Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Cell, Vol. 47, 527–535, November 21, 1986, Copyright © 1986 by Cell Press

Evidence for *Trans* Splicing in Trypanosomes

Richard E. Sutton and John C. Boothroyd
Department of Medical Microbiology
Stanford University School of Medicine
Stanford, California 94305

Summary

The 5′ ends of trypanosome mRNAs consist of an identical sequence of 35 nucleotides. This "mini-exon" sequence is derived from the 5′ end of a 137 nucleotide RNA (medRNA). The remainder of each mRNA is derived from a protein-coding exon that is not linked to the mini-exon. We propose that medRNA is spliced in *trans* to de-novo-initiated transcripts of protein-coding genes. This *trans* splicing model predicts that the downstream portion of medRNA will be part of a branched structure and then be released as a free product (minRNA). We demonstrate that significant levels of minRNA exist in trypanosome RNA. Furthermore, minRNA can be released from high molecular weight RNA by a HeLa cell S100 "debranching" extract. We conclude that *trans* splicing is the physiological process by which mature mRNA molecules are synthesized in trypanosomes.

Introduction

All examined mRNAs of Trypanosoma brucei, including those coding for the antigenically distinct variant surface glycoproteins, contain the same 5′ untranslated leader of 35 nucleotides (Boothroyd and Cross, 1982; De Lange et al., 1984a; Van der Ploeg et al., 1982; Dorfman and Donelson, 1984; Parsons et al., 1984). The "mini-exon" encoding this spliced leader is part of a 1.35 kb segment that is tandemly repeated approximately 200 times (in one or more clusters) in the trypanosome genome (De Lange et al., 1983; Michiels et al., 1983; Nelson et al., 1983). Homologous sequences have been detected in the genomes and on mRNAs of related kinetoplastid organisms (De Lange et al., 1984b; Nelson et al., 1984).

Recently it was found that the mini-exon repeat directs the synthesis of a discrete 137 nucleotide transcript, termed the mini-exon–derived RNA (medRNA) or spliced-leader RNA (Campbell et al., 1984; Kooter et al., 1984; Milhausen et al., 1984). The 5′-most 35 nucleotides of medRNA, corresponding to the mini-exon sequence itself, are immediately followed by a consensus 5′ splice signal (GUAUGA). Similarly, immediately upstream of all protein-coding exons examined (Bernards et al., 1984; Boothroyd and Cross, 1982; Clayton, 1985; Gonzalez et al., 1985; Kimmel et al., 1985; Liu et al., 1983; Michels et al., 1986; Pays et al., 1982; Sather and Agabian, 1985; Swinkels et al., 1986; Tschudi et al., 1985; Van der Ploeg et al., 1982), there is a consensus 3′ splice site ($[C/U]_nNNAG$). Together with the fact that the mini-exon repeats and at least some structural genes are on different chromosomes (Van der Ploeg et al., 1984: Guyaux et al., 1985), these observations have led to the suggestion that mRNAs in trypanosomes are generated by a novel process termed "discontinuous transcription" to reflect the fact that a chimeric RNA is generated by transcription of two unlinked loci (reviewed in Boothroyd, 1985; Borst, 1986).

Several models have been put forward to describe how chimeric mRNAs are generated (Campbell et al., 1984; Kooter et al., 1984; Milhausen et al., 1984). In the transcription-reinitiation model, medRNA is synthesized first and then used as a primer for reinitiation of transcription upstream of the structural gene. In the second model, the two RNA molecules are independently transcribed and then ligated end-to-end. In both of these models, a contiguous precursor RNA is thus formed that, as described above, would have consensus splice-site signals at the exon–intron boundaries. Both of these models predict, therefore, that a conventional intramolecular splicing event is responsible for processing of the precursor RNAs into a mature mRNA. A third model proposes that the two RNA molecules are transcribed separately, but that the two exons are then combined by an intermolecular splice (i.e., in *trans*).

A unique prediction of the *trans* splicing model is the generation of an RNA species of 102 nucleotides corresponding to the 3′ portion of the medRNA. If we ignore the possible addition of one or more nucleotides during the splicing process itself (as in the splicing of group I introns in other systems [Cech and Bass, 1986]), the 5′ end of such a molecule would be at the 5′ splice site, and its 3′ end would correspond to the 3′ end of medRNA. Here, we demonstrate that precisely such a molecule exists in substantial amounts in the trypanosome. We further show that debranching of size-fractionated RNA from trypanosomes releases the 102 nucleotide species. These data strongly suggest that intermolecular or *trans* splicing is a physiological process in trypanosomes.

Results

Detection of a Second RNA Species from the Mini-Exon Repeats

Using a synthetic oligonucleotide probe, we previously reported the detection of a unique 137 nucleotide transcript from the mini-exon repeat, the first 35 nucleotides of which correspond to the mini-exon sequence itself (Campbell et al., 1984). In subsequent analysis using hybridization probes with substantially higher specific activity, we consistently observed an additional, minor product of approximately 100 nucleotides (data not shown). The RNA used in these latter experiments was prepared from purified trypanosomes by a procedure involving many manipulations, and thus was subject to possible artifacts. We therefore chose to repeat these experiments using RNA prepared by the most rapid method possible. To do this, infected rats with high parasitemia were exsanguinated, and the RNA was rapidly extracted with hot phenol

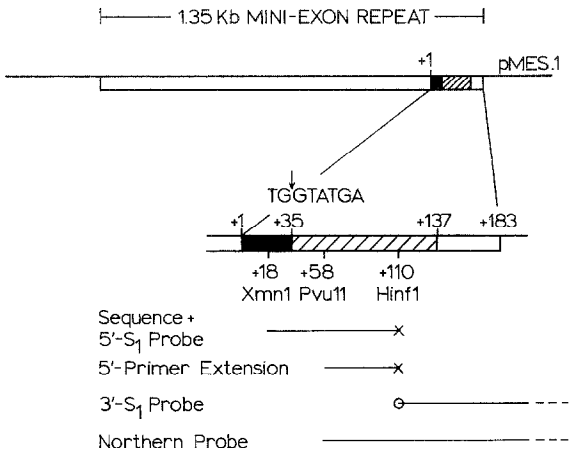


Figure 1. Map of pMES.1 and End-Labeled Probes Used for RNA Analysis

The boxed region of pMES.1 represents the 1.35 kb Sau3A–Sau3A mini-exon repeat, which has been completely sequenced (Campbell et al., 1984). The filled-in portion of the insert corresponds to the 35 bp mini-exon, and position +1 is the start site of transcription in vivo. The hatched portion represents nucleotides +36 to +137 of medRNA, which we have termed the mini-intron (see text). The sequence of the mini-exon repeat spanning the splice junction is shown; the arrow indicates the position of the 5′ splice site. The positions of relevant restriction sites are also shown. Refer to Experimental Procedures for description of the construction of the probes. A 5′-^{32}P-labeled end is indicated by an "x", and a 3′-^{32}P-labeled end is indicated by an "o".

