Volume 9 Number 7 1981 Nucleic Acids Research

The nucleotide sequence of the initiation region of the ribosomal transcription unit from mouse

René Bach[1], Ingrid Grummt[2+] and Bernard Allet[3]

[1]Départment de Biologie Animale, 154, Route de Malagnou, CH-1224 Chêne-Bougeries, Switzerland, [2]Max-Planck-Institut für Biochemie, D-8033 Martinsried, GFR, and [3]Biogen, S.A. Route de Troinex 3, CH-1227 Caruge, Switzerland

Received 17 February 1981

ABSTRACT

The 5' end of 45S pre-rRNA has been located on a cloned rDNA fragment from mouse by r-loop mapping and the nuclease S1 protection technique. 45S pre-rRNA could be shown to represent the primary transcript of the ribosomal genes because 5' polyphosphate termini have been detected by an enzymatic assay. The sequence of about 1100 nucleotides surrounding the initiation site for ribosomal RNA transcription has been determined. Features of this region of the ribosomal DNA will be discussed. A comparison of the nucleotide sequence with corresponding areas of ribosomal genes from other eukaryotes does not reveal significant homology in the region of transcription initiation.

INTRODUCTION

The transcription of ribosomal genes from eucaryotes is very efficiently regulated. Though several investigations have shown that most of this transcriptional control acts at the level of initiation (1,2), the molecular mechanisms which modulate the initiation frequency of RNA polymerase I according to the physiological state of the cell are still unknown.

In order to study the mechanism and regulation of rRNA transcription the rDNA of *Mus musculus* including the 5' flanking sequences has been cloned and characterized (3,4). In this paper we have mapped the 5' end of the 45S pre-rRNA on this rDNA fragment and determined the nucleotide sequence surrounding the initiation site of transcription where control signals for initiation of transcription are likely to reside. Up to now increasing information about sequence organization of eukaryotic genes transcribed by RNA polymerase II or III has been accumulated. Comparison of these sequence data revealed a conservation of certain sequences at the flanking regions of the gene which may act as recognition and/or regulatory sites for transcription

© IRL Press Limited, 1 Falