequence of trypanosome derived hsp 70 mRNA. The tracks labeled AGTC are the TCAG di-deoxy primer extension sequence incubations with a 30 nt hsp 70 specific primer. As a result, the hsp 70 mRNA and mini-exon sequence can be directly read from the bands marked with asterisks (schematically indicated in the right hand sequence). The track labeled 0 is a primer extension with no added di-deoxy nucleotides and the band at 78 nt corresponds to the 5' end of the hsp 70 derived mini-exon cDNA. Panel D. Primer extension analysis of hsp 70 mRNA as indicated in panel C. S100 extracts were supplemented with 20 ug of trypanosome variant 118 a RNA

and incubated for various periods of time (as indicated on top of each of the panels) after which the RNAs were purified and analyzed by primer extension. The fate of the degradation at the 5' end of the hsp 70 mRNA was thus determined by primer extension with an hsp 70-specific oligonucleotide. Lane M contains a PBR 322/ Hpa II marker, lane S100 a single aliquot of S100 RNA, and lane 118a the extension products derived from 20 ug of 118a RNA.

schematic in Figure 4 for primer location). Primer extensions on input B-globin pre-RNA mapped the predicted 5' end of the input RNA molecule at 323 nt (Figure 4B, lane 1). However, the extension product generated from the B-globin circular high molecular weight RNA species was much longer and measured 485 nt. 485 nt corresponds to the size of the input SP6 RNA molecule minus about 10 nt. This finding indicates that a circular RNA molecule was generated with a regular 3'-5' phosphodiester bond at the site of ligation. If a 2'-5' phosphodiester bond were formed, reverse transcriptase would have recognized this bond, leading to a strong stop in the primer extension, and the accumulation of a 313 nt cDNA (39). The 485 nt cDNA thus results from extension around the circle until reverse transcriptase stops at the 5' end of its own primer (see schematic in Figure 4B). Also note that the band which runs as the nicked circle in Figure 4A (lanes 4 and 6) is most likely a randomly nicked circle since no additional bands can be detected in the primer extension in which this purified material was used.

We next determined whether different substrates yielded similar cleavage and ligation products, as suggested by the accumulation of shortened hsp 70 RNA in figure 3C. Figure 4C illustrates the S1 mapping of SP6 RNA derived from the heat shock 70 gene construct (figure 1), which had been incubated for two hours with trypanosome nuclear extract or S100 extract. S1 mapping was performed using a ³²P end labeled Eco RI-Nea I restriction fragment derived from the hsp 70 gene construct (Figure 4C). First, at the zero time point, the size of S1 nuclease protected DNA measured 118 nt (Figure 4C, lane 2), which corresponds to the location of the intact 5' end of hsp 70 pre-RNA. Second, the S1 protection pattern of SP6 RNA incubated in nuclear and S100 extracts was identical, each showing 105 nt and 101 nt S1 resistant bands (figure 4C lanes 3 and 4).

Therefore, after a two hour incubation in either S100 or nuclear extract, about 90% of the input RNA is converted into two major cleavage products, which are shortened at their 5' ends by 13 and 18 nt, respectively (Figure 4C, lanes 3 and 4).

Cleavage appears to be a progressive reaction, involving several intermediates (indicated with arrowheads in Figure 4D) that, accumulate in two major shortened products. That these shortened RNAs are end-products of the exo-ribonuclease degradation was shown by their purification from the extract, followed by their reintroduction into an S100 extract. These already shortened SP6 RNAs now remain resistant to further degradation (Figure 4E, lanes labeled 0 through 2 hours). In comparing the B-globin construct, hsp 70 constructs 1 and 2 (data not shown) and the SP6 mini-exon constructs, which are all shortened similarly (see next section), we have not been able to determine a consensus sequence which defines the site of termination of cleavage.

The trypanosome derived mini-exon protects RNA from cleavage by the 5' exo-ribonuclease.

Having tested a number of SP6 transcripts in trypanosome nuclear and S100 extracts, 5' exonucleolytic cleavage was invariably seen. Panel A of Figure 5 illustrates the time course of digestion of capped SP6 medRNA after incubation with a trypanosome S100 extract. Again, the exonuclease activity trimmed the molecule from its 5' end, resulting in the accumulation of several shortened RNA fragments. Only minor amounts of circular RNA were formed which do not reproduce well at these exposures. We assume that sequence specificity may play a role in the circularization.

