Volume 14 Number 22 1986 Nucleic Acids Research

The U2 RNA analogue of *Trypanosoma brucei gambiense:* implications for a splicing mechanism in trypanosomes

Christian Tschudi*, Frank F.Richards and Elisabetta Ullu

Yale MacArthur Center for Molecular Parasitology, Department of Internal Medicine, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510, USA

Received 12 August 1986; Revised and Accepted 21 October 1986

ABSTRACT

We have isolated the gene coding for the U2 analogue in trypanosomes. The 148 nucleotide long U2 RNA is capped and transcribed from a single copy gene. The 5' half of the molecule is highly homologous to mammalian U2 RNA, while the 3' half does not show any significant sequence homology with the mammalian counterpart. Nevertheless, the trypanosome U2 RNA can be folded into a secondary structure resembling the one proposed for U2 RNA. The presence of a U2 analogue and most likely other U RNAs in trypanosomes suggests that splicing is involved at some point in the maturation of mRNA. Possible interactions of the U2 RNA with the spliced leader RNA are considered.

INTRODUCTION

Every eukaryotic cell examined so far contains a set of small nuclear RNAs (snRNAs) designated U RNAs (1). These RNA species, ranging in size from about 60 to several hundred nucleotides, are metabolically stable and are associated with several different proteins to form small ribonucleoprotein particles (snRNPs). Unlike other small RNAs (tRNA, 5S and 7SL RNA) which are synthesized by RNA polymerase III, U RNAs, except U6, are synthesized by RNA polymerase II and have a 5'-terminal trimethylguanosine cap structure (1).

Recently, experimental data have provided evidence that U snRNPs are essential components for the splicing of nuclear mRNA precursors in vitro (2-5) and in vivo (6,7). The U1 and U2 snRNP interact with two of the three conserved sequence elements found in all introns, the 5' splice site and the site of lariat formation, respectively (8). The third conserved element, the 3' splice site, is probably recognized by a particle containing U5 RNA (9).

The process of RNA splicing is required to remove intervening sequences that are present in the primary transcripts of protein-coding genes of eukaryotes. Although the number of such genes analysed in trypanosomes is small, evidence is accumulating that these parasitic protozoa do not contain introns. Instead, the structure of the mature transcripts suggests a rather unusual mechanism for the synthesis of mRNA (reviewed in refs. 10, 11). Every mRNA yet examined contains

the same 35 nucleotide (nt.) sequence at its 5' end. This sequence, called the spliced leader (SL), is not encoded in the same transcription unit as is the main body of the mRNA but is found at the 5' end of a 140 nt. long RNA (SL RNA) which is encoded by a repeated gene family. Transcription primed with the SL RNA or intermolecular splicing between the SL RNA and a transcript from the protein-coding gene are two models that could explain these observations.

In this paper, we describe the U2 RNA gene isolated from Trypanosoma brucei gambiense. U2 RNA is 148 nt. long and probably contains a trimethylguanosine cap. Our data provide evidence that trypanosomes contain U RNAs which, in mammalian systems, participate in the processing of mRNA precursors.

MATERIALS AND METHODS

Trypanosomes. The trypanosomes used in this study are from the TXTat variant antigen type of Trypanosoma brucei gambiense and the YTat variant antigen type of Trypanosoma brucei rhodesiense.

Isolation of nucleic acids. High molecular weight trypanosome DNA and total cytoplasmic RNA was prepared as described (12). For the isolation of small RNAs, a cytoplasmic S100 fraction from procyclic trypanosomes (YTaT 1.1) grown in culture was used. The S100 fraction was prepared essentially as described by Dignam et al. (13). In our hands, this procedure gave an RNA preparation which was mostly depleted of ribosomal RNAs and was substantially enriched for small non-ribosomal RNAs. RNA samples were fractionated on a 6% polyacrylamide gel containing 7M urea and individual RNA species were isolated by soaking the appropriate gel slice in 50mM Tris (pH 8.0), 1mM EDTA, and 500mM KCl.

