Volume 10 Number 5 1982 Nucleic Acids Research

# The nucleotide sequence at the transcription termination site of the ribosomal RNA gene in Tetrahymena thermophila

Nanni Din*, Jan Engberg+ and Joseph G.Gall*

*Department of Biology, Yale University, New Haven, CT 06511, USA, and +Department of Biochemistry B, Panum Institute, University of Copenhagen, Copenhagen, Denmark

Received 23 December 1981; Accepted 5 February 1982

ABSTRACT

The sequence of 415 nucleotides surrounding the transcription termination site for ribosomal RNA in Tetrahymena thermophila has been determined. The positions of the 3'-ends of mature 26S rRNA, pre-26S rRNA and 35S pre-rRNA were localized within this sequence by hybridization of the purified RNA species to selected DNA fragments, followed by S1 nuclease treatment of the hybrid and a precise sizing of the RNA-protected DNA fragments on sequencing gels. The 35S pre-rRNA population contained molecules with two distinct 3'-ends, one of which is identical to the end of pre-26S and 26S rRNA, while the other corresponds to a position 15 nucleotides further downstream, which is assumed to be the transcription termination site. The non-coding DNA strand contains a cluster of T's at the putative termination site, and several other T clusters are found further downstream. A short inverted repeat sequence is located near the putative termination site within the transcribed region. The possible role of these structures for transcription termination is discussed.

## INTRODUCTION

In the ciliated protozoan, Tetrahymena, ribosomal RNA is transcribed from two coding regions located as inverted repeats on extrachromosomal palindromic DNA molecules with a size of approximately 20 kb (1-3). Mature 5.8S, 17S and 26S rRNA are generated from a common 35S pre-rRNA precursor by several steps of nucleolytic cleavage (4-7), including one splicing event (8,9). We have recently presented a complete transcription map of Tetrahymena thermophila rDNA, in which we located the map position of the 3'-end of 35S pre-rRNA, using a combination of R-loop mapping, Southern-blotting and, to a limited extent, S1-nuclease mapping (11). Here we report the nucleotide sequence in this region and the precise localization of the 3'end of the 35S pre-rRNA within this sequence, thus identifying the putative transcription termination site for T. thermophila rDNA.





0305-1048/82/1005-1503$2.00/0

## MATERIALS AND METHODS

Isolation of rDNA restriction fragments. The recombinant plasmid pRP14 used in this study has been described previously (11). Plasmid DNA was isolated by the cleared lysate method (12), and after restriction enzyme digestion performed as recommended by the suppliers (BRL and New England Biolabs), DNA fragments were isolated from low-melting agarose gels (Seaplaque, Marine Colloids) (13) or from polyacrylamide gels (14).

Isolation of RNA. For isolation of total nuclear RNA or the nuclear precursors to mature 26S rRNA, 35S pre-rRNA and pre-26S rRNA, Tetrahymena macronuclei were isolated as described by Cech & Rio (8). The RNA was extracted from the isolated nuclei, and, for preparation of the purified precursors, separated by sucrose gradient centrifugation, as described by Eckert et al.(6). Mature 26S rRNA was isolated from whole cells using the same method.

DNA labelling and sequence analysis. 5'-ends of DNA restriction fragments were labelled by the aid of $\gamma$-$^{32}$P-ATP (NEN or Amersham) and T4 polynucleotide kinase (PL-biochemicals) (14). 3'-end labelling of recessed 3'-ends was achieved by using the Klenow fragment of E. coli DNA polymerase I (Boehringer) and the appropriate $\alpha$-$^{32}$P-dNTP under conditions as described (15). DNA sequencing was carried out by the base specific chemical cleavage method and the cleaved fragments were electrophoresed in 0.4 mm thick 20%, 10% or 8% polyacrylamide gels containing 7 M urea (14).

S1 nuclease mapping. This was done as described in detail recently (9), except that varying amounts (5-250U) of S1 nuclease were used, and that the protected fragments were analysed on thin 7 M urea-containing polyacrylamide gels, as used for DNA sequence analysis.

## RESULTS

Mapping of the 3' ends of pre-rRNA, pre-26S and 26S rRNA. Fig. 1 (a) shows the transcription map of T. thermophila rDNA with the recognition sites for the restriction enzyme Hind III indicated. Previously, we had mapped the 3'-ends of pre-rRNA, pre-26S and 26S rRNA at or very close to the Hind III site 8 kb from the symmetry axis of the rDNA molecule (11). The following S1-mappings localized the 3'-ends of all three RNA species within the .42 kb Hind III fragment situated immediately to the right of the Hind III site at map pos. 8 kb. The .42 kb Hind III fragment was 3'-end labelled and samples of the labelled fragment were hybridized separately to 26S, pre-26S and 35S rRNA. The RNA-DNA hybrids were treated with S1 nuclease and the protected fragments were electrophoresed in a urea-containing polyacrylamide gel and

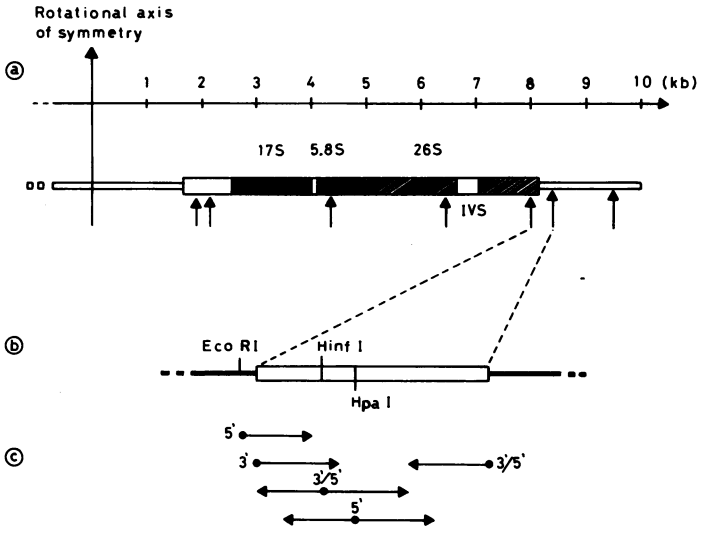


*Fig. 1. Sequencing strategy for the transcription termination region (a): Transcription map of rDNA. The thin box represents the non-transcribed part of rDNA, and the wide box the transcribed part with the coding regions for the mature rRNA species (17S, 5.8S and 26S rRNA) indicated by hatching. IVS: intervening sequence. The arrows indicate Hind III restriction sites. Only half of the palindromic rDNA molecule is shown. (b): The .42 kb Hind III rDNA fragment (open box) containing the transcription termination region, cloned into pBR322 (solid line). Vertical lines indicate restriction sites used for end-labelling and subsequent sequence analysis as indicated in (c).*

visualised by autoradiography. The results are shown in Fig. 2. 26S (lanes a & b) and pre-26S (lane c) rRNA both protect only one fragment of about 140 bp, which locates the 3'-ends of these two RNA species about 140 bp downstream from the Hind III site 8 kb from the symmetry axis of the rDNA molecule, while 35S pre-rRNA (lane e), as well as total nuclear RNA (lane f) gave rise to two protected bands, one of the same size as the 26S rRNA protected fragment and another one slightly larger. This shows that a major part of the pre-rRNA molecules contains about 10-20 extra nucleotides at the 3'-end compared to mature 26S rRNA. The fact that the S1 experiment using the total nuclear RNA preparation gave rise only to the same protected bands as isolated 35S pre-rRNA and no larger ones, shows that no transcripts located further downstream that those found in 35S pre-rRNA could be detected.

A more precise mapping of the 3'-end of the three different isolated RNA species was obtained by analysis of the protected DNA fragments on a sequencing gel alongside the sequence ladder of the corresponding non-S1 treated

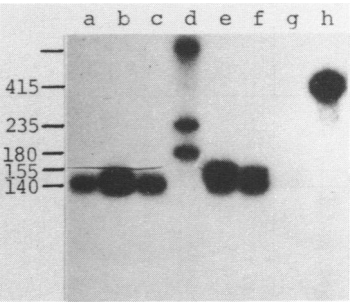


*Fig. 2. S1 nuclease mapping of the 3'-ends of pre-rRNA, pre-26S and 26S rRNA within the .42 kb Hind III rDNA fragment. The 3'-end labelled .42 kb Hind III fragment was hybridized to the purified rRNA species or to total nuclear RNA, treated with S1 nuclease and the protected DNA fragments separated on an 8% polyacrylamide - 7M urea gel. Lane a: 26S rRNA, 10 U S1, lane b: 26S rRNA, 50 U S1, lane c: pre 26S rRNA, 50 U S1, lane d: Hind III/HpaI digested pRP14 DNA used as size marker, lane e: 35S pre-rRNA, 50 U S1, lane f: total nuclear RNA, 50 U S1, lane g: no RNA, 50 U S1, lane h: untreated .42 kb Hind III fragment.*

DNA fragment. The result of this is shown in Fig. 3. In panel a, the sequence ladder was obtained by chemical cleavage using the 180 bp Hind III/Hpa I fragment (see Fig. 1b) of the .42 kb Hind III fragment, 3'-end labelled only at the Hind III site. In panel b a more spread-out sequence ladder in the region of the 3'-end unique for the 35S pre-rRNA is shown. This sequence ladder, as well as the S1-protected fragments shown alongside, was obtained using the 60 bp Hinf I/Hpa I fragment (Fig. 1b), 3'-end labelled only at the Hinf I site. It is seen that the S1 treatment usually gave rise to a short ladder of fragments, differing in size by one nucleotide, instead of only one band. The failure of S1 to generate one single band corresponding exactly to the protected region probably reflects the fact that S1 may stop one or two bases from the hybrid region or may nibble into the RNA-protected DNA strand, possibly due to "breathing" of the hybrid. The position of the S1 ladder relative to the actual end of the protected region is dependent on the amount of S1 nuclease used. In the experiments shown in panel a of Fig. 3, 50 U S1 were used throughout under conditions as described in methods, except in lane e where only 10 U S1 were used. In panel b, 5, 50 and 250 U were used in lanes a, b and c resp. When a correction is made for the different mobilities of an S1 generated fragment and one generated by chemical cleavage (which is one nucleotide shorter), it can be deduced that the last nucleotide protected by 26S rRNA is one of the six nucleotides between $A_{140}$ (indicated by the lower arrow in panel a) and $G_{145}$, where the numbering refers to the

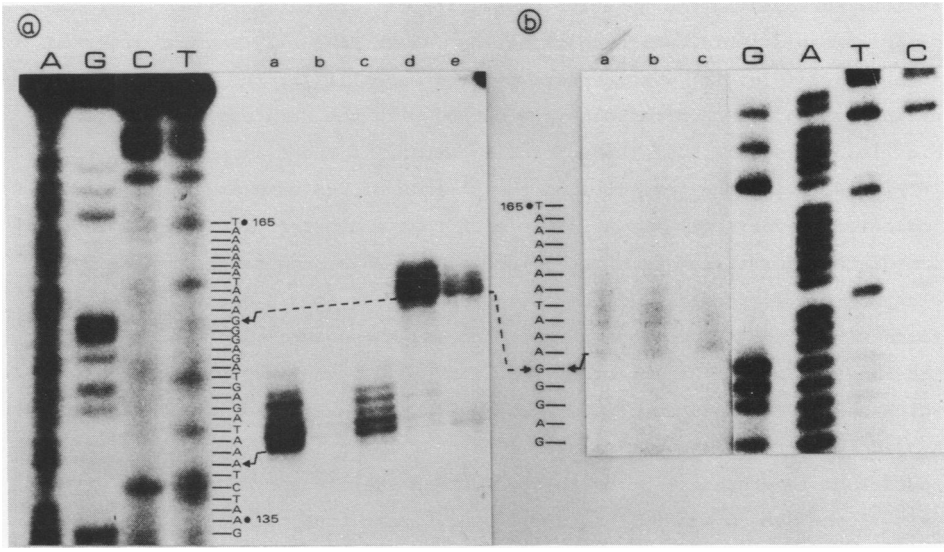


*Fig. 3. Precise mapping of the 3'-ends of pre-rRNA, pre-26S and 26S rRNA by sizing the S1 protected fragments on a sequencing gel. (a): The 180 bp Hind III/HpaI fragment of the cloned .42 kb Hind III rDNA fragment was labelled at the 3'-end of the Hind III site using α-$^{32}$P-dATP and E. coli DNA pol I (Klenow fragment), and sequenced. In parallel, part of the terminally labelled fragment was hybridized to the purified rRNA species and treated with S1 nuclease. The protected fragments were denatured and electrophoresed in an 8% sequencing gel, along with the fragments generated by the 4 base-specific chemical cleavage reactions (A,G,C,T). Lane a: 26S rRNA, 50 U S1, lane b: no RNA, 50 U S1, lane c: pre-26S rRNA, 50 U S1, lane d: 35S pre-rRNA, 50 U S1, lane e: 35S pre rRNA, 10 U S1. (b): The same exp. as in a, except that the 60 bp HinfI/HpaI fragment 3'-end labelled at the HinfI site was used, and the DNA fragments analysed on a 20% sequencing gel. Lane a: 35S pre-rRNA, 5 U S1, lane b: 35S pre-rRNA, 50 U S1, lane c: 35S pre-rRNA, 250 U S1.*

position of the nucleotide counting from the start of the Hind III site. Pre-26S rRNA protects the same range of fragments, as expected from the results shown in Fig. 2, although the band corresponding to $A_{140}$ is very faint (panel a, lane c). The most prominent bands protected by 35S pre-rRNA terminate with the nucleotides between $G_{154}$ and $T_{158}$, with fainter but quite distinctive bands around pos. 141-145, corresponding to the 26S and pre-26S rRNA protected bands (panel a, lanes d & e). Increasing amounts of S1 nuclease, as used for the experiments shown in panel b, results in the appearance or only one band in the upper region, namely at pos. $G_{154}$.

Nucleotide sequence of the transcription termination region. The sequence of the entire .42 kb Hind III fragment was determined in order to examine the

regions surrounding the putative transcription termination site. The recombinant plasmid pRP14, which contains the .42 kb Hind III fragment inserted into pBR322 (Fig. 1b) was used as starting material for the sequence determination, using the chemical cleavage method with the strategy depicted in Fig. 1c. The sequence of the non-coding strand of the fragment is presented in Fig. 4., with the positions of the 3'-ends of 26S rRNA and 35S pre-rRNA indicated by vertical arrows. It is seen that a cluster of 6T's is located immediately downstream from the 3'-end of pre-rRNA, and several other long T clusters are found still further downstream. Such long T clusters are not found within the transcribed part of the .42 kb sequence, nor indeed within the RNA-like strand of any other region of the transcribed part of rRNA, which has so far been sequenced (about 40%, N. Din; N. Kan, unpublished data). A dyad symmetry region with a stem length of 13 bp (allowing 4 mismatches) and a loop length of 11 bp is located within the transcribed region, with the axis of the dyad symmetry located 26 bp upstream from the putative transcription termination site. The symmetry region is marked by horizontal arrows above the sequence line in Fig. 4. Another symmetry region, in which about half the nucleotides in pos. 130-155 can form a stem with nucleotides in pos. 170-195, is marked by underlining the sequence. In the region downstream from the 3'-end of pre-rRNA several symmetry regions can be found, in which the stems are formed mostly by basepairing between T and A clusters.

