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. 38, 721-729, October 1984, Copyright © 1984 by MIT 0092-8674/84/100721-09 $02.00/0

Identification of a Small RNA Containing the Trypanosome Spliced Leader: A Donor of Shared 5′ Sequences of Trypanosomatid mRNAs?

Michael Milhausen, Richard G. Nelson, Susan Sather, Murray Selkirk, and Nina Agabian
Department of Biochemistry SJ-70
University of Washington
Seattle, Washington 98195

Summary

The 35 nucleotide spliced leader (SL) sequence is found on the 5′end of numerous trypanosome mRNAs, yet the tandemly organized reiteration units encoding this leader are not detectably linked to any of these structural genes. Here we report the presence of a class of discrete small SL RNA molecules that are derived from the genomic SL reiteration units of Trypanosoma brucei, Trypanosoma cruzi, and Leptomonas collosoma. These small SL RNAs are 135, 105, and 95 nucleotides, respectively, and contain a 5′-terminal SL or SL-like sequence. S1 nuclease analyses demonstrate that these small SL RNAs are transcribed from continuous sequence within the respective SL reiteration units. With the exception of the SL sequence and a concensus donor splice site immediately following it, these small RNAs are not well conserved. We suggest that the small SL RNAs may function as a donor of the SL sequence in an intermolecular process that places the SL at the 5′ terminus of many trypanosomatid mRNAs.

Introduction

Many mRNAs of Trypanosoma brucei contain the same 5′-terminal 35 nucleotide leader sequence (Parsons et al., 1984a; DeLange et al., 1984a). This untranslated leader was first identified (Boothroyd and Cross, 1982; Van der Ploeg et al., 1982) as a sequence shared among mRNAs encoding the antigenically distinct variant surface glycoproteins (VSGs), which are sequentially expressed during antigenic variation in the mammalian host (for review see Vickerman, 1978; Parsons et al., 1984b). The leader sequence is not encoded by the DNA immediately flanking VSG structural genes and has therefore been termed the spliced leader (SL), to indicate that it is transcribed from another region of the genome (Nelson et al., 1983).

In the genome of Trypanosoma brucei, sequences encoding the SL are reiterated approximately 200 times and, for the most part, are organized in a large tandem array(s) of directly repeated 1.4 kb units, which each containing a single SL exon (Nelson et al., 1983; DeLange et al., 1983). A few of these units, however, are orphons (Childs et al., 1981)—that is, they are dispersed from the tandem array (Nelson et al., 1983; Parsons et al., 1984c). Given the occurrence of the SL on all VSG transcripts and the observation that sequences flanking the 5′ side of the SL exon in the 1.4 kb repeat unit resemble eucaryotic polymerase II promoters, it has been proposed that the tandem array defines the 5′ boundary of a unique VSG gene expression locus and serves as a multiple promoter, allowing high levels of VSG gene transcription (DeLange et al., 1983). The presence of the SL on VSG transcripts, however, is but one example of its larger and more fundamental role in trypanosomatid gene expression. First, sequences homologous to the T. brucei SL are highly reiterated not only in the genomes of other species of Trypanosoma, for example the intracellular New World parasite T. cruzi, but also in other genera of the family Trypanosomatidae, notably the evolutionarily primitive insect parasite, Leptomonas collosoma (Nelson et al., 1984; DeLange et al., 1984b). Neither T. cruzi nor L. collosoma undergo antigenic variation, yet Northern analysis of RNA from these organisms reveals that their SL-like sequences are transcribed and are present on RNAs of diverse size (R. G. N. and M. S., unpublished data). In addition, the SL is heavily transcribed in T. brucei procyclic culture form cells (analogous to the insect midgut stage of the trypanosome life cycle) where VSG transcripts and VSGs are undetectable (DeLange et al., 1983; Parsons et al., 1984c). Second, we have isolated and characterized many different T. brucei non-VSG cDNA clones that contain the SL (Parsons et al., 1984a). Individual clones identify discrete SL-containing transcripts; some of these are abundant or rare, constitutive or regulated, and are transcribed from single or multicopy genes. As is the case with VSG genes, the SL sequence is not encoded contiguously with, or detectably linked to, these structural genes. Based on the frequency of SL cDNA clones in bloodstream and procyclic culture form libraries, it seems likely that many, if not all, trypanosome mRNAs share this 35 nucleotide leader sequence. Thus the SL sequence is unique neither to African trypanosomes nor to VSG gene expression.

The hypothesis that explains the presence of the SL on VSG mRNA by intramolecular splicing of a very large primary transcript seems untenable, simply because more SL-utilizing genes have already been identified than there are possible SL-encoding tandem arrays and orphons. We favor the hypothesis that the SL is added to these numerous transcripts by an intermolecular process, i.e., in trans. The two simplest mechanisms for such a process are primed transcription and post-transcriptional splicing/ligation. In the first case, an SL transcript initiated in the SL reiteration unit would serve as a primer for the initiation of transcription from SL-utilizing genes. The second mechanism suggests that a similar SL transcript might serve as a substrate for intermolecular ligation to appropriate structural gene transcripts. Both models predict that a discrete SL-containing RNA is transcribed not only from the SL reiteration units of T. brucei, but from those of other trypanosomatid species as well.

In this report we demonstrate the presence of small SL RNAs that are transcribed from the SL reiteration units of T. brucei, T. cruzi, and L. collosoma. These RNAs contain an SL sequence at their 5′ ends and extend 3′ an additional 100, 70, and 60 nucleotides, respectively. With the exception of the SL itself, the sequences of these

small RNAs are not well conserved. However, a consensus mRNA donor splice site is present in each small RNA immediately 3′ to the SL sequence. Similarly, a consensus mRNA acceptor splice site is present in SL-utilizing genes at a position corresponding to the location of the 3′ end of the SL in the mature SL mRNA. These and other observations suggest that the small SL RNAs serve as donors of the SL sequence found on trypanosomatid mRNAs.

The SL Reiteration Unit Encodes a Small SL RNA

Total RNA, extracted from T. brucei, L. collosoma, and Crithidia fasciculata and $LiCl_2$ fractionated RNA from T. cruzi, were electrophoresed on 10% polyacrylamide-7 M urea gels. Ethidium bromide staining (Figure 1A) revealed a distinct pattern of small RNA species ranging in size from 100 to 250 nucleotides. Each of these trypanosomatid organisms has a similar profile of abundant small RNAs although individual RNA sizes vary. These small RNA profiles are comparable to those of C. fasciculata (Gray, 1981), Leishmania tarantolae (Simpson and Simpson, 1978), and T. brucei (Cordingly and Turner, 1980), described previously. Two of these RNAs are analogous to the 5S (MacKay et al., 1980) and 5.8S (Schnare and Gray, 1982) ribosomal RNAs found in other eucaryotes.