Since all trypanosome mRNAs share the same capped 5' 39 nt mini-exon sequence, we wished to determine whether or not this RNA is also a substrate for the exo-ribonuclease. A direct comparison of SP6 medRNA and the authentic trypanosome derived medRNA indicates that, at the nucleotide sequence level, the differences in the RNAs are at the 5' end of the RNA molecules (Figure 5 bottom panel). SP6 medRNA has an extra G residue adjacent to 7mG, and three base changes (from CUA in wild type to UAC in the SP6 RNA). In addition, a possibly more significant

difference between the two types of RNA molecules is that the trypanosome-derived mini-exon RNA has a heavily methylated 5' end (nucleotides marked with asterisks in Figure 5). These secondary base modifications are not present in the SP6 transcripts. It is thus possible that the secondary base modifications confer resistance of the trypanosome derived medRNA to the 5' exonuclease.

To test this assumption, we incubated S100 extracts, containing abundant amounts of the medRNA, under the standard conditions, taking samples at different time points. RNA was recovered and resolved on 5% denaturing polyacrylamide urea (8M) gels, which were subsequently electro-blotted onto gene-screen plus membranes. To detect the 140 nt mini-exon precursor, we probed the filter with a ³²P end-labeled 24 nt oligonucleotide specific for the medRNA intron. As shown in Figure 5B, the hybridization signal of the medRNA (marked at 140 nt) remained stable throughout the 3-hour incubation in S100 extracts. This result is in sharp contrast to the degradation of SP6 medRNA transcripts which were incubated in an identical parallel assay (Panel 5A). The hybridization pattern of endogenous RNA in the S100 extract turned out to be drastically different from that of the total trypanosome RNA (Figure 5B) in that several other small RNA species, in addition to the 140 nt mini-exon precursor, hybridized with the mini-exon specific probe. These additional RNAs are more abundant in the S100 extract (compare figure 5B, lane 118a RNA with S100 RNA in lane 0). In addition, the S100 extract contained ample amounts of endogenous RNA estimated to be at concentrations of up to 160 ug per ml of extract. It is conceivable that the post-nuclear extract enriches small RNA species, thus explaining the difference in hybridization intensities of mini-exon precursors when compared between total RNA and S100 RNA.

It is important to determine whether purified trypanosome RNA added back to the extracts, is also resistant to the exonuclease. This can exclude the possibility that proteins, already bound to endogenous medRNA but not to the newly added, naked, SP6-derived medRNA, confer resistance. We therefore performed similar experiments in which we added previously

purified trypanosome RNA containing the medRNA, back to the S100 extracts and obtained identical results (data not shown). However, it is still possible that the medRNA is rapidly assembled into an RNP particle (40) which is resistant to the exo-ribonuclease. We, therefore, also carried out primer extension analysis using the hsp 70 mRNA as the target molecule for mini-exon identification to assure that the mini-exon moiety only is sufficient to confer resistance to the exo-ribonuclease. In addition, the advantage in studying protein coding mRNAs, instead of the medRNA, is that most of these molecules have been eliminated from the S100 extract during preparation due to their association with polysomes. Most of the signal that results in the primer extension will therefore be derived from the newly added previously purified mRNAs. Hsp 70 mRNA in T.brucei is known to have a mini-exon spliced onto its 5' end (27; Lee and Van der Ploeg, in preparation). Figure 5C shows the direct dideoxy nucleotide sequence of the T.brucei hsp 70 mRNA using a synthetic 30-mer as a primer, located 12 nt downstream of the 3' splice site. The nucleotide sequence up to the strong stop indeed represents the trypanosome mini-exon sequence derived from the T.brucei hsp 70 mRNA. The band indicated with an asterisk at 78 nt thus demarcates the 5' end of the cDNA derived from the mini-exon at hsp 70 mRNA. We then incubated 20 ug of total trypanosome variant 118a RNA, or trypanosome variant 118a RNA which was pretreated with Tobacco Acid Pyrophosphatase (data not shown) with S100 extract for various periods of time, ranging from zero to three hours. RNA samples were recovered and the fate of the 5' end of the hsp 70 mRNA was determined by primer extension with the hsp 70-derived oligonucleotide. Again, the major extension product was the 78 nt fragment, identifying the mini-exon, whether native or decapped (data not shown) hsp 70 mRNA was used. Note that the intensity of the signal in the S100 and 118a RNA mixture exceeds that of the control primer extensions performed with each sample separately (Figure 5D). Several minor extension products are also detected, some of which represent mini-exon addition at minor alternate 3' splice sites (Lee and Van der Ploeg, in preparation). However, the major signal in the primer extension, which identifies the mini-

Figure 6.

Mg²⁺ and ATP requirement for 5' exonuclease and RNA ligase. **Panel A:** By varying the magnesium concentration from 0 to 6 mM by 1 mM increments, the requirements of the exonuclease and ligase for Mg²⁺ was determined. Hsp 70 3' exon capped SP6 RNA was used as a substrate. The insert shows a shorter exposure of the autoradiogram. **Panel B:** Both the ATP and phosphocreatine concentrations of the extract were varied as indicated on top of each of the panels. The hsp 70 3' exon construct was used as a substrate.

exon at the 5' end of the hsp 70 mRNA is unchanged throughout the three-hour incubation in the S100 extract. These results show that, like the medRNA, the newly introduced hsp 70 mRNA is resistant to the 5' exo-ribonuclease. Control experiments in which SP6 RNA and 118a RNA were added simultaneously showed the predicted exonuclease sensitivity of the SP6 RNA (data not shown). Although the nucleotide changes in the sequence may thus render SP6 medRNA sensitive to the 5' exo-ribonuclease, it seems more plausible to attribute the difference in sensitivity of SP6 RNA compared to endogenous RNA to the absence of the unusual modified cap structure.

Titration of Mg^{2+} and ATP requirement of the enzymes in S100 extracts.

The exonuclease and ligase activities are similar in that they both require Mg^{2+} (Figure 6) but not the presence of Ca^{2+} (data not shown). This discriminates the exo-ribonuclease from previously identified Ca^{2+} dependent endo- and exo-ribonucleases in *T.brucei* (41). Figure 6A shows the Mg^{2+} titration assay, with 1 mM increments from 0 to 6 mM. In the absence of Mg^{2+} , neither cleavage nor ligation of hsp 70 SP6 RNA was observed (panel, 0). At a concentration of 1 mM Mg^{2+} , full activity of the exonuclease is restored (also see short exposure) while the ligation products are not yet detectable. With subsequent increases in the Mg^{2+} concentration circular molecules accumulate, without a further change in the cleavage pattern.

We finally tested the requirement of the two enzymatic activities for ATP and phosphocreatine (Figure 6B). In the absence of ATP and phosphocreatine, ligation does not proceed, whereas the exonuclease cleavage event occurs efficiently (panels labeled 0). After addition of ATP, the formation of circular hsp 70 RNA molecules resumes, to an extent comparable to the standard assay conditions in which both ATP and phosphocreatine are included. Note that the difference in the relative migration of circular high molecular weight fragments in panels A and B results from the fact that gels with different percentages of polyacrylamide are compared. Replacing ATP with phosphocreatine gives rise to a lag period of about one hour before the production of circular hsp 70 RNA (Figure 6B). This

one-hour lag period presumably corresponds to the phosphocreatine mediated ATP regeneration process in the trypanosome S100 extract. Hence, Mg^{2+} is required for both cleavage and ligation activities, but only the activity of the RNA ligase is ATP-dependent. We had previously noted that even though the nuclear extracts efficiently cleave the SP6 RNA, ligase activity seems absent or reduced (Figure 3C). We tested whether or not this could be due to a more enhanced ATPase activity in the nuclear extract than in the S100 extract. Indeed the ATPase activity immediately degrades all ATP in nuclear extracts, thus accounting for the reduced ligase efficiency in the nuclear extract (data not shown).

