Volume 14 Number 23 1986 Nucleic Acids Research

# Three small RNAs within the 10 kb trypanosome rRNA transcription unit are analogous to Domain VII of other eukaryotic 28S rRNAs

Theodore C.White, Gloria Rudenko and Piet Borst

Division of Molecular Biology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Received 8 October 1986; Accepted 5 November 1986

ABSTRACT

We have localized the six ribosomal RNAs (rRNAs) which encode the 28S rRNA region of Trypanosoma brucei. These six rRNAs include two large rRNAs, 28S α (approx. 1840 nt) and 28S β (approx. 1570 nt), and four small rRNAs of approximate sizes 220, 180, 140 and 70 nt. Three of these four small rRNAs (180, 70 and 140) are found at the 3' end of the 28S rRNA region. Sequence analysis of this area shows that these three small rRNAs encode Domain VII, the last domain of secondary structure in the 28S rRNAs of eukaryotes. Hybridization of labeled nascent RNA to the cloned repeat unit and S1 nuclease protection analysis of putative precursors show that transcription initiates approximately 1.2 kb upstream of the 18S rRNA and terminates after the last small rRNA (140) at the 3' end of the 28S rRNA region. Analysis of three putative rRNA precursors suggests that the small rRNAs are not processed from the primary transcript until after the usual processing of the 5.8S rRNA region.

## INTRODUCTION

In the nucleus of most eukaryotes, the ribosomal repeat unit includes three mature rRNA species, the 18S, 5.8S and 28S rRNAs. These three rRNAs are processed from one large primary transcript (reviewed in 1). In certain organisms, other processing steps occur which result in unusual versions of these three rRNAs. For instance, the 28S rRNA of certain members of the dipteran insects (2, 3, 4), vertebrates (5, 6) and protozoa (7, 8, 9) can be split into two separate rRNAs, which are called 28S α and 28S β. Furthermore, the 5.8S rRNA of dipteran insects is present as two separate species, a shortened 5.8S rRNA and a 2S rRNA (10, 11).

The rRNA repeat unit which requires the most complex processing system is the rRNA repeat unit of trypanosomes and other kinetoplastidae. In these organisms, the 28S rRNA unit is split into 28S α and β (7, 12, 13, 14, 15, 16, 17). In addition, four small rRNAs (srRNAs) have been observed (17, 18, 19, 20). They are approximately 220, 180, 140 and 70 nucleotides long, although the exact size differs among organisms (cf. 20). In trypanosomes, the srRNAs are present in stoichiometric amounts in isolated polysomes (18). They have

been localized in the trypanosome rRNA repeat unit and were tentatively mapped in the area of the 28S rRNA (19).

While investigating an *in vitro* circularization reaction involving the 180 srRNA, we found that the published map positions of the srRNAs were not correct. In this paper, we present the correct localizations and sequences of these srRNAs and propose a functional role for these srRNAs in the ribosome. Our analysis of the transcription of this repeat unit has localized the areas of initiation and termination of transcription and identified three putative precursors possibly involved in the processing of this rRNA repeat unit.

## MATERIALS AND METHODS

**DNA manipulations.** A clone containing two ribosomal repeats was originally obtained from a cosmid library (21). A subclone containing one ribosomal RNA repeat unit, pR4, has already been described (22) and is shown in Fig. 1. DNA fragments from pR4 were isolated from low melting point agarose by phenol extraction (23) or by using Elutip-d columns (Schleicher & Schuell). Fragments A, F, O, P, R, X and Y (see Figs. 1 and 4) were subcloned to facilitate analysis of the repeat unit. Restriction enzyme digestions, size separation of fragments by electrophoresis, and the cloning of DNA fragments were performed using standard techniques (23).

Purified fragments were 5' end labeled by treatment with calf intestinal alkaline phosphatase, followed by T4 polynucleotide kinase in the presence of $\gamma$-$^{32}$P ATP (3000 Ci/mmol; 23). Fragments were 3' end labeled by treatment with the Klenow fragment of *E. coli* DNA polymerase I or with T4 DNA polymerase in the presence of $\alpha$-$^{32}$P dATP or dCTP (3000 Ci/mmol) and unlabeled nucleotides (23). The S1 nuclease protection procedure (24) was performed as previously described (25).

DNA sequence analysis was performed with the chemical degradation procedure (26) using 5' or 3' end labeled fragments; or with the dideoxy sequencing method (27) with modifications (28) using fragments subcloned into the pEMBL9 vector (29).

**RNA manipulations.** RNA was prepared from both the cultured insect form and the bloodstream form of *Trypanosoma brucei brucei*, strain 427 using lithium chloride precipitation (30). RNA blots were prepared after electrophoretic size fractionation of glyoxylated RNA in agarose gels (31). RNA blots were hybridized with nick translated $\alpha$-$^{32}$P labeled probes (32, 33). Post hybridizational washes were to 0.1 x SSC at 65°C.

The nuclear run-on assay involves the extension of nascent RNA in isolated nuclei (22) in the presence of $\alpha$-$^{32}$P UTP (3000 Ci/mmol). After purification, the RNA was hybridized to DNA blots as previously described (34).

## RESULTS

### Localization of mature rRNAs

We have localized the mature rRNAs and their putative precursors within the ribosomal repeat unit. The restriction map of one complete cloned repeat unit, pR4, is shown in Fig. 1. The approximate positions of the 18S, 5.8S and 28S $\alpha$ and $\beta$ rRNAs were previously determined (35; T. Berkvens, P. Sloof and P. Borst, unpublished data) and, more recently, the entire sequences of the 18S rRNA and the 5.8S rRNA were determined (36,37). From the restriction enzyme

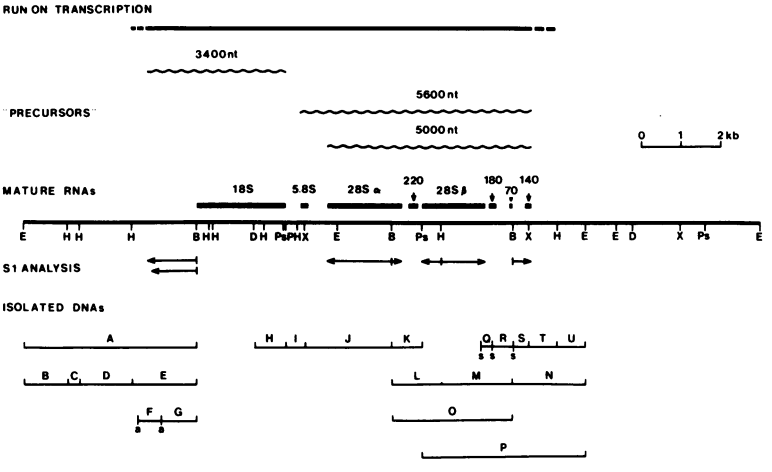


Figure 1. Map of the cloned ribosomal repeat unit. The rDNA was inserted into the Pst I site of pBR328. The vector interrupts the repeat unit at the Pst I site at the end of fragment K. The restriction enzyme cleavage sites that have been mapped include: B = Bgl II, D = Hind III, E = Eco RI, H = Hinc II, P = Pvu II, Ps = Pst I, and X = Xho I. The DNA probes A - U used in this study are shown below the restriction map. The exact location of the mature rRNAs (including the srRNAs) is shown above the map. The three putative precursors detected by RNA blot analysis (see Fig. 2) and the extent of run on transcription (see Fig. 7) are shown above the mature RNAs. Several S1 protection experiments are diagrammed below the map (see Table 1). Fragments Q and R are the last two adjacent Sau 3A fragments from fragment M and their ends are designated s. Fragments F and G are the last two adjacent Alu I fragments from fragment E and their ends are designated a.