The RNAs shown in Figure 1A were electroblotted to a Zeta-probe membrane and hybridized with a ^{32}P-labeled synthetic oligodeoxynucleotide complementary to 22 of the 35 nucleotide SL (22-mer/SL probe, see Figure 5). As shown in Figure 1B, the probe hybridizes strongly with RNA species of approximately 133, 108, and 95 nucleotides in RNA from T. brucei, T. cruzi, and L. collosoma, respectively. In T. brucei, higher molecular weight small RNAs often hybridize with the SL probe (lanes 2 and 3, Figure 1B); these have not been further characterized. The RNA from T. brucei bloodstream forms (Figure 1B, lane 2) and from procyclic culture forms (lane 3) shows identical patterns of hybridization. C. fasciculata genomic DNA does not contain sequences with significant homology to the 22-mer (Nelson et al., 1984; however, see DeLange et al., 1984b) and, as expected, RNA from this trypanosomatid does not hybridize with the probe (lane 5).

The hybridization observed in the high molecular weight region of the Northern blot (Figure 1B, lanes 2 and 3) is due to T. brucei mRNA transcripts that contain the SL. $LiCl_2$ precipitation of total RNA separates small RNA species, including those containing the SL, from the rest of the larger RNA (data not shown). Therefore, there is no hybridization of the 22-mer to mRNA in T. cruzi samples prepared by $LiCl_2$ fractionation (Figure 1B, lane 6). In this experiment hybridization of the 22-mer prode to L. collosoma mRNA was not detected; the reasons for this are unclear, however the 22-mer does hybridize to larger molecular weight RNA under less stringent hybridization conditions.

An SL-like sequence is shared by most trypanosomatids and in each organism the SL is a small part of a larger reiteration unit (Nelson et al., 1984). In T. cruzi and L. collosoma, the unit sizes are approximately 0.6 kb as compared to 1.4 kb in T. brucei. The SL reiteration units are directly and tandemly repeated in the genomes of each of these trypanosomatids and homology among units is limited to the SL sequence itself. The remainder of the 600–1400 bp is essentially nonhomologous and species-specific (Nelson et al., 1984; unpublished data; DeLange et al., 1984).

In order to determine whether these small SL RNAs are transcribed from the reiteration unit, the RNA blot shown in Figure 1B was hybridized with probes prepared from the cloned reiteration units of each organism under con-

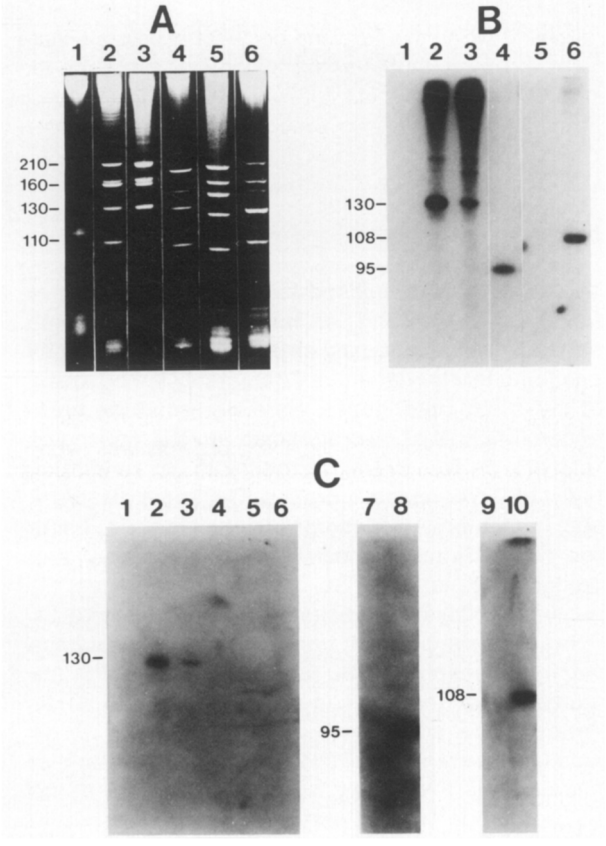

Figure 1. Analysis of Trypanosomatid Small RNAs by Gel Electrophoresis and Hybridization

Total RNA samples were isolated, electrophoresed, visualized, and electroblotted as described in Experimental Procedures. (A) Ethidium bromide stain of RNA (10 μg/slot); lane 1, E. coli; lane 2, T. brucei bloodstream form; lane 3, T. brucei procyclic culture form; lane 4, L. collosoma; lane 5, C. fasciculata; lane 6, T. cruzi. The markers indicate the sizes in nucleotides of T. brucei small RNAs determined by calibration with 3′-end-labeled and denatured Hpa 2 restriction fragments from pBR 322. (B) A blot of the gel shown in A was hybridized to a ^{32}P 5′-end-labeled synthetic 22 nucleotide oligomer (see Experimental Procedures) complementary to the SL. (C) The blot of (B) and blots of adjacent lanes of the same gel (not shown in A) were hybridized with ^{32}P-labeled probes prepared by Klenow synthesis of M13 (sense strand) SL reiteration unit clones from T. brucei, mpTb-1.4M(1); T. cruzi, mpTc-0.6S(1); and L. collosoma, mpLc-0.6T(1) as described in Experimental Procedures. Lanes 1–6 as in (A) and (B), hybridized with T. brucei probe. Lanes 7 and 8 T. brucei and L. collosoma RNA, respectively, hybridized with L. collosoma probe. Lanes 9 and 10, T. brucei and T. cruzi, respectively, hybridized to T. cruzi probe.

ditions in which the SL alone does not hybridize (Figure 1C). The T. brucei SL reiteration unit hybridized with the 133 bp SL RNA but did not hybridize with the small SL RNAs from heterologous species. Similarly, the T. cruzi and L. collosoma reiteration unit probes hybridized only with the small SL RNA from the homologous organism (Figure 1C). These experiments demonstrate that the SL and additional sequences specific to each respective reiteration unit are transcribed to yield a discrete class of small SL RNA molecules in each of these organisms.

