Apparent generation of a segmented mRNA from two separate tandem gene families in *Trypanosoma cruzi*

Antonio Gonzalez¹, Terry J. Lerner², Maria Huecas, Beatriz Sosa-Pineda, Nadia Nogueira³ and Paul M. Lizardi*

Laboratory of Cell Biology, The Rockefeller University, New York, NY 10021, and ³Department of Medical and Molecular Parasitology, New York University Medical Center School of Medicine, New York, NY 10010, USA

Received 26 April 1985; Revised and Accepted 19 July 1985

ABSTRACT

Using a cDNA for an abundant *Trypanosoma cruzi* mRNA as probe, we have cloned and sequenced a gene which is organized in at least 20 nearly perfect tandem repeats of 940 base pairs. The 5' end of the mRNA has been sequenced by primer extension and found to contain a 35 nucleotide mini-exon (or spliced-leader) sequence that is ubiquitous in trypanosome mRNAs. This sequence, however, is not present in the tandem genomic repeats which encode the exon containing the major portion of the mRNA. Previous studies have shown that the 35-nucleotide sequence is encoded by a separate tandem gene family. One model to explain the formation of a segmented mRNA invokes priming of transcription by a small RNA which contains the leader sequence at its 5' end. However, northern blot analysis of total trypanosome RNA reveals a ladder of molecules larger than the mature mRNA, which appear to be faithful multimeric copies of the tandem gene. The discrete sizes of these RNAs correspond to those expected for partially processed precursors. These observations lend credence to the possibility of an alternative model where segmented mRNAs are generated by inter-molecular splicing.

INTRODUCTION

Recent studies on the structure of *Trypanosoma brucei* messenger RNA's have revealed an unusual sequence organization where the majority, if not all mRNAs share identical 35-nucleotide termini. These short 5'-terminal sequences, originally termed mini-exons (and, more recently, spliced leaders or SL sequences) were discovered in *Trypanosoma brucei* variable surface glycoprotein mRNAs (1,2). Subsequently SL sequences were found to be present in non-VSG gene products of *T. brucei* (3,4,5) as well as in genomic DNA of other species of trypanosomes including *T. cruzi* (6,7). In the case of VSG genes all attempts to map SL-coding DNA immediately upstream of the sequences coding for the main body of the mRNA have been unsuccessful. This analysis has been carried to its logical limit for at least one specific VSG genotype, and indeed the SL sequences can not be detected in the entire chromosome which carries the expression-linked copy of the VSG sequence (8).

Genes encoding the SL sequence have been characterized and sequenced in

T. brucei, T. vivax, T. cruzi, and L. collosoma (3,6,9,10,12).

The genes are clustered in tandem arrays, with a few scattered orphans, and their overall reiteration frequency is about 200 copies per genome. The repeating units are of uniform length, but their size varies among different trypanosome species. These tandem genes are abundantly transcribed (11-13), yielding a small RNA product whose length (98-140 nucleotides, approximately) is also species-specific. An SL sequence resides at the 5' end of this small RNA, and is apparently utilized to generate mRNA 5' ends via a trans-acting mechanism which remains to be elucidated.

In this report we describe a tandemly organized gene from Trypanosoma cruzi whose structure and transcriptional expression is of considerable interest with regard to the mechanism of mRNA formation in these organisms.

MATERIALS AND METHODS

DNA Extraction and Construction of Genomic Clones

Total DNA from T. cruzi strain Y epimastigotes cultured in LIT (14) was isolated as described (15). The DNA was incubated with BamHI to generate a partial digest, and electrophoresed in a preparative lane of a 1% agarose gel. The material was electroblotted for 2 hours at 4.5 volts/cm. to two

adjacent membranes. On one side was a 1 cm. wide vertical strip of Gene Screen Plus (New England Nuclear, Boston, Mass.) and next to it a 7 cm. wide strip of NA-45 DEAE (Schleicher & Schuell, Keene, N.H.). After electroblotting the strip of Gene Screen Plus was hybridized with labeled plasmid containing the LP8 cDNA insert (16) in order to localize the ladder of gene fragments generated by the partial digest (refer to figure 1B). The Gene Screen membrane was then aligned with the DEAE membrane to locate the dimer, and trimer fragments. Horizontal DEAE membrane strips of about 2.5 mm. X 70 mm. were cut with a razor blade, and the bound DNA was recovered as described (17). This DNA was inserted in the BamHI site of PUC-13 (18), and clones were isolated by high density colony hybridization (19).

DNA Analysis and Sequencing

Southern blot hybridization was carried out using standard procedures, except that in some cases we used gels of high agarose content (2.5% - 4.0%) in the presence of 40% formamide (20). To sequence the genomic clones we generated a series of deletion subclones using exonuclease III treatment followed by S1 nuclease digestion (21) and Klenow filling. Individual plasmid preparations that had been subjected to increasing times of ordered deletion were pooled and electrophoresed in an agarose gel. Size cuts

separated by intervals of 300 base pairs were isolated by band interception with NA-45 DEAE membrane (17). The ordered deletion fragments were subcloned in M13MP11 (22) and sequenced using the Sulfur-35 dideoxy method (23).

RNA Analysis and Primer Extension Sequencing

Total cellular RNA was isolated from epimastigotes using the proteinase K and sodium perchlorate method (24). Contaminating DNA was removed by digestion with RNase-free DNase (25). Northern blot hybridization analysis was carried out by standard methods (26), except that the agarose gels contained 50% percent formamide and 1.5M formaldehyde. We have observed that in the presence of these two reagents the mobility plot for denatured DNA markers shows relatively good agreement with the mobility of ribosomal RNA's of known size (27). Primer extension sequencing of the 5' terminus of LF8 mRNA was carried out by standard methods (28). The DNA primer fragment was labeled at the 5' ends with polynucleotide kinase and hybridized to total epimastigote RNA under R-looping conditions. The reverse transcriptase product was sequenced using the chemical method (29).

RESULTS

Isolation of a Tandemly Repeated Genomic Clone Using the LF8 cDNA Probe

We have described elsewhere (16) the isolation and sequencing of a cDNA clone corresponding to a highly abundant messenger RNA from *T. cruzi*. This cDNA, called LF8 codes for a polypeptide of unknown function, whose predicted amino acid sequence shows strong homology to a number of calcium-binding proteins such as troponin and myosin. Using labeled LF8 cDNA as probe, we examined the genomic organization of this sequence by Southern blot hybridization. The pattern obtained after cleavage of epimastigote DNA with four different restriction enzymes was identical, and consisted of a single band at about 940 bp (Fig 1A). One explanation for this unusual result is the presence of multiple gene copies organized in a tandem head-to-tail fashion. If this were the case, all enzymes that cleave only once in the repeat unit should cut out fragments of the same unit length. Subsequent analysis of the time course of digestion with HindIII (Fig 1B) revealed a ladder pattern of bands in the partially digested genomic DNA, with all sizes being multiples of 940. Thus the genes which presumably code for the LF8 mRNA are indeed organized in a tandem array which contain at least 20 adjacent repeats as judged by the longest ladders observed on partial digests. Semi-quantitative signal comparisons in Southern blots containing known amounts of pTC1F8 DNA as copy-number standards yielded a copy number of 30 genes per haploid

Figure 1. Analysis of 1F8 gene organization by Southern blotting.

A Total *T. cruzi* DNA was cleaved with the indicated restriction enzymes (B, BamHI; H, HindIII; P, PstI; S, SalI), fractionated in an agarose gel and transferred to 32 P-labeled nitrocellulose. The blot was hybridized with probe obtained by nick-translation of the pTCLF8 cDNA clone. The size markers on the left side are HaeIII fragments of phi-X 174 DNA, whose length in base pairs is indicated at the right side of the figure.

B Partial digest analysis: 5 micrograms of total DNA was digested with HindIII (0.1 unit per microgram DNA) for [1] 20 min; [2] 40 min; [3] 60 min; [4] 120 min; [5] overnight. Electrophoresis in a 1% agarose gel and Southern blot hybridization were as described in experimental procedures.

genome, (data not shown). These experiments have been repeated with DNA from blood form trypomastigotes and the organization, as well as the copy-number of the 1F8 gene cluster has been found to be the same. Null digests generated by restriction enzymes which do not cleave within the 1F8 tandem

```

*****
GTGTGTGTGTGTGTGTGTGGGCTCCTGTGCCCTTGTGTGTGTGTCTTCTTTGCTGCTGTGGGCTTGG      70
=====
TGCCACCTCTCTCTTTCTGTTGTGGGTTTGTGTTGTTGCGCGCAGGTTGTGCAACAGTCGATAATAAGC      140
=====
                                met
CGAGTCAAGGAGAGCAACGAGATGGGTGCTTGTGGGTGCAAGGGCTCGACGAGCGACAAGGGGTTGGCGA      210

GCGATAAGGACGGCAAGAAGGCCAAGGACCGCAAGGAAGCGTGGGAGCGCATTGCGCAGGCGATTCTCG      280

TGAGAAGACCGCCGAGSCAAAACAGCGCCGCATCGAGCTCTTCAAGAAGTTCGACAAGAACGAGACTGGG      350

AAGCTGTGCTACGATGAGGTGCACAGCGGCTGCCTCGAGGTGCTGAAGTTGGACGAGTTTACGCCGCGAG      420

TGCGCGACATCACGAAGCGTGCCTTTGACAAGGCGAGGGCCCTGGGCAGCAAGCTGGAGAACAAGGGCTC      490
              (C)                      (GT)
              ---                      ^
CGAGGACTTTGTTGAATTTCTGGAGTTCGGTCTGATGCTGTGCTACATCTACGACTTCTTCGAGCTGACG      560

GTGATGTTTCGACGAGATTGACGCCTCCGGCAACATGCTGGTCGACGAGGAGGAGCTCAAGCGCGCCGTGC      630

CCAAGCTTGAGGCGTGGGGCGCCAAGGTCGAGGATCCCGCGCGCTGTTCAAGGAGCTCGATAAGAACGG      700

CACTGGGTCCGTGACGTTTCGACGAGTTTGTCTGCGTGGGCTTCTGCAGTCAAAGTGGACGCCGACGGCGAC      770

                                ter
CCGGAACAAGTGCAGGAGAGCGCGTGAAGCTCGCGGCCGAGAGTGACGTGCCCGCGCCGTTGCCAATGT      840

#####
AAACACCGTGTCCGCATGCGGGCAGGCACTTGTGTTTTGCTGGGCGGAGCACATTCTCCACTTTATGTAGT      910
*****
                                940
TGATGGTCATTTTACCTTCTACTGTGCTT
=====