Immunoprecipitation. RNA III was purified as described above and ^{32}P-labelled at the 3' end in a reaction mixture containing 100µCi (α-^{32}P)pCp, 5µM ATP, 50mM HEPES (pH 8.3), 10mM $MgCl_2$, 3mM DTT, 10% DMSO, and 1 unit of T4 RNA ligase. Immunoprecipitations with antibodies directed against either 2,2,7-trimethylguanosine or 7-methylguanosine (kindly provided by Douglas Black) were performed as decribed by Chabot et al. (9).

Isolation of genomic clones. U2 genes were isolated from a phage library consisting of TXTat 1 DNA partially digested with Sau3A and inserted into the BamHI site of the bacteriophage vector EMBL 4 (12). Approximately 15,000 recombinants were screened with pCp labelled RNA III and hybridizations were carried out as described below for Southern and Northern blots.

Southern and Northern hybridizations. Genomic and recombinant DNAs were digested with restriction enzymes, separated on agarose gels and transferred to nitrocellulose filters by the method of Southern (14). Low molecular weight RNAs

were run on 6% acrylamide gels containing 7M urea and transferred electrophoretically to zeta-probe membranes (Bio-Rad). Hybridizations were done at 37°C for 20-48 hrs in a solution containing 50% formamide, 5xDenhardt's (0.1% Ficoll, 0.1% bovine serum albumin, 0.1% polyvinylpyrrolidone), 5xSET (0.1M Tris pH 7.8, 5mM EDTA, 0.75M NaCl), 50µg/ml tRNA, and 0.2% SDS. Membranes were rinsed briefly at room temperature in 2xSSC and then washed two times for 60 min. each at 65°C in 2xSSC and 0.2% SDS.

Miscellaneous. Primer extension and DNA sequence analysis were done as described previously (12). RNA transcripts were synthesized with SP6 RNA polymerase following the procedure of Melton et al. (15).

RESULTS

Identification of a U RNA candidate in trypanosomes

We first analysed the small RNAs present in a ribosome-free cytoplasmic extract of trypanosome cells (see Materials and Methods). As shown in Figure 1A, this fraction contains as major RNA components tRNA, 5S RNA, a small ribosomal RNA (sRNA 4 according to ref. 16) and 7SL RNA (E. Ullu, unpublished observation).

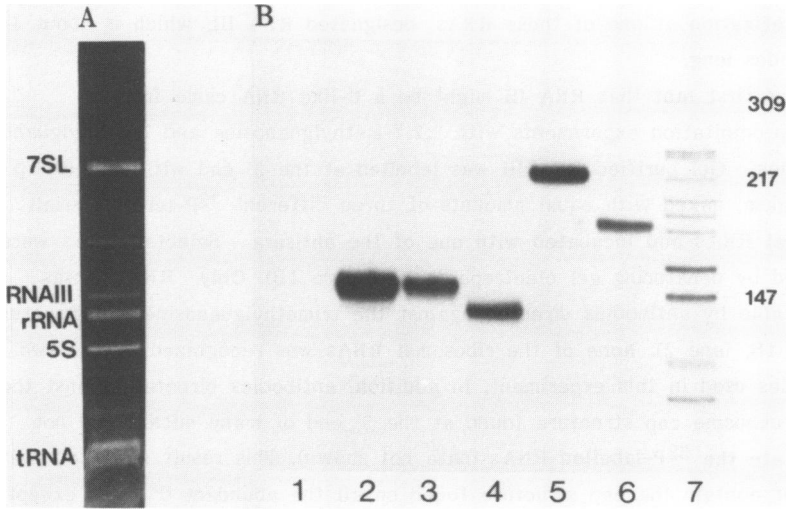

Figure 1. Polyacrylamide gel electrophoresis and immunoprecipitation of small RNAs. (A) Ethidium bromide stained 6% acrylamide-7M urea gel of small cytoplasmic RNAs prepared as described in Materals and Methods. (B) Autoradiograph of a 6% acrylamide-7M urea gel of (^{32}P) pCp-labelled RNAs. Lanes 1 and 2 show the immunoprecipitation of a mixture of RNA III (lane 3) plus three small ribosomal RNAs (lanes 4 to 6) with non-immune and anti-trimethylguanosine sera, respectively. Lane 7: ^{32}P-labelled fragments from a HpaII digest of pBR322.

```
-110  CGGGCCATTACAGAGCCGGGGCCCGCCCGAATATATAATTACTTATTGAA   -61

 -60  TTTATTCTTTGGTATTTCCTAGCTTGTTGCAGCCTACGGAACTTTTGGAC   -11

                +1
 -10  AAGGCACTGCATATCTTCTCGGCTATTTAGCTAAGATCAAGTTATTAAAC    39

  40  TGTTCTTATCAGAGTAACTCCTGATACGGGCCTTTGGCCCAAGGATCAAA    89

  90  ACTGTTGCCTGTCCCGCGTTCTTCCGGGGTTCCACTTGTCCGGACGGAGC   139

 140  GCGACGGTCGCGAACATTTTCCGACAAAACCAACCACTGAACGTTGTAAG   189

 190  GCAGTTATTGGAACGACCCGCATTTCCCATCACTTTATTCTCGCTTGCTG   239

 240  TATACTTTAACGCCTCGGATCC  261
```

Figure 2. DNA sequence of the trypanosome U2 gene. The non-coding strand is shown and the sequence is numbered with respect to the first nucleotide of the coding region. The position of the presumptive 3' end of the U2 RNA is indicated by a bar.

In addition, we noticed several minor RNA species of various sizes. These RNAs were gel purified and used for further analysis. Here we report the characterization of one of these RNAs, designated RNA III, which is about 148 nucleotides long.

Our first hint that RNA III might be a U-like RNA came from immunoprecipitation experiments with 2,2,7-methylguanosine and 7-methylguanosine antibodies. Gel purified RNA III was labelled at the 3' end with (α-^{32}P)pCp and RNA ligase, mixed with equal amounts of three different ^{32}P-labelled small ribosomal RNAs and incubated with one of the antisera. Selected RNAs were then analyzed by denaturing gel electrophoresis (Figure 1B). Only RNA III was precipitable by antibodies directed against the trimethylguanosine cap structure (Figure 1B, lane 2). None of the ribosomal RNAs was recognized by the two antibodies used in this experiment. In addition, antibodies directed against the 7-methylguanosine cap structure found at the 5' end of many mRNAs did not precipitate the ^{32}P-labelled RNAs (data not shown). This result indicated that RNA III might contain the cap structure found on all the abundant U RNAs except U6.

The RNA III gene

To isolate the gene coding for the 148 nt. RNA we screened a trypanosome genomic library (12) by hybridization with ^{32}P-labelled RNA. Several positive recombinant phage were analyzed in detail and a combination of restriction enzyme mapping and Southern hybridization revealed that the clones cover the same genomic locus to different extents. The unique region complementary to RNA III was