## DISCUSSION

We have mapped the 3'-ends of 35S pre-rRNA, pre-26S and mature 26S rRNA on

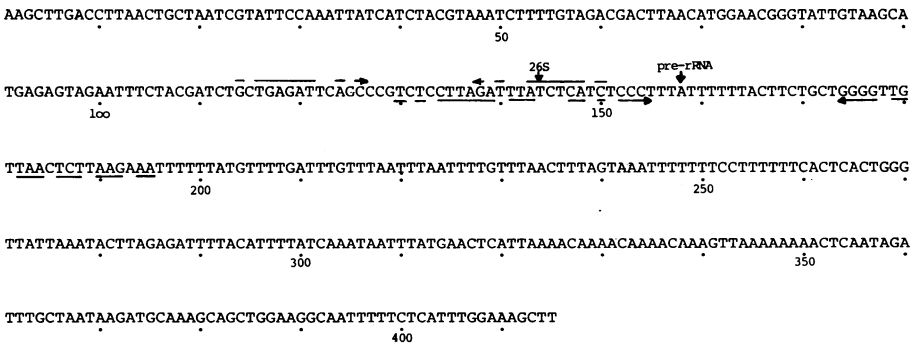


*Fig. 4. Nucleotide sequence of the non-coding strand around the transcription termination site in T. thermophila rDNA. The 3'-terminal nucleotides of 26S rRNA and pre-rRNA are indicated by vertical arrows. The horizontal arrows indicate dyad symmetry regions.*

the macronuclear ribosomal RNA genes of T. thermophila, and determined the nucleotide sequence of a 415 bp region surrounding these sites. The data show that 35S pre-rRNA is 15 nucleotides longer than pre-26S and mature 26S rRNA, which imply that 3' end processing is a necessary step in the maturation of pre-rRNA in Tetrahymena. 3'-end processing of pre-rRNA has also been shown to occur in the yeast S. carlsbergensis, where 7 nucleotides are removed (20), and in mouse and rat hepatoma cells, where the exact extent of the processing is undetermined but is less than 100-200 bp (21). In contrast, no 3'-terminal processing seems to occur during the maturation of pre-rRNA of two other species, Drosophila melanogaster (22) and Xenopus laevis (22).

In itself, S1 mappings of pre-rRNA 3'-ends do not ensure that the sites for transcription termination have been identified. To allow such a conclusion, it is essential to show that the pre-rRNA used is a true primary transcript and not a processing intermediate. Rigorous proof to that effect does not exist for Tetrahymena pre-rRNA, nor, to the best of our knowledge, for any other examined pre-rRNA species. However, the following observations strongly indicate that the Tetrahymena 35S pre-rRNA is indeed a true primary transcript. i.e. a direct product of transcription termination. Leer et al (10) have identified a protein factor, which is essential for the production of 35S pre-rRNA in isolated Tetrahymena nucleoli. If this protein is removed (e.g. by salt washing or detergent treatment), a longer pre-rRNA molecule, which extends far into the region 3' distal to the 35S pre-rRNA, is produced. Addition of the purified protein factor to salt-washed nucleoli engaged in transcription, results in the production of the normal 35S pre-rRNA, while addition of the protein factor to completed extra-long pre-rRNA has no effect. Thus, the protein factor is not a processing enzyme, but seems to be a transcription termination factor, and these results further imply that 35S pre-rRNA is a direct product of transcription termination. In the present paper, we have tried to detect in vivo transcribed sequences 3' distal to those found in 35S pre-rRNA by hybridisation of total nuclear RNA to the relevant rDNA fragments, but with negative result. These two lines of evidence lead us to believe that the map position of the distal 3'-end of 35S pre-rRNA coincides with the transcription termination site.