```

Figure 2. Nucleotide sequence of 1F8 genomic repeat unit. The sequence shown here is derived from one of the two consecutive genomic repeat units (1880 bp, flanked at either end by the *Bam*HI site located at position 662) that were sequenced in their entirety, except for a stretch of 40 nucleotides in one of the two repeats. Those 40 nucleotides which were sequenced only once in the genomic clone were contained within the presumed protein coding region in the original 1F8 cDNA clone. The T at position 446, which agrees with the cDNA sequence (16) was substituted for a C in one the genomic repeats. The TG at positions 475-476 is in disagreement with the cDNA sequence, which contains GT at those positions. The asterisks at the top left mark the oligo GT sequence found at every repeat. The "#" symbols near the bottom indicate a candidate site for poly-A addition, based on its precise homology to the sequence AACACC preceding the poly A in VSG mRNAs. An even stronger similarity to the 3' end of IaTat 1.2 VSG mRNA (30) is indicated by asterisks. The markings "met" and "ter" demarcate the beginning and end of the only large open reading frame in the sequence. The fine arrows at position 116 indicate the site of a discontinuity in the mRNA produced by this gene. The doubly underscored regions mark the 160 bp domain used as an "inter-exon" probe in Figure 5.

array show a single band of very high molecular weight (greater than 25 kb) and no hybridization signals on smaller DNA fragments, hence there is no evidence for the existence of orphon (unlinked) copies of the gene. Long radioautographic exposures of Southern blots which display the 940 bp monomer reveal two faint bands whose size varies according to the enzyme being used. These may represent flanking fragments (data not shown).

The high abundance of the 1F8 genomic sequences allowed us to isolate genomic clones directly in plasmids, after partial BamHI digestion of total *T. cruzi* DNA. Clones were obtained with either two or four tandem repeats in the insert. The clone containing a dimeric insert (1880 bases) was sequenced in its entirety, except for a stretch of 40 base pairs in the protein coding region of one of the repeats. The sequence of a single repeating unit is shown in figure 2. The genomic sequence diverged from the cDNA sequence at two positions (see bases in parenthesis in the middle of the figure) one of which was located at a wobble nucleotide in the open reading frame (position 446 in the sequence). The change of GT for TG at positions 475 and 476 would introduce a substitution of a leucine instead of an arginine in the presumed amino acid sequence. A notable feature of the genomic sequence is the occurrence of 9 consecutive repeats of the dinucleotide GT (see bases 1-18) every 940 base pairs. The presumed initiator AUG is found approximately 150bp downstream of the oligo-GT sequence. The sequence AACACCGTGT, which occurs about 50bp downstream of the translation terminator TGA, closely resembles the sequence AACACCTGT which immediately precedes the poly A sequence at the 3' end of the IaTat 1.2 VSG mRNA (30). The sequence AACACC is present near the 3' end of virtually all VSG mRNAs (30) and must therefore be considered a likely candidate site for polyadenylation in 1F8 mRNA. The sequence AAAATTTT, which often precedes the poly A addition site in VSG genes of *T. brucei* does not occur anywhere in the untranslated regions of the 1F8 gene, and therefore is probably not a good candidate for a general trypanosome poly-A addition signal.

Nucleotide Sequence at the 5' End of 1F8 mRNA

To determine the RNA sequence at the 5' end of the 1F8 RNA we carried out a primer extension experiment using reverse transcriptase and a DNA fragment beginning at a HinfI site which is found 24 bp downstream of the 5' end of the original 1F8 cDNA clone. The results of this sequencing experiment are shown in Figure 3. The RNA sequence is found to extend about 40 nucleotides upstream from the 5' end of the cDNA clone. The first 35 nucleotides of the RNA do not show any sequence homology to the 1F8 genomic

Figure 3. Nucleotide sequence at 5' end of LF8 mRNA deduced by primer extension. A cDNA restriction fragment of 94 base pairs (bounded by the *Hinf*I site at position 144 and the *Ava*II site at position 238 in the LF8 sequence shown in figure 2) was labeled at the 5' ends, hybridized to total RNA and used for primer extension sequence analysis with reverse transcriptase. The top of the figure shows the experimental data and a partial reading of the gel. At the bottom of the figure is shown the deduced sequence for the 5' end of LF8 mRNA flanked by the sequence of the genomic clone (from Fig.2) and the predicted sequence of *T. cruzi* SL RNA (6,12). The curved arrow marks the 5' boundary of the original pTC1F8 cDNA clone. The fat arrow demarcates the position where the 35-nucleotide spliced leader sequence is joined to the main body of the mRNA. The asterisks (*) mark identities in nucleotide sequence. The vertical lines underscore the 35-nucleotide spliced leader sequence in the mRNA. A dash over a base marks the positions of pyrimidines preceding the candidate acceptor splice site, and a double slash (=) indicates identities with the consensus splice junction sequences of higher organisms (45).

sequence. Instead, they correspond exactly to the predicted sequence of a trypanosome spliced leader, based on the sequence of genomic clones coding for *T.cruzi* SL RNA (6,12). The site where the genomic sequence diverges from

Figure 4. Size distribution of RNA's hybridizing to LF8 cDNA probe. Total RNA was extracted from *T. cruzi* and treated with DNA'se as described in Materials and Methods. The RNA, together with suitable DNA size markers, was electrophoresed in a 2.5% agarose gel containing 1.2M formaldehyde and 50% formamide, and subsequently blotted to nitrocellulose. The hybridization probe was clone pTCLF8-550 which bears a LF8 cDNA insert. Hybridization was carried out as described in Materials and Methods, at a temperature of 53 degrees Centigrade. Lane 1: phage lambda DNA cleaved with HindIII; Lane 2: RNA from epimastigotes; Lane 3: Total genomic DNA from *T. cruzi* cleaved partially with BamHI; Lane 4: phage phi X-174 DNA cleaved with HaeIII. The large arrows point to the two bands of LF8 genomic DNA monomer and dimer generated by the partial digest. Numbers are approximate lengths, in bases.

the mRNA sequence contains the sequence CAG/GT and is preceded by a fairly pyrimidine-rich domain. These sequence features are reminiscent of the consensus sequence for a 3' splice site in higher eukaryotes.

Size Distribution of LF8 Transcripts

The size of RNA molecules containing the LF8 sequence was investigated by northern blot hybridization using nick-translated cDNA as probe. Size markers were provided by *T. cruzi* genomic DNA fragments generated by partial cleavage with BamHI (as in figure 1B, lane 4). The agarose gel system used in this experiment contained 50% formamide and 1.5M formaldehyde, which together provide a strong denaturation criterion. The pattern of hybridization shows a major band at approximately 900 nucleotides in the RNA obtained from the insect form of the parasite (Fig.4, lane 2). A less abundant hybridizing RNA migrates at approximately 1900 nucleotides, very close to the 1880 nucleotide DNA dimer band generated by the partial BamHI digest. This RNA band has been

Figure 5. Hybridization of putative LF8 mRNA precursor molecules with a coding DNA probe versus an "inter-exon" DNA probe. RNA from epimastigotes was electrophoresed and blotted as in figure 4, except that the DNase treatment was omitted to maximize the integrity of large RNA molecules. Samples 1,2,3 were hybridized with the pTCLF8-550 cDNA probe, while samples 4,5,6 were hybridized with an "inter-exon" DNA probe (refer to Figure 2) encompassing 160 nucleotides from an HgiAI site (map position 889 in fig 2) found 38 bp downstream of the presumed poly A addition site to a BssHII site (map position 109 in fig. 2) found 7 bp upstream of the splice site. Lane 1: Total genomic DNA cleaved partially with BamHI; Lane 2: Total RNA (8ug); Lane 3: Poly A - containing RNA selected by oligo dT-cellulose chromatography (3ug); Lane 4: Total RNA (25ug); Lane 5: Poly A - containing RNA (5ug); Lane 6: Total genomic DNA cleaved partially with HindIII. The RNA preparations used in lanes 4 & 5 are different from those in lanes 2 & 3. The radioautographs were exposed for 14 days with two intensifying screens. The letter "a" indicates the major LF8 mRNA band. Letters "b" through "i" indicate larger LF8-related RNAs.

observed in several experiments, and its abundance is variable, perhaps being influenced by the speed with which cultures are harvested. Experiments where the concentration of formaldehyde in the gel loading buffer was varied have ruled out the possibility that the band is the result of a cross-linking artefact. Melting of the hybridized RNA in 50% formamide, 2XSSC at increasing temperatures (data not shown) results in concomitant release of label from both the 900 and 1880 nucleotide RNA bands, suggesting that both hybrids are equally stable.