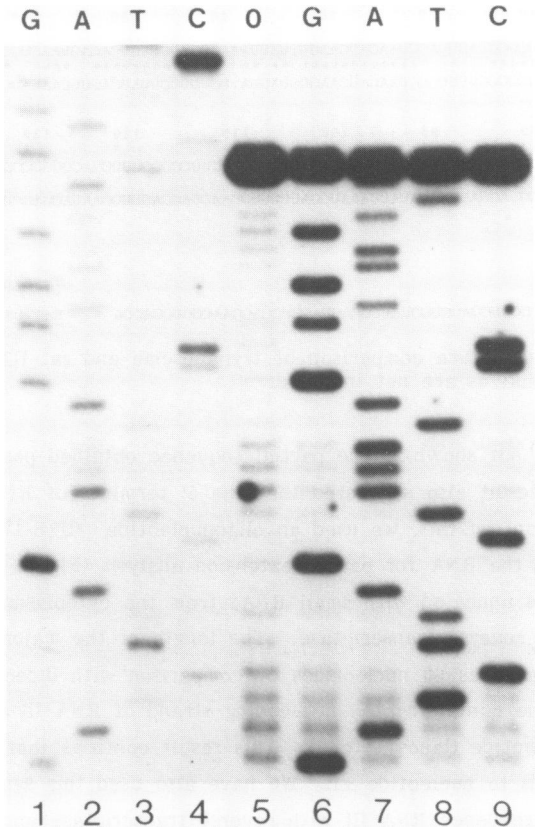

Figure 3. DNA sequence analysis of U2 RNA. 10% acrylamide-7M urea gel of dideoxy sequence determination. Lanes 1-4: The U2 coding region subcloned into M13 was sequenced with the oligonucleotide U2B-17. Lanes 5-9: Small cytoplasmic RNAs were primed with U2B-17 in the absence of dideoxynucleotide triphosphates (lane 5) or appropriate dideoxynucleotide/deoxynucleotide triphosphate mixtures were included in the reaction (lanes 6-9).

localized in a 375 bp SmaI - BamHI restriction fragment. The DNA sequence of this fragment was determined and is shown in Figure 2.

In order to locate the sequences that specify RNA III we performed the following experiments. First, we determined the coding strand for RNA III. The 375 bp SmaI-BamHI fragment was subcloned into SP6 vectors in both orientations, single stranded ^{32}P-labeled RNA was synthesized and hybridized to a Northern blot of small RNAs. The sequence shown in Figure 2 represents the non-coding strand for RNA III, since an SP6 transcript complementary to this sequence detects the 148 nt. RNA (data not shown). Next we analysed the 148 nt. RNA by direct RNA

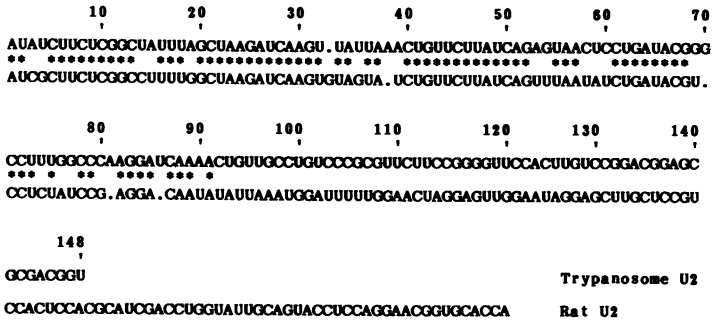

Figure 4. Primary structure comparison of trypanosome and rat U2 RNA. Occasional matches in the 3' halves are not included.

sequencing (results not shown). The partial sequence obtained positioned the 3' end to nt. 148. This result also suggested that the 5' terminus of RNA III is at about position +1. To confirm this, we used an oligonucleotide (U2B-17) complementary to nt. 44 to 60 of the RNA for primer extension analysis (Figure 3). The 5' end labelled primer was annealed with small RNAs from the cytoplasmic S100 fraction and extended with reverse transcriptase. The length of the major extension product (lane 5) was estimated as 60 nucleotides by comparison with dideoxy-sequence ladder of the m13 subclone containing the non-coding strand of RNA III and primed with the same oligonucleotide (lanes 1 to 4). This result confirms that the 5' end of RNA III corresponds to nucleotide +1. We have also used this synthetic oligonucleotide to sequence RNA III with reverse transcriptase and dideoxynucleotide

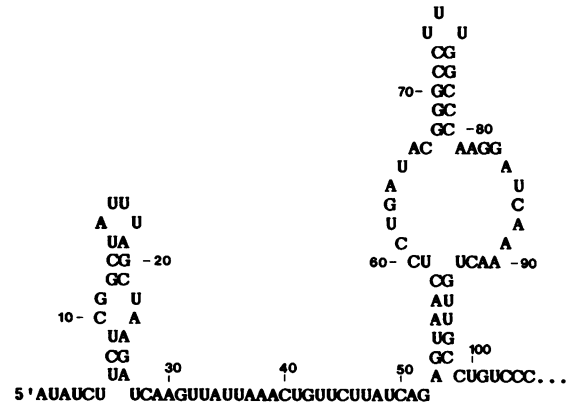

Figure 5. A model for the secondary structure of the 5' half of the trypanosome U2 RNA.