The SL Is Located at the 5′ Termini of Small SL RNAs

In order to determine the location of the SL within the small SL-RNA the synthetic 22 nucleotide oligomer was hybridized to size-selected RNA fractions enriched for the small SL RNA and primer-extended with reverse transcriptase. The 22-mer is complementary to nucleotides 10 to 31 of both the T. brucei and the T. cruzi 35 nucleotide SL sequences (see Figure 5). If the SL is located at the 5′ end of the small SL RNAs of these organisms, one would expect the primer to be extended by 9 nucleotides in both cases. Small SL RNA-containing fractions were prepared by elution from polyacrylamide-urea gels and annealed with unlabeled 22-mer. The primer was extended with reverse transcriptase in the presence of all four deoxyribonucleotide triphosphates, one of which was ^{32}P-labeled deoxycytidine triphosphate. Analysis of the reaction products is shown in Figure 2. The major extension product obtained from the enriched small SL RNA fractions of T. brucei and T. cruzi indicated that the SL is located at the 5′ end of the small SL RNAs of these organisms. The T. brucei small SL RNA extension reveals one major product of 31 nucleotides, the size expected for a 5′-terminal location of the SL. Other less abundant extension products were also detected; these either represent size heterogeneity within the small SL RNA fraction or artifacts produced in vitro. Primer extension of the T. cruzi small SL RNA fraction also revealed one major extension product, indicating that the SL is located precisely at the 5′ terminus of the small SL RNA (see Figure 2B). Minor extension products were again detected, suggesting the existence of small SL RNA species that start either 5 nucleotides within the SL sequence or contain additional nucleotides 5′ to the SL.

In the 37 nucleotide L. collosoma SL-like sequence a region of 17 contiguous nucleotides is complementary to the 22-mer, but both the 3′ and 5′ 22-mer termini are mismatched. Therefore a 23 nucleotide oligomer complementary to positions 15–37 of the L. collosoma SL (see Figure 5) was synthesized (P. Barr, Chiron Corporation) for use as a primer. The 23-mer was ^{32}P 5′-end-labeled, annealed with the L. collosoma small SL RNA fraction, and extended with reverse transcriptase in the presence of unlabeled deoxyribonucleotide triphosphates. Two major extension products of 37 and 41 nucleotides were observed (Figure 2C). The former product is consistent with a 5′-terminal location of the SL, while the latter suggests

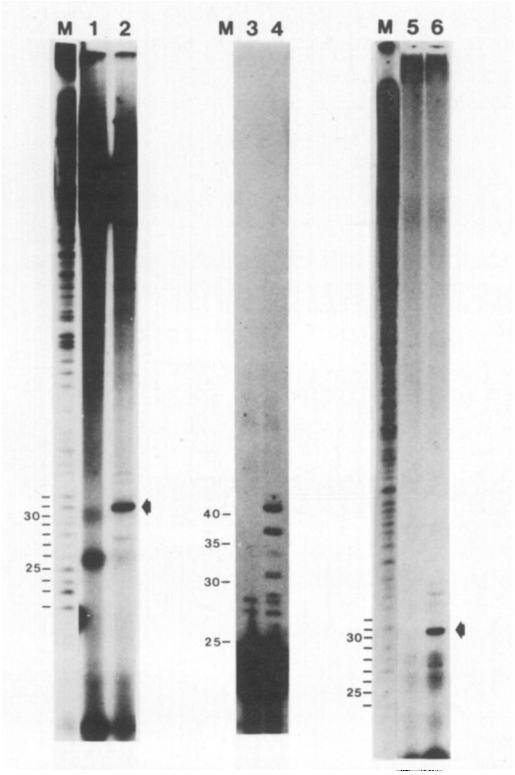

Figure 2. Primed Extensions Using the 22-mer and 23-mer Oligonucleotides and Size-Fractionated Small RNA

The small SL RNA fractions from T. brucei, L. collosoma, and T. cruzi were isolated by elution from separate preparative polyacrylamide gels. Lanes 1 and 2 are experiments with T. brucei and lanes 5 and 6 are with T. cruzi small SL RNA fractions, respectively. Each fraction (~10 ng) was incubated without (lanes 1 and 5) and with (lanes 2 and 6) unlabeled 22-mer (200 ng) and extended with AMV reverse transcriptase as described in Experimental Procedures. Lanes 3 and 4 represent primed extension of L. collosoma small RNA fraction using the 23 nucleotide oligomer specific for the L. collosoma SL sequence (see Experimental Procedures; Figure 5). In this case the 23-mer was ^{32}P-labeled at its 5′ end, hybridized, and extended in the absence (lane 3) and presence (lane 4) of L. collosoma small RNA. Following denaturation the samples were sized on 10% polyacrylamide-urea sequencing gels using combined M13 sequencing reactions, M, as ladders to determine the extension sizes. Arrows indicate the major extension product observed for each small SL RNA. Extension products seen in both experimental and control lanes apparently represent self-primed extensions from the sRNA fractions.

that a substantial fraction of the L. collosoma small SL RNA population contains four additional nucleotides 5′ to the SL.

The SL RNAs Are Transcribed from Continuous Sequences with the Reiteration Units

The sequences encoding the small SL RNAs were examined for the presence of introns by S1 nuclease protection experiments. The reiteration unit sense strand from each organism was cloned into the M13 bacteriophage vector mp8 (Nelson et al., 1984). The complementary strand was synthesized by Klenow extension from the annealed M13 sequencing primer in the presence of one ^{32}P-labeled nucleotide triphosphate. These probes were then hybridized with total RNA and treated with S1 nuclease to degrade single-stranded DNA. The labeled DNA fragments

protected by RNA from S1 nuclease digestion were sized by polyacrylamide-urea gel electrophoresis. If each small SL RNA is transcribed from continuous DNA sequence, a fragment equivalent in size to the small SL RNA (as determined by Northern analysis) should be protected. As shown in Figure 3, T. brucei total RNA protected labeled fragments ranging between 130 and 135 nucleotides from S1 degradation. In addition to this major class, several small labeled fragments were also protected; most of these were also present in the control lane (no RNA) and presumably result from annealing of probe to the unlabeled complementary M13 strand (Figure 4, lanes 1 and 3). However, the appearance of the two sets of fragments centered at 115 and 100 nucleotides was RNA-dependent. It is unclear whether the latter fragments are protected by small RNA species that exist in vivo or whether they are artifacts of the S1 analysis. However, in Northern blots, an RNA species of about 100 nucleotides is sometimes observed to hybridize with the T. brucei reiteration unit probe but not the 22-mer probe (data not shown). This RNA species may be derived from the 100 nucleotides 3′ to the SL sequence in the small SL RNAs.

