The EMBO Journal vol.7 no.2 pp.455-463, 1988

Polygene transcripts are precursors to calmodulin mRNAs in trypanosomes

Christian Tschudi and Elisabetta Ullu

Yale MacArthur Center for Molecular Parasitology, Department of Internal Medicine, Yale University School of Medicine, New Haven, CT 06510, USA

Communicated by N.J.Proudfoot

In African trypanosomes, calmodulin is encoded by a small family of tandemly repeated genes consisting of three to four units. We show that all the members of the calmodulin cluster of *Trypanosoma brucei gambiense* are expressed. In addition to mature mRNAs, steady-state RNA contains a small percentage of polygene transcripts which comprise at least two and probably all calmodulin genes. The 5′ ends of a portion of these molecules appear to be indistinguishable from those of mature calmodulin mRNAs. Polygene transcripts are not polyadenylated and have discrete ends which map in the intergenic regions downstream from the polyadenylation sites. Using biotinylated hybridization probes and selection of the hybrids on streptavidin–agarose, we further show that calmodulin polygene transcripts are the most abundant RNA species detected in pulse-labelled RNA of cultured procyclic trypanosomes. Our data strongly imply that polygene transcripts are authentic precursors to mature calmodulin mRNAs.

Key words: calmodulin gene cluster/3′ end formation/primary transcript/*Trypanosoma brucei*/biotin–streptavidin chromatography

Introduction

One of the many novel aspects of the biology of trypanosomes concerns the biosynthesis of mRNA (reviewed in Boothroyd, 1985; Borst, 1986). Every mRNA molecule yet examined is composed of two parts. A 35 nucleotide sequence lies at the very 5′ end of the mRNAs. This sequence, called the spliced leader (SL) or mini-exon sequence, is joined in a covalent fashion to the mRNA which carries the information for the synthesis of different proteins (e.g. see Boothroyd and Cross, 1982; Van der Ploeg *et al.*, 1982; De Lange *et al.*, 1984; Parsons *et al.*, 1984). The SL sequence initially forms the 5′ end of a discrete RNA of ~ 135 nucleotides, the SL RNA (Campbell *et al.*, 1984; Kooter *et al.*, 1984; Milhausen *et al.*, 1984). This SL RNA is encoded by a set of genes which is located on one or two chromosomes (De Lange *et al.*, 1983; Michiels *et al.*, 1983; Nelson *et al.*, 1983), while the protein coding genes are presumably dispersed throughout the genome. Recent evidence suggests that the addition of the SL sequence to the mRNA takes place by splicing *in trans* (Murphy *et al.*, 1986; Sutton and Boothroyd, 1986). At present we do not know the structure of the mRNA primary transcripts which are the substrates for *trans*-splicing and what genetic signals, protein cofactors and ribonucleoprotein particles are required for the process.

Another unknown aspect of mRNA biosynthesis in trypanosomes is the location and structure of promoter sequences. Most protein coding genes are present in multiple copies which are organized in tandem arrays and are separated from each other by short intergenic sequences of a few hundred nucleotides (Thomashow *et al.*, 1983; Clayton, 1985; Gonzalez *et al.*, 1985; Tschudi *et al.*, 1985; Michels *et al.*, 1986). Recently, Gonzalez *et al.* (1985) have presented evidence that a set of tandemly repeated genes in *Trypanosoma cruzi* generates transcripts larger than mature mRNA and which could conceivably encompass more than one coding region. Such polygene transcripts could originate from initiation of transcription at promoter sequences located outside the gene cluster. Monomeric transcripts would then be produced by *trans*-splicing and subsequent processing.

We have previously reported that *Trypanosoma brucei gambiense* contains three tandemly repeated calmodulin genes, which we refer to as genes A, B and C in a 5′ to 3′ direction (Tschudi *et al.* , 1985). The nucleotide sequences of the three translated regions are identical, while the sequences separating the protein coding regions differ only at three positions. In the present study, we describe the detailed analysis of the transcripts derived from the calmodulin locus of *T.brucei*. Our results show that all the members of the calmodulin gene family are transcribed. We present evidence that transcription of the calmodulin genes generates polygene transcripts. These molecules have discrete 3′ ends which extend downstream from the polyadenylation sites and might derive from a termination or an RNA processing event within the intergenic regions.

Results

The basic structure of the calmodulin genes in *T.brucei* is illustrated in Figure 1. Using partial genomic digestions and Southern hybridization with sequences flanking the calmodulin genes, we have found that *T.b.gambiense* variant TXTaT 1.0 has two chromosomal loci encoding calmodulin (C.Tschudi, in preparation). The two gene clusters differ only in the number of genes: locus I and II contain four and three calmodulin coding regions, respectively (Figure 1A). This organization is not present in an isolate of the *T.brucei* subspecies, *Trypanosoma brucei rhodesiense*, which contains four genes at each locus. The physical map of the fourth gene, B′ is indistinguishable from that of gene B, but we have not established the identity of the B′ gene at the nucleotide level. Throughout this manuscript, we will refer to both as B genes. Although most of the experiments presented here were carried out with RNA isolated from procyclic forms of *T.b.rhodesiense* (which can be easily grown and labelled in culture), we have obtained similar results using bloodstream forms of *T.b.gambiense* as a source of