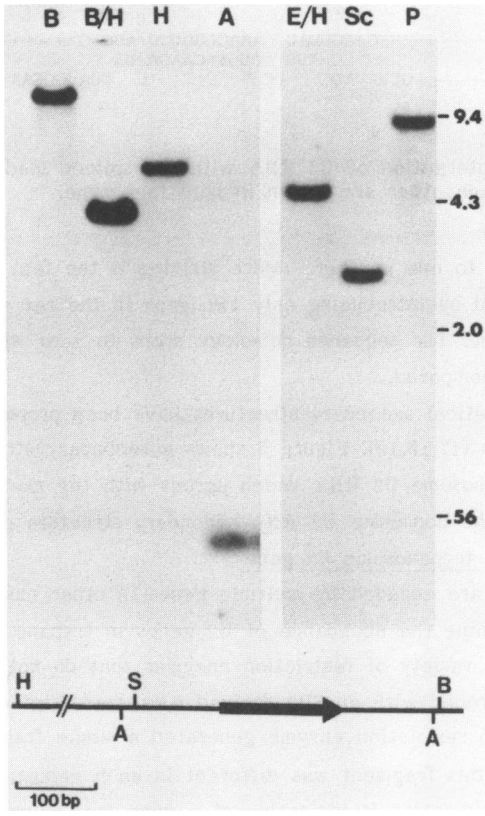

Figure 6. Genomic Southern hybridization. 0.8% agarose gels of various digests of TXTat1 DNA were transfered to nitrocellulose and probed with sequences coding for U2 RNA. Lambda digested with HindIII served as a molecular weight marker and sizes are indicated in kilobase pairs. A partial restriction map of the cloned U2 gene is shown at the bottom. [The HindIII and BamHI restriction sites are 3.9 kb apart.] Abbreviations: A, AvaI; B, BamHI; E, EcoRI; H, HindIII; P, PstI; S, SmaI; Sc, ScaI.

triphosphates (Figure 3, lanes 6 to 9). This result showed that the RNA sequence is colinear with the cloned DNA sequence to the very 5' end revealing that RNA III does not contain the spliced leader at the 5' end.

RNA III is analogous to U2

We have compared the sequence of RNA III with that of snRNAs from mammals. The best homology was found between U2 RNA and the trypanosome sequence. The alignment of the trypanosome sequence with that of rat U2 RNA is shown in Figure 4. The 5' halves of the two molecules show extensive regions of highly conserved primary structure. In this region (nt. 1 to 73) the two sequences

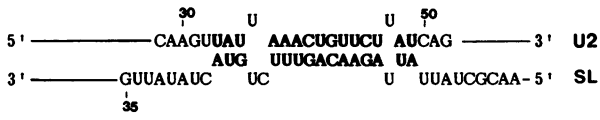

Figure 7. Possible interaction of U2 RNA with the spliced leader RNA. Nucleotides base pairing with each other are shown in bold face type.

are 78% homologous to one another. More striking is the fact that an optimal alignment is obtained by introducing only two gaps in the rat and one gap in the trypanosome sequence. The sequence homology drops to zero when the 3' halves of the two RNAs are compared.

Several hypothetical secondary structures have been proposed for the U2 RNA of higher eukaryotes (17,18,19). Figure 5 shows a secondary structure prediction for the 5' half of trypanosome U2 RNA which agrees with the model proposed by Keller and Noon (17) and the consensus U2 RNA secondary structure derived by Ares (20).