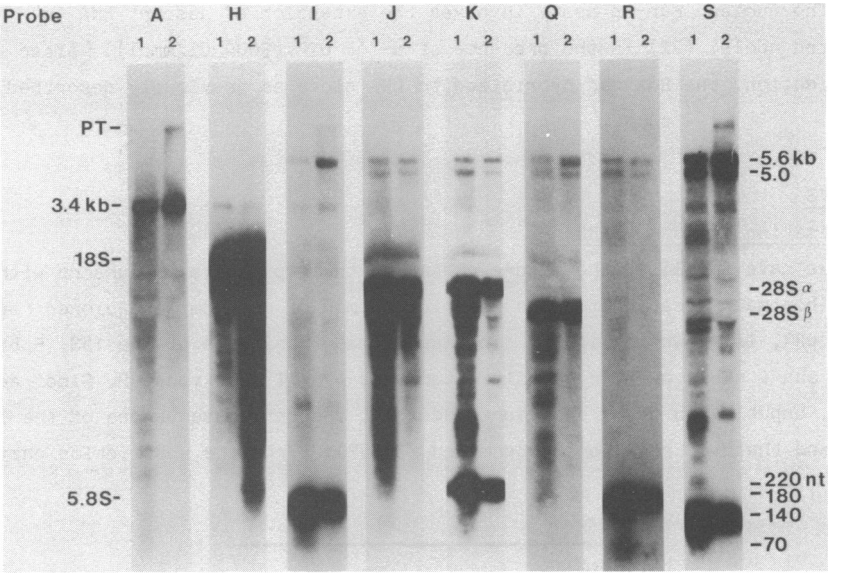


Figure 2. Hybridization of RNA blots with various restriction fragments from pR4. 40 µg of total RNA (1 = culture form, 2 = blood stream form) was glyoxylated, electrophoresed on agarose gels and blotted to nitrocellulose. Identical strips were hybridized with nick translated DNA fragments A, H, I, J, K, Q, R, and S from pR4. Precursor sizes are estimates based on denatured fragments of λ DNA digested with Hind III. PT = putative primary transcript. Several blots are overexposed to allow detection of the putative precursors.

cleavage sites predicted by these sequences, we can locate the 18S and 5.8S rRNAs on our map. To determine the approximate position of the srRNAs, we hybridized DNA fragments from the rRNA repeat unit (Fig. 1) to RNA blots of total RNA and washed at high stringency (65°C, 0.1 X SSC, see Fig. 2). Fragment K from the center of the 28S unit hybridizes to the 220 srRNA. Fragment R near the 3' end of the unit hybridizes to the 180 and 70 srRNAs. Fragment S at the 3' end of the unit hybridizes to the 140 srRNA (as does fragment T, data not shown). The 70 srRNA was only faintly detectable by standard RNA blot analysis. To verify the location of the 70 srRNA, fragment R was hybridized to an electroblot of total RNA which had been size fractionated on a 6% polyacrylamide gel. This confirmed that the 70 srRNA sequence is contained within fragment R (data not shown).

The precise position of the srRNAs relative to the 28S α and β rRNAs was determined by S1 nuclease protection analysis. Examples of these S1 nuclease protection experiments are shown in Fig. 3. The results of all the S1

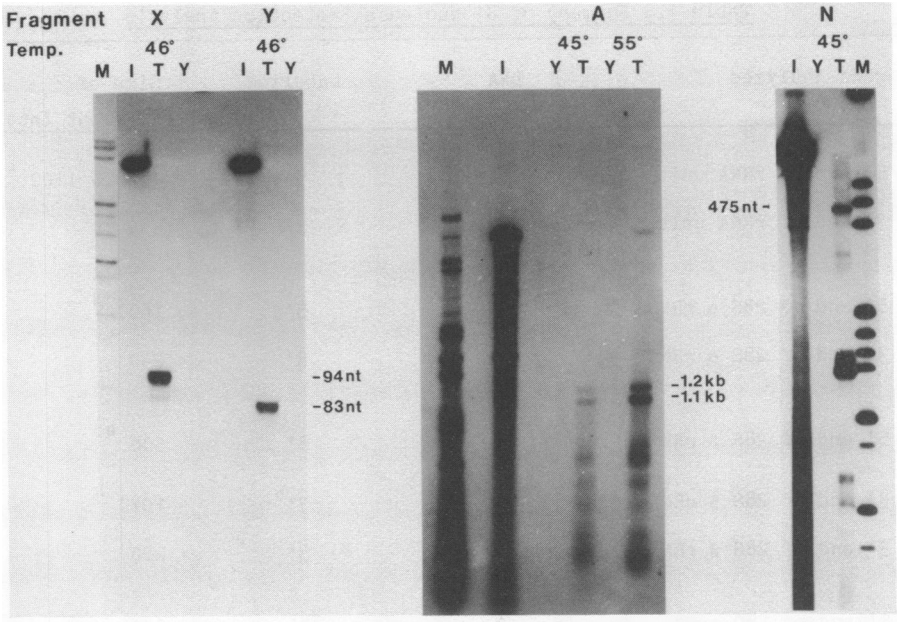


Figure 3. S1 analysis of the 180 srRNA and the 5' and 3' ends of the transcription unit. Fragments A and N (Fig. 1) and fragments X and Y (Fig. 4) were used in these experiments as described in Table 1. Lanes are designated as follows: I = input DNA, Y = yeast RNA (control RNA), T = trypanosome RNA, M = marker DNA. The temperature of hybridization is shown for each fragment. Fragments A and N show several smaller protected fragments in addition to the labeled fragments. We interpret these as melting out of the (AT)-rich spacer and/or polymorphisms within the spacer. Fragment A was hybridized at two different hybridization temperatures as shown.

experiments are summarized in Table 1 and are diagrammed in Figs. 1 and 4. For each RNA, the 5' and 3' ends of the rRNA were determined from one internal restriction enzyme cleavage site. Fragments containing the 220 and 70 srRNAs did not contain suitable restriction enzyme cleavage sites for S1 analysis. The 220 srRNA was not investigated further. The position of the 70 srRNA was confirmed by sequencing.