Since the primer-extension analysis suggested that the 5′ termini of the vast majority of T. brucei small SL RNAs was precisely at the beginning of the SL, we hypothesized that the variability in length seen in the S1 protection experiments was localized to the 3′ end. This was confirmed by further S1 nuclease analysis using a fragment of the reiteration unit that extends from the Rsa I site within the SL (between nucleotides 26 and 27) through the Mbo I site at position 182 (see Figure 5) into the flanking vector sequence. The fragment was labeled at its 3′ end, hybridized with total RNA, and treated with S1 nuclease. Fragments protected from degradation ranged from 103 to 109 nucleotides in length (Figure 3, lanes 3 and 4), which is consistent with the predicted fragment sizes based on the position of the Rsa I site and the distribution of sizes observed for small SL RNA. Size heterogeneity thus occurs at the 3′ end of the small SL RNA transcript.

S1 analyses using uniformly labeled reiteration unit probes from T. cruzi and L. collosoma indicated that RNA preparations from these organisms protected fragments whose sizes were consistent with those of the small SL

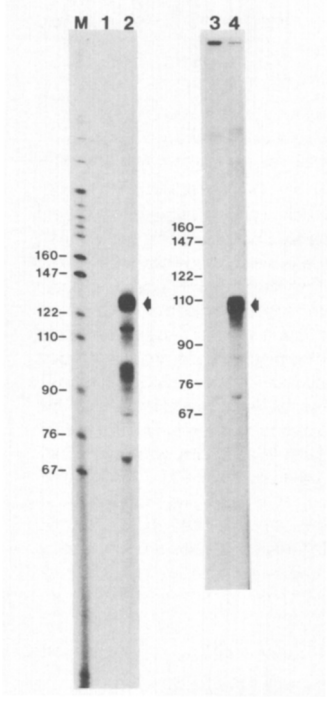

Figure 3. S1 Nuclease Analysis of T. brucei Small SL RNA

A ^{32}P-labeled probe complementary to M13 sense strand SL reiteration unit clone mpTb-1.4M(1) was synthesized as described in Experimental Procedures. This probe was hybridized in the absence (lane 1) and presence (lane 2) of 20 μg of T. brucei total RNA, S1 nuclease-treated, denatured, and electrophoresed on an 8% polyacrylamide-urea gel as described in Experimental Procedures. Markers, lane M, were 3′-end-labeled DNA fragments of an Hpa 2 digestion of pBR 322, which were denatured prior to electrophoresis. Lanes 3 and 4 show an S1 nuclease analysis of T. brucei total RNA with a ^{32}P 3′-end-labeled Rsa I fragment from the reiteration unit. The labeled Rsa site is within the SL sequence and the fragment extends beyond the poly(T) stretch observed in the T. brucei sequence (see Figure 5). The labeled fragment was hybridized without (lane 3) and with (lane 4) T. brucei total RNA. Hybridization and S1 nuclease treatment were as described in Experimental Procedures. The arrow indicates the major protected fragments.

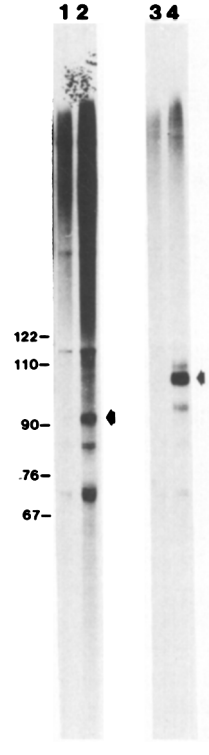

Figure 4. S1 Nuclease Analysis of T. cruzi and L. collosoma Small SL RNAs

^{32}P-labeled probes complementary to the M13 sense strand reiteration unit clones of T. cruzi and L. collosoma were synthesized as described in Experimental Procedures. These probes were hybridized, S1 nuclease-treated, and analyzed by electrophoresis as in Figure 3 except hybridization was at 37°C. Lane 1, L. collosoma probe; lane 2, L. collosoma probe plus 20 μg of L. collosoma RNA; lane 3, T. cruzi probe; lane 4, T. cruzi probe plus 20 μg of T. cruzi RNA. Markers are as in Figure 3 and arrows indicate the major protected fragments.

RNAs determined in Figure 1. The major protected fragments were somewhat heterogeneous in size, ranging from 90 to 94 nucleotides for L. collosoma, and 101 to 105 nucleotides for T. cruzi (Figure 4). For the most part the heterogeneity is localized to the 3′ termini of these small SL RNAs as determined by S1 analysis using end-labeled probes as described above (data not shown).

The combined data demonstrate that the predominant small SL RNA from each organism starts at or near the 5′ terminus of an SL-like sequence, extends 3′ without interruption, and ends in a region spanning several nucleotides.

Sequence of the Small SL RNAs

The DNA sequences surrounding the SL in the reiteration units of T. brucei, L. collosoma, and T. cruzi are shown in Figure 5; the complete sequence of these units will be reported in detail elsewhere. In general, the sequences that encode the small SL RNAs are not well conserved. However, the sequences that are conserved are centered around the SL sequence itself and the region just downstream from the 3′ end of the small SL RNAs. Within each of the SL sequences, there is a region of 14 contiguous nucleotides, the position and sequence of which are maintained in each of these organisms. The remaining nucleotides of the SL sequences are also conserved to various degrees: the SL sequence from T. cruzi differs from that of T. brucei by only three base residues, while the L. collosoma SL has 29 nucleotides in common with the T. brucei SL. Hybridization experiments performed with the 22 nucleotide T. brucei SL probe at reduced stringency indicate that most other genera of the Trypanosomatidae also contain repeated genomic sequences with homology to the SL (Nelson et al., 1984; and unpublished data). It is clear from these experiments however, that in each case sequence conservation in the SL region spanned by the 22-mer is less than that observed between the T. brucei and L. collosoma sequences.