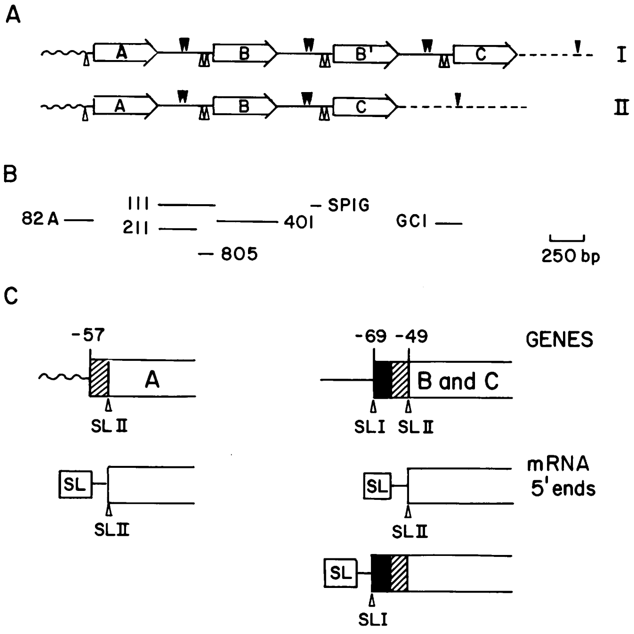

Fig. 1. (**A**) Schematic structure of the calmodulin locus in *T.brucei*. Two chromosomal loci are shown which contain four (I) and three (II) calmodulin genes, respectively. The protein coding regions are indicated by open boxes and the arrows denote the direction of transcription. The repeating unit of the A and B genes is 843 bp and begins 57 nt upstream from the AUG of gene A (Tschudi *et al.*, 1985). In contrast, the gene C coding region is present on a truncated version of this repeating unit. This is because immediately downstream of the C gene termination codon the nucleotide sequence is different from that of the corresponding region of the A and B genes. Sequences upstream of gene A are shown as wavy lines and sequences downstream of the termination codon of gene C are represented as broken lines. Filled arrowheads indicate poly(A) addition sites; sites for the addition of the spliced leader are indicated by open arrowheads. The gene cluster is shown in the same scale as the probes in (B). (**B**) The probes used for Northern hybridization and RNase protection experiments. The position of the various subcloned regions are represented by solid lines and are aligned with the lower one of the two chromosomal loci shown in (A). Refer to Materials and methods for a description of the construction of the probes. (**C**) Comparison of the upstream regions of the calmodulin genes and structure of the mRNA 5′ ends. Homologous sequences are represented by identical shading. Sequences upstream of the calmodulin genes are homologous up to nucleotide −57 from the AUG initiation codon. This position marks the beginning of the repeating unit of the calmodulin genes. The calmodulin B and C genes have two sites for the addition of the spliced leader sequence, one at position −69 (SL I) and the other at position −49 (SL II) relative to the AUG initiation codon; the A gene has only one such site at position −49 (SL II). Therefore, the mRNAs derived from gene A have the SL sequence joined only at the SL site II, while the mRNAs from the B and C genes have the SL sequence joined either at the SL site I or at the SL site II. SL indicates the 35-nt SL sequence.

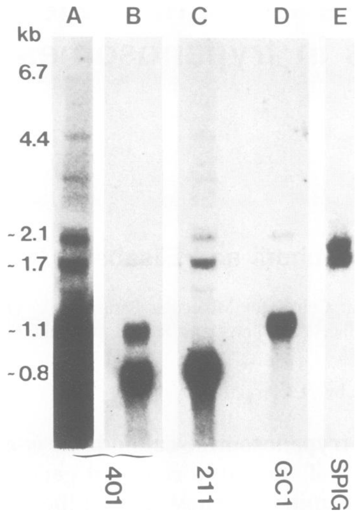

Fig. 2. Northern hybridization of trypanosome RNA. Total RNA (20 μg) was fractionated by formaldehyde−agarose gel electrophoresis and transferred to a nitrocellulose filter. Except for **lane E**, identical strips of the same filter were hybridized to gene-specific ^{32}P-labelled RNA probes. **Lanes A** and **B** show the hybridization with 401 which is complementary to the translated region of all calmodulin mRNAs. Two different exposure times are shown. **Lane C** was hybridized with probe 211 which is specific for the A and B genes. **Lane D** was hybridized with the gene C specific probe GC1. **Lane E** shows the hybridization with the intergenic probe SPIG. The two hybridizing bands correspond to the 2.1- and 1.7-kb transcripts. The approximate sizes in kilobases of the hybridizing bands were calculated using λ/*Hin*dIII DNA fragments and the trypanosome rRNAs as size markers. The positions of two λ/*Hin*dIII DNA fragments (6.6 and 4.4 kb) are also indicated.

RNA. Unfortunately, we have not yet been able to obtain procyclic culture forms with our *T.b.gambiense* strain.

All the members of the calmodulin gene cluster are expressed

We used two observations to distinguish between the transcripts derived from the various genes. First, the sequence immediately downstream of the termination codon of gene C is different when compared to that of the corresponding region of genes A and B (Tschudi *et al.*, 1985; Figure 1A). Thus, the 3′ untranslated sequence of gene C was expected to be diagnostic for the expression of this gene. Second, the nucleotide sequences separating the translated region of genes A, B and C differ at three positions (Tschudi *et al.*, 1985) and preliminary S1 mapping experiments indicated that the 3′ ends of mature calmodulin mRNAs included these variable nucleotides. Using this information, we analysed the sequence of calmodulin cDNA clones from *T.b.gambiense* and found that all three genes are expressed (see Materials and methods). Two cDNA clones were derived from gene A, one from gene B and one from gene C. Since the four cDNA clones ended with a long poly(A) track, we were able to map precisely the poly(A) addition sites used by the various genes. Two such sites were identified for the A and B genes at positions 645 and 663, while the 3′ end of the gene C mRNA mapped at position 910 relative to the respective AUG initiation codons. Thus, we predict that the mRNA from gene C should be ~250 nucleotides longer than the A and B mRNAs. In addition, these structural data showed that the intergenic region separating two adjacent calmodulin genes is a mere 110 nucleotides (Figure 1A).

The size of the calmodulin transcripts was analysed by Northern blotting using as probes antisense RNAs specific for various portions of the gene cluster (see Figure 1B). Probe GC1, which is derived from the 3′ untranslated region of gene C, identifies a major RNA species of ~1.1 kb (Figure 2, lane D). This is in good agreement with the mol. wt we calculated from the corresponding cDNA clone. The mature mRNAs from the A and B genes are ~800 nucleotides long and are specifically identified by probe 211 (Figure 2, lane C). Both the 1.1-kb and 0.8-kb RNA species hybridize to the calmodulin coding region probe 401, thus confirming their identity (Figure 2, lanes A and B).

The intergenic region is transcribed

The Northern blot analysis, shown in Figure 2, also reveals several higher mol. wt RNA species. In particular, two RNAs of ~2.1 kb and 1.7 kb caught our attention. The fact that the 2.1-kb RNA hybridizes to the gene-C-specific probe (lane D) as well as to the gene-A- and B-specific probes (lane C) suggested the possibility that these transcripts contain two calmodulin coding regions. This is further supported by the observation that probe SPIG, which contains sequences from the intergenic region not present in mature mRNA, specifically hybridizes to the 2.1-kb as well as to the 1.7-kb RNA (Figure 2, lane E).

To obtain a detailed structure of transcripts derived from the calmodulin locus and to confirm that the intergenic region is expressed, we used RNase protection assays in combination with S1 nuclease mapping. Figure 3B displays the results of an RNase mapping experiment using total trypanosome RNA and the three different antisense RNAs diagrammed in Figure 3A. When the 5′ ends of calmodulin RNAs are analysed with probe 805, two prominent RNA fragments, labelled SL I (110 nt) and SL II (90 nt), are protected (lanes 2 and 5). This is in agreement with our previous finding that mature mRNAs have heterogeneous 5′ ends due to the use of two different sites for the joining of the SL sequence. Gene A has one site at position −49 (SL II) and genes B and C have two sites, one at position −69 (SL I) and the other at position −49 (SL II) relative to the respective AUG initiation codons (Tschudi *et al.*, 1985; see also Figure 1C). A third prominent RNA, labelled X, is also protected by the 805 probe. By priming poly(A)$^+$ RNA with a calmodulin-specific oligonucleotide, we find an mRNA with several nucleotide substitutions in the 5′ untranslated sequence of the calmodulin mRNA, between nucleotides −35 to −55, relative to the AUG initiation codon (Tschudi *et al.*, 1985 and unpublished observation). This mRNA could give rise to the protected RNA fragment labelled X. At present we do not know whether this mRNA represents a true calmodulin mRNA or the product of a non-calmodulin gene with partial sequence homology. In addition, the 805 probe generates two protected RNA species (one indicated by an arrow and the other by 100) derived from transcripts with a lower abundance than mature calmodulin mRNA. The longer of the two fragments has the size expected from full protection of the probe. The size of the other one (100 nt), together with the observation that this RNA is also generated by probe 111 (lanes 7 and 11) suggests that there are RNA molecules with 5′ ends mapping between the two SL sites of the B and C genes. Since the 5′ ends of these RNA molecules coincide with the 5′ border of the calmodulin repeating unit (see Figure 1C), the 100 nt protected fragment might also derive from transcripts initiating upstream of gene A. Given the repetitious nature of the calmodulin genes, we cannot distinguish between these two possibilities. In summary, most calmodulin transcripts have the 5′ ends characteristic of mature mRNA and a minority of transcripts have extensions at the 5′ ends.