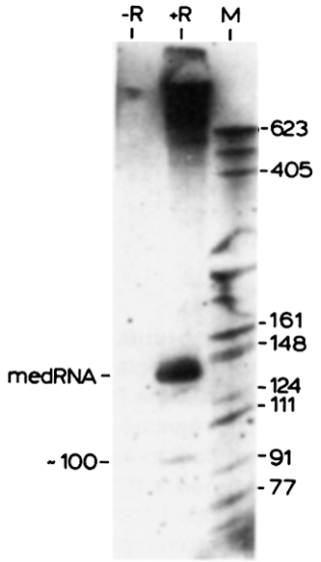


Figure 2. Northern Blot of Low Molecular Weight RNA

Lane 1 (−R), 100 μg of yeast tRNA; lane 2 (+R), 40 μg trypanosome total RNA; lane 3 (M), 3′-end-labeled MspI fragments of pAT153 (sizes in nucleotides indicated at right). The yeast tRNA was purchased from Sigma and was extracted with phenol. The trypanosome total RNA was made by the hot-phenol method directly from infected rat blood. RNA samples were denatured and size-fractionated as described in Experimental Procedures. The nucleic acid was electrophoretically transferred to Genetrans paper and was probed with antisense RNA corresponding to nucleotides +58 to +180 of the mini-exon repeat (see Figure 1) as described in Experimental Procedures. The positions of medRNA and the ~100 nucleotide RNA species are indicated.

as described in Experimental Procedures. Forty micrograms of this RNA was fractionated on a denaturing polyacrylamide gel, transferred to Genetrans paper, and probed with an antisense RNA probe corresponding to the sequence from +58 through +183 relative to the start site of transcription of medRNA (see Figure 1).

The resulting autoradiogram (Figure 2) reveals two bands: one species of 137 nucleotides (medRNA, as previously reported) and one species of about 100 nucleotides, present at about 12% of the medRNA level (as determined by densitometric analysis of the autoradiogram shown). The hybridization at the very top of the gel is probably nonspecific since it is also observed when total yeast RNA is electrophoresed and similarly probed (data not shown). With such probes, the smaller species was consistently detected in this RNA preparation, in total RNA extracted by guanidine isothiocyanate or guanidine-HCl, and in nuclear RNA (data not shown). The size of this smaller RNA species is approximately that predicted for a by-product of *trans* splicing (see above). We chose, therefore, to determine if it had the exact termini predicted by such a model.

Precise Characterization of the 100-mer

S1 nuclease and primer extension analyses were used to identify the exact 5′ end of the 100-mer. The fragment used in the primer extension experiment was 5′-^{32}P-labeled at the HinfI site at position +110 of pMES.1 and then recut at the PvuII site at position +58 (Figure 1). This fragment was hybridized to 10 μg of total RNA and then extended with reverse transcriptase. Three extension products were observed: two at about 110 nucleotides and one extending to position +36 (Figure 3).

The extension product predicted for medRNA is 110 nucleotides; therefore, medRNA is the probable source of at least one of the two bands in this region. We cannot explain the presence of two bands (differing by about 4 nucleotides in relative mobility), but we have found their relative amounts to vary from experiment to experiment (compare, for example, the autoradiograms of Campbell et al. [1984] and Boothroyd et al. [1985]). The difference could be due to some sort of microheterogeneity in the medRNA or to capping or some other posttranscriptional modification. The product extended to position +36 corresponds to an RNA species whose 5′ end is precisely at the 5′ splice site of medRNA, as would be expected for the putative by-product of *trans* splicing.

Using an S1 probe, also 5′-end-labeled at the HinfI site at position +110 but extending up to the XmnI site at position +18, substantial full-length protection from S1 nuclease was observed (lane 7, Figure 3) in addition to protected bands corresponding to species terminated at positions +35, +37, and +38. This pattern of three protected species terminated near +36 was consistently observed (see, for example, Figure 4) even when the S1 nuclease concentration was varied over two orders of magnitude (data not shown). This pattern of protection indicates that there is an RNA species whose 5′ end is within two nucleotides of the 5′ splice site, which is consistent with the primer extension result above.

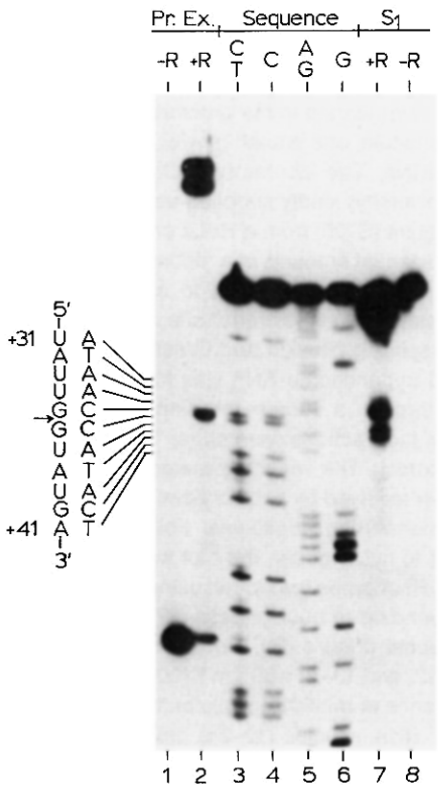


Figure 3. S1 Nuclease Protection and Primer Extension of Total RNA

Lanes 1 and 2 show primer extension using the 52 nucleotide 5′-end-labeled probe spanning nucleotides +58 to +110 (see Figure 1). Lane 1 (−R) shows extension products when the hybridization included 100 μg yeast tRNA alone; Lane 2 (+R) shows extension products in the presence of 10 μg of trypanosome total RNA in addition to the carrier tRNA. Lanes 3–6 show a chemical sequence ladder of the S1 probe, with the specificities of the reactions indicated at the top of each lane. Lanes 7 and 8 show the pattern of S1 nuclease protection with the 5′-end-labeled probe spanning +18 to +110. Lane 7 (+R) shows protection in the presence of 10 μg of total trypanosome RNA and 100 μg of yeast tRNA as carrier; lane 8 (−R) shows protection when the hybridization included 100 μg of yeast tRNA alone. The sequences of the antisense probe and the corresponding transcript from position +31 to +41 are indicated at the left, with an arrow indicating the 5′ splice site in the RNA sequence.

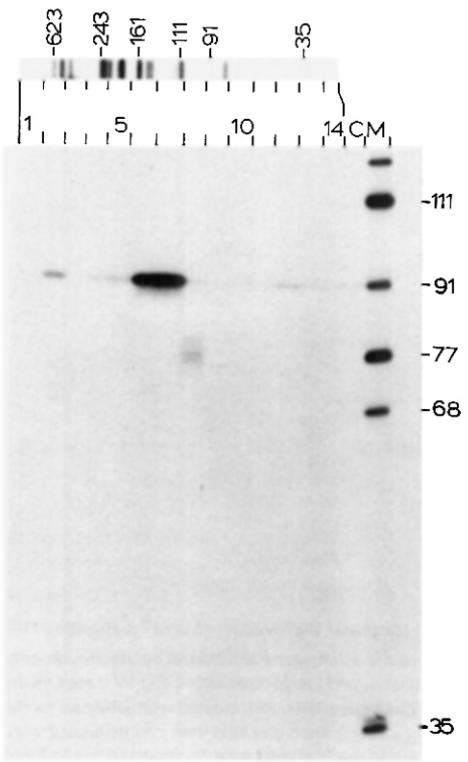


Figure 4. S1 Nuclease 5′ Protection of Size-Fractionated RNA

Total trypanosome RNA (100 μg) was denatured in formamide and then fractionated on a 5% polyacrylamide–7 M urea gel. The gel was cut into 14 pieces (13 of 1.0 cm and the fourteenth of 0.5 cm), and the RNA was electroeluted as described in Experimental Procedures. S1 nuclease protection using the 5′-end-labeled probe corresponding to nucleotides +18 to +110 (see Figure 1) was performed as outlined in Experimental Procedures. The resulting DNA was denatured and then electrophoresed on an 8% polyacrylamide–7 M urea sequencing gel. An autoradiogram of radiolabeled markers (pAT153 cut with MspI) run in parallel with the fractionated RNA is shown above the lanes; markings indicate the limits of each gel slice. The numbering of the lanes (1–14) refers to the gel slices, beginning at the top of the gel. In lane C, only 100 μg of yeast tRNA carrier was included in the hybridization reaction. Lane M contains DNA size markers (MspI fragments of pAT153), whose lengths are indicated at right.

It is our experience that primer extension and S1 nuclease analyses are complementary but that primer extension is more reliable than S1 nuclease protection as an indicator of the exact 5′ terminus of an RNA molecule. S1 nuclease can produce incomplete hydrolysis in the immediate vicinity of the transition point from double-stranded to single-stranded nucleic acid. We conclude, therefore, that a RNA species exists within the trypanosome whose 5′ terminus is at position +36.

To determine whether the 100-mer identified in the Northern blot (Figure 2) is the same molecule as that in total RNA that has its 5′ end at position +36, we repeated the S1 nuclease mapping and primer extension studies on size-fractionated RNA. To do this, total RNA was fractionated on a polyacrylamide gel, which was then sliced, and the RNA from individual slices was electroeluted. The eluant of each fraction was coprecipitated with the same 5′-end-labeled probe used in the experiment described in Figure 3, and standard S1 nuclease protection reactions were performed (Figure 4). The major protection of full-length probe observed in fractions 6 and 7 in Figure 4 represents hybridization of the probe to medRNA. Background protection of full-length probe in other fractions probably represents back-hybridization between the two strands of the input probe. The only fraction to show protected species of less than full length is fraction 8, which comprises RNA molecules of 90–110 nucleotides (Figure 4). This fraction showed three protected species in the size range 74–78 nucleotides. This is the same pattern of bands observed in S1 nuclease protection experiments on total RNA described above, and thus confirms that an RNA molecule of about 100 nucleotides is responsible for the protection at position +36 noted in total RNA. When primer extension experiments were performed using size-fractionated RNA and the 5′-end-labeled primer described

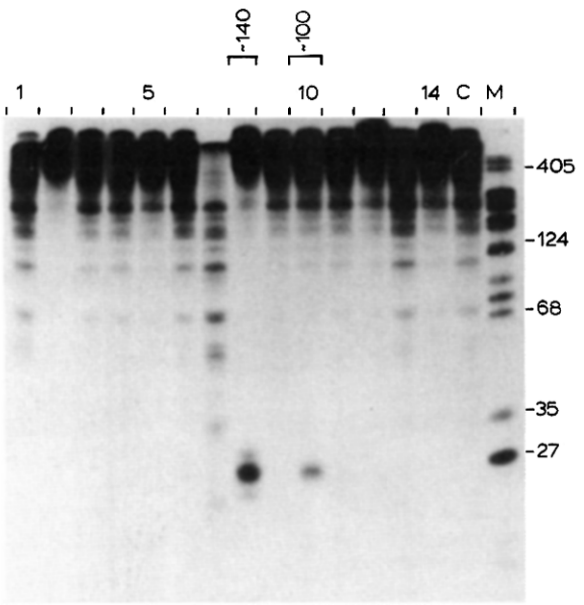


Figure 5. S1 Nuclease 3′ Protection of Size-Fractionated RNA

Total trypanosome RNA was fractionated by electrophoresis as described in Figure 4, and the gel was sliced into 14 pieces as described in Figure 4. The eluted RNA was treated with DNAase as described in Experimental Procedures and was then coprecipitated with 100 μg of yeast tRNA and the 3′-end-labeled S1 probe shown in Figure 1. S1 nuclease protection was performed as described in Experimental Procedures. The resulting DNA was denatured and then electrophoresed on a 10% polyacrylamide–7 M urea gel. Lanes 1–14 refer to the gel slices, beginning at the top of the gel. Fractions containing medRNA (~140) and the mini-intron RNA (~100) are indicated; this was determined by ethidium bromide staining of DNA markers run in parallel (not shown). In lane C, only 100 μg of yeast tRNA was included in the hybridization reaction. Markers (lane M) are MspI fragments of pAT153 whose lengths are indicated at right. The partial degradation of the probe in fraction 7 observed in this experiment was not reproducible.

in Figure 1, a product extended to position +36 was found in only those fractions corresponding to the 100-mer, whereas the full-length extension product (to position +1) was present only in fractions containing medRNA (data not shown).

To establish that the 100-mer has the same 3′ terminus as medRNA, we used for S1 nuclease protection analysis a probe 3′-end-labeled at the HinfI site at position +110 (see Figure 1), and prepared more size-fractionated RNA in the manner described above (Figure 5; note, however, that the gels were not run identically, and therefore the size range of RNA in a given fraction is not exactly comparable in the two experiments). A major protected species of 27 nucleotides was observed in fractions 8 and 10 (Figure 5), indicating that the 3′ ends of the RNAs in these fractions are at the same position, i.e., at position +137 (see Figure 1). Fraction 10 corresponds to the 100-mer, and fraction 8 to medRNA (Figure 5). Other, higher bands were apparent in these and the remaining lanes, but, since they were also present in the control lane (Figure 5, lane C), which received only carrier tRNA, they do not reflect protection due to specific trypanosome transcripts.

To maintain a consistent nomenclature, we will refer to this 102 nucleotide RNA as mini-intron RNA or minRNA.

The Mini-Intron RNA Also Exists As Part of a Branched Structure

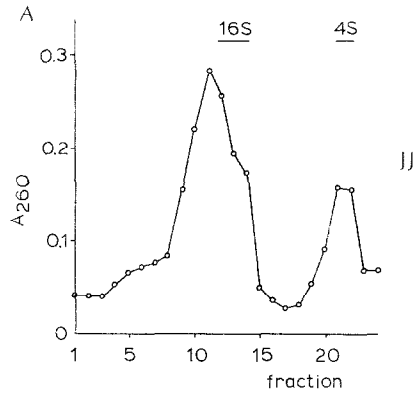
If *trans* splicing occurs analogously to *cis* splicing, an expected intermediate in the process would be a Y-shaped RNA molecule, one branch of which would correspond to the minRNA. The laboratory of Dr. Michael Green (Harvard University) kindly supplied us with the 100,000 × g supernatant (S100) from a HeLa cell extract. This extract has substantial amounts of a "debranching" activity which has been previously shown to cleave the 5′–2′ phosphodiester bond at the branch site of lariats produced during *cis* splicing (Ruskin and Green, 1985).

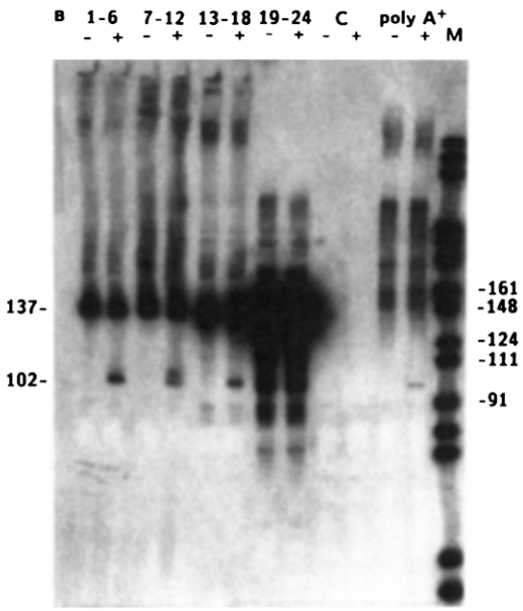
Total trypanosome RNA was fractionated by centrifugation through a sucrose gradient (Figure 6A), and aliquots of the fractions were either treated or not with this S100 extract. The resulting material was deproteinized and then resolved by polyacrylamide gel electrophoresis under denaturing conditions. Following electrophoretic transfer to nylon paper, the blot was probed with an antisense RNA probe (produced using T7 RNA polymerase) corresponding to nucleotides −275 to +110 of the mini-exon repeat (Figure 6B). Treatment of pools of fractions 1–6, 7–12, and 13–18 with the S100 extract resulted in the appearance of minRNA, while in the fractions containing smaller RNA species (19–24), treatment caused an increase in the amount of minRNA detected (Figure 6B).

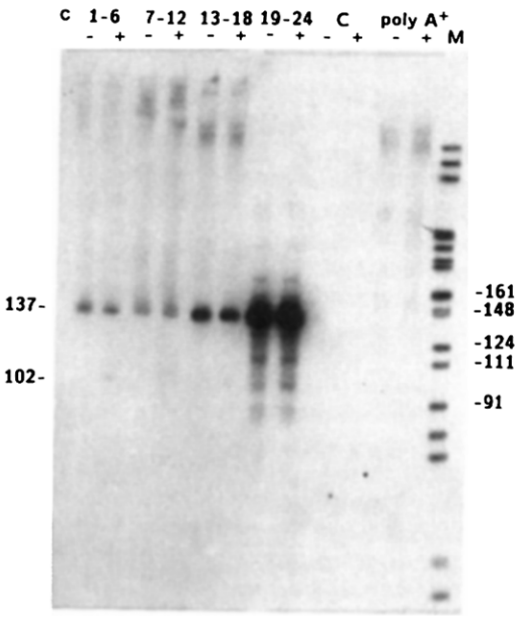
The most likely interpretation of these results is that a bond, similar to the 5′–2′ branch produced in *cis* splicing (Ruskin et al., 1984; Padgett et al., 1984), links the minRNA to the high molecular weight RNA and that the HeLa cell extract is specifically cleaving at this site, resulting in the release of the minRNA. The only plausible alternative to this explanation is that the extract is specifically recognizing the 5′ splice site of medRNA and cleaving it to release the minRNA. Although medRNA does contaminate the high molecular weight fractions, there is no correlation between its amount and that of the minRNA generated by treatment with the extract. For example, as shown in Figure 6C, the amount of medRNA in pooled fractions 13–18 is about five times that in fractions 1–6. Yet, as seen in Figure 6B, the opposite applies to minRNA: the amount in pooled fractions 13–18 is substantially less than that in pooled fractions 1–6. Figure 6B further demonstrates that the HeLa cell extract can release minRNA from poly(A)$^+$ RNA that lacks medRNA. We therefore discount the possibility that the HeLa cell extract is cleaving medRNA, and conclude that the minRNA is associated with high molecular weight RNA as part of a branched structure. This branching is most likely a result of a 5′–2′ phosphodiester bond between the guanosine at the 5′ end of minRNA (position +36; see Figure 1) and some nucleotide (adenosine, by analogy to *cis* splicing) within the intron portion of a precursor RNA. A more detailed model for this is presented and discussed below.

Discussion

We demonstrate here that, in addition to medRNA, trypanosomes possess a discrete and relatively abundant RNA of 102 nucleotides corresponding to nucleotides







+36 through +137 of medRNA. We do not believe this is an artifact of the RNA preparation, since several different RNA preparations, including RNA rapidly prepared from trypanosomes in whole rat blood, gave the same result. This molecule has the precise characteristics predicted of one of the products of *trans* splicing in that its 5′ end exactly coincides with the 5′ splice site of medRNA, and its 3′ terminus corresponds precisely to the 3′ terminus of medRNA. We have termed this product minRNA (mini-intron RNA). We have also shown that minRNA can be released from poly(A)$^+$ RNA (which lacks medRNA) and from high molecular weight RNA by incubation with the debranching S100 extract from HeLa cells.

The results presented here are consistent with a mechanism of *trans* splicing that is closely analogous to the type II *cis* splicing of nuclear-encoded RNAs in yeast and vertebrates (reviewed in Padgett et al., 1986). The same consensus sequences are present at the 5′ and 3′ splice sites for both types of splicing, and the S100 extract from HeLa cells (cells in which *trans* splicing is not known to occur in vivo) is capable of recognizing and cleaving the putative branched structures.

In Figure 7 we present a model for *trans* splicing that uses the results presented here and adapts them to the *cis* splicing models developed for other systems. This model proposes that the first step in *trans* splicing is cleavage at the 5′ splice site of medRNA coupled to branch formation at a site upstream of the 3′ splice site in the acceptor RNA (aRNA). The mini-exon sequence is then linked to the protein-coding exon of the aRNA to yield a mature mRNA and a branched structure composed of the two introns. The final step is the debranching of the two introns, resulting in three products: a mature mRNA including the 5′-most 35 nucleotides of medRNA, a segment of unknown length corresponding to the intronic sequences upstream of the protein-coding exon, and a mini-intron comprising the 3′-most 102 nucleotides of medRNA. An alternative mechanism involving a simple crossover be-

Figure 6. Debranching of Size-Fractionated RNA

Total trypanosome RNA was fractionated on a sucrose gradient as described previously (Campbell et al., 1984). (A) The absorbance profile at 260 nm of the gradient. (B) Fractions were pooled in groups of six, and a sample representing 15% of each pool was precipitated with 25 μg of Escherichia coli tRNA. A sample (2 μg) of Poly[A]$^+$ RNA prepared as described in Experimental Procedures was also coprecipitated with 25 μg E. coli tRNA. Aliquots of each pool and the poly(A)$^+$ RNA were treated or not (+ or −) with 5 μl of HeLa cell S100 debranching extract exactly as described previously (Ruskin and Green, 1985). Following treatment, all samples were treated with proteinase K (50 μg per reaction) in 1% SDS, 12.5 mM EDTA, 150 mM NaCl, and 0.1 M Tris–HCl (pH 7.5). The RNA was then extracted with phenol and chloroform, precipitated with ethanol, and resolved by electrophoresis on a 5% polyacrylamide–7 M urea gel. The RNA was transferred to Genetrans paper and was probed with an antisense RNA probe corresponding to nucleotides −275 to +110 of the mini-exon repeat as described above. The filter was washed extensively in 0.1× SSC, 0.1% SDS at 65°C. The fractions in each pool are indicated at the top of each pair of lanes. Poly(A)$^+$ RNA was used in lanes marked poly A$^+$. Control reactions (receiving carrier tRNA only) are marked with a C. DNA size markers (lengths indicated at right) are MspI fragments of pAT153. (C) Shorter exposure of the autoradiogram shown in (B) in order to indicate the relative amounts of medRNA in each pool.

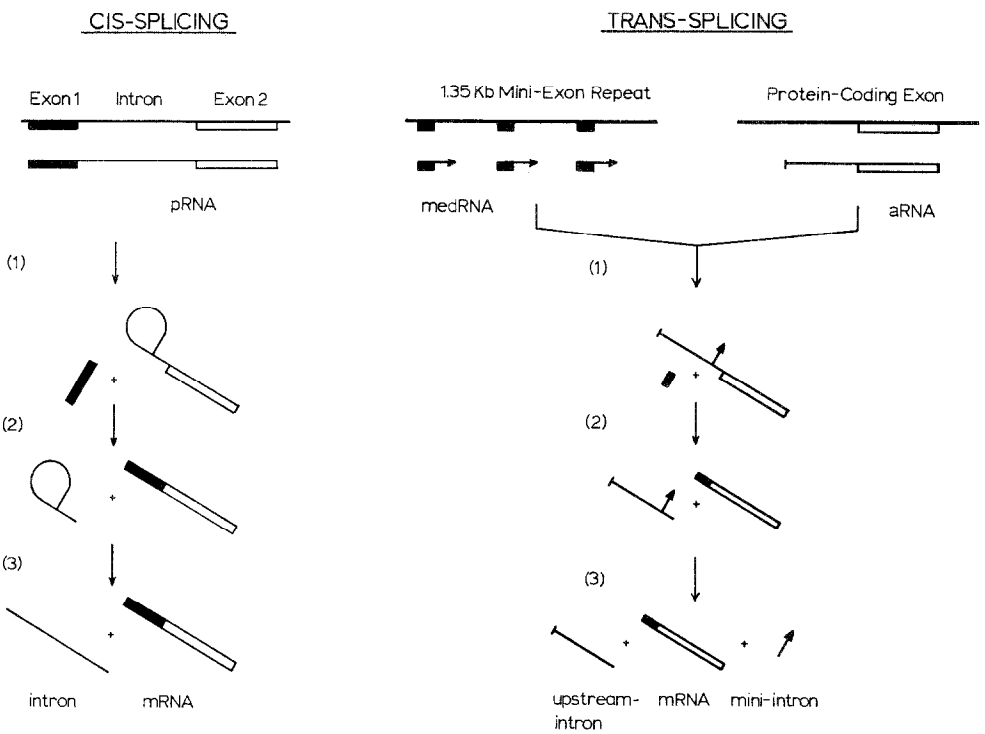


Figure 7. Comparison of *Cis* and *Trans* Splicing

A model of "conventional" *cis* splicing is shown at left. This is based on the published work of others on *cis* splicing in yeast and higher organisms (for review, see Padgett et al., 1986). Step 1 involves breakage of the 5′–3′ phosphodiester bond at the 5′ exon–intron boundary, and the generation of a 5′–2′ branched molecule (or "lariat"). The upstream exon 1 is then linked to exon 2 with release of free lariat intron in step 2. Step 3 in the debranching of the lariat to yield a linear molecule. At right is shown a model of *trans* splicing in trypanosomes. The top left line represents the tandem, genomic mini-exon repeats, each of which directs the synthesis of medRNA, composed of the 35 nucleotide mini-exon sequence (shown as a filled-in box) and a 3′ tail of 102 nucleotides (the arrow; not to scale). The top right line represents the DNA of a protein-coding gene, with the exon as the open box. A precursor RNA (which we have termed the aRNA) is transcribed from this region. The precise 5′ and 3′ limits for any aRNA are not known. The steps in *trans* splicing are analogous to those of *cis* splicing, shown at left. There are three final products of this reaction, however: the upstream intron of aRNA, the mini-intron, and the mature mRNA.

tween the medRNA and aRNA (i.e., exchange between the two phosphodiester bonds) would not be expected to result in the minRNA being linked as a 5′–2′-bonded branch on higher molecular weight RNA. It would, rather, predict a conventional 5′–3′ linkage with no obvious signal for release of the minRNA. We do not, therefore, consider a model involving such reciprocal crossing-over a likely alternative.

The model in Figure 7 predicts that two types of branched molecules should exist: those possessing the attached protein-coding exon and those generated after displacement of the branched intron by the mini-exon sequence. These latter molecules would clearly be smaller than the former, and this may explain the finding that the putative branched molecules are detected in the four pooled size fractions (Figure 6B).