Sequence Analysis of the srRNAs

We have sequenced the 1.5 kb of DNA encoding the three srRNAs at the 3' end of the rRNA unit (Figs. 4 and 5). The sequence is compared to the published sequences of yeast 26S rRNA sequence (38) and of the srRNAs of Crithidia obtained by RNA sequencing of the isolated srRNAs (39). Models for the secondary structure of the T. brucei srRNAs were constructed based on the models proposed for yeast (40, 41) and Xenopus (42)(Fig. 6).

Table 1. Summary of S1 Nuclease Protection Analysis

| rRNA Analyzed | DNA Probe[1] | Labeling[2] | Size of Fragment (nt)[3] |
|---|---|---|---|
| 5' end of rRNA unit | A | 5' | 1200 & 1100 [4] |
| 3' end of rRNA unit | N | 3' | 475 |
| 5' end of 28S α rRNA | J | 5' | 1600 [4] |
| 3' end of 28S α rRNA | O | 3' | 240 |
| 5' end of 28S β rRNA | L | 5' | 500 [4] |
| 3' end of 28S β rRNA | M | 3'[5] | 1070 [4] |
| 3' end of 28S β rRNA | X | 3' | 220 |
| 5' end of 18O srRNA | X | 5' | 94 |
| 3' end of 18O srRNA | Y | 3' | 83 |
| 5' end of 14O srRNA | S | 5' | 70 |
| 3' end of 14O srRNA | T | 3' | 63 |

[1] See Figures 1 and 4.

[2] Fragments were labeled at the 5' end by polynucleotide kinase. Fragments were labeled at the 3' end by the Klenow large fragment of DNA polymerase I of E. coli.

[3] Fragments were usually sized on denaturing polyacrylamide gels using plasmid pAT 153 cut with Msp I as a size marker.

[4] Fragments were sized on alkaline agarose gels instead of denaturing polyacrylamide gels, using λ DNA cut with Eco RI and Hind III as a size marker.

[5] Fragment was 3' end labeled with T4 DNA polymerase.

rRNA transcription

Labeled nascent RNA produced by "run-on" transcription in isolated nuclei (22) was hybridized to Southern blots of restriction enzyme digests of DNA from pR4 and two subclones. The labeled RNA hybridizes to many restriction fragments of pR4 and the intensity of most of the bands is proportional to the size of the respective fragments. Some fragments do not hybridize and thus

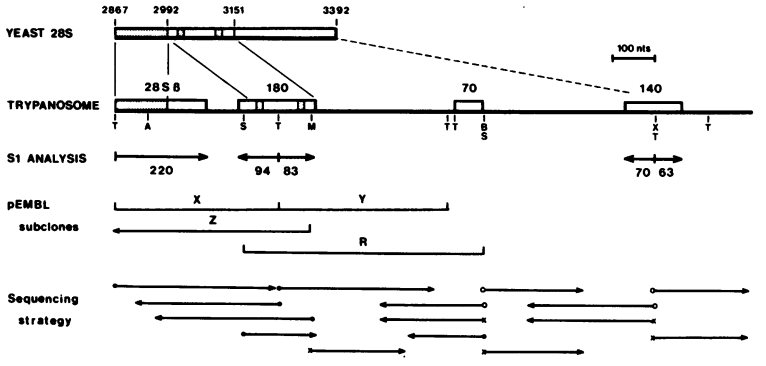


Figure 4. The region extending 3' of the 28S ß rRNA. Restriction enzyme cleavage sites that are shown include A = Alu I, B = Bgl II, M = Msp I, S = Sau 3A, T = Taq I and X = Xho I. The region is shown compared to the 3' end of the yeast 28S rRNA (38) with the numbering from the yeast sequence. Stretches of complete homology greater than 15 nucleotides are shown as hatched areas. The results of several S1 protection experiments described in Table 1 are diagrammed below the restriction map.

The entire region was sequenced using both Maxam and Gilbert (26) and dideoxy (27) sequencing. The sequencing strategy is shown at the bottom of the figure. Arrows which begin with • were sequenced from pEMBL subclones by dideoxy sequencing. Regions that were determined by Maxam and Gilbert sequencing are shown as arrows that begin with either a O (5' end labeling ) or X (3' end labeling).

are not part of the transcription unit. The interpretation of these data is summarized in Fig. 1. The results which clearly define the 5' and 3' ends of the transcription unit are shown in Fig. 7. The 5' end of transcription is defined in lanes 1 and 2. Of the four fragments in lane 1 (B, C, D, and E), only the 1.6 kb fragment E hybridizes with nascent RNA, suggesting that transcription starts within this fragment. Since nascent RNA hybridizes with both fragments F (500 nts) and G (900 nts) in lane 2, both fragments must be transcribed. The stoichiometry of these fragments suggests that transcription starts within fragment F. Lanes 3 to 5 of Fig. 7 define the 3' end of transcription. Nascent RNA hybridizes to fragments S and T but does not hybridize to fragment U, suggesting that transcription terminates within fragment T. The transcription unit as determined by this analysis is approximately 10 kb, leaving an 8 kb non-transcribed spacer.

A putative full-length primary transcript of the rRNA unit is barely detectable in the RNA blot analysis (PT in Fig. 2). More pronounced are three other putative precursors from this transcription unit. The 3.4 kb RNA is detected by fragment A at the beginning of the transcription unit and also by