Small regions of sequence flanking both sides of the SL, and both sides of the small SL RNA sequence, are also shared among these three trypanosomatids. The nine nucleotides 5′ to the SL are identical in T. brucei and T. cruzi while L. collosoma shares seven of these nucelotides. On its 3′ side, the SL sequence is immediately followed by a GT nucleotide pair in each of these organisms. This

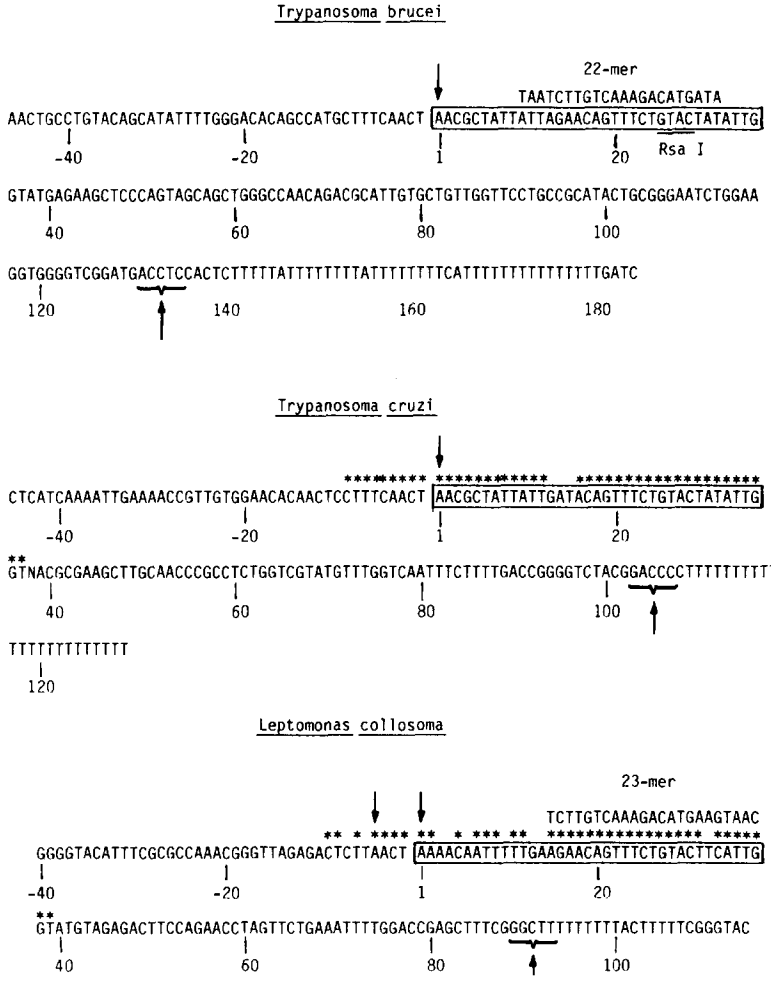

Figure 5. DNA Sequence of the Small SL RNA Regions in the Reiteration Units from T. brucei, T. cruzi, and L. collosoma

Each reiteration unit was cloned into M13 and sequenced as previously described. Nucleotide number 1 indicates the first base in the spliced leader sequence (enclosed in a box). The synthetic 22-mer and 23-mer oligonucleotides are aligned with their respective complementary sequence in the T. brucei and L. collosoma SLs. Arrows indicate the 5′ and 3′ termini of each small SL RNA as determined in Figures 2, 3, and 4 and as discussed in the text. The asterisks above nucleotide residues in the T. cruzi and L. collosoma sequences indicate identity with the T. brucei sequence (see text).

dinucleotide is identical with the invariant dinucleotide present in the concensus sequence of intron 5′ splice sites (Lerner et al., 1980; Rogers and Wall, 1980) and, as discussed below, appears to be a functional donor splice junction. Putative transcription initiation and termination regions for each small SL RNA are also shown in Figure 5. Approximately 30 bp 5′ to the SL in T. brucei, an AT-rich region has been noted as a possible RNA polymerase II concensus promoter (DeLange et al., 1983; Campbell et al., 1984a, 1984b; Dorfman and Donelson, 1984). This concensus is not maintained in either of the other reiteration units, although an AT-rich region at −30 to −40 bp is found in each sequence. Immediately following the 3′ end of the small SL RNAs in all three cases is a sequence very rich in thymidine residues. The role of this punctuated poly(T) tract is unknown but its conserved features and location suggest that it is a functional sequence element.

Discussion

The SL was first identified as a 5′-terminal sequence shared by VSG transcripts (Boothroyd and Cross, 1982; Van der Ploeg et al., 1982) but it is now clear that many, if not all, T. brucei mRNAs contain this 35 nucleotide leader (Parsons et al., 1984a; DeLange et al., 1984a). Several structural genes corresponding to SL mRNAs have been examined and in each of the cases studied, no linkage relationship has been demonstrated between the SL and the structural genes (Nelson et al., 1983; DeLange et al., 1983; Parsons et al., 1984a; Myler et al., 1984). Although there are approximately 200 SL exons in the genome (DeLange et al., 1983), their organization in one or at most a few large tandem array(s) and three to four orphons (Nelson et al., 1983) precludes the possibility that each of the many SL-utilizing genes is located in cis to an SL exon. We have therefore hypothesized that SL and structural gene transcripts are joined by a mechanism that operates in trans (Nelson et al., 1984; Parsons et al., 1984a).

How Is the SL Added to mRNAs?

Eight T. brucei genes encoding mRNAs that employ the SL have been sequenced, five VSG (Boothroyd and Cross, 1982; Murphy et al., 1984; Michiels et al., 1983; Bernards et al., 1984, Van der Ploeg et al., 1982) and three non-VSG genes (S. Sather and K. Watkins, unpublished data). Each of these genes contains a consensus mRNA 3′ intron splice site precisely at the point of divergence between the structural gene and mRNA sequence, i.e., at the position where the SL is found in the mature transcript. Given the conservation (and apparent utilization) of an acceptor splice site in all eight genes, it appears that conventional mRNA splicing is involved in the process by which the SL is added to target mRNAs, either in the intermolecular reaction itself or during subsequent intramolecular RNA maturation.

In this report we have demonstrated the presence of small RNAs that contain the SL, or SL-like sequences, at or near their 5′ ends in each of three trypanosomatid organisms. With the exception of conserved features centered around the SL sequence itself, these small SL RNAs are quite divergent and appear to be the only transcripts derived from their respective SL reiteration units. The conservation of a 5′ mRNA intron splice site immediately flanking the 3′ end of the SL in each of these three small SL RNAs suggests that these molecules serve as SL donors in an intermolecular RNA joining reaction.

We propose two simple models for the addition of the SL to target mRNAs in *trans* (Figures 6A and 6B). First, the small SL RNA (Figure 6A, left) or a cleavage product containing the intact SL and 5′ splice site (Figure 6A, right) might be used to prime transcription of SL-utilizing genes. Alternatively, the small SL RNA and target mRNAs may be independently transcribed from their own promoters and subsequently joined by either end-to-end ligation (Figure 6B, right) or an intermolecular version of conventional RNA splicing (Figure 6B, left). In several of these alternatives the initial intermolecular joining event produces a precursor SL mRNA containing an intervening sequence. Provided both the SL donor and target mRNA acceptor splice sites are retained in the precursor, it could be processed normally.