This is the first biological system in which *trans* splicing has been suggested to have an important and routine physiological function. Prior to this report, analogies have been drawn between the 5′ leader of trypanosome mRNAs and the common 5′ leader of coronavirus mRNAs (see, for example, Boothroyd, 1985). However, it is now clear that in this latter system, "discontinuous transcription" describes the use of a common RNA primer in the reinitiation of transcription upstream of many protein-coding regions (Makino et al., 1986), and thus is quite distinct from the model proposed here.

Trans splicing in vitro has been observed in the presence of HeLa cell nuclear extracts in experiments with synthetic donor (containing a 5′ splice site) and acceptor (containing a 3′ splice site) RNAs (Konarska et al., 1985; Solnick, 1985). However, this process is extremely inefficient (at best only a few percent of the level of *cis* splicing in vitro) and may require base pairing, although the latter is the subject of some debate (Solnick, 1986; Sharp and Konarska, 1986). We find no significant complementarity between the mini-intron region and the introns located upstream of trypanosome protein-coding exons sequenced to date.

One important question raised by these findings concerns the function of the mini-exon sequence and *trans* splicing. Most possibilities that have been suggested center around the stabilization, transport, or translation of mRNAs (for review, see Boothroyd, 1985, and Borst, 1986). Until a class of mRNAs lacking the mini-exon sequence is found, it will continue to be difficult to choose between these various possible functions. Yet, whatever its function, the mini-exon has apparently been important

enough to be retained (even at the level of the mini-exon sequence) throughout the Kinetoplastida (De Lange et al., 1984b; Nelson et al., 1984).

Before we consider the possible evolutionary relationship between *cis* and *trans* splicing, three important pieces of information must be recalled. First, trypanosomes are ancient organisms thought to have diverged from the eukaryotic line before the divergence of the plants, animals, and fungi (Sogin et al., 1986). Second, "conventional" introns (i.e., those immediately flanked by adjacent exons) have not yet been observed in any trypanosome gene so far examined nor, to our knowledge, in any gene from an equally primitive eukaryote. Third, *cis* and *trans* splicing must have a related ancestry because of their use of the same consensus sequences at the 5′ and 3′ splice sites.

Given all this, the question that must be addressed is whether one of *cis* and *trans* splicing is the progenitor of the other, or whether they in fact have arisen independently from a common ancestor. Based on the above, the simplest interpretation would be that *cis* splicing evolved from *trans* splicing at some point after the kinetoplastids split away from the main line of the eukaryotic evolutionary tree. This would be in conflict with recent suggestions that *cis* splicing of the sort discussed here existed in bacteria, but disappeared in these and other fast-growing organisms (introns are very rare in yeast) because of selection against such energetically expensive pathways (discussed in Gilbert et al., 1986; Sharp, 1985). An alternative explanation for the apparent absence of *cis* splicing in trypanosomes might be that *cis* splicing cannot coexist with *trans* splicing because of the molecular chaos that would result if independent protein-encoding exons were randomly spliced together in *trans* without regard to their coding potential. Obviously, one way around this latter problem would be for the *trans* splicing machinery to recognize only the mini-exon and the most upstream exon of an aRNA as substrates. Such a constraint could explain the strong conservation of the mini-exon sequence in the different genera of the Kinetoplastida, although the presence of the same consensus sequence directly at the 5′ splice sites of medRNA and the *cis* splicing substrates of higher organisms does not suggest any obvious specificity.

Despite the above, it should be noted that *trans* splicing could have an important evolutionary role in allowing the occasional generation of chimeric mRNAs (resulting from aberrant splicing between acceptor RNAs). If reintegrated back into the genome (through reverse transcription), these chimeric mRNAs might lead to the evolution of new genes.

Determination of the precise relationship between *cis* and *trans* splicing will require further work on the detailed mechanisms underlying both phenomena, and comparative studies on other primitive eukaryotes. However, whatever the outcome, two things are clear. First, *trans* splicing adds a new level of complexity to gene regulation in trypanosomes, and it will require elucidation before other important phenomena, such as antigenic variation, can be fully understood. Second, the existence of *trans* splicing and the apparent absence of *cis* splicing in these primitive eukaryotic organisms will need to be incorporated into any new models of when, how, and why *cis* splicing evolved.

Experimental Procedures

Reagents

Restriction endonucleases and other DNA-modifying enzymes were purchased from New England Biolabs. T7 RNA polymerase was obtained from Pharmacia, and S1 nuclease was from Bethesda Research Laboratories. Polynucleotide kinase was purchased from U.S. Biochemicals, and [γ-^{32}P]ATP and [α-^{32}P]UTP were from ICN Biochemicals. Genetrans paper was purchased from Plasco (Woburn, MA), and reverse transcriptase was from Life Sciences. HeLa cell S100 extract was provided by Dr. Michael Green's laboratory (Harvard University).

Construction of Probes and Primers

Construction and nucleotide sequencing of the recombinant plasmid pMES.1 (Figure 1) have been reported previously (Campbell et al., 1984). This was the source of probes for all experiments described below. The probe used in the S1 nuclease 5′ protection experiments was 5′-end-labeled at the HinfI site at position +110 (relative to the start site of transcription of medRNA in vivo), recut at the XmnI site at position +18, and gel-isolated. The probe used in the S1 nuclease 3′ protection experiments was 3′-end-labeled by filling-in at the HinfI site at position +110 according to Maniatis et al. (1982) and then recutting at the PstI site in the vector. The primer extension probe was 5′-end-labeled at the HinfI site at position +110, recut at the PvuII site at +58, and gel-isolated. For the experiment shown in Figure 2, the 500 bp fragment of pMES.1, extending from the PvuII site (position +58) through the end of the insert at position +183 to the EcoRI site in the vector, was cloned into SmaI- and EcoRI-digested pGEM-1 (Promega Biotec) such that antisense RNA was made under the direction of the phage T7 RNA polymerase promoter used under the conditions recommended by Promega Biotec. For the experiment shown in Figure 6, the 400 bp HinfI fragment of pMES.1, extending from −275 to +110, was cloned into the SmaI site of pGEM-2 (Promega Biotec) in the antisense orientation for the T7 promoter.

Preparation of RNA

T. brucei brucei (strain 427, MITat 1.2) was grown in Sprague-Dawley rats as described previously (Cross, 1975). Where necessary, trypanosomes were purified on a DEAE Sephadex column (Lanham and Godfrey, 1970). Total RNA from whole infected rat blood was prepared by the hot-phenol method (Boothroyd and Cross, 1982). RNA for use in the S1 nuclease 3′ protection experiments was treated with RNAase-free DNAase (Cooper Biomedical) at 20 μg/ml for 10 min at 37°C. Poly(A)$^+$ RNA was prepared by batch elution at 60°C off oligo(dT)–cellulose (Nakazato and Edmonds, 1974).