Copy number of the trypanosome U2 gene

Since U RNAs are encoded by multiple genes in other eukaryotic organisms (1), we wished to determine the abundance of U2 genes in trypanosomes. Genomic DNA was digested with a variety of restriction enzymes that do not cut within the U2 coding region and probed with an SP6 derived RNA transcript containing the region specifying U2. Each restriction enzyme generated a single fragment hybridizing to U2 and the size of this fragment was different in each genomic digest (Figure 6). This pattern of hybridization is indicative of a gene that is present once in the genome. Furthermore, digestion of trypanosome DNA with AccI or with a combination of BamHI and HindIII yielded bands we predicted from the restriction map of the isolated genomic region (Figure 6). We conclude therefore that trypanosome U2 RNA is encoded by a single copy gene.

DISCUSSION

Genes coding for U2 RNA have been isolated from a variety of eukaryotic organisms (1). In all cases, they are found in multiple copies in the genome and the transcript of these genes is approximately 189 nucleotides long. In addition, mammalian genomes contain a large number of U2 pseudogenes (1). We have identified the U2 analogue in African trypanosomes and found that a single copy gene encodes an RNA of 148 nucleotides.

We have chosen to analyze small RNAs in a cytoplasmic fraction of trypanosome cells. However, it is important to point out that the substantial enrichment of U2 RNA in our preparations does not reflect the actual distribution

of small RNAs in the cell. Using standard cell fractionation techniques we found that U2 RNA is predominantly localized in the cell nucleus (unpublished results) which agrees with the subcellular localization of U RNAs in other organisms (1). There are a number of other features common to U RNAs and the trypanosome U2 analogue. As indicated by their name, U RNAs are rich in uridylic acid, usually in the order of about 30%. Trypanosome U2 has a U content of 30% which is in good agreement with this general rule. A 5' terminal trimethylguanosine cap structure is another characteristic of U RNAs. We were able to precipitate trypanosome U2 RNA with an antibody directed against this cap structure (Figure 1B) and suspect therefore that the cap structure has been conserved in U RNAs of protozoa. Furthermore, the level of sensitivity of U RNA synthesis to α-amanitin indicates that these genes are transcribed by RNA polymerase II (1). Preliminary experiments in our laboratory showed that the level of inhibition of U2 transcription by the antibiotic α-amanitin is comparable to that of the genes coding for the calcium binding protein calmodulin. However, there is at least one major difference in the biosynthesis of these two RNAs. Calmodulin mRNAs have the spliced leader at their 5' ends, whereas U2 RNA does not contain this sequence. Whether different polymerases with a similar α-amanitin sensitivity, different specialized transcription complexes, or post-transcriptional events generate mRNAs and U RNAs remains to be determined.

When comparing the trypanosome sequence with that of rat U2 RNA (Figure 4) it is evident that the U2 RNA molecule can be divided into two domains. The 5' half, between positions 1 and 73, defines the U2 domain, while the 3' half is heterologous. Such an asymmetric distribution of sequence homology was already noted between Drosophila and rat U2 RNA (21), although to a much lesser extent (86% 5' homology versus 63% 3' homology). In addition, the U2 RNA homologue of the yeast Saccharomyces cerevisiae contains a U2 domain which has a size and location similar to the one we find in trypanosome U2 RNA (20). Taken together, these observations have several implications for the evolution and function of U2 RNA. The extensive degree of sequence homology between the 5' ends of the U2 RNA from organisms so distantly related in evolution must reflect strong functional constraints on this region of the molecule. This supports the notion that the "U2 domain", as defined in the trypanosome and yeast RNAs, is the "business end" of the U2 RNA molecule. Black et al. (8) have shown that the U2 snRNP interacts with the branch point of the pre-messenger RNA during splicing in in vitro extracts. Whether U2 RNA itself or a protein component of the RNP mediates this interaction remains to be determined.