To investigate further the nature of the larger RNA the northern blotting experiments were repeated using as probe an "inter-exon" restriction fragment which encompasses 160 bp between a HgiAI site downstream of the

presumed poly A addition site and a BssHII site just upstream of the splicing site in the lF8 map (see doubly underscored region in Figure 2). This probe should not hybridize with mature lF8 mRNA, but should detect any transcripts of the lF8 gene which contain sequences flanking the 736-nucleotide exon. Figure 5, lanes 2 and 3, shows a control blot containing total RNA and poly A+ RNA, respectively, hybridized with the cDNA probe. A faint RNA band, marked "e", is seen at an apparent size of approximately 1950 nucleotides; other large RNAs give weaker signals. Lanes 4 and 5 contain relatively larger RNA inputs and were hybridized with the "inter-exon" probe. The total RNA (lane 3) shows bands with apparent sizes of 1000 (b), 1600 (c), 1900 (d), 2050 (e), and even larger molecules (g,h,i). At least three of these bands (b,c,d) are also visible in the poly A+ RNA (lane 5). This experiment shows that the "inter-exon" domain of the lF8 gene is transcribed together with the large exon, and gives rise to RNA molecules of discrete size classes, some of which are polyadenylated. The simplest interpretation of this data is that the ladder of larger RNA bands represents transcripts containing more than one repeat of the lF8 tandem gene array. These larger RNAs may be metabolic precursors of the 900 nucleotide mature mRNA, but this remains to be proven.

DISCUSSION

There are other published examples of tandemly organized protein-coding genes in trypanosomes. For instance, the genes coding for alpha and beta tubulins have been reported to occur as alternating tandem arrays in both T. brucei (31,32) and Leishmania (33,34). In T. cruzi, however, the beta tubulin genes are not organized in perfect tandem repeats (our unpublished observations). The number of repeats of the lF8 gene in T. cruzi (approximately 30) is greater than that of the tubulin genes of T. brucei (8-17). This frequency may reflect the requirement for a high transcriptional output of the messenger, since the intensity of lF8 mRNA in northern blots indicates that it is roughly twice as abundant on a molar basis as beta tubulin mRNA (our unpublished results, obtained using a T. cruzi beta tubulin cDNA probe). An important question which remains to be answered is whether the hypothetical lF8 promoter occurs once every repeating unit, as opposed to being located at a unique site upstream of the tandem array. Our present lack of knowledge about the sequence of trypanosome promoters precludes the tentative identification of putative promoter sequences in the non-coding spacer of the lF8 tandem array. An interesting structural feature of the lF8 tandem gene array is the presence of perfect

oligo GT repeats between the open reading frames. The oligo-GT sequence may be a potential site for the formation of Z-DNA structure (35), and could conceivably play a role in the establishment of a transcriptionally active chromatin domain (36,37).

Since we have not determined the nucleotide sequences flanking the 1F8 tandem array, we can not exclude the possibility that a 35 nucleotide SL coding domain would be present upstream of the 25-30 copies of the 940 bp repeating unit. However, if transcription were to start at such a site, it would provide a spliced leader for only one repeat of the tandem array. Given the high abundancy of the mRNA, it seems unlikely that this gene cluster contains such a large number of perfect, yet non-functional repeats. We mentioned earlier that studies on variable surface glycoproteins in T. brucei have essentially ruled out the hypothesis that the small leader sequence found in VSG mRNA is coded for by an upstream DNA element (see 7). Rather, the available evidence suggests that the small leader sequence present in VSG's and in a variety of other mRNAs is derived in trans from the so-called SL RNA via a mechanism that is not yet understood (11-13). It seems likely that T. cruzi, just like T. brucei, has evolved a mechanism which allows for the addition of an SL sequence in trans, obviating the requirement for genetic linkage. The fact that the genes coding for the SL small RNA in T. cruzi, like those of African trypanosomes, are organized in tandem arrays (6,7) and abundantly transcribed (11-13) suggests that the operant mechanisms for spliced leader expression and ligation to mRNA are identical among all trypanosome species. Thus we tentatively conclude that the 1F8 mRNA of T. cruzi is a segmented RNA molecule produced by two (probably unlinked) families of tandem genes, the 1F8 array, and the SL array.

The 900 nucleotide RNA detected in northern blot experiments is just about the right size for a product of the 940 bp tandem gene array which we have sequenced. The CAG/GT junction where the spliced leader is added is separated from the candidate poly A addition site by 736 bp. Adding 35 nucleotides contributed by the spliced leader and 50-100 nucleotides (estimated) for the poly A sequence we arrive at a predicted size of approximately 821-871 nucleotides for the 1F8 mRNA. As yet we have not demonstrated a precursor-product relationship between the ladder of larger RNAs and the 900 nucleotide mRNA. Figure 6 shows two models which, in our view, could most reasonably explain the experimental observations. These models are analogous to two of the four hypothetical schemes previously

Figure 6. Alternative models for the generation of 1F8 mRNA. (A) Primed transcription followed by conventional splicing and (B) Intermolecular splicing. The symbols used are as follows: dotted area = oligo GT sequence; shaded area = main body of mRNA sequence; a = consensus polyadenylation signal; s = consensus splice junction; ME = mini-exon; PRI = hypothetical RNA polymerase priming site; TER = hypothetical RNA polymerase termination site; SPL = splicing enzyme (or ribozyme); A+ = polyadenylation enzyme; AAA = poly A tail of mRNA. The numbers at the left of the different RNA molecules are estimated sizes, in nucleotides, assuming an average poly-A tail length of 50 nucleotides. The numbers joined by large brackets indicate the fact that such molecules would not be resolved as different size classes in agarose gels.

outlined by Milhausen et. al. (12). In the scheme at the left of the Figure the SL RNA is proposed to serve as a primer for initiation by a trypanosome RNA polymerase at sites which could be located either downstream or upstream of the oligo(GT) sequence domain. The primary transcript would then undergo cleavage and poly A addition at the appropriate site, and finally a conventional RNA splicing reaction would place the 35-nucleotide spliced leader at the correct position in the mRNA, with removal of a short intervening sequence. This model could be reconciled with the existence of larger size classes of 1F8 RNA by proposing that transcriptional termination is leaky and produces read-through transcripts, or alternatively, by proposing that there are multiple promoter-terminator combinations that can be used simultaneously. Further processing of multimeric transcripts would yield a single functional mRNA molecule because only one product would have a spliced leader.

The second hypothetical scheme proposes that the primary transcription product is a multimeric RNA molecule initiated at a site which could lie either inside or outside of the tandem DNA cluster. This long precursor molecule would be subsequently processed by two types of reactions, one a cleavage and polyadenylation event and the other a novel intermolecular splicing reaction. In this reaction the donor site would be in the SL small RNA, just downstream of the 35-nucleotide spliced leader, and the acceptor site would be the CAG/GT junction in the multimeric precursor. The two molecules may interact by themselves or may be brought together by the trypanosome equivalent of U1 small ribonucleoprotein particles (the SPL entities in the diagram), which would presumably contain the splicing activity. This model predicts the possible existence of a variety of partially processed precursor molecules, some of which are outlined in Figure 6B. While some of the mRNA precursor molecules schematized in the diagram are shown with a spliced leader sequence at the 5' end, we have no evidence whatsoever as to the presence or absence of a spliced leader in the precursor-sized RNA detected in the northern blot experiments. It is remarkable, however, that the predicted size classes shown in figure 6B (975-990, 1760, 1915-1930, 2085) are in relatively good agreement with the estimated sizes of the molecules detected in northern blots with the "inter-exon" probe (1000, 1600, 1900, 2050). The intermolecular splicing model also predicts the existence of even larger precursor molecules, and indeed a set of three faint RNA bands migrating in the vicinity of LF8 trimer DNA has been observed (Figure 5, "g,h,i"). Unfortunately their size could not be measured reliably in this experiment.

The available evidence is not sufficient to decide between the two models outlined above. There are well documented examples of primer-dependent transcriptional initiation in RNA viruses (38-41). A relatively minor problem with the priming model would seem to be the absence of a consensus sequence at the 3' end of SL RNA among various trypanosome species. However, in the case of host RNA-primed transcription of influenza virus there is no apparent requirement for sequence homology between primer and template RNA's (42). The intermolecular splicing model has no precedent in other biological systems but in fact it provides the most reasonable explanation for the various size classes of LF8 transcripts which we have observed. Zimmerm (43) has hypothesized that intermolecular RNA recombination is an attractive model for the expression of viroid genetic information during infection and could also provide a general mechanism for

expanding the coding potential of genes in a combinatorial way. This latter phenomenon may be of adaptative value for organisms such as trypanosomes, where such somatic recombination events would increase genetic plasticity and possibly facilitate changes in patterns of gene expression. While the 35-nucleotide spliced leader sequence does not code for protein, the possibility of post-transcriptional combinatorial joining of other (coding) exons remains to be explored. Such a situation could be diagnosed by looking for trypanosome cDNA clones in which adjacent restriction fragments within an open reading frame map at totally different loci on genomic Southern blots.