In the priming model, sequence complementarity between the SL, or remainder of the small SL RNA, and a region 5′ to target genes could direct the primer to a site from which transcription of downstream mRNA sequences is initiated by extension from the 3′ end of the SL primer. However, any such sequence homology must be quite limited as neither SL nor reiteration unit probes hybridize to genomic clones of SL target genes, even though these clones contain large amounts of 5′-flanking sequences (Parsons et al., 1984a). A small region of homology to the SL is shared among the 5′-flanking regions of the three (non-VSG) target genes recently sequenced in our laboratory (S. Sather and K. Watkins, unpublished data). Blocks of 5, 6, and 10 contiguous nucleotides, complementary to all or part of the SL sequence pACAGTTTCTG, occur on the template strands just 5′ to the acceptor splice sites of these three genes. It is not clear if this observation is pertinent to priming since the sites of transcription initiation have not been determined. Clearly if these or other SL target genes are initially transcribed as precursors containing excess 5′ sequence, and preliminary evidence suggests this may be the case for one of the above mentioned genes (Parsons et al., 1984a) as well as for one VSG gene (Van der Ploeg et al., 1982), the putative primer recognition sequence should be located upstream from the 5′ end of the primary transcript, not adjacent to the structural gene as in the above cases.

Sequence complementarity is not a mandatory requirement of the priming model since a protein, e.g., RNA polymerase, associated with either the SL primer or the target gene could also serve to align the two sequences for transcription initiation. Both corollaries of the priming model raise the same question: if transcription of all SL mRNAs requires priming, how is the small SL RNA itself transcribed? The simplest answer is that these genes are

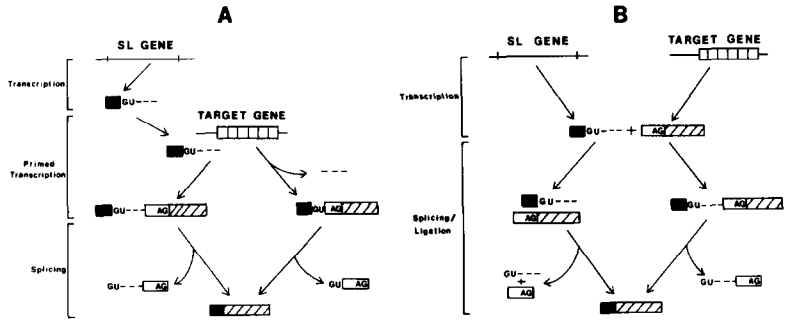

Figure 6. Models of SL Addition to Target mRNAs in Trans

(A) primed transcription and (B) intermolecular splicing/ligation. The symbols used are as follows: solid line (—), small SL RNA gene (i.e., the reiteration unit); solid box (■), the 35 nucleotide SL; GU, the 5′ splice site immediately following the SL in the small SL RNA; dashed line (---), the remainder of the small SL RNA beginning immediately 3′ to the GU pair; vertical hatched box (▮), SL target gene; open box (□), putative 5′ precursor RNA transcribed from the target gene; AG 3′ splice site present in target gene nucleotides and hypothesized to be present on a target gene mRNA precursor prior to maturation; angled hatched box (▨), mature SL mRNA. Each model presents two pathways of SL RNA maturation from several alternatives. The pathways illustrated favor utilization of the "apparent" intron/exon splice junctions present on these RNAs. In all cases intermediates and by-products are shown primarily to aid in visualization of each step; these intermediates may or may not be detectable.

transcribed by different polymerases, one of which requires the product of the other, i.e., the small SL RNA, as a cofactor. Alternatively, a single polymerase might first transcribe the small SL RNA and, with this primer bound, alter its promoter recognition properties to transcribe SL mRNAs. Campbell et al. (1984a, 1984b) have suggested a related discontinuous transcription model for VSG gene expression.

The second model for the addition of the SL to target mRNAs involves an intermolecular interaction between a small RNA and a separately transcribed mRNA precursor. Phrased in this manner, this statement might also describe models proposed for the function of the small U1 (Lerner et al., 1980; Rogers and Wall, 1980) and U4 (Berget, 1984) RNAs in mRNA splicing and polyadenylation, respectively. However, unlike U RNAs the small SL RNA would function as a substrate consumed during the course of the reaction. The small SL RNA and precursor mRNA could be ligated end-to-end to produce a precursor containing a hybrid intron. This precursor could then be processed normally by splicing at the SL donor and target mRNA acceptor splice sites. On the other hand, the 3′ end of the small SL RNA and the 5′ end of the precursor mRNA may be thought of as parts of a discontinuous intron, which, if aligned properly, could undergo intermolecular splicing by conventional intramolecular mechanisms. Both variations of the splicing model require that the target mRNA is transcribed as a precursor containing at the least two additional 5′ nucleotides (AG/) that comprise the acceptor splice site (Figure 6B). As noted above, DNA sequences 5′ to each of eight SL-utilizing structural genes contain the AG at the appropriate splice junction.

There are no compelling precedents that favor either of these models; both mechanisms have been invoked in the past to explain related phenomena. Transcription of the influenza (Krug et al., 1979) and coronavirus (Spaan et al., 1983; Lai et al., 1984) RNA genomes, and production of defective interfering particles (Lazzirini et al., 1981) of negative-strand RNA viruses have been explained by RNA priming of transcription reinitiation (and other variations of a jumping polymerase model). In the case of influenza virus, small 5′ capped oligonucleotides derived from host-cell mRNAs are used by a viral polymerase complex to prime viral transcription; no sequence homology between primer and template is involved in this process (Plotch et al., 1981). Intermolecular RNA recombination (i.e., splicing/ligation) occurring post-transcriptionally has been hypothesized in the putative origin and evolution of viroid RNAs from plant cell introns (Zimmern, 1982).

Both models suggest that the small SL RNA resides in the nucleus, most likely as an RNP complex. In general, the priming model predicts that the SL (or small SL RNA) sequence is present on nascent target mRNA transcripts, whereas in the splicing/ligation model the SL is added post-transcriptionally. Splicing intermediates are for the most part different, depending on the function of the SL in either priming or post-transcriptional ligation. It may be possible to characterize some of the excised sequences (perhaps the 100 bp fragment, Figure 3) to resolve between alternative models. Our efforts to reconstitute these activities in vitro, patterned on recent developments in mammalian systems (Krainer et al., 1984; Grabowski et al., 1984), if successful, should define the mechanism of SL addition to trypanosomatid mRNAs.