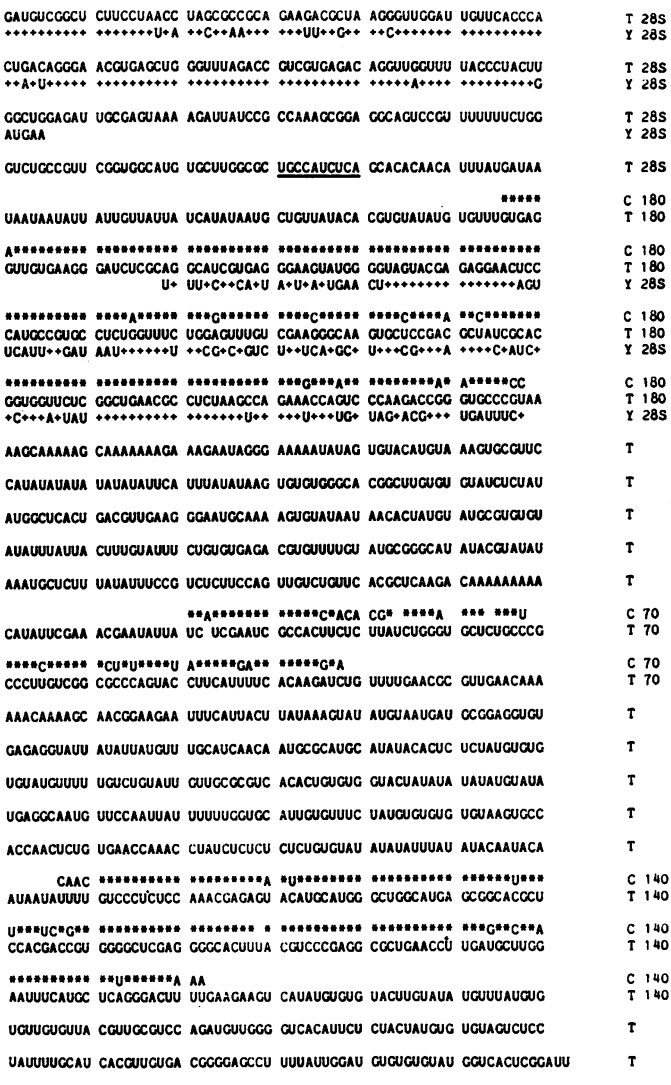

Figure 5. Sequence of the DNA region encoding the trypanosome 28S β (3' end) and 180, 70 and 140 srRNAs. The sequence (1502 nts), presented as the RNA of the putative precursor, is shown (T) compared to yeast (Y, homology is shown as +; 38) and to Crithidia srRNAs (C, homology is shown as *; 39). The yeast sequence was also determined from the DNA and is presented as RNA. The 3' end of the 28S β has not been determined exactly. The S1 analysis predicts that it is within the underlined region.

fragments H and I from within the 18S rRNA. Two other putative precursors (5.6 and 5.0 kb) are detected by fragments J, K, Q, R, and S which contain the 28S rRNAs. However, only the larger of the two 28S rRNA putative precursors

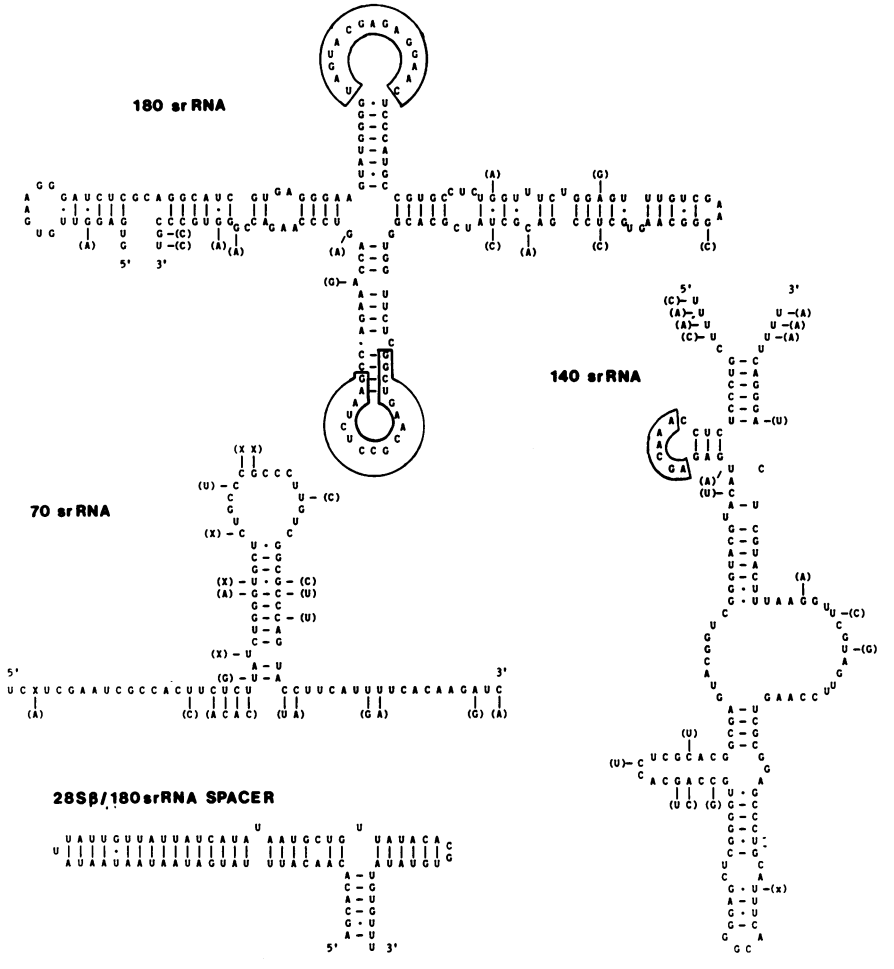


Figure 6. The secondary structures of the trypanosome 180, 70 and 140 srRNAs and the 75 nt spacer. The 180 and 140 srRNA models were constructed using the published 28S rRNA models as guides (40, 41, 42). The 70 srRNA and the 75 nt spacer models were constructed with the aid of a computer to include the largest regions of base pairing. Differences in the corresponding Crithidia sequences (39) are shown in parentheses adjacent to the structure. An x means that the nucleotide is not present in that sequence but is present in the other srRNA sequence. Regions of sequence homology to other eukaryotic RNAs are blocked in this figure and discussed in the text.

(5.6 kb) is detected by fragment I containing the 5.8S rRNA. The probes that hybridize to these large RNAs also hybridize with a hetergeneous collection of smaller RNAs, especially in culture form RNA (lanes 1). These do not comigrate with specific RNA bands in the ethidium stained gel and we attribute

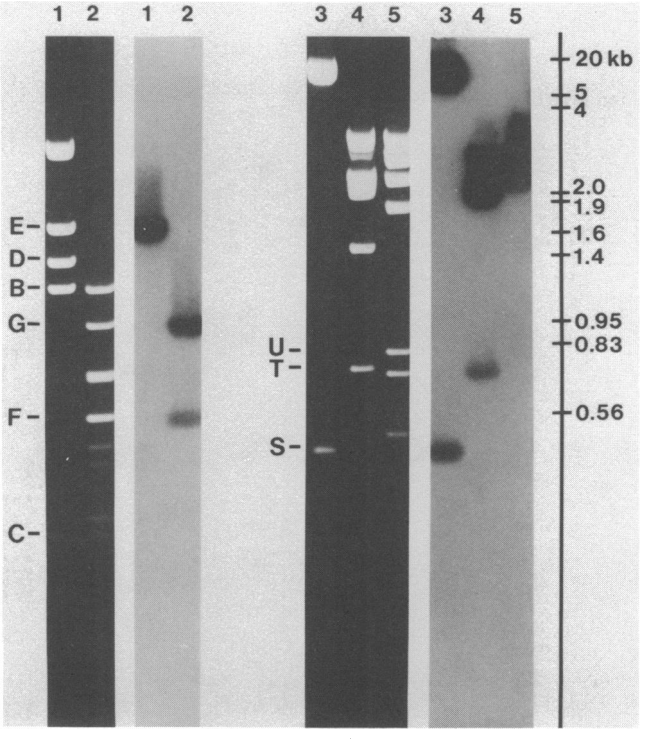


Figure 7. Hybridization of labeled nascent transcripts to restriction fragments of pR4 subclones. Lanes 1 and 2 contains 2 µg of a subclone containing fragment A (see Fig. 1). Lanes 3 to 5 contain 2 µg of a subclone containing fragment P (see Fig. 1). Digests include: lane 1 = Hinc II and Eco RI, lane 2 = Alu I, lane 3 = Bgl II and Xho I, lane 4 = Xho I and Hinc II and lane 5 = Hinc II and Eco RI. The results are summarized in Fig. 1. Specific fragments from Fig. 1 which are discussed in the text are labeled next to the ethidium bromide stain.

them to the presence of degradation products of the rRNAs and their precursors.