The proposed transcription termination site of T. thermophila rDNA exhibits features which are present in termination sites for other RNA polymerases. Thus, the rho-independent transcription termination sites of procaryotic RNA polymerases are characterized by the presence of T clusters preceeded by a dyad symmetry region (16). As pointed out above (Fig. 4) these features are

also found in Tetrahymena rDNA. The eucaryotic RNA pol III which transcribes 5S and tRNA genes as well as genes for other small RNA's, also terminates transcription in a stretch of T's (17-19); no other structural features have been shown to be involved in the pol III termination. However, these characteristics do not apply to all putative rRNA transcription termination sites. Fig. 5 shows a comparison of the 3'-terminal sequences of pre-rRNA and/or mature L-rRNA and the neighbouring non-transcribed spacer regions among 6 species (Tetrahymena thermophila, E. coli (24), Neurospora crassa (25), Saccharomyces carlsbergensis (20), D. melanogaster (22) and X. laevis (23)). Among all the listed sequences a relatively high degree of homology within the mature rRNA sequence is evident but right at the putative transcription termination sites the homology is limited, with one exception. Yeast and Tetrahymena both exhibit an almost identical dodecanucleotide sequence, TTTTTTA$^{T}/_{C}$TTCT, around the 3'-end of pre-rRNA.

In yeast, this sequence is located immediately next to the mature rRNA sequence (in fact the first T of this sequence is preserved as the last nucleotide of mature 26S rRNA of yeast) while in Tetrahymena the identical sequence (except for a substitution of $T_8$ by C) is found 15 bp downstream from the 3'-end of mature rRNA. More than half of the dodecanucleotide sequence (TTTTTTAT) is present in the pre-rRNA molecule of yeast, while, so far as our data go, the pre-rRNA of Tetrahymena terminates with the nucleotide immediately in front of the common sequence. However, since our data stems from S1 nuclease mapping, the nibbling effect of S1 nuclease may have led us to map the 3'-end a few nucleotides too far upstream. If the pre-rRNA is indeed longer than estimated by our S1 mappings, it must contain a stretch of U's at the end, which would not base-pair very strongly with the A-stretch of the coding DNA strand, and this might therefore be susceptible to S1 nuclease. In any case, it is striking that the same dodecanucleotide is present at the presumptive termination site in two different species,, and it seems likely that it is involved as a recognition site for transcription termination. In both the Tetrahymena and yeast rRNA gene (20), and to a lesser extent also in the Xenopus rRNA gene (23), an extended dyad symmetry region is found immediately surrounding and downstream from the termination site, and this has lead to the proposal that sequences downstream from the termination site are involved in the termination process (20). In Tetrahymena this dyad symmetry region is formed by bases in pos. 130-155 and 170-195 (numbered according to Fig. 4).