Size Fractionation of RNA

About 100 μg of total RNA was size-fractionated by gel electrophoresis under denaturing conditions as described elsewhere (Maniatis et al., 1982). In brief, RNA samples were dissolved in 80% formamide, 0.1% bromophenol blue, 0.1% xylene cyanol FF, and 0.11× tris–borate–EDTA buffer, and were heated for 3 min at 90°C. After being quickly cooled on ice, samples were loaded on a 5% polyacrylamide gel (acrylamide:bisacrylamide ratio of 29:1) containing 7 M urea and were electrophoresed at 250V. Radiolabeled DNA markers in the appropriate size range were run in parallel. The entire strip of gel containing the electrophoresed RNA was cut into 13 equal slices plus one (fraction 14) containing about half the amount of gel as the others. The RNA contents of each slice were purified by electroelution out of the gel into a dialysis bag and were concentrated by precipitation with ethanol in the presence of 100 μg of carrier yeast tRNA.

Sucrose gradient fractionation of RNA was identical to that previously described (Campbell et al., 1984).

RNA Blot Analysis

RNA was resolved by gel electrophoresis as described for size fractionation above. Following electrophoresis the gel was washed in water, and the RNA was electrophoretically transferred to Genetrans paper (Campbell et al., 1984). After a 30 min prehybridization in 30% formamide, 50 mM Pipes (pH 6.4), 0.1% SDS, 0.3 M NaCl, and 1× Denhardt's solution, the filter was hybridized in the same buffer overnight with the antisense RNA probe described above at 10^7 cpm/ml (spec. act.: $\sim 10^9$ cpm/μg RNA). Washes were at 55°C for 2 hr in 0.1× SSC and 0.1% SDS with two changes of buffer, or at 65°C for 4 hr in 0.1× SSC and 0.1% SDS. Size markers were end-labeled fragments of pAT153 cut with MspI.

S1 Nuclease and Primer Extension Analyses

RNA and probes for S1 nuclease mapping (described above) were coprecipitated, redissolved in 80% formamide, 400 mM NaCl, 40 mM Pipes (pH 6.7), and 1 mM EDTA, denatured by heating to 75°C for 15 min, and hybridized overnight at 37°C. Duplexes were treated with 300 units/ml of S1 nuclease for 30 min at 37°C as described previously (Berk and Sharp, 1977). For the primer extension, RNA and the primer were annealed as described above, precipitated with ethanol, and dried. Primers were extended for 1 hr at 37°C using reverse transcriptase (300 units/ml) in 50 mM Tris (pH 8.3), 10 mM dithiothreitol, 60 mM KCl, 6 mM $MgCl_2$, and 2 mM each of the deoxynucleotide triphosphates. For both the S1 nuclease treatment and primer extension, recovered labeled DNA was analyzed by gel electrophoresis on 8% polyacrylamide–7 M urea gels (Sanger and Coulson, 1978). Size markers were as described above or were sequence ladders (Maxam and Gilbert, 1980) of the homologous probe.

Acknowledgments

We thank Drs. Ed Mocarski and Larry Burg for critical reading of the manuscript, Dr. David Campbell for many useful comments and suggestions, Phillip Zamore and Dr. Michael Green of Harvard University for the generous gift of lariat RNA and HeLa cell S100 extract, and Ms. Lorrain Sargent for secretarial help. This work was supported in part by grants from the National Institutes of Health (AI21023) and the MacArthur Foundation. R. E. S. was supported in part by the Medical Scientist Training Program of the NIH (GM07365) and by a training grant from the Johnson and Johnson Co. J. C. B. holds a Special Fellowship from the Burroughs-Wellcome Fund.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 3, 1986; revised September 25, 1986.

References

Berk, A. J., and Sharp, P. A. (1977). Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell *12*, 721–732.

Bernards, A., De Lange, T., Michels, P. A. M., Liu, A. Y. C., Huisman, M. J., and Borst, P. (1984). Two modes of activation of a single surface antigen gene of Trypanosoma brucei. Cell *36*, 163–170.

Boothroyd, J. C. (1985). Antigenic variation in African trypanosomes. Ann. Rev. Microbiol. *39*, 475–502.

Boothroyd, J. C., and Cross, G. A. M. (1982). Transcripts coding for different variant surface glycoproteins in Trypanosoma brucei have a short identical exon at their 5′-end. Gene *20*, 279–287.

Boothroyd, J. C., Campbell, D. A., and Sutton, R. E. (1985). Expression of surface antigen genes in Trypanosoma brucei involves a novel system of discontinuous transcription. In Modern Approaches to Vaccines, R. A. Lerner, R. M. Channock, and F. Brown, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 61–66.

Borst, P. (1986). Discontinuous transcription and antigenic variation in trypanosomes. Ann. Rev. Biochem. *55*, 701–732.

Campbell, D. A., Thornton, D. A., and Boothroyd, J. C. (1984). Apparent discontinuous transcription of Trypanosoma brucei surface antigen genes. Nature *311*, 350–355.

Cech, T. R., and Bass, B. L. (1986). Biological catalysis by RNA. Ann. Rev. Biochem. *55*, 599–629.

Clayton, C. E. (1985). Structure and regulated expression of genes encoding fructose biphosphate aldolase in Trypanosoma brucei. EMBO J. *4*, 2997–3003.

Cross, G. A. M. (1975). Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of Trypanosoma brucei. Parasitology *71*, 393–417.

De Lange, T., Liu, A. Y. C., Van der Ploeg, L. H. T., Borst, P., Tromp, M. C., and Van Boom, J. H. (1983). Tandem repetition of the 5′ miniexon of variant surface glycoprotein genes: a multiple promoter for VSG gene transcription? Cell *34*, 891–900.

De Lange, T., Michels, P. A. M., Veerman, H. J. G., Cornelissen, A. W. C. A., and Borst, P. (1984a). Many trypanosome messenger RNAs share a common 5′ terminal sequence. Nucl. Acids Res. *12*, 3777–3789.

De Lange, T., Berkvens, T. M., Veerman, H. J. G., Carlos, A., Frasch, C., Barry, J. D., and Borst, P. (1984b). Comparison of the genes coding for the common 5′ terminal sequence of messenger RNAs in three trypanosome species. Nucl. Acids Res. *12*, 4431–4443.

Dorfman, D., and Donelson, J. (1984). Characterization of the 1.35 kb DNA repeat unit containing the conserved 35 nucleotides at the 5′-termini of VSG mRNAs in Trypanosoma brucei. Nucl. Acids Res. *12*, 4907–4920.

Gilbert, W., Marchionni, M., and McKnight, G. (1986). On the antiquity of introns. Cell *46*, 151–154.

Gonzalez, A., Lerner, T. J., Huccas, M., Sosa-Pineda, B., Nogueira, N., and Lizardi, P. M. (1985). Apparent generation of a segmented mRNA from two separate tandem gene families in Trypanosoma cruzi. Nucl. Acids Res. *13*, 5789–5804.

Guyaux, M., Cornelissen, A. W. C. A., Pays, E., Steinert, M., and Borst, P. (1985). Trypanosoma brucei: a surface antigen mRNA is discontinuously transcribed from two distinct chromosomes. EMBO J. *4*, 995–998.

Kimmel, B. E., Samson, S., Hirschberg, R., and Yarbrough, L. R. (1985). Tubulin genes of the African trypanosome Trypanosoma brucei rhodesiense: nucleotide sequence of a 3.7-kb fragment containing genes for alpha and beta tubulins. Gene *35*, 237–248.

Konarska, M. M., Padgett, R. A., and Sharp, P. A. (1985). *Trans* splicing of mRNA precursors in vitro. Cell *42*, 165–171.

Kooter, J., De Lange, T., and Borst, P. (1984). Discontinuous synthesis of mRNA in trypanosomes. EMBO J. *3*, 2387–2392.

Lanham, S. M., and Godfrey, D. G. (1970). Isolation of salivarian trypanosomes from man and other mammals using DEAE-cellulose. Exp. Parasitol. *28*, 521–534.

Liu, A. Y. C., Van der Ploeg, L. H. T., Rijsewijk, F. A. M., and Borst, P. (1983). The transposition unit of VSG gene 118 of Trypanosoma brucei: presence of repeated elements at its border and absence of promoter-associated sequences. J. Mol. Biol. *167*, 57–75.

Makino, S., Stohlman, S. A., and Lai, M. M. C. (1986). Leader sequences of murine coronavirus mRNAs can be freely reassorted: evidence for the role of free leader RNA in transcription. Proc. Natl. Acad. Sci. USA *83*, 4204–4208.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Maxam, A. M., and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. *65*, 499–560.

Michels, P. A. M., Poliszczak, A., Osinga, K. A., Misset, O., van Beeumen, J., Werenga, R. K., Borst, P., and Opperdoes, F. R. (1986). Two tandemly linked identical genes code for the glycosomal glyceraldehyde-phosphate dehydrogenase in Trypanosoma brucei. EMBO J. *5*, 1049–1056.

Michiels, F., Matthyssens, G., Kronenberger, P., Pays, E., Dero, B., Van Assel, S., Darville, M., Cravador, A., Steinert, M., and Hamers, R. (1983). Gene activation and re-expression of a Trypanosoma brucei variant surface glycoprotein. EMBO J. *2*, 1185–1192.

Milhausen, M., Nelson, R. G., Sather, S., Selkirk, M., and Agabian, N. (1984). Identification of a small RNA containing the trypanosome spliced leader: a donor of the shared 5′ sequences of trypanosomatid mRNAs? Cell *38*, 721–729.

Nakazato, H., and Edmonds, M. (1974). Purification of messenger RNA and heterogeneous nuclear RNA containing poly(A) sequences. Methods Enzymol. *29*, 431–443.

Nelson, R. G., Parsons, M., Barr, P. J., Stuart, K., Selkirk, M., and Agabian, N. (1983). Sequences homologous to the variant antigen mRNA spliced leader are located in tandem repeats and variable orphons in Trypanosoma brucei. Cell *34*, 901–909.

Nelson, R. G., Parsons, M., Selkirk, M., Newport, G., Barr, P. J., and Agabian, N. (1984). Sequences homologous to the variant antigen mRNA spliced leader are present in Trypanosomatidae which do not undergo antigenic variation. Nature *308*, 665–667.

Padgett, R. A., Konarska, M. M., Grabowski, P. J., Hardy, S. F., and Sharp, P. A. (1984). Lariat RNAs as intermediates and products in the splicing of messenger RNA precursors. Science *225*, 898–903.

Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., and Sharp, P. A. (1986). Splicing of messenger RNA precursors. Ann. Rev. Biochem. *55*, 1119–1150.

Parsons, M., Nelson, R. G., Watkins, K. P., and Agabian, N. (1984). Trypanosome mRNAs share a common 5′ spliced leader sequence. Cell *38*, 309–316.

Pays, E., Lheureux, M., and Steinert, M. (1982). Structure and expression of a Trypanosoma brucei gambiense variant specific antigen gene. Nucl. Acids Res. *10*, 3149–3163.

Ruskin, B., and Green, M. R. (1985). An RNA processing activity that debranches RNA lariats. Science *229*, 135–140.

Ruskin, B., Krainer, A. R., Maniatis, T., and Green, M. R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. Cell *38*, 317–331.

Sanger, F., and Coulson, A. R. (1978). The use of thin acrylamide gels for DNA sequencing. FEBS Lett. *87*, 107–110.

Sather, S., and Agabian, N. (1985). A 5′ spliced leader is added in trans to both alpha- and beta-tubulin transcripts in Trypanosoma brucei. Proc. Natl. Acad. Sci. USA *82*, 5695–5699.

Sharp, P. A. (1985). On the origin of RNA splicing and introns. Cell *42*, 397–400.

Sharp, P. A., and Konarska, M. M. (1986). Does *trans* splicing in vitro require base pairing between RNAs? Response to Solnick. Cell *44*, 211.

Sogin, M. L., Elwood, H. J., and Gunderson, J. H. (1986). Evolutionary diversity of eukaryotic small-subunit rRNA genes. Proc. Natl. Acad. Sci. USA *83*, 1383–1387.

Solnick, D. (1985). *Trans* splicing of mRNA precursors. Cell *42*, 157–164.

Solnick, D. (1986). Does *trans* splicing in vitro require base pairing between RNAs? Cell *44*, 211.

Swinkels, B. W., Gibson, W. C., Osinga, K. A., Kramer, R., Veeneman, G. H., van Boom, J. H., and Borst, P. (1986). Characterization of the gene for the microbody (glycosomal) triosephosphate isomerase of *Trypanosoma brucei*. EMBO J. *5*, 1291–1298.

Tschudi, C., Young, A. S., Ruben, L., Patton, C. L., and Richards, F. F. (1985). Calmodulin genes in trypanosomes are tandemly repeated and produce multiple mRNAs with a common 5′ leader sequence. Proc. Natl. Acad. Sci. USA *82*, 3998–4002.

Van der Ploeg, L. H. T., Liu, A. Y. C., Michels, P. A. M., De Lange, T., Borst, P., Majumder, K., Weber, H., and Veeneman, G. H. (1982). RNA splicing is required to make the messenger RNA for a variant surface antigen in trypanosomes. Nucl. Acids Res. *10*, 3591–3604.

Van der Ploeg, L. H. T., Cornelissen, A. W. C. A., Michels, P. A. M., and Borst, P. (1984). Chromosome rearrangement in Trypanosoma brucei. Cell *39*, 213–221.