DISCUSSION

In several eukaryotes, 5' exo-ribonucleases have been identified that are capable of degrading uncapped but not capped RNAs (33, 48-50). The function of these exo-ribonucleases is unclear. It has been proposed that these enzymes serve a function in nuclear RNA processing or turn-over. In contrast, we have identified a 5'-3' exo-ribonuclease in S100 and nuclear extracts of trypanosomes which mediates the degradation of uncapped and capped SP6 RNAs but which cannot degrade the trypanosome derived mini-exon cap, in the in vitro assays. We attribute this resistance to the presence of four modified bases located at the 5' end of the in vivo-generated mini-exon. Since the SP6 medRNA differed from the trypanosome medRNA by four nucleotides of its primary sequence we cannot exclude that nucleotide specificity exists. Since the exo-ribonuclease is present in large quantities in nuclear as well as post-nuclear S100 extracts, we assume that during extraction the exo-ribonuclease leaks from the nucleus, as it occurs with medRNA. The exo-ribonuclease may, therefore, be a nuclear enzyme.

It is unclear how cap removal, shortening via exo-ribonuclease digestion and RNA circularization are brought about. Our initial protein purifications show that the exo-ribonuclease and RNA ligase are separate enzymes (data not shown). Obviously, RNA ligation first requires cap removal. We therefore tested capped RNAs as well as transcripts with triphosphates or monophosphates at their 5' ends (data not

shown). Even though each of these substrates is shortened by the exo-ribonuclease, RNA ligase activity was only observed if the SP6 RNA, capped or uncapped, was first trimmed by the exo-ribonuclease. In total three events are thus required: The first involves cap removal, presumably by the exo-ribonuclease, the second requires exo-ribonuclease mediated shortening of the SP6 RNA and the third involves ligation of the RNA.

We consider it unlikely that the exo-ribonuclease and its associated RNA ligase activity are enzymes of the tRNA processing pathway, since the nucleases involved in tRNA processing are highly specific endonucleases (51). The fact that the progressive shortening from the 5' end of SP6 RNAs stops at a specific point where the enzyme no longer proceeds, and the fact that the enzyme cannot degrade trypanosome derived capped RNAs indicate a function for the enzymes in RNA processing. Since the exo-ribonuclease terminates after about 20 nt only, it is most likely not responsible for the degradation of mRNA molecules. We have not been able to find structural elements in the different RNA substrates, which would explain at which site the exo-ribonuclease stops. Alternatively, proteins bound to the SP6 RNA template or secondary base modifications could protect the RNA from further exo-ribonuclease action (41, 49).

Micrococcal nuclease sensitivity of the exo-ribonuclease.

The exo-ribonuclease is sensitive to pretreatment of the S100 or nuclear extracts with micrococcal nuclease (MNase; data not shown). Control experiments, which confirmed that the enzyme is sensitive to MNase, included: EGTA + Ca^{2+} and MNase, zero time incubations; control for Ca^{2+} inhibition of the enzyme; control for DNase I sensitivity of the enzyme; and mix experiments of MNase treated and untreated extracts, showing that exo-ribonuclease activity can be restored by the addition of as much as 5% of an untreated extract. However, after purification and separation of the exo-ribonuclease and RNA ligase by DEAE-sephadex ion-exchange column chromatography the partially purified enzyme lost its MNase sensitivity. It is thus unclear whether in vivo the exo-ribonuclease requires an RNA co-factor.

The role of the cap in pre-mRNA processing is unclear. In vivo, the cap increases pre-mRNA stability and translational

efficiency (25, 42-44). In addition, the presence of a 5' cap is important in RNA splicing since its presence increases the efficiency of the RNA splicing reaction (33, 45, 46). The involvement of the cap in splicing is also illustrated by the fact that micromolar amounts of cap analogues in vitro inhibit splicing (45). In trypanosomes, the 5' end of every mRNA is peculiar and consists of the same 39 nt mini-exon with a 7mGpppA cap (7, 47). This mini-exon cap is unusual in that it is flanked by four bases that have undergone secondary base modifications by methylation (7). The function of this modified cap, which is found on every mRNA of trypanosomes, is unclear.

It is tempting to speculate that the *exo-ribonuclease* can discriminate between mini-exon containing mRNAs and primary transcripts of protein coding genes, which do not yet have a mini-exon spliced onto their 5' ends. A possible function of the *exo-ribonuclease* could thus be in the protection of intermediates of trans-splicing at the peculiar polycistronic pre-mRNAs of trypanosomes (35, 52). These predicted capless intermediates may be trimmed by the *exo-ribonuclease* and subsequently circularized. They could thus be protected from degradation by nucleases that degrade uncapped nuclear RNAs.