The 3' terminal region of the 26S and 35S pre-rRNA coding region and the

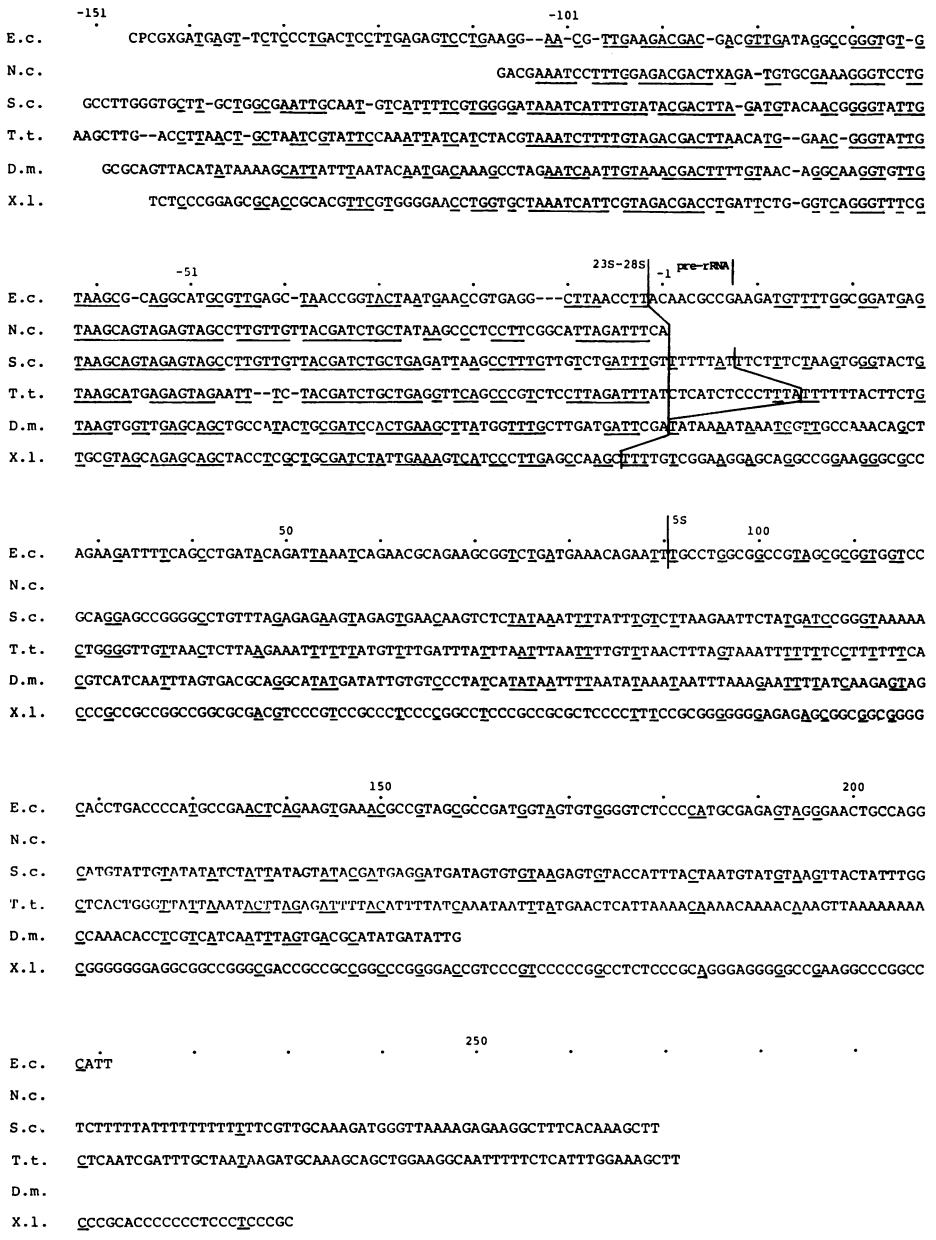


*Fig. 5. Comparison of the DNA sequences around the 3'-ends of the coding regions for L-rRNA among E. coli (24), N. crassa (25), S. carlsbergensis (20), T. thermophila, D. melanogaster (22) and X. laevis (23). The non-coding strands are shown. Dashes are put in to allow maximal alignment of homologous sequences, while empty spaces denote lack of sequence information.*

3'-distal spacer region of another Tetrahymena species, T. pyriformis, has also been sequenced and the 3'-ends of mature and pre-rRNA mapped (29). The two species show identical processing schemes, and the DNA sequences are virtually identical (96% homologous) within the terminal 158 bp of the pre-rRNA coding region and 80% homologous within the next 40 nucleotides further downstream, whereupon the sequences diverge completely. The conservation of the sequence some 40 bp outside what we have mapped as transcribed into 35S pre-rRNA is suggestive of an essential function of this region and supports the proposal that it is involved in transcription termination, possibly because of the hairpin structure which may be formed using bases in this region.

The availability of a purified protein which is involved in transcription termination (10) opens up the possibility of performing binding studies between this protein and the DNA sequences around the termination site. If the "footprints" (30) of this protein can be found on the DNA, this may help reveal which sequences are involved in the termination of rRNA transcription.

ACKNOWLEDGEMENTS

We wish to thank Dr. Werner Eckert for providing purified pre-rRNA for initial experiments in this work, and Drs. Nancy Kan and Ed Niles for sharing their sequence data with us prior to publication. The expert typing of the manuscript by Ms. E. Høyer is gratefully acknowledged. This work was supported by the European Molecular Biology Organization (grant to ND), the Danish Research Council for Natural Sciences (grants to ND & JE) and NATO (grant 1292).

N.D. present address:
Department of Biochemistry B, Panum Institute, University of Copenhagen, Denmark

REFERENCES

1. Engberg, J., Andersson, P., Leick, V. and Collins, J. (1976). J. Mol. Biol. 104, 455-470.
2. Karrer, K.M. and Gall, J.G. (1976). J. Mol. Biol. 104, 421-453.
3. Din, N. and Engberg, J. (1979). J. Mol. Biol. 134, 555-574.
4. Kumar, A. (1970). J. Cell Biology 45, 623-634.
5. Prescott, D.M., Bostock, C., Gamow, E. & Lauth, M. (1971). Exp. Cell Res. 67, 124-128.
6. Eckert, W.A., Kaffenberger, W., Krohne, G. & Franke, W.W. (1978). Eur. J. Biochem. 87, 607-161.
7. Niles, E.G. (1978). Biochemistry 17, 4839-4844.
8. Cech, T.R. and Rio, D.C. (1979). Proc. Natl. Acad. Sci. USA 76, 5051-5055.
9. Din, N., Engberg, J., Kaffenberger, W. and Eckert, W.A. (1979). Cell 18, 525-532.

10. Leer, J.C., Tiryaki, D. and Westergaard, O. (1979). Proc. Natl. Acad. Sci. USA 76, 5563-5566.
11. Engberg, J., Din, N., Eckert, W.A., Kaffenberger, W. and Pearlman, R.E. (1980). J. Mol. Biol. 142, 289-313.
12. Clewell, D.B. & Helinski, R.R. (1969). Proc. Nat. Acad. Sci., USA 62, 1159-1166.
13. Wieslander, L. (1979). Anal. Biochem. 98, 305-309.
14. Maxam, A.M. and Gilbert, W. (1980). Meth. in Enzymol. 65, 499-560.
15. Smith, M., Leung, D.W., Gillam, S., Astell, C.R., Montgomery, D.L. and Hall, B.D. (1979) Cell 16, 753-761.
16. Rosenberg, M. and Court, D. (1979). Ann. Rev. Genet. 13, 319-353.
17. Brown, D.D. and Gurdon, J.B. (1977). Proc. Natl. Acad. Sci. USA 74, 2064-2068.
18. Bogenhagen, D.F., Sakonju, S., and Brown, D.D. (1980). Cell 19, 27-35.
19. Koski, R.A., Clarkson, S.G., Kurjan, J., Hall, B.D. and Smith, M. Cell 22, 415-425.
20. Veldman, G.M., Klootwijk, J., DeJonge, P., Leer, R.J., Planta, R.J. (1980). Nucleic Acids Res. 8, 5179-5192.
21. Hamada, H., Kominami, R. & Muramatsu, M. (1980). Nucleic Acids Res. 8, 889-903.
22. Mandal, R.K. & Dawid, I.B. (1981). Nucleic Acids Res. 9 (8), 1801-1811.
23. Sollner-Webb, B. and Reeder, R.H. (1979). Cell 18, 485-499.
24. Braun, R.J., Young, R.A. & Steitz, J.A. (1980). Cell 19, 393-401.
25. Kelly, J.M. & Cox, R.A. (1981). Nucleic Acids Res. 9 (5) 1111-1121.
26. Samols, D.R., Hagenbuchle, O. & Cage, L.P. (1979). Nucleic Acid Res. 7, 1101-1119.
27. Azad, A.A. & Deacon, N.J. (1980). Nucleic Acids Res. 8, 4365-4377.
28. Rae, P.M.M., Kohorn, B.D. & Wade, R.P. (1980). Nucleic Acids Res. 8, 3491-3504.
29. Niles, E.G., Sutiphong, J., Haque, S. & Jain, R.K. (1981). Manuscript submitted.
30. Galas, D.J. & Schmitz, A. (1978). Nucleic Acids Res. 5, 3157-3170.