The end points of the putative precursors were determined by end-labeling fragments A and N, hybridizing them to total RNA and digesting the hybrids with S1 (see Table 1). The results are shown in Fig. 3 and are diagramed in Fig. 1. Total RNA yields protected fragments of 1.2 and 1.1 kb of fragment A suggesting that the 5' end of the 3.4 kb putative precursor is staggered. At the 3' end of the unit, total RNA protects 475 bp of fragment N, suggesting

**A**

CTTTCCACCC AGCGCGGGTG CATTCTGGCT CTTATATATA CTTATTGTCA TGACAGAGTA

TATTGTACTG TGTTGATAAG GGACGGGTAA CTGTATTGAA GAGCCGATGC TTTTGACATG

TTAGATATAA TATGTTTTAT TGTAAAGTCA ATACAACACA CAATAGGATA ATAATGATAA

AGTTAAAAAA GTATATATAG TAATAGAAAT ATATCTTATA TAGGAAAGAT TAAGCAGTAA

AAGTAGCGCT TACGGCGTAC GGAGCAGGAG AGCAACTGAC CGCTCTCAGA CCGATGGCGA

CGACATAAAC GCGCCTCTCG GGCTTCTGCT CTGTAAACTT TGCGCAATGC GATACGGGAA

CAGTTTTTGC GCTTGTTCAG TAGGTTGCCG AATGGCTGTG GGTATGACAC TTCCTCAGTT

AAGGGAAGGA GCGAGTGTGT GTAATACGTG TGTTGCTCTG ACAAAACGAA AGGGCACATC

GGCACGCTGC TCGTAAAGGT TTTTAATTAC CTTCACAG

**B**

| | -30 | -20 | -10 | +1 | +10 |
|---|---|---|---|---|---|
| Well conserved nts | | | G | PyPu | (Pu-rich) |
| Trypanosoma brucei | ATTAAGCAGT | AAAAGTAGCG | CTTACGGCGT | ACGGAGCAGGA | |
| Saccharomyces cerevisiae | tgtgaggAA | AAGTAGttgg | gaggtacttT | Atgcgaaagc | |
| Physarum polycephalum | atgcttctta | AAAAGaaacc | caagatacaT | Atagggggg | |
| Tetrahymena pyriformis | aaagtatcag | ggggGTAaaa | atgcatattT | Aagaagggga | |
| Acanthamoeba castellani | aaactggtcg | gaccGTccga | aagtatataT | Aaagggacgg | |

Figure 8. A. Sequence of fragment F (see Fig. 1; 518 nts) which includes the probable rRNA promoter. The 5' ends of the S1 protected fragments are indicated by arrows.

B. Sequence comparison of the putative trypanosome pol I promoter with other eukaryotic pol I promoters (43, 44).

that the 3' end of the 5.6 and 5.0 kb putative precursors is just 3' of the Xho I site, at the end of the 140 srRNA.

Fragment F, which spans the two staggered ends at the beginning of the transcription unit (see Figs. 1 and 3) was sequenced. The ends of the putative precursors were precisely located by running S1 protection analysis and sequencing analysis in adjacent lanes of a polyacrylamide gel (data not shown). The sequence is shown in Fig. 8A, with the ends of the protected fragments indicated. The 5' end of the longest protected fragment has limited homology to promoters for RNA polymerase I in other eukaryotes (Fig. 8B; 43, 44). The end of the other protected fragment has no obvious homology to pol I promoters.

Table 2. Size of Domain VII in various eukaryotes.

| Organism | Size of Domain VII[1] |
|---|---|
| Yeast | 400 |
| Physarum | 463 |
| Xenopus | 410 |
| Mouse | 487 |
| Rat | 400 |
| Trypanosome - 180,70 and 140 srRNAs (183, 77 and 135 nts) | 395 |

[1] for references, see text.

## DISCUSSION

### rRNA repeat unit

We have localized all six rRNAs which make up the 28S region of trypanosomes (see Fig. 1). Four of the rRNAs (28S α, 28S β, 180 srRNA and 140 srRNA) were localized by S1 protection analysis from an internal restriction site (see Table 1). Three of the four srRNAs (180, 70 and 140) were localized by sequence analysis of the 3' end of the 28S region (Fig. 5). The remaining 220 srRNA is located between the 28S α and 28S β (Fig. 2 and our preliminary sequence data). The incorrect localization of three of the srRNAs by Hasan et. al. (19) can be attributed to two factors. First, the clones used did not contain the true location of the 140 srRNA. Second, the low hybridization stringency employed by these authors led to hybridization of the srRNAs with several regions within the rDNA repeat. This cross hybridization may be due to long range base pairings with which the srRNAs interact with the other rRNAs in the ribosome.

### Primary sequence and secondary structure

We have compared the sequence at the 3' end of the 28S unit to the sequence of the 26S rRNA of yeast (38). As shown in Figs. 4 and 5, there is a large block of homology between the trypanosome 28S β and the yeast sequence. The homology stops six nucleotides from the end of Domain VI, a domain of secondary structure in yeast (40, 41) and Xenopus (42). Continuing 3' of this region, the only significant homology between the trypanosome and yeast 26S rRNA sequences is two blocks of homology within the 180 srRNA. The yeast sequence which is compared to the 180 srRNA starts at the beginning of Domain VII of yeast. Thus, the two yeast regions shown in Fig. 5 are contiguous to

each other, as diagrammed in Fig. 4. The junction between Domains VI and VII is a minor expansion region or region of variability in eukaryotic 28S rRNAs (45, 46). The largest variation at this junction in other eukaryotes is 27 additional nucleotides in Physarum. In trypanosomes, 192 extra nucleotides are present in the DNA including 95 nucleotides at the 3' end of the 28S β, 75 nucleotides of transcribed spacer and 22 nucleotides at the 5' end of the 180 srRNA. It is also of interest to note that the overall length of Domain VII varies between 400 and 500 nucleotides in most eukaryotes (Table 2). The sum of the three srRNAs is 395 nucleotides, close to the other eukaryotic values. The sizes of these srRNAs are based on primary sequence analysis and homology to the Crithidia sequence. We have not determined the exact ends of the trypanosome srRNAs.

We have also compared the trypanosome sequence to the Crithidia srRNAs. The percent homology between trypanosome and Crithidia are: 180 srRNAs = 91%, 70 srRNAs = 71% and 140 srRNAs = 86%. The 70 srRNAs, which are the least homologous, correspond to the central nonconserved region of Domain VII and are part of an expansion segment within the 28S rRNAs of eukaryotes (45, 46).

The primary sequence shown in Fig. 5 includes three spacers between the rRNAs, all of which are (AT)-rich, including 75 nucleotides between the 28S β and the 180 srRNA (23% GC), 322 nucleotides between the 180 and 70 srRNAs (33% GC) and 328 nucleotides between the 70 and 140 srRNAs (34% GC). The rRNAs themselves are 54-59% GC. The 75 nucleotide spacer can be folded into a stem/loop structure, potentially bringing the 28S β and the 180 srRNA into close proximity during processing (Fig. 6).

Of interest are the secondary structures of the srRNAs (Fig. 6). The 180 srRNA can be folded into a clover-leaf structure identical to the proposed secondary structure for the first part of Domain VII of many eukaryotes (40, 41, 42, 47, 48). The blocks of homology between the 180 srRNA and the yeast 26S rRNA occur in the top and bottom loops of the structure (boxed regions in the figure) and these regions are conserved even in E. coli (49, 50, 51). In yeast, the top loop is cleaved by α sarcin which suggests that it has a role (along with Domain VI) in the ribosomal A site (40). It is also of interest to note the similarity of this secondary structure (and the homology of the top loop) to the U1 RNA (52). The first 23 nucleotides of the 180 srRNA do not have a counterpart in Domain VII of yeast and may be involved in long range interactions with other rRNAs. Alternatively, these 23 nucleotides can form a stem/loop structure which brings the 5' and 3' ends of the 180 srRNA into close proximity. This closed structure may exist, at least in vitro

since we have observed an *in vitro* RNA ligase activity which specifically forms circular 180 srRNA (unpublished results).

The 70 srRNA forms a stem-loop structure but there is no obvious homology to the middle section of Domain VII of other organisms, which is not unexpected since this region is an expansion segment in eukaryotic 28S rRNAs. The 140 srRNA can be folded into a secondary structure which resembles the structures at the 3' end of Domain VII of several eukaryotes (40, 42, 48) including a stem-loop structure in which the loop sequence includes 5'-A-N-A-C-G-A-3' (boxed region) which is conserved even in *E. coli* (49) and in chloroplast 4.5S rRNA (see below).

Further support for these proposed secondary structures comes from the corresponding *Crithidia* sequences (shown in parentheses in Fig. 6). Most of the substitutions are within unpaired regions of the structures. There are pairs of compensating base changes in the right arm of the 180 srRNA structure and in the stem-loop of the 70 srRNA structure. The most remarkable set of substitutions is at the 5' and 3' ends of the 140 srRNA in which 7 of 8 U's are replaced by 5 A's and 2 C's; the significance of these differences is not obvious.

Transcription unit

The transcription unit of the rRNA of trypanosomes has been defined both by "run-on" transcription in isolated nuclei (Fig. 1 and 7) and by S1 protection analysis (Fig. 3). Transcription appears to start 1.2 kb in front of the 18S rRNA. The S1 analysis of this region (Fig. 3) shows two S1 protected fragments. These could result from polymorphisms in the ribosomal repeat units, duplicate promoters or processing of the primary transcript. It is unlikely that two promoters are active since sequence analysis of this region (Fig. 8A) does not show a duplication. The sequence analysis does show a region with limited homology to other eukaryotic promoters surrounding the 5' end of the putative precursor 1.2 kb upstream of the 18S rRNA (Fig. 8B). Since the 5' end of the putative precursor 1.1 kb upstream of the 18S rRNA has no homology to other eukaryotic promoters, it could be an S1 artifact, a DNA polymorphism or a processing site. Processing sites located several hundred nucleotides downstream of the promoter are found in many organisms including mouse (53), human (54), *Physarum* (55) and maize (56).

Transcripts extending 3' of the 140 srRNA are not detectable in stable RNA. S1 protection of fragment N shows that the 5.6 and 5.0 kb putative precursors end at the 3' end of the 140 srRNA. Furthermore, nascent RNA hybridizes to fragment T which contains the last 70 nts of the 140 srRNA and

does not hybridize to the adjacent fragment U. The stoichiometry of the hybridization to fragments S and T in Fig. 7 suggest that transcription terminates somewhere near the center of fragment T. From the S1 results of Fig. 3, we conclude that rapid processing must be occurring at the 3' end of the 140 srRNA, generating the 3' ends of the 28S rRNA putative precursors.

Processing

The entire transcription unit from promoter to terminator is approximately 10 kb (Fig. 1). A possible primary transcript of about this size is present in RNA blots (Fig. 2). The first processing of this primary transcript apparently occurs 5' of the 5.8S rRNA, generating putative precursors of 3.4 kb (containing the 18S rRNA) and 5.6 kb (containing the 5.8S rRNA and the 28S region). The 3.4 kb precursor is presumably processed to yield the 18S rRNA, while the 5.6 kb precursor is processed to yield the 5.8S rRNA and another putative precursor of 5.0 kb containing the 28S region. These initial processing steps are similar to rRNA processing in many other organisms (1). The processing of the 28S region into six different rRNAs must occur after these initial processing steps since all probes from the 28S region hybridize to the 5.6 and 5.0 kb putative precursors (Fig. 2). Furthermore, the Crithidia srRNAs all have a 5' monophosphate and a 3' hydroxyl, suggesting that srRNAs are processed from larger precursors (39). More direct evidence for this tentative conclusion has recently come from two types of experiments: in pulse-labeling experiments, the mature srRNAs (but not the 5S RNA) are only labeled after a lag period (P. Laird, J. Zomerdijk and P. Borst, unpublished results); during UV-inactivation of transcription, the srRNAs (but not the 5S RNA) behave as processing products of a single 10 kb transcription unit (P. Johnson and P. Borst, unpublished results).