White and Borst (53) previously described an RNA ligase which circularizes small rRNAs in vitro. Subsequently Laird (personal communication) also observed an activity of this ligase on other RNAs. We thus assume that the enzymatic activity of their extract is similar to the activity we described. Our attempts to identify small (<2500 nt) circularized pre-mRNAs with in vivo ³²P labelled total and nuclear RNA have thus far been unsuccessful. It is, however, possible that larger circular RNAs exist, since these would not have been detected.

The finding of a 5' *exo-ribonuclease* that degrades capped RNAs for about 20 nt only, but which cannot degrade RNAs that have the common mini-exon cap trans-spliced onto their 5' ends, may allow a more detailed analysis of the function of the cap in pre-mRNA processing and mRNA turnover.

ACKNOWLEDGEMENTS

We thank all colleagues from the laboratory for critical reading of the manuscript and for discussion. We thank Mary Gwo-

Shu Lee for performing the dideoxy sequencing of hsp 70 mRNA and Argiris Efstratiadis for a gift of HeLa cell nuclear extracts. This work was supported by a grant from the National Institutes of Health (AI 21784) and by a grant from the John D. and Catherine T. MacArthur foundation.

REFERENCES

1. Van der Ploeg, L.H.T., Liu, A.Y.C., Michels, P.A.M., De Lange, T., Borst, P., Majumder, H.K., Weber, H., Veeneman, G.H. and Van Boom, J. (1982b) *Nucl. Acids Res.* 10, 3591-3604.
2. Boothroyd, J. and Cross, G.A.M. (1982) *Gene*, 20, 281-289.
3. Van der Ploeg, L.H.T., Cornelissen, A.W.C.A., Michels, P.A.M., and Borst, P. (1984) *Cell*, 39, 213-221.
4. Guyaux, M., Cornelissen, A.W.C.A., Pays, E., Steinert, M. and Borst, P. (1985) *EMBO J.* 4, 995-998.
5. Walder, J.A., Eder, P.S., Engman, D.M., Brentano, S.t., Walder, R.Y., Knutzon, D.S., Dorfman, D.M. and Donelson, J.E. (1986) *Science*, 233, 569-571.
6. Cornelissen A.W.C.A., Verspieren M.P., Toulme, J.J. Swinkels B.W. and Borst, P. (1986) *Nucleic Acids Res.*, 14, 5605-5614.
7. Freistadt, M.S., Cross, G.A.M., Branch, A.D. and Robertson, H.D. (1987) *Nucleic Acids Res.*, 15, 9861-9879.
8. Perry, K.L., Watkins, K.P. and Agabian, N. (1987) *Proc.Natl.Acad.Sci.*, 84, 8190-8194.
9. Sutton, R.E. and Boothroyd, J.C. (1988) *Mol. Cell. Biol.*, 8, 494-496.
10. Sutton, R. and Boothroyd, J.C. (1986) *Cell*, 47, 527-535.
11. Murphy, W.J., Watkins, K.P. and Agabian, N. (1986) *Cell*, 47, 517-525.
12. Ralph, D., Huang, J. and Van der Ploeg, L.H.T. (1988) *EMBO*, 7, 2539-2545.
13. Van der Ploeg, L.H.T. (1986) *Cell*, 47, 479-480.
14. Borst, P. (1986) *Ann. Rev. Biochem.*, 5, 701-732.
15. De Lange, T., Liu, A.Y.C., Van der Ploeg, L.H.T., Borst, P., Tromp, M.C. and Van Boom, J.H. (1983) *Cell*, 34, 891-900.
16. Nelson, R.G., Parsons, M., Bar, Ph.J., Stuart, K., Selkirk, M. and Agabian, N. (1983) *Cell*, 34, 901-909.
17. Kooter, J. M., De Lange, T. and Borst, P. (1984) *EMBO J.* 3, 2387-2392.
18. Campbell, D.A., Thornton, D.A. and Boothroyd, J.C. (1984) *Nature*, 311, 350-355.
19. Milhausen, M., Nelson, R.G., Sather, S., Selkirk, M. and Agabian, N. (1984) *Cell*, 38, 721-729.
20. Mount, S.M. (1982) *Nucl. Acids Res.*, 10, 459-472.
21. Shapiro, M.B. and Senapathy, S. (1987) *Nucl. Acids Res.* 15, 7155-7174.
22. Michels, P.A.M., Van der Ploeg, L.H.T., Liu, A.Y.C. and Borst, P. (1984) *EMBO J.*, 3, 1345-1351.
23. Lee, G.-S. M. and Van der Ploeg, L.H.T. (1987) *Mol. Cell. Biol.* 7, 357-364.
24. Lacy, E., Hardison, R.C., Quon, D. and Maniatis, T. (1979) *Cell*, 18, 1273-1283.

-
25. Green, M.R., Maniatis, T. and Melton, D.A. (1983) *Cell*, 32, 681-694.
 26. De Lange, T., Michels, P.A.M., Veerman, H.J.G., Cornelissen, A.W.C.A. and Borst, P. (1984) *Nucl. Acids Res.*, 12, 3777-3791.
 27. Glass, D., Polvere, R.I. and Van der Ploeg, L.H.T. (1986) *Mol. Cell. Biol.*, 6, 4657-4666.
 28. Lockard, R.E., Rieser, L. and Vournakis, J.N. (1981) In: *Structural analysis of Nucleic Acids*. (eds. J.C. Chirikjian and T.S. Papas; Elsevier N. Holland; Gene Amplification and Analysis, Vol 2).
 29. Grabowski, P.J., Seller, S.R. and Sharp, P.A. (1985) *Cell*, 42, 345-353.
 30. Hernandez, N. and Keller, W. (1983) *Cell*, 35, 89-99.
 31. Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) *Nucl. Acids Res.*, 11, 1475-1489.
 32. Levine, R. and Braner, R.W. (1951) *J. Lab. Clin. Med.* 38, 474-480.
 33. Krainer, A.R., Maniatis, T., Ruskin, B. and Green, M.R. (1984) *Cell*, 36, 993-1005.
 34. Dudler, R. and Travers, A.A. (1984) *Cell*, 38, 391-398.
 35. Shea, C. Lee, M. G.-S. and Van der Ploeg, L.H.T. (1987) *Cell*, 50, 603-612.
 36. Auffray, C., and Rougeon, F. (1980) *Eur. J. Biochem.* 107, 303-314.
 37. Zeitlin and Efstratiadis, (1984) *Cell*, 39, 589-602.
 38. Kole, R. and Weissman, S.M. (1982) *Nucleic Acids Res.*, 10, 5429-5445.
 39. Ruskin, B., Krainer, A.R., Maniatis, T. and Green, M.R. (1984) *Cell*, 36, 993-1005.
 40. Miller, S.I. and Wirth, D.F. (1988) *Mol. Cell. Biol.* 8, 2597-2603.
 41. Ghenle, G.O. (1985) *Mol. Biochem. Parasitol.* 15, 37-47.
 42. Banerjee, A.K. (1980) *Microbiol. Rev.*, 44, 175-205.
 43. Rhoads, R.E. (1985) *Prog. Mol. Subcell. Biol.*, 9, 104.
 44. Furuichi, Y., Lafandra, A. and Shatkin, A.J. (1977) *Nature*, 266, 235-239.
 45. Konarska, M.M., Padgett, R.A and Sharp, P.A. (1984) *Cell*, 38, 731-736.
 46. Patzelt, E., Blass, D. and Kuechler, E. (1983) *Nucleic Acids Res.*, 11, 5821-5835.
 47. Laird, P.W., Kooter, J.M., Loosbroek N. and Borst, P. (1985) *Nucleic Acids Res.*, 13, 4253-4266.
 48. Ruskin, B. and Green, M.R. (1985) *Cell*, 43, 131-142.
 49. Noble, J.C.S., Prives, C. and Manley, J.L. (1986) *Nucleic Acids Res.*, 14, 1219-1235.
 50. Stevens, A. and Maupin M.K. (1987) *Nucleic Acids Res.*, 15, 695-708.
 51. Peebles, C.L., Gegenheimer, P. and Abelson, J. (1983) *Cell*, 32, 525-536.
 52. Johnson, P.J., Kooter, J.M. and Borst, P. (1987) *Cell*, 51, 273-281.
 53. White, T.C and Borst, P. (1987) *Nucleic Acids Res.*, 15, 3275-3290.
 54. Laird, P.W., Zomerdijs, J.C.B.M., De Korte, D. and Borst, P. (1987) *EMBO J.* 6, 1055-1062.
-