Probe 211 identifies the 3′ ends of mature mRNAs derived from the A and B genes (Figure 3B, lanes 12 and 15). The sizes of the two major clusters of protected RNA fragments (185 and 165 nucleotides) are in agreement with the positions of the two major polyadenylation sites deduced by sequencing calmodulin cDNA clones (Figures 1A and 3A). In addition, there are transcripts extending downstream from the polyadenylation sites of the A and B genes, since we observe full protection of the 211 probe (indicated by an arrow).

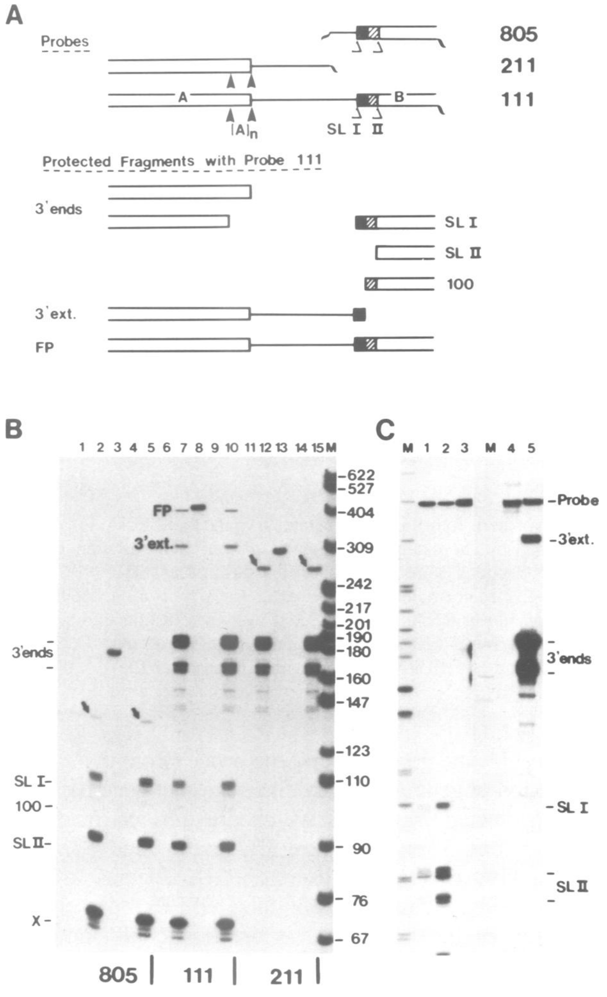

Fig. 3. RNase protection and nuclease S1 mapping of total trypanosome RNA. (**A**) Structure of the three antisense RNA probes 805, 211 and 111, and of the protected fragments with probe 111. Calmodulin coding regions of the A and B genes are shown as boxes and the solid line indicates the intergenic region. Solid arrowheads indicate the poly(A) addition sites of the calmodulin A and B genes. The sites for spliced leader addition (SL I and II) are represented as open arrowheads. The sequences between the SL I and SL II sites are diagrammed as in Figure 1C. The wavy lines represent vector sequences. (**B**) RNase mapping of calmodulin transcripts. **Lanes 1−5**, probe 805; **lanes 6−10**, probe 111; **lanes 11−15**, probe 211. 10 μg of total RNA were hybridized to 15 000 c.p.m. (**lanes 2, 7, 12**) or 45 000 c.p.m. (**lanes 5, 10, 15**) of probe and digested with ribonuclease as described in Materials and methods. Control lanes without trypanosome RNA for the two probe concentrations are shown in **lanes 1, 6** and **11**, and in **lanes 4, 9**, and **14**, respectively. **Lane 3, 8**, and **13** show the undigested probes. Solid arrows point to fragments derived from full protection of the 805 and 211 probes. 3′ ext. indicates the 300-nt RNA. **Lane M**, 3′ end-labelled *Hpa*II digested pBR322 fragments. SL I and SL II, fragments extending up to the SL addition sites I and II, respectively; 3′ ends, fragments originating from the mature 3′ ends of the A and B genes; 100, protected fragment from the upstream region common to the A, B and C genes. (**C**) S1 analysis of trypanosome total RNA. 111 DNA was labelled at the 5′ end (**lanes 1−3**) or at the 3′ end (**lanes 4** and **5**), annealed to 10 μg of RNA (**lane 1**) or 20 μg of RNA (**lanes 2** and **5**), and digested with S1 nuclease as described in Materials and methods. **Lanes 3** and **4** show the control reactions with no RNA added. Size marker and symbols as defined in (B).

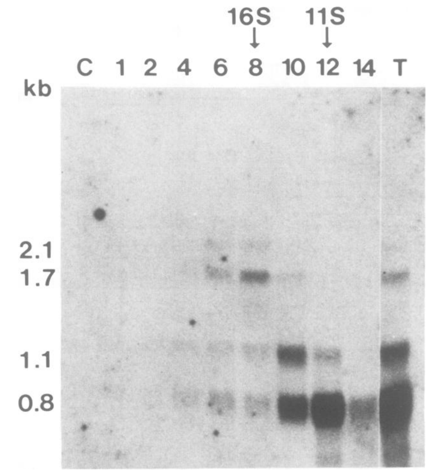

Fig. 4. Northern hybridization of total trypanosome RNA fractionated by sucrose-gradient centrifugation. RNA aliquots of the indicated fractions were electrophoresed through a 1.2% agarose−2.2 M formaldehyde gel. After transfer to a nitrocellulose filter, calmodulin RNAs were detected by hybridization to antisense 401 RNA. **Lane T**, 20 μg unfractionated total RNA; **lane C**, 20 μg unfractionated total RNA digested with RNase prior to electrophoresis.

Probe 111, the gene-spacer-gene probe (from the untranslated region of gene A to the coding region of gene B), shows major protected fragments which are derived from both mature 5′ and 3′ ends (Figure 3B, lanes 7 and 10). Most interestingly, two other discrete RNA species are also detected. The longest fragment (labelled FP) represents full protection of the probe and is diagnostic of calmodulin transcripts connecting two gene units. To map the ends of the other protected RNA species (labelled 3′ ext), we carried out S1 nuclease protection experiments (Figure 3C) with 111 DNA labelled either at the 5′ end (lanes 1 and 3) or at the 3′ end (lanes 4 and 5). After hybridization to total trypanosome RNA and S1 digestion, the 5′ end-labelled probe generates the protected fragments diagnostic of the 5′ ends of mature calmodulin mRNAs (SL I and SL II). (S1-protected fragments derived from mRNAs initiating at the SL II site are heterogeneous in size, probably because of trimming of the ends of S1 nuclease.) S1 analysis with 3′ end-labelled 111 DNA (lane 5) generates the fragments characteristic of mature 3′ ends (labelled 3′ ends) and also gives rise to a protected fragment (indicated 3′ ext.) whose length (300 nt) agrees precisely with the length of the RNA fragment observed with 111 RNA in the RNase protection experiments (Figure 3B, lanes 7 and 10). This showed that a proportion of the calmodulin transcripts extend beyond the positions of the 3′ ends of the A and B mRNAs and have discrete ends. By comparing a number of different experiments and using the non-sequence-specific ribonuclease T_2 for RNase mapping, we can position the 3′ end of these transcripts 115 to 120 nucleotides downstream of the nearest polyadenylation site, that is between the two SL joining sites of the B and C genes.

Calmodulin polygene transcripts

The data presented so far show that in trypanosome RNA there exist transcripts with 3′ extensions and transcripts comprising two calmodulin gene units. To obtain conclusive evidence that these RNAs are longer than mature mRNAs and have the structure of polygene transcripts, total trypanosome RNA was fractionated in two dimensions; first by sucrose density gradient centrifugation and second by analysing individual fractions by Northern blotting and RNase mapping. Figure 4 displays the results of the Northern blot analysis using the calmodulin coding region probe 401, which will visualize the products of transcription of all genes. By this analysis, mature calmodulin mRNAs (1.1 kb and 0.8 kb) can be effectively separated from longer transcripts. In particular, the 1.7- and 2.1-kb RNA species, previously detected in total RNA (see Figure 2), are specifically enriched in fractions with a sedimentation value of ~ 16S (peak at fraction 8). When the RNA from the same gradient fractions is analysed by RNase mapping with the gene-spacer-gene probe 111 (Figure 5A), two protected fragments can be seen which are specifically enriched in the heavier portions of the gradient. These are derived from polygene transcripts (labelled FP) and from transcripts with 3′ extensions (labelled 3′ ext.). In addition, the original autoradiogram also showed enrichment for the 100 nt RNA species which did not reproduce in Figure 5A, but can be seen in Figure 6, lane 2. In the experiment shown, the amount of the fully protected RNA is lower than that of the 3′ extended RNA fragment. However, we believe that this does not reflect the actual abundance of these transcripts in the cell. Since we have observed that the relative amount of fully protected fragments is somewhat variable from experiment to experiment, we argue that this is an artefact due to some variability in the conditions employed for RNase mapping.

To detect transcripts initiating upstream of calmodulin gene A, we assayed the gradient fractions with probe 82A which is complementary to the 5′ end of the gene A coding region and to sequences upstream to position −200 (Figure 5B). The RNA species labelled FP in Figure 5B represents full protection of the 82A probe and is indicative of transcripts initiating at least 200 nt upstream of the coding region of gene A. Again these transcripts are specifically enriched in the heavier region of the gradient demonstrating that they are larger than mature mRNA. Similar results were obtained with the GC1 probe (data not shown) indicating that transcription proceeds downstream from the polyadenylation site of gene C.

In addition, RNase mapping with the 82A probe generates fragments characteristic of the 5′ ends of mature calmodulin mRNAs (labelled SL I* and SL II). The SL II fragments represent the 5′ ends of mRNA molecules with the SL at site II (see Figures 3B and 5A). On the other hand, the SL I* RNA fragments are ~ 10 nt shorter than the SL I fragments obtained with probe 805 and 111 (Figures 3B and 5A). This is because the calmodulin mRNAs with the SL at site I are entirely derived from the B and C genes and probe 82A is complementary to the 5′ ends of these mRNAs only up to position −57 (see Figure 1C). The fact that in this experiment the SL I* fragments are heterogeneous in size could be due to differential trimming of the ends by RNase A, since this problem can be alleviated in part by digesting the RNA duplexes at 0°C (unpublished observation).

Taken together these observations strongly argue that the calmodulin genes are part of a large transcription unit. By excising and counting the various protected fragments from