Experimental Procedures

Cell Culture and RNA Methods

Procedures for cell culture and RNA isolation from E. coli, T. brucei, L. collosoma, and Crithidia fasciculata were as previously described (Milhausen et al., 1983; Parsons et al., 1983). High-salt-fractionated T. cruzi RNA was kindly provided by Drs. P. Lizardi and N. Nogueira. RNA was electrophoresed on a 10% polyacrylamide-7 M urea gel in the presence of TBE (80 mM Tris-CL, pH 8.3, 80 mM Na borate, and 2.5 mM NaEDTA). RNA bands were visualized by ethidium bromide. RNA gels were electroblotted to a Zeta-probe membrane (BioRad Laboratories) in 40 mM Tris-Cl, pH 7.8, 20 mM NaAcetate, and 2 mM NaEDTA for 1 hr at 0.3 V/cm and 1 hr at 0.7 V/cm, baked at 80°C for 2 hr under vacuum, and then hybridized (see below). The small SL RNA fractions from T. brucei, L. collosoma, and T. cruzi were isolated by elution from separate preparative polyacrylamide gels. Strips of polyacrylamide from the small SL RNA-containing regions of

each gel were sliced and incubated at 4°C for 16 hr and 65°C for 2 hr in TBE. The buffer was decanted, extracted with iso-amyl alcohol, and ethanol-precipitated.

Hybridization to RNA Blots

A synthetic 22 nucleotide oligomer complementary to the SL (22-mer; see Figure 5) was ^{32}P 5′-end-labeled as described previously (Nelson et al., 1983). Conditions for hybridization of this probe to RNA blots were; 18 hr at 34°C in 5× SSPE (0.9 M NaCl, 0.05 M NaH_2PO_4, 0.005 M Na_2EDTA, pH 7.4), 0.25% sarkosyl, 200 μg/ml denatured salmon testis DNA and with 5 × 10^6 cpm of probe (specific activity ca. 5 × 10^6 cpm/pmole) per ml of hybridization solution. After hybridization the blot was washed twice at 20°C and twice at 34°C in the fresh hybridization solution.

M13 (sense strand) SL-reiteration unit clones from T. brucei, mpTb-1.4M(1); T. cruzi, mpTc-0.6S(1); and L. collosoma, mpLc-0.6T(1) were previously described (Nelson et al., 1984). ^{32}P-labeled probes were prepared by Klenow synthesis using the sequencing primer in the presence of ^{32}P-dCTP and dATP, dGTP, dTTP. Hybridization to RNA blots was for 18 hr at 42°C in 5× SSPE, 0.1% sarkosyl, 200 μg/ml salmon testis DNA, 50% formamide, 25 mM sodium phosphate, and with 4 × 10^6 cpm of probe per ml of hybridization solution. After hybridization the blots were washed three times at 20°C in 1× SSPE, 0.1% sarkosyl, and once at 50°C in 0.3× SSPE, 0.1% sarkosyl.

Primer Extensions

Small RNA fractions (~10 ng) obtained by elution as described above were incubated with unlabeled 22-mer (200 ng) in a final volume of 5 μl. All samples were heated for 15 min at 65°C, then 45 min at 37°C for T. brucei and 25°C for T. cruzi small RNA. The following were then added to the samples; 1 μl of 10× buffer (500 mM Tris-Cl, pH 8.3, 500 mM KCl, 80 mM $MgCl_2$ and 400 mM DTT), 1 μl of 10 mM dATP, 10 mM dGTP, and 10 mM TTP and 10 μCi of α-^{32}P-dCTP (400 Ci/mmole, New England Nuclear). Extension reactions, in a final volume of 10 μl, were initiated by the addition of 14 U AMV reverse transcriptase. Following incubation at 42°C for 30 min the reaction was stopped by the addition of SB buffer (90% formamide, 0.1% bromphenol blue), and the samples boiled and loaded onto a 10% polyacrylamide urea sequencing gel. Primer extension of the L. collosoma small RNA fraction was as above for T. brucei with the following exceptions: the 23 nucleotide ^{32}P 5′-end-labeled oligomer (see Figure 5) was used and 1 μl of 10 mM dCTP was substituted for the radioactive dCTP.

S1 Nuclease Analysis

^{32}P-labeled probes complementary to the M13 sense strand reiteration unit clones of T. brucei, T. cruzi, and L. collosoma were synthesized as described above. Each probe was hybridized in the absence (lane 1) and presence (lane 2) of 20 μg of total RNA. The probe and RNA-probe samples were boiled for 10 min and then incubated at 50°C for 3 hr. Hybridization conditions in 30 μl were; 80% formamide, 40 mM PIPES, pH 6.4, 1 mM NaEDTA, and 400 mM NaCl. The samples were diluted by the addition of 300 μl of 0.28 NaCl, 0.048 M NaAcetate, 4.5 mM $ZnSO_4$, 20 μl/ml single-stranded salmon testis DNA, and 14 U/ml S1 nuclease (PL Biochemicals). Following a 30 min incubation at 37°C the samples were phenol-extracted and ethanol-precipitated in the presence of 25 μg of carrier tRNA. The redissolved pellets were electrophoresed on 8% polyacrylamide-urea gels. 3′ S1 nuclease analysis of T. brucei total RNA was with a ^{32}P 3′-end-labeled Rsa I fragment from the reiteration unit. The labeled Rsa site is with the SL sequence and the fragment extends beyond the poly(T) stretch observed in the T. brucei sequence (see Figure 5). Hybridization and S1 nuclease treatment were as described above.

DNA Cloning and Sequencing

The reiteration units from T. brucei, T. cruzi, and L. collosoma were cloned in M13 phages and sequenced as previously described (Parsons et al., 1984a).

Acknowledgments

The authors thank Dr. Philip Barr for synthesis of the L. collosoma 23-mer primer, Dr. Marilyn Parsons for critical discussions and readings of the manuscript, Kenneth Watkins for useful discussions and unpublished data, Drs. Paul Lizardi and Nadia Noqueira for T. cruzi RNA samples, Dr. Jean Feagin for T. brucei procyclic culture from RNA, Dr. K. Stuart for T. brucei cells, and Dr. John Boothroyd and his colleagues for the exchanges of information prior to publication. This work was supported by the Rockefeller Foundation Great Neglected Disease Program, National Institutes of Health Grant AI 17309 and NSF Grant PCM8021551. N. A. is a Burroughs Wellcome Scholar of Molecular Parasitology.