The processing of the 28S rRNA region into six different rRNAs appears to be unique to the trypanosomatids. Some other organisms split their 28S rRNA into two halves (2, 3, 4, 5, 6, 57). However, none of these organisms have an srRNA encoded within this gap, as is the case with the 220 srRNA in trypanosomes. Even more unusual is the processing of three srRNAs from their surrounding spacers at the 3' end of the region, resulting in a functional Domain VII. The structure of this domain supports the theory that the expansion segments or variable regions within the rRNAs are evolutionarily related to the transcribed spacers (58, 59). This suggests that either large rRNAs were constructed from smaller rRNAs each possessing a specific function or that larger rRNAs can be broken apart into smaller rRNAs each able to function separately from the remaining large rRNA. Several types of rRNA

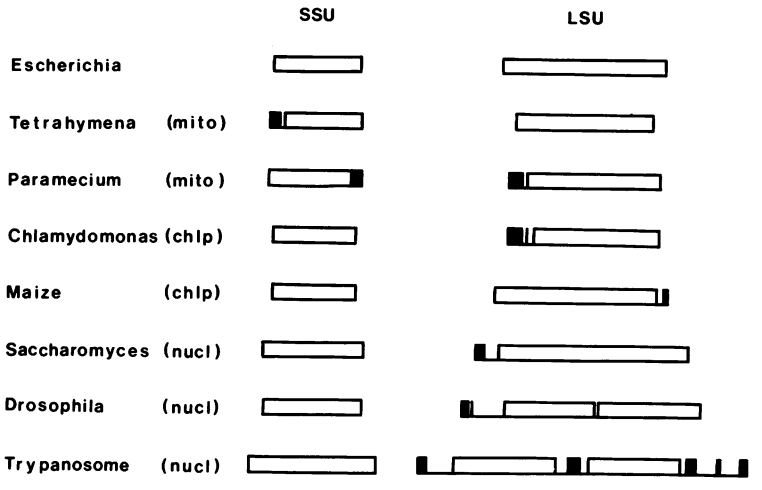


Figure 9. Schematic diagram of different rRNA processing systems. The small subunit rRNA (SSU, "16S"-like rRNA) and large subunit rRNA (LSU, "23S"-like rRNA) are shown for various organisms (for references, see text). Small rRNAs are shown as black boxes. Transcribed spacers are shown as lines. The figure does not attempt to show the size of the spacer between the SSU and LSU rRNAs.

processing systems are compatible with these theories (Fig. 9). In eukaryotic nuclei (e.g. Saccharomyces), the 5.8S rRNA, which is separated from the 28S rRNA by a transcribed spacer, corresponds to the 5' end of the 23S rRNA of prokaryotes (60, 61). Furthermore, the 5.8S rRNA region of dipteran insects (e.g. Drosophila) is composed of a short 5.8S rRNA and a 2S rRNA separated by a short (AT)-rich spacer which is not part of the 5.8S rRNA structure (see Fig. 9; 10, 11).

Analysis of the rRNAs of eukaryotic organelles also supports the theory that the transcribed spacers and the expansion segments are related (see Fig. 9). The small subunit rRNA (SSU rRNA or "16S like") of Tetrahymena mitochondria is composed of two rRNAs, α (200 nts) and β (14S rRNA) (59), separated by a 54 nucleotide (AT)-rich spacer. Comparison of the secondary structure models shows that the spacer is present in other SSU rRNAs as an expansion segment. A similar situation may exist at the 3' end of the SSU rRNA in Paramecium mitochondria (62).

The rRNAs of organelles do not contain a 5.8S rRNA. However, the rRNA unit of chloroplasts in Chlamydomonas contains two srRNAs, 7S (282 nts) and 3S (47 nts) which are encoded between the SSU and LSU (large subunit rRNA or "23S like"). The 7S and 3S rRNAs are separated by an (AT)-rich spacer and are

homologous to the 5.8S rRNA of eukaryotic nuclei and the 5' end of LSU rRNAs (63). Paramecium mitochondria also contain an srRNA (283 nts) between the SSU and LSU rRNAs which is separated from the LSU rRNA by a 19-26 nucleotide (AT)-rich spacer (64).

Finally, the 3' end of the chloroplast LSU rRNA of higher plants is present as a separate 4.5S rRNA (see Fig. 9; 65, 66). This processing system has several analogies to the trypanosome system. In maize, the 4.5S rRNA is separated from the rest of the LSU rRNA by a 93 nucleotide spacer which can be folded so that the LSU rRNA and the 4.5S rRNA are brought into close proximity, possibly for processing (67). A similar folding of the spacer occurs between the 28S β and 18S srRNA as shown in Fig. 6. The 4.5S rRNA also has a secondary structure similar to the last part of Domain VII of the LSU rRNA, and is highly analogous in size and secondary structure to the 14S srRNA of trypanosomes. Although all the processing systems discussed above have similarities to the trypanosome system, the processing of the 28S rRNA region of trypanosomes into six different rRNAs is by far the most complex processing system described to date.

Acknowledgements

We thank Dr. Wendy Gibson for providing RNA blots and Dr. Paul Sloof, Section for Molecular Biology, Laboratory of Biochemistry, University of Amsterdam, for providing the cosmid cR1 and the subclone pR4. We thank Drs. P. Sloof and C. Klootwijk, University of Amsterdam, for critical reading of the manuscript and our colleagues for their advice and comments during this work. This work was supported by The Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for the advancement of Pure Research (ZWO). T.C.W. was supported by a fellowship from the National Institute of Allergy and Infectious Diseases, N.I.H., U.S.A.

References

1. Perry, R.P. (1976) Ann. Rev. Biochem. 45, 605-629.
2. Shine, J. and Dalgarno, L. (1973) J. Mol. Biol. 75, 57-72.
3. Jordan, B.R., Jourdan, R. and Jacq, B. (1976) J. Mol. Biol. 101, 85-101.
4. Ware, V.C., Renkawitz, R. and Gerbi, S.A. (1985) Nucl. Acids Res. 13, 3581-3597.
5. Lava-Sanchez, P.A. and Puppo, S. (1975) J. Mol. Biol. 95, 9-20.
6. Leipoldt, M. and Schmidtke, J. (1982) in Genome Evolution. Dover, G.A. and Flavell, R.B. (eds) Academic Press, N.Y., N.Y., pp. 219-236.
7. Reijnders, L., Sloof, P., Sival, J. and Borst, P. (1973) Biochim. Biophys. Acta. 324, 320-333.
8. Stevens, A.R. and Pachler, P.F. (1972) J. Mol. Biol. 66, 225-237.
9. Eckert, W.A., Kaffenberger, W., Krohne, G. and Franke, W.W. (1978) Eur. J. Biochem. 87, 607-616.