Since no intron-containing genes have been isolated in trypanosomes, the

substrates for the function of trypanosome U2 RNA remain to be identified. An attractive possibility is that the process of discontinuous transcription which is characteristic of trypanosome protein-coding genes, requires U2 RNA for the splicing of the spliced leader sequence onto the 5' end of all trypanosome mRNA molecules. In support of this hypothesis we find a striking complementarity between the U2 RNA and the 5' portion of the spliced leader RNA (Figure 7). 15 out of 17 nucleotides of the U2 RNA can be base paired with the SL RNA sequence from position 10 to 27. The formation of a hybrid structure between these two molecules could represent a molecular intermediate in the generation of 5' termini of trypanosome mRNAs. Although this possibility is only a matter of speculation, the identification of a U2 analogue in trypanosomes [which according to the unrooted evolutionary tree proposed by Sogin et al. (22), diverged from the main branch of eukaryotic evolution much earlier than yeast] gives credit to the hypothesis that splicing of pre-messenger RNA is a very primitive function of the eukaryotic organism.

ACKNOWLEDGEMENTS

We would like to thank Valerie Asher for performing the experiment shown in Figure 1B. We are grateful to Doug Black for providing the antibodies and for valuable comments on the manuscript. We thank Aino Ruusala for synthesizing oligonucleotides, Sue Wait for technical assistance and Charlene Sullivan for typing the manuscript. This work was supported by the MacArthur Foundation and by USPHS grants #AI08614 and GM12607.

*To whom correspondence should be addressed

REFERENCES

1. Reddy, R. and Busch, H. (1983) Prog. Nucleic Acids Res. Mol. Biol. 30, 127-162.
2. Yang, V.W., Lerner, M.R., Steitz, J.A. and Flint, S.J. (1981) Proc. Natl. Acad. Sci. USA 78, 1371-1375.
3. Padgett, R.A., Mount, S.M., Steitz, J.A. and Sharp, P.A. (1983) Cell 35, 101-107.
4. Krämer, A., Keller, W., Appel, B. and Lührmann, R. (1984) Cell 38, 299-307.
5. Krainer, A.R. and Maniatis, T. (1985) Cell 42, 725-736.
6. Bozzoni, I., Annesi, F., Beccari, E., Fragapane, P., Pierandrei-Amaldi, P. and Amaldi, F. (1984) J. Mol. Biol. 180, 1173-1178.
7. Fradin, A., Jove, R., Hemenway, C., Keiser, H.D., Manley, J.C. and Prives, C. (1984) Cell 37, 927-936.
8. Black, D.L., Chabot, B. and Steitz, J.A. (1985) Cell 42, 737-750.
9. Chabot, B., Black, D.L., LeMaster, D.M. and Steitz, J.A. (1985) Science 230, 1344-1349.
10. Boothroyd, J.C. (1985) Ann. Rev. Microbiol. 39, 474-502.

11. Borst, P. (1986) Ann. Rev. Biochem. 55, 701-732.
12. Tschudi, C., Young, A.S., Ruben, L., Patton, C.L. and Richards, F.F. (1985) Proc. Natl. Acad. Sci. USA 82, 3998-4002.
13. Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) Nucl. Acids Res. 11, 1475-1489.
14. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
15. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) Nucl. Acids Res. 12, 7035-7056.
16. Hasan, G., Turner, M.J. and Cordingley, J.S. (1984) Gene 27, 75-86.
17. Keller, E.B. and Noon, W.A. (1985) Nucl. Acids Res. 13, 4971-4981.
18. Reddy, R., Henning, D., Epstein, P. and Busch, H. (1981) Nucl. Acids Res. 9, 5645-5658.
19. Branlant, C., Krol, A., Ebel, J.-P., Lazar, E., Haendler, B. and Jacob, M. (1982) EMBO J. 1, 1259-1265.
20. Ares, M.Jr. (1986) Cell, in press.
21. Alonso, A., Jorcano, J.L., Beck, E. and Spiess, E. (1983) J. Mol. Biol. 169, 691-705.
22. Sogin, M.L., Elwood, H.J. and Gunderson, J.H. (1986) Proc. Natl. Acad. Sci. USA 83, 1383-1387.