The availability of plasmid systems for the generation in vitro of functional RNA transcripts (44) makes it possible to generate multimeric RNA's from cloned lF8 tandem arrays. Using such transcripts one could test whether cell-free extracts from trypanosomes can modify a multimeric precursor RNA to generate spliced products via intermolecular ligation. The design of suitable in vitro transcription systems for the study of trypanosome polymerases should permit the analysis of the hypothetical priming reaction. Thus, the tools are at hand to test the validity of the two alternative models discussed here.

An important question which remains to be addressed concerns the existence of trypanosome nuclear RNP particles analogous to the so called U1 snurps of higher eukaryotes. These particles, which contain U1 small nuclear RNA have been implicated as central components of the mRNA splicing machinery of eukaryotic cells (45-49). We have recently identified three trypanosome small RNA molecules which bear a 2,2,7-trimethylguanosine cap structure at their 5' end. One of these RNA's shows considerable primary sequence homology to human U1 snRNA (manuscript in preparation), thus it seems reasonable to assume that U1-snurps do exist in trypanosomes. Whether these particles play a role in the generation of discontinuous mRNAs, possibly mediating an intermolecular RNA splicing reaction remains an exciting subject for future work.

ACKNOWLEDGMENTS

We thank David J. Anderson and Fred R. Kramer for critical reading of the manuscript. This work was supported by grants from the Carl J. Herzog Foundation (awarded to P.M.L. and N.N.), the UDNP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (T16/18/23 6A, awarded to N.N.), the Rockefeller Foundation (N.N.), and the N.I.H. (Biomedical Research Support Grant No. F07-RR05399-23, awarded to N.N.). N. N. is the recipient of a Research Career Award from the New York Heart Association.

Present addresses: ¹Centro de Biología Molecular, Universidad Autónoma de Madrid, Canto Blanco, Madrid-34, Spain, and ²Integrated Genetics, 51 New York Avenue, Framingham, MA 01701, USA

*To whom correspondence should be addressed

REFERENCES

1. Van der Ploeg, L.H.T., Liu, A.Y.C., Michels, P.A.M., De Lange, T., Borst, P., Majumder, H.K., Weber, H., Veeneman, G.H., and Van Boom, J. (1982) Nucl. Acids Res. 10, 3591-3604.
2. Boothroyd, J.C., and Cross, G.A.M. (1982) Gene 20, 281-289.
3. DeLange, T., Liu, A.Y., Van der Ploeg, L.H.T., Borst, P., Tromp, M.C., and Van Boom, J.H. (1983) Cell 34, 891-900.
4. Parsons, M., Nelson, R.G., Watkins, K.P., and Agabian, N. (1984) Cell 38, 309-316.
5. De Lange, T., Michels, P.A.M., Veerman, H.J.G., Cornelissen, W.C.A., and Borst, P. (1984) Nucl. Acids Res. 12, 3777-3790.
6. De Lange, T., Berkvens, T., Veerman, H.J.G., Frasch, C.C., Barry, J.D., and Borst, P. (1984) Nucl. Acids Res. 12, 4431-4443.
7. Nelson, R.G., Parsons, M., Selkirk, M., Newport, G., Barr, P.J., and Agabian, N. (1984) Nature 308, 665-667.
8. Van der Ploeg, L.H.T., Cornelissen, A.W.C.A., Michels, P., and Borst P. (1984) Cell 39, 213-221.
9. Nelson, R.G., Parsons, M., Barr, P., Stuart, K., Selkirk, M., and Agabian, N. (1983) Cell 34, 901-909.
10. Dorfman, D.M., and Donelson, J.E. (1984) Nucl. Acids Res. 12, 4907-4920.
11. Campbell D.A., Thornton, D.A., and Boothroyd, J.C. (1984) Nature 311, 350-355.
12. Milhausen, M., Nelson, R.G., Sather, S., Selkirk, M., and Agabian, N. (1984) Cell 38, 721-729.
13. Kooter, J.M., De Lange, T., and Borst P. (1984) EMBO Jour. 3, 2387-2392.
14. Nogueira, N., Bianco C., and Cohn, Z.A. (1975) J. Exp. Med. 142, 224-227.
15. Gonzalez, A., Prediger, E., Huecas, M., Nogueira, N., and Lizardi, P.M. (1984) Proc Natl. Acad. Sci. USA, 3356-3360.
16. Lizardi, P.M., Lerner, T.J., Gonzalez, A., and Nogueira, N. (1985) In Cold Spring Harbor Vaccines 85, R.A. Lerner, R.M. Chanock, and F. Brown, eds, pp. 67-70, Cold Spring Harbor Laboratory, New York.
17. Lizardi, P.M., Binder, R., and Short, S.A. (1984) Gene Anal. Tech. 1, 33-39.
18. Messing, J. and Vieira, J. (1982a) Gene, 19, 259-268.
19. Hanahan D. and Messelson, M. (1983) In Methods in Enzymology. R. Wu, L. Grossman, and K. Moldave, eds, Vol. 100, pp. 333-342, Academic Press, New York.
20. Sun, Y.L., Xu, Y.Z., and Chambon, P. (1982) Nucl. Acids Res. 10, 5753-5763.
21. Sakonju, S., Bogenhagen, D.F., and Brown, D.D. (1980) Cell 19, 13-25.
22. Messing, J. and Vieira, J. (1982b) Gene 19, 269-276.
23. Biggin, M.D., Gibson, T.J., and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3965.
24. Lizardi, P.M. (1983) In Methods in Enzymology, J.S. Fleischer and B. Fleischer, eds, Vol. 96, pp. 24-38, Academic Press, New York.
25. Tullis, R.H., and Rubin, H. (1980) Analyt. Biochem. 107, 260-264.
26. Thomas, P. (1980) Proc. Natl. Acad. Sci. USA 77, 5201-5205.
27. Lizardi, P.M. (1979) Cell 18, 581-589.
28. Ghosh, P.K., Reddy, V.B., Piatak, M., Lebowitz, P., and Weissman, S.M. (1980) In Methods in Enzymology, L. Grossman and K. Moldave, eds., Vol. 65, pp. 580-595, Academic Press, New York.

29. Maxam, A. and Gilbert, W. (1980) In *Methods in Enzymology*, L. Grossman and K. Moldave, eds, Vol. 65, pp. 499-560, Academic Press, New York.
30. Murphy, W.J., Brentano, S.T., Rice-Fitch, A.C., Dorfman, D., and Donelson, J. (1984) *J. Protozool.* 31, 65-73.
31. Thomashow, L.S., Milhausen, M., Rutter, W. and Agabian, N. (1983). *Cell* 32, 35-43.
32. Seebeck, T., Whittaker, P.A., Imboden, M.A., Hardman, N., and Braun, R. (1983) *Proc. Natl. Acad. Sci. USA*, 80, 4634-4638.
33. Landfear, S.M., McMahon-Pratt, D., and Wirth, D.F. (1983) *Mol. Cell. Biol.* 3, 1070-1076.
34. Huang, P.L., Roberts, B.E., McMahon-Pratt, D.M., David, J.R., and Miller, J.S. (1984) *Mol. Cell. Biol.* 4, 1372-1383.
35. Wang, A.H.-J., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G., and Rich, A. (1979) *Nature* 282, 680-686.
36. Nordheim, A., Pardue, M.L., Lafer, E.M., Moller, A., Stollar, B.D., and Rich, A. (1981) *Nature* 294, 417-422.
37. Lipps, H.J., Nordheim, A., Lafer, E.M., Ammerman, D., Stollar, D., and Rich, A. (1983) *Cell* 32, 435-441.
38. Krug, R.M., Broni, B.B., and Bouloy, M. (1979) *Cell* 18, 329-334.
39. Lazzirini, R.A., Keene, J.D., and Schubert, M. (1981) *Cell* 26, 145-154.
40. Spaan, W., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S. van der Zeijst, B., and Siddell, S.G. (1983) *Eur. Mol. Biol. J.* 2, 1839-1844.
41. Lai, M.M.C., Baric, R.S., Brayton, P.R., and Stohlman, S.A. (1984) *Proc. Natl. Acad. Sci. USA* 81, 3626-3630.
42. Plotch, S.J., Buoloy, M., Ulmanen, I., and Krug, R.M. (1981) *Cell* 23, 847-858.
43. Zimmerm, D. (1982) *Trends Biochem. Sci.* 7, 205-207.
44. Krieg, P.A., and Melton, D.A. (1984) *Nucl. Acids Res.* 12, 7057-7070.
45. Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L., and Steitz, J.A. (1980) *Nature* 283, 220-224.
46. Yang, V.W., Lerner, M.R., Steitz, J.A., and Flint, S.J. (1981) *Proc. Natl. Acad. Sci. USA* 78, 1371-1375.
47. Mount, S.M., Pettersson, I., Hinterberger, M., Karmas, A., and Steitz, J.A. (1983) *Cell* 33, 509-518.
48. Padgett, R.A., Mount, S.M., Steitz, J.A., and Sharp, P.A. (1983) *Cell* 35, 101-107.
49. Kramer, A., Keller W., Appel, B., and Luhrman, R. (1984) *Cell* 38, 299-307.