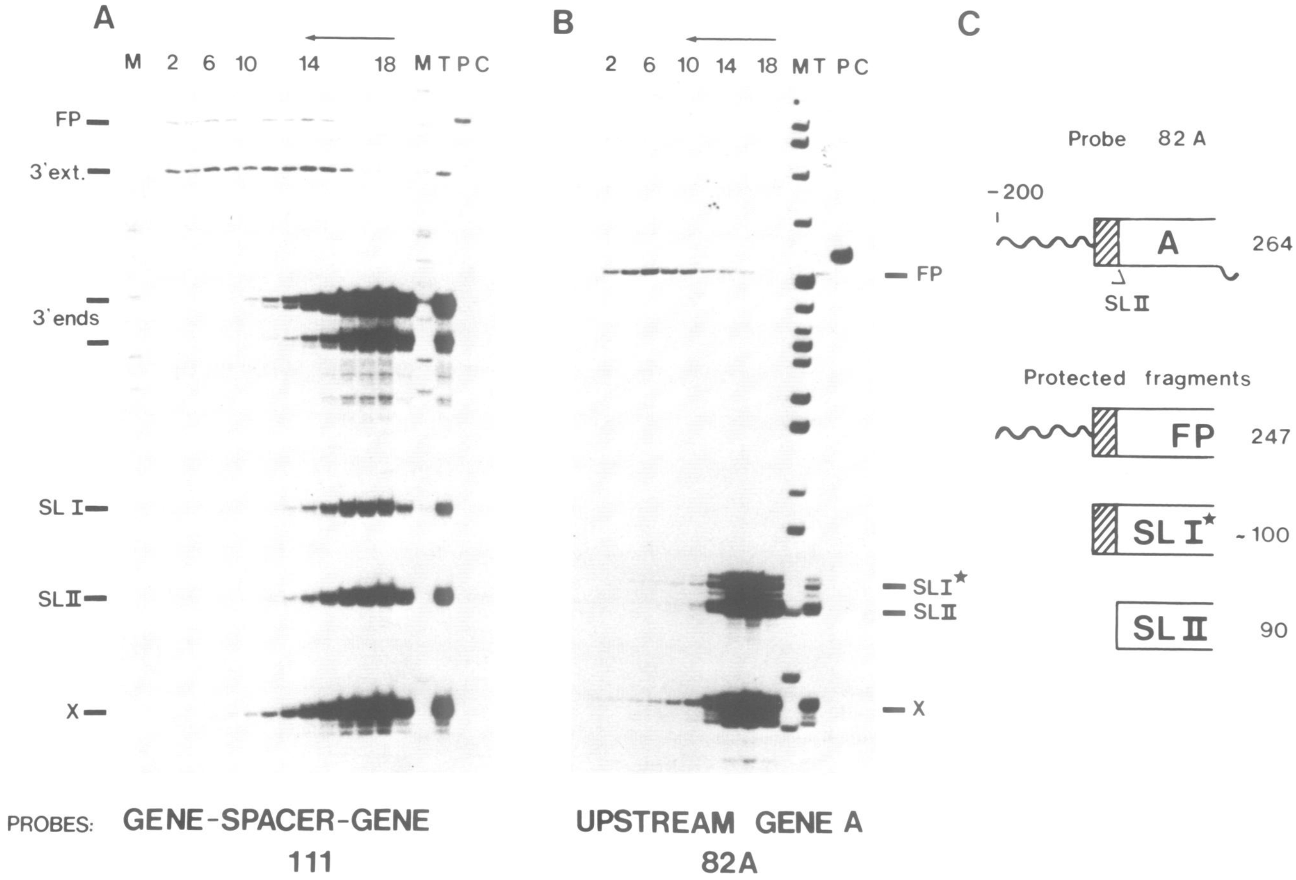

Fig. 5. RNase mapping of total trypanosome RNA fractionated by sucrose-gradient centrifugation. RNAs from selected fractions as indicated were analysed with probe 111 (**A**) and probe 82A (**B**). Panel **C** shows a schematic representation of probe 82A and the protected fragments observed in panel B. Graphic symbols are as described in the legend in Figure 1C. Bars indicate the various protected fragments described in Figure 3. FP represents full protection of the probe; **lane T**, 10 μg unfractionated total RNA; **lane P**, undigested probe; **lane C**, control lane with no RNA added. The arrows indicate the direction of sedimentation.

a number of different RNase mapping experiments, we estimate that in steady-state RNA ~3–5% of the calmodulin transcripts are larger than mature mRNA.

Do polygene transcripts have mature 3′ and 5′ ends?

The RNase protection analysis shown in Figure 5A suggests that polygene transcripts end with 3′ extensions and do not have mature 3′ ends. This is further supported by the finding that the majority of these RNAs fail to bind to oligo(dT)–cellulose (data not shown). To confirm this point and to obtain information about the structure of the 5′ ends, we prepared an RNA fraction enriched for molecules larger than mature mRNAs by three cycles of sucrose density gradient centrifugation combined with oligo(dT)–cellulose chromatography. These RNAs were then analysed by RNase mapping using the gene-spacer-gene probe (Figure 6 but see Figure 3A for the structure of the probe and of the protected fragments) and the relative amounts of the various protected fragments obtained with the enriched RNA fraction (lane 2) and with total RNA (lane 3) were compared. Similar to what we observe in the heavier portion of the sucrose gradient shown in Figure 5A, the following RNAs are specifically enriched in the size-selected poly(A)$^-$ RNA fraction: polygene transcripts (as indicated by full protection of the probe, FP), transcripts with 3′ extensions (labelled 3′ ext.) and transcripts which give rise to the 100-nt RNA species (100). As indicated by the presence of low amounts of mature 3′ end fragments (labelled 3′ ends), the RNA preparation still contains some residual calmodulin mRNA. Therefore, the mature 5′ end fragments (labelled SL I and SL II) observed with the selected RNA fraction could be derived either exclusively or only in part from the mature mRNA. To estimate the proportion of 5′ end fragments contributed by calmodulin mRNA, we asked what would be the intensity of 5′ end fragments derived from the mRNA present in our selected RNA fraction. To this end we obtained an exposure of lane 3 (which displays the RNase protected fragments of total RNA) that would show the mature 3′ end fragments with an intensity similar to that of the corresponding fragments in the selected RNA fraction (lane 2). In this exposure (lane 1) the 5′ end fragments are much less intense than in lane 2. This shows that relative to the total RNA the ratio of the signal of mature 5′ ends to mature 3′ ends increases in the enriched RNA sample. This is best accounted for by the existence of polygene transcripts with mature 5′ ends, but not with mature 3′ ends.

Newly made calmodulin transcripts

To support the view that polygene transcripts are authentic intermediates in the biosynthesis of calmodulin mRNAs, we

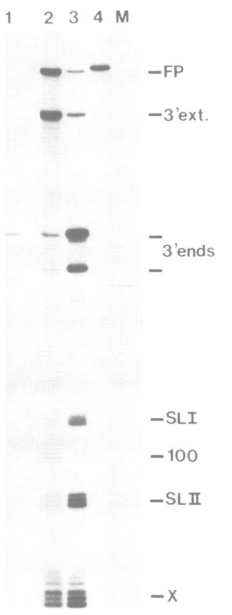

Fig. 6. RNase mapping with probe 111 of poly(A)$^-$ RNA enriched for trypanosome sequences longer than mature mRNA. **Lane 2**, 10 μg enriched RNA; **lane 3**, 10 μg total RNA; **lane 4**, undigested 111 probe; **lane 1**, shorter exposure of lane 3.

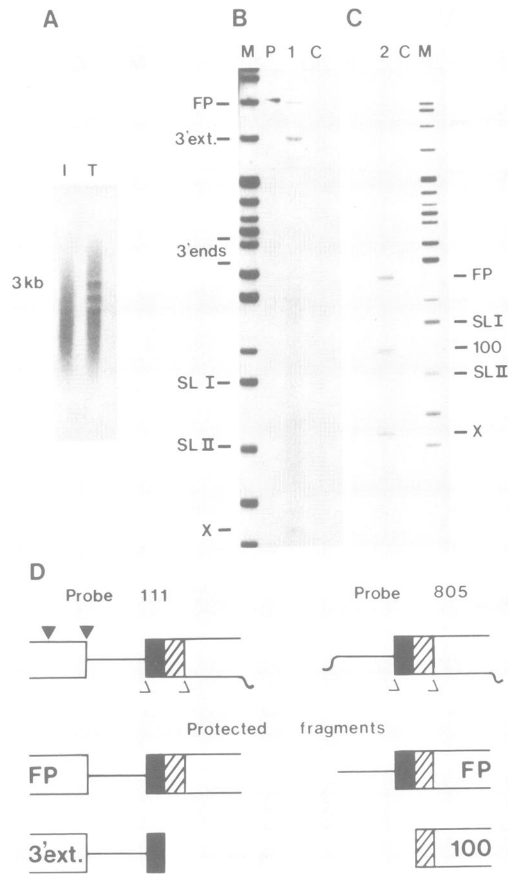

Fig. 7. Analysis of ^{32}P-labelled newly made trypanosome RNA. (**A**) Size distribution of *in vivo* labelled RNA fractionated by agarose–formaldehyde gel electrophoresis. **Lane T**, 10 000 c.p.m. of total RNA; **lane I**, 10 000 c.p.m. of RNA enriched for calmodulin sequences with biotinylated 111 RNA. (**B**) and (**C**) RNase mapping of calmodulin-enriched [^{32}P]RNA with biotinylated 111 probe (**lane 1**) and biotinylated 805 probe (**lane 2**). **Lane P**, ^{32}P-labelled 111 RNA; **lane C**, the same hybridization mixture used in **lanes 1** and **2** but denatured prior to RNase digestion. The expected positions of protected fragments characteristic of mature 3′ ends (3′ ends) and mature 5′ ends (SL I and SL II) are indicated by bars. Panel **D** shows a diagram of the structures of probes 111 and 805 and the corresponding protected fragments observed in panels B and C. Graphic symbols are identical to those used in Figure 1C but the sizes of the probes are not on scale.