10. Pavlakis, G.N., Jordan, B.R., Wurst, R.M. and Vournakis, J.N. (1979) Nucl. Acids Res. 7, 2213-2238.
11. Jordan, B.R., Latil-Damotti, M. and Jourdan, R. (1980) Nucl. Acids Res. 8, 3565-3573.
12. Castro, C., Hernandez, R. and Castaneda, M. (1981) Mol. Biochem. Parasitol. 2, 219-233.
13. Garvin, R.T., Hill, R.C. and Weber, M.M. (1978) Archiv. Biochem. Biophys. 191, 774-781.
14. Leon, W., Fouts, D.L., and Manning, J. (1978) Nucl. Acids Res. 5, 491-504.
15. Villalba, E. and Ramirez, J.L. (1982) J. Protozool. 29, 438-443.
16. Cordingley, J.S. and Turner, M.J. (1980) Parasitology 81, 537-551.
17. Gray, M.W. (1981) Mol. Cell. Biol. 1, 347-357.
18. Cordingley, J.S. and Turner, M.J. (1980) Mol. Biochem. Parasitol. 1, 91-96.
19. Hasan, G., Turner, M.J. and Cordingley, J.S. (1984) Gene 27, 75-86.
20. Hernandez, R., Nava, G., and Castaneda, M. (1983) Mol. Biochem. Parasitol. 8, 297-304.
21. Van der Ploeg, L.H.T., Valerio, D., de Lange, T., Bernards, A., Borst, P. and Grosveld, F.G. (1982) Nucl. Acids Res. 10, 5905-5923.
22. Kooter, J.M. and Borst, P. (1984) Nucl. Acids Res. 12, 9457-9472.
23. Maniatis, T., Fritsch, E.F. amd Sambrook, J. (eds) (1982) in: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor N.Y.
24. Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721-732.
25. Bernards, A., van der Ploeg, L.H.T., Frasch, A.C.C., Borst, P., Boothroyd, J.C., Coleman, S. and Cross, G.A.M. (1981) Cell 27, 497- 505.
26. Maxam A.M. and Gilbert, W. (1980) In Methods in Enzymology, vol. 65. L. Grossman and K. Moldave, eds. (New York: Academic Press) pp.499-560.
27. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
28. Biggins, M.D., Gibson, T.J., and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3965.
29. Dente, L., Cesareni, G. and Cortese, R. (1983) Nucl. Acids Res. 11, 1645-1655.
30. Auffray, C. and Rougeon, F. (1980) Europ. J. Biochem. 107, 303-314.
31. Thomas, P.S. (1980) Proc. Natl. Acad. Sci. USA. 77: 5201-5205.
32. Rigby, P.W.J., Dieckman, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
33. Jeffreys, A.J. and Flavell, R.A. (1977) Cell 12, 429-439.
34. Kooter, J.M., de Lange, T. and Borst, P. (1984) EMBO Journal 3, 2387-2392.
35. Hasan, G., Turner, M.J., and Cordingley, J.S. (1982) Nucl. Acids Res. 10, 6747-6762.
36. Sogin, M.L., Elwood, H.J. and Gunderson, J.H. (1986) Proc. Natl. Acad. Sci. USA 83, 1383-1387.
37. Dorfman, D.M., Lenardo, M.J., Reddy, L.V., Van der Ploeg, L.H.T. and Donelson, J.E. (1985) Nucl. Acids Res. 13, 3533-3549.
38. Georgiev, O.I., Nikolaev, N., Hadjiolov, A.A., Skryabin, K.G., Zakharyev, V.M. and Bayev, A.A. (1981) Nucl. Acids Res. 9, 6953-6958.
39. Schnare, M.N., Spencer, D.F. and Gray, M.W. (1983) Can. J. Biochem. Cell. Biol. 61, 38-45.
40. Veldman, G.M., Klootwijk, J., de Regt, V.C.H.F., Planta, R.J., branlant, C., Krol, A., and Ebel, J-P. (1981) Nucl. Acids Res. 9, 6935-6952.
41. Hogan, J.J., Gutell, R.R. and Noller, H.F. (1984) Biochemistry 23, 3330-3335.
42. Clark, C.G., Tague, B.W., Ware, V.C. and Gerbi, S.A. (1984) Nucl. Acids Res. 12, 6197-6220.

43. Balzi, E., DiPietro, A., Goffeau, A., van Heerikhuizen, H. and Klootwijk, J. (1985) Gene 39, 165-172.
44. Kownin, P., Iida, C.T., Brown-Shimer, S. and Paule, M.R. (1985) Nucl. Acids Res. 13, 6237-6248.
45 Ware, V.C., Tague, B.W., Clark, C.G., Gourse, R.L., Brand, R.C. and Gerbi, S.A. (1983) Nucl. Acids Res. 11, 7795-7817.
46. Hassouna, N., Michot, B. and Bachellerie, J-P. (1984) Nucl. Acids Res. 12, 3563-3583.
47. Hadjiolov, A.A., Georgiev, O.I., Nosikov, V.V., and Yavachev, L.P. (1984) Nucl. Acids Res. 12, 3677-3693.
48. Michot, B., Hassouna, N. and Bachellerie, J-P. (1984) Nucl. Acids Res. 12, 4259-4279.
49. Branlant, C., Krol, A., Machatt, M.A., Pouyet, J., Ebel, J-P., Edwards, K. and Kossel, H. (1981) Nucl. Acids Res. 9, 4303-4324.
50. Noller, H.F. (1984) Ann. Rev. Biochem. 53, 119-162.
51. Maly, P. and Brimacombe, R. (1983) Nucl. Acids Res. 11, 7263-7286.
52. Steitz, J.A., Wolin, S.L., Rinke, J., Pettersson, I., Mount, S.M., Lerner, E.A., Hinterberger, M. and Gottlieb, E. (1983) Cold Spring Harbor Symp. Quant. Biol. 47, 893-900.
53. Miller, K.G. and Sollner-Webb, B. (1981) Cell 27, 165-174.
54. Financsek, I., Mizumoto, K., Mishima, Y. and Muramatsu, M. (1982) Proc. Nat. Sci. U.S.A. 79, 3092-3096.
55. Blum, B., Pierron, G., Seebeck, T. and Braun, R. (1986) Nucl. Acids Res. 14, 3153-3166.
56. McMullen, M.D., Hunter, B., Phillips, R.L. and Rubenstein, I. (1986) Nucl. Acids Res. 14, 4953-4968.
57. Wellauer, P.K. and Dawid, I.B. (1977) Cell 10, 193-212.
58. Cox, R.A. and Kelly, J.M. (1981) FEBS Letters 130, 1-6.
59. Schnare, M.N., Heinonen, T.Y.K., Young, P.G., and Gray, M.W. (1986) J. Biol. Chem. 261, 5187-5193.
60. Nazar, R.N. (1980) FEBS Letters 119, 212-214.
61. Jacq, B. (1981) Nucl. Acids Res. 9, 2913-2932.
62. Seilhamer, J.J., Olsen, G.J. and Cummings, D.J. (1984) J. Biol. Chem. 259, 5167-5172.
63. Rochaix, J.D. and Darlix, J.L. (1982) J. Mol. Biol. 159, 383-395.
64. Seilhamer, J.J., Gutell, R.R. and Cummings, D.J. (1984) J. Biol. Chem. 259, 5173-5181.
65. MacKay, R.M. (1981) FEBS Letters 123, 17-18.
66. Edwards, K. and Kossel, H. (1981) Nucl. Acids Res. 9, 2853-2869.
67. Strittmatter, G. and Kossel, H. (1984) Nucl. Acids Res. 12, 7633-7647.