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 August 3, 1984

References

Berget, S. M. (1984). Are U4 small nuclear ribonucleoproteins involved in polyadenylation? Nature *309*, 179–181.

Bernards, A., DeLange, T., Michiels, P., Liu, A., Huisman, M., and Borst, P. (1984). Two modes of activation of a single surface antigen gene of Trypanosoma brucei. Cell *36*, 136–170.

Boothroyd, J. C., and Cross, G. A. M. (1982). Transcripts coding for variant surface glycoproteins of *Trypanosoma brucei* have a short identical exon at their 5′ end. Gene *20*, 281–289.

Campbell, D. A., van Bree, M. P., and Boothroyd, J. C. (1984a). The 5′-limit of transposition and upstream barren region of a trypanosome VSG gene: tandem 76 base-pair repeats flanking $(TAA)_{90}$. Nucl. Acids Res. *12*, 2759–2773.

Campbell, D. A., Thornton, D. A., and Boothroyd, J. C. (1984b). The spliced-leader of a trypanosome mRNA exists as a short, discrete transcript indicating discontinuous transcription. Nature, in press.

Childs, G., Maxson, R., Cohn, R. H., and Kedes, L. (1981). Orphons: dispersed genetic elements derived from tandem repetitive genes of eucaryotes. Cell *23*, 651–663.

Cordingly, J. S., and Turner, M. J. (1980). 6.5S RNA; preliminary characterization of small RNAs in *Trypanosoma brucei*. Mol. Biochem. Parasitol. *1*, 91–96.

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

DeLange, 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*, 3777–3790.

DeLange, T., Berkvens, T. M., Veerman, H. J. G., Frasch, A. C. 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. M., and Donelson, J. E. (1984). Characterization of the 1.35 kilobase DNA repeat unit containing the conserved 35 nucleotides at the 5′-termini of variable surface glycoprotein mRNAs in *Trypanosoma brucei*. Nucl. Acids Res. *12*, 4907–4920.

Grabowski, P. J., Padgett, R. A., and Sharp, P. A. (1984). Messenger RNA splicing in vitro: an excised intervening sequence and a potential intermediate. Cell *37*, 415–427.

Gray, M. W. (1981). Unusual patterning of ribonucleic acid components in the ribosome of *Crithidia fasciculata*, a trypanosomatid protozoa. Mol. Cell. Biol. *1*, 347–357.

Krainer, A. R., Mariatis, J., Ruskin, B., and Green, M. R. (1984). Normal and mutant human β-globin pre-mRNAs were faithfully and efficiently spliced in vitro. Cell *36*, 993–1005.

Krug, R. M., Broni, B. B., and Bouloy, M. (1979). Are the 5′ ends of influenza virus mRNAs synthesized in vivo donated by host mRNAs? Cell *18*, 329–334.

Lai, M. M. C., Baric, R. S., Brayton, P. R., and Stohlman, S. A. (1984). Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus, a cytoplasmic RNA virus. Proc. Nat. Acad. Sci. USA *81*, 3626–3630.

Lazzirini, R. A., Keene, J. D., and Schubert, M. (1981). The origins of defective interfering particles of the negative-strand RNA viruses. Cell *26,* 145–154.

Lerner, M. R., Bayle, J. A., Mount, S. M., Wolin, S. L., and Steitz, J. A. (1980). Are snRNPs involved in splicing? Nature *283,* 220–224.

MacKay, R. M., Gray, M. W., and Doolittle, W. F. (1980). Nucleotide sequence of *Crithidia fasciculata* cytosol 5S ribosomal ribonucleic acid. Nucl. Acids Res. *8,* 4911–4917.

Michiels, F., Mathyssens, 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. *7,* 1185–1192.

Milhausen, M., Nelson, R. G., Parsons, M., Newport, G., Stuart, K., and Agabian, N. (1983). Molecular characterization of initial variants from the lsTat 1 serodeme of *Trypanosoma brucei*. Mol. Biochem. Parasitol. *9,* 241–254.

Murphy, W. J., Brentano, S. T., Rice-Ficht, A. C., Dorfman, D. M., and Donelson, J. E. (1984). DNA rearrangement of the variable surface antigen genes of the trypanosomes. J. Protozool. *31,* 65–73.

Myler, P., Nelson, R. G., Agabian, N., and Stuart K. (1984). Two mechanisms of expression of a predominant variant antigen gene of *Trypanosoma brucei*. Nature *309,* 282–284.

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

Parsons, M., Nelson, R. G., Newport, G., Milhausen, M., Stuart, K., and Agabian, N. (1983). Genomic organization of *Trypanosoma brucei* variant antigen gene families in sequential parasitemias. Mol. Biochem. Parasitol. *9,* 255–269.

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

Parsons, M., Nelson, R. G., and Agabian, N. (1984b). Antigenic variation in African trypanosomes: DNA rearrangements program immune evasion. Immun. Today *5,* 43–50.

Parsons, M., Nelson, R. G., Stuart, K., and Agabian, N. (1984c). Variant antigen genes of *Trypanosoma brucei*: genomic alteration of a spliced leader orphon and retention of expression-linked copies during differentiation. Proc. Nat. Acad. Sci. USA *81,* 684–688.

Plotch, S. J., Bouloy, M., Ulmanen, I., and Krug, R. M. (1981). A unique cap (m^7) (m GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. Cell *23,* 847–858.

Rogers, J., and Wall, R. (1980). A mechanism for RNA splicing. Proc. Nat. Acad. Sci. USA *77,* 1877–1879.

Schnare, M. N., and Gray, M. W. (1982). Nucleotide sequence of an exceptionally long 5.8S ribosomal RNA from *Crithidia fasciculata*. Nucl. Acids Res. *10,* 2085–2092.

Simpson, L., and Simpson, A. M. (1978). Kinetoplast RNA of Leishmania tarentolae. Cell *14,* 169–178.

Spaan, W., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., van der Zeijst, B., and Siddell, S. G. (1983). Corona virus mRNA synthesis involves fusion of non-contiguous sequences. Eur. Mol. Biol. J. *2,* 1839–1844.

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

Vickerman, K. (1978). Antigenic variation in trypanosomes. Nature *273,* 613–617.

Zimmern, D. (1982). Do viriods and RNA viruses derive from a system that exchanges genetic information between eukaryotic cells? Trends Biocher Sci. *7,* 205–207.