decided to analyse the amount of readthrough transcription using pulse-labelled RNA in RNase protection experiments. To obtain [^{32}P]RNA of high sp. act. we made cultured trypanosome cells permeable by brief exposure to the detergent lysolecithin, followed by incubation for 10 min in transcription buffer containing [α-^{32}P]GTP as described by Miller *et al.* (1978) and Contreras and Fiers (1981). The size distribution of the ^{32}P-labelled RNA is shown in Figure 7A. Because the use of total ^{32}P-labelled RNA in RNase protection experiments produced a very high background, we first enriched for newly made calmodulin transcripts by hybridization to a biotinylated antisense RNA probe followed by affinity chromatography on streptavidin–agarose. In the experiment shown in Figure 7B, we used the gene-spacer-gene probe for the enrichment step and subsequent RNase mapping. Three major labelled RNA fragments are protected from RNase digestion (Figure 7B, lane 1). Full protection of the probe (labelled FP) demonstrates that transcription of the calmodulin genes proceeds uninterrupted through the intergenic region. The protected RNA of 300 nucleotides (labelled 3′ ext.) has the same electrophoretic mobility as the 3′ extension product we analysed in steady-state RNA. The diffuse RNA band of ~270 nucleotides has not yet been analysed further. Most interestingly, no RNA fragments with the sizes expected from mature 5′ and 3′ ends are detected. However, since there is some background in the region of the gel where the mature 5′ end fragments would migrate, we repeated the experiment with the 5′ end-specific probe 805 (Figure 7C). Two major calmodulin-specific RNA fragments of 135 nt (labelled FP) and 100 nt are detected in approximately equimolar amounts. The larger one represents full protection of the probe (see for a comparison, Figure 3B, lane 2), while the smaller one is most likely equivalent to the 100-nt RNA detected in size-selected poly(A)$^-$ RNA (see Figure 6, lane 2). Also by this analysis we fail to detect any protected fragments diagnostic of mature 5′ ends.

Discussion

We report here on the structure of the transcripts from the calmodulin locus. First we show that all genes of the calmodulin cluster are expressed. Second, we show that the calmodulin genes are part of a large transcription unit. We have identified transcripts which connect at least two genes and transcripts which extend upstream from gene A and downstream from the polyadenylation site of gene C. The 1.7- and 2.1-kb RNA species which are detected by Northern blotting (Figure 2) are likely to represent dimeric RNAs. The 1.7-kb RNAs can conceivably extend from gene A to gene B or connect the B genes. On the other hand, the structure of the 2.1-kb transcript is consistent with it being a dimer of the B and C genes. We also observe calmodulin transcripts several kilobases in length: such molecules could span the entire gene cluster. These long calmodulin RNAs have defined ends, suggesting that they represent specific products of transcription, of processing or of both. The 3′ ends of the transcripts encompassing the A and B genes map downstream from the polyadenylation sites between the two SL addition sites of the B and C genes. Instead, the 5′ ends of at least a proportion of these molecules appear to be indistinguishable from those of mature mRNA, suggesting that they have the SL sequence already joined to them. Are these molecules authentic precursors to mature calmodulin mRNAs or are they dead-end products of transcription and/or processing? Our claim that polygene transcripts are indeed precursors to mature mRNA is based on the observations that newly made RNA contains a great proportion of molecules spanning the A and B genes and that no mature 5′ and 3′ ends are detected. Unfortunately, we have been unable to perform kinetic studies in detergent-treated trypanosome cells because transcription ceases after 10 min due to the high activity of endogenous ATPases which hydrolyse ATP to AMP and deplete the ATP pool even in the presence of an ATP regeneration system (unpublished observation).

The RNase mapping experiments with newly made RNA failed to detect any protected fragments diagnostic of mature 5′ and 3′ ends suggesting that the 5′ and 3′ ends of mature calmodulin mRNAs are generated by post-transcriptional events. Splicing *in trans* of the SL sequence onto the pre-mRNA body most likely generates the 5′ end of calmodulin mRNAs. The mechanism that forms 3′ ends, on the other hand, remains unknown. No canonical eukaryotic polyadenylation signal, AAUAAA (Proudfoot and Brownlee, 1976), or any other recognizable conserved sequence appears to be present upstream or downstream from the polyadenylation sites of trypanosome mRNAs.

An intriguing observation common to both steady-state and newly made calmodulin RNA is the presence of transcripts extending downstream from the poly(A) addition site of the A and B genes. In steady-state RNA the majority of these molecules appear to be longer than calmodulin mRNA. This conclusion is based on the Northern hybridization with the intergenic probe SPIG (Figure 2, lane E), which predominantly detects putative calmodulin dimers and on the RNase mapping of gradient-fractionated RNA with probe 111 (Figure 5A). The 3′ ends of these 3′-extended transcripts define the primary site of 3′ end formation in the calmodulin gene cluster, since no other 3′ ends can be identified in newly made calmodulin RNA. We can think of two possible mechanisms which might generate transcripts with 3′ extensions: (i) transcription termination of RNA polymerase II at the site of 3′ end formation; (ii) endonucleolytic cleavage of larger transcripts that extend downstream from the site of 3′ end formation. Our data do not allow us to distinguish between these possibilities. This is because of our inability to identify precisely the origin of the 100-nt RNase-protected fragment detected in newly made RNA. At least part of this RNA species is derived from transcripts initiating upstream from calmodulin gene A. However, we cannot exclude the possibility that this fragment also originates from transcripts with the 5′ ends within the intergenic region, a few nucleotides from the site of 3′ end formation. Such molecules might result from endonucleolytic cleavage of polygene transcripts at the primary site of 3′ end formation. Alternatively, termination and subsequent re-initiation of transcription might occur in this region. The latter situation would be similar to what is observed in the transcription of the mouse and frog RNA genes (Grummt *et al.*, 1986; Henderson and Sollner-Webb, 1986; McStay and Reeder, 1986). Whatever the mechanism, it is puzzling that the primary site of 3′ end formation in the calmodulin A and B genes lies almost precisely at the border of the repeating unit. It is possible that the signals responsible for this phenomenon were part of the ancestral calmodulin gene and were subsequently amplified in the formation of the cluster.

In African trypanosomes, primary transcripts and putative promoter elements for housekeeping genes have been elusive so far. Recent evidence suggests that trypanosome genes are part of transcription units much larger than a single gene. This conclusion is based on the hybridization of nuclear run-on transcripts to cloned genes which revealed that the transcribed area can cover several kilobases upstream from the structural gene region (for review see Borst, 1986). However, these studies do not provide evidence for a single continuous transcript. Rather they show that these regions are actively transcribed. The results presented here are consistent with a transcription mechanism that generates primary transcripts comprising more than one gene unit. Although other interpretations are possible, we favor the possibility that transcription begins upstream of calmodulin gene A and proceeds through the cluster. Using polygene calmodulin transcripts as substrates, *trans*-splicing with the leader sequence would then generate the 5′ ends of mature calmodulin mRNAs. In this scenario, we can envisage *trans*-splicing as a mechanism to edit mRNA coding sequences from complex transcription units. The use of RNA processing rather than transcription initiation to generate the 5′ end of mRNA might have spared trypanosome cells from evolving unique promoter sequences upstream from each coding region.

Materials and methods

Trypanosomes

The cloned *T.b.gambiense* antigenic variant type 1 was used in this study (Merritt *et al.*, 1983). The procyclic trypanosome forms of the YTaT strain *T.b.rhodesiense* were obtained from L.Ruben and grown at 28°C in SM medium (Cunningham, 1977), supplemented with 20% (v/v) heat-inactivated fetal calf serum and 25 mM Hepes.

Construction of plasmids

The previously described calmodulin genomic clone λCM51 (Tschudi *et al.*, 1985) was used to generate convenient restriction fragments for hybridizations and nuclease protection experiments. All plasmids were constructed in the vectors pSP64 and pSP65 (Melton *et al.*, 1984) and the restriction fragments were inserted in such an orientation that SP6 transcription gives an RNA which is complementary to calmodulin RNA.

To construct plasmid 82A the *Apa*I site at position −210 relative to the AUG initiation codon of gene A was changed into a *Bam*HI site. The *Bam*HI−*Eco*RI fragment of 250 bp, containing 210 bp upstream of gene A and 38 bp of translated sequence, was then inserted into the corresponding sites of pSP65. Plasmid 401 contains the 452-bp *Eco*RI−*Hin*dII fragment, encoding 417 bp of calmodulin translated sequence as well as 35 bp of 3′-untranslated sequence. Plasmid 211, obtained by subcloning a *Hin*dIII−*Bgl*II fragment, contains 185 bp of 3′-untranslated region of gene A and 85 bp of intergenic sequence. Plasmid 111 contains gene A coding sequence, the intergenic region and gene B coding sequence inserted into pSP65 between the *Eco*RI and *Hin*dIII sites. The insert in plasmid SPIG encompasses 70 bp of the intergenic region between an *Rsa*I site, 10 bp downstream of the second poly(A) addition site, and a *Bgl*II site located 30 bp upstream of the first site of SL addition. The insert in GC1 spans the polyadenylation site of gene C between a *Sca*I site and a *Sna*BI restriction site and contains 176 bp of 3′-untranslated sequence as well as 35 bp of flanking sequence. Construction of the plasmids was verified by DNA sequence analysis. More detailed information about the various constructs can be obtained upon request.

RNA manipulations

RNA was isolated from trypanosome cells using the urea−LiCl method (Auffray and Rougeon, 1980). To remove contaminating DNA, RNA samples were digested with 100 μg/ml RNase-free DNase in the presence of placental ribonuclease inhibitor for 30−60 min at 37°C. For Northern blot analysis, RNAs were fractionated by electrophoresis through a 1.2% agarose gel in the presence of 2.2 M formaldehyde. The same concentration of formaldehyde was added to the electrophoresis buffer. Prior to electrophoresis, samples were incubated at 70°C in sample buffer (Maniatis *et al.*, 1982) for 15 min. Under these conditions, and using formaldehyde solutions free of any precipitate, we observe complete denaturation of lambda DNA restriction fragments run in parallel. After electrophoresis the RNA was transferred to a nitrocellulose filter. Hybridizations were carried out at 50−56°C in the presence of 50% formamide plus standard components. After hybridization the nitrocellulose filter was washed at 65°C in 300 mM NaCl, 30 mM Na citrate and 1% SDS. To remove background due to unspecific binding of the probe to the large rRNAs, the nitrocellulose filter was treated with a solution containing 20 μg/ml RNase A in 300 mM NaCl, 30 mM Na citrate at room temperature for 5−10 min and then washed several times in the same solution minus the RNase A.

Total RNA was fractionated by centrifugation through 5−20% sucrose density gradients containing 20 mM Na acetate (pH 5.0) and 1 mM EDTA. Prior to centrifugation, the RNA samples were incubated in 1 mM EDTA at 65°C for 10 min in order to dissolve aggregates. Gradients were centrifuged in the Sorvall rotor TH641 for 16 h at 30 000 r.p.m. at 4°C. Cytoplasmic RNA from HeLa cells was run through a parallel gradient to provide sedimentation standards.

Synthesis of ^{32}P-labelled RNA probes and RNase protection analysis was done essentially according to Melton *et al.* (1984). Incubations with RNase A at 20 μg/ml and RNase T_1 at 2 μg/ml were carried out at 30°C for 60 min. At the end of the incubation, samples were treated with proteinase K (100 μg/ml) in the presence of 1% SDS for 10 min at 30°C. After ethanol precipitation, protected fragments were fractionated on 6% polyacrylamide/7 M urea gels.

For S1 nuclease analysis plasmid 111 was labelled at the *Hin*dIII site using T4 DNA polymerase and [α-^{32}P]dCTP (Maniatis *et al.*, 1982) or at the *Eco*RI site using T4 polynucleotide kinase and [γ-^{32}P]ATP (Maniatis *et al.*, 1982). After end-labelling, the insert DNA was digested with *Eco*RI or *Hin*dIII, respectively. Double-stranded fragments were purified by electrophoresis through a native 4% polyacrylamide gel. Hybridization to total RNA and digestion with S1 nuclease were carried out as described by Hernandez (1985) except that the hybridization temperature was 45°C.

Isolation of cDNA clones

A cDNA library was constructed from poly(A)$^+$ RNA isolated from *T.b.gambiense* cells following the protocol of Huynh *et al.* (1985). Briefly, RNA was copied into double-stranded cDNA with avian myeloblastosis virus reverse transcriptase. Blunt-ended molecules were ligated to synthetic *Eco*RI linkers and size selected cDNA fragments were cloned into the *Eco*RI site of λ*641* (Murray *et al.*, 1977). The cDNA library was screened with a ^{32}P-labelled SP6 RNA probe (Benton and Davis, 1977) which is complementary to the translated region of the calmodulin gene (plasmid 401). Positive recombinants were characterized by restriction enzyme mapping and DNA sequence analysis using the dideoxy chain-termination method of Sanger *et al.* (1977).

RNA synthesis in permeable trypanosome cells

Procyclic trypanosomes were made permeable following the method of Miller *et al.* (1978). Mid-logarithmic phase trypanosomes (5 × 10^6 cells/ml) were washed three times in 1% glucose, 0.9% NaCl, 10 mM Tris−HCl (pH 8.0), 1.5 mM $MgCl_2$ and resuspended in 1/10th of the original volume in transcription buffer: 10 mM Hepes−KOH (pH 7.9), 80 mM KCl, 5 mM $MgCl_2$, 0.5 mM EDTA, 1 mM DTT and 10% glycerol. Cells were chilled on ice for 15 min and L-α-lysophosphatidylcholine, palmitoyl, (lysolecithin) was added to a final concentration of 500 μg/ml. After incubation on ice for 5 min, cells were diluted with 2 vol of transcription buffer, harvested by centrifugation and resuspended in transcription buffer in 1/100th of the original culture volume. The labelling was initiated by adjusting the cell suspension to 1.5 mM ATP, 0.2 mM CTP and UTP, 0.04 mM GTP, 8 mM creatine phosphate, 0.4 mg/ml creatine kinase (Boehringer Mannheim) and 400 μCi/ml [α-^{32}P]GTP (400 Ci/mmol, Amersham International). Labelling was done at 28°C for 10 min and cells were lysed by adding an equal volume of 2% SDS, 1 mg/ml proteinase K and 40 mM EDTA. After phenol/chloroform extraction, nucleic acids were precipitated with ethanol, treated with RNase-free DNase (Worthington) and the ^{32}P-labelled RNA was again precipitated with ethanol. This procedure yields ~5−10 × 10^7 c.p.m./mg of total RNA.

RNase mapping of in vivo ^{32}P-labelled calmodulin RNA

^{32}P-labelled calmodulin RNAs were selected by hybridization to biotinylated 111 or 805 antisense RNAs followed by affinity chromatography with streptavidin−agarose (BRL). Biotinylated probes were synthesized using SP6 polymerase in a standard reaction containing 0.5 mM each ATP, GTP, CTP and UTP plus 0.05 mM biotin-11-UPT (BRL). After synthesis the template DNA was digested with RNase-free DNase. The mixture was then treated with 100 μg/ml proteinase K plus 1% SDS and 10 mM EDTA (pH 8.0) at 50°C for 30 min. RNAs were recovered by three consecutive ethanol precipitations with 2 M NH_4 acetate (pH 6.0) and finally washed with 70% ethanol. As judged by the precipitation of free [α-^{32}P]GTP added to a parallel mixture, this procedure removes most of the unincorporated nucleotides from the biotinylated RNA. 0.5−1 × 10^8 c.p.m. of total [^{32}P]RNA were resuspended in 1 ml hybridization buffer (80% formamide, 400 mM NaCl, 30 mM Pipes, pH 6.4, 1 mM EDTA) and heated at 80°C till the RNA was completely dissolved. Approximately 5 μg of biotinylated RNA in hybridization buffer was added to the [^{32}P]RNA and hybridization was carried out at 55°C for 16−24 h. At the end of the incubation the mixture was diluted with 1 vol of water and ethanol precipitated. RNAs were resuspended in 1 ml binding buffer (0.3 M NaCl, 5 mM EDTA, 50 mM Tris, pH 7.5) and added to 0.5 ml streptavidin−agarose equilibrated in the same buffer. The slurry was mixed for 1 h at room temperature and the resin recovered by low-speed centrifugation and washed several times with binding buffer till no counts were present in the supernatant. With certain batches of strepavidin−agarose more counts were eluted by washing in low-salt binding buffer (30 mM NaCl). Bound RNAs were eluted in 1% SDS, 10 mM EDTA by heating for 5 min at 95−100°C. Selected RNAs were ethanol precipitated and the hybridization-selection procedure repeated one more time. Since the biotinylated RNA is also eluted from the resin, no additional probe was added to the second hybridization reaction. Approximately 2.5−5 × 10^5 c.p.m. [^{32}P]RNA were recovered at the end of the procedure. Aliquots were annealed again as described above and digested with ribonuclease A and T_1 as described by Melton *et al.* (1984).

Acknowledgements

Our thanks go to Don Crothers and Frank Richards for support and continued interest in this work. We thank Nouria Hernandez and David Solnick for critical reading of the manuscript. The technical assistance of Sue Wait is gratefully acknowledged. This work was supported by the MacArthur Foundation and by USPHS grant AI08614 to Frank F.Richards. C.T. and E.U. are investigators of the MacArthur Foundation.

References

Auffray,C. and Rougeon,F. (1980) *Eur. J. Biochem.*, **107**, 303−314.

Benton,W.D. and Davis,R.W. (1977) *Science*, **196**, 180−182.

Boothroyd,J.C. (1985) *Annu. Rev. Microbiol.*, **39**, 475−502.

Boothroyd,J.C. and Cross,G.A.M. (1982) *Gene*, **20**, 279−287.

Borst,P. (1986) *Annu. Rev. Biochem.*, **55**, 701−732.

Campbell,D.A., Thornton,D.A. and Boothroyd,J.C. (1984) *Nature*, **311**, 350−355.

Clayton,C.E. (1985) *EMBO J.*, **4**, 2997−3003.

Contreras,R. and Fiers,W. (1981) *Nucleic Acids Res.*, **9**, 215−236.

Cunningham,I. (1977) *J. Protozool.*, **24**, 325−329.

De Lange,T., Liu,A.Y.C., Van der Ploeg,L.H.T., Borst,P., Tromp,M.C. and Van Boom,J.H. (1983) *Cell*, **34**, 891−900.

De Lange,T. Michels,P.A.M., Veerman,H.J.G., Cornelissen,A.W.C.A. and Borst,P. (1984) *Nucleic Acids Res.*, **12**, 3777−3790.

Gonzalez,A., Lerner,T.J., Huecas,M., Sosa-Pineda,B., Nogueira,N. and Lizardi,P.M. (1985) *Nucleic Acids Res.*, **13**, 5789−5804.

Grummt,I., Kuhn,A., Bartsch,I. and Rosenbauer,H. (1986) *Cell*, **47**, 901−911.

Henderson,S. and Sollner-Webb,B. (1986) *Cell*, **47**, 891−900.

Hernandez,N. (1985) *EMBO J.*, **4**, 1827−1837.

Huynh,T.V., Young,R.A. and Davis,R.W. (1985) In Glover,D. (ed.), *DNA Cloning Techniques: A Practical Approach.* IRL Press, Oxford, pp. 49−78.

Kooter,J., De Lange,T. and Borst,P. (1984) *EMBO J.*, **3**, 2387−2392.

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

McStay,B. and Reeder,R.H. (1986) *Cell*, **47**, 913–920.

Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.

Merritt,S.C., Tschudi,C., Konigsberg,W.H. and Richards,F.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1536–1540.

Michels,P.A.M., Poliszczak,A., Osinga,K.A., Misset,O., Van Beeumen,J., Wierenga,R.K., Borst,P. and Opperdoes,F.R. (1986) *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) *EMBO J.*, **2**, 1185–1192.

Milhausen,M., Nelson,R.G., Sather,S., Selkirk,M. and Agabian,N. (1984) *Cell*, **38**, 721–729.

Miller,M.R., Castellot,J.J.,Jr and Pardee,A.B. (1978) *Biochemistry*, **17**, 1073–1080.

Murphy,W.J., Watkins,K.P. and Agabian,N. (1986) *Cell*, **47**, 517–525.

Murray,N., Brammer,W.J. and Murray,K. (1977) *Mol. Gen. Genet.*, **150**, 53–61.

Nelson,R.G., Parsons,M., Barr,P.J., Stuart,K., Selkirk,M. and Agabian,N. (1983) *Cell*, **34**, 901–909.

Parsons,M., Nelson,R.G., Watkins,K.P. and Agabian,N. (1984) *Cell*, **38**, 309–316.

Proudfoot,N.J. and Brownlee,G.G. (1974) *Nature*, **252**, 359–362.

Sanger,F., Nicklen,S. and Coulsen,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

Sutton,R.E. and Boothroyd,J.C. (1986) *Cell*, **47**, 527–535.

Thomashow,L.S., Milhausen,M., Rutter,W.J. and Agabian,N. (1983) *Cell*, **32**, 35–43.

Tschudi,C., Young,A.S., Rubin,L., Patton,C.L. and Richards,F.F. (1985) *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) *Nucleic Acids Res.*, **10**, 3591–3604.

Received on October 21, 1987; revised on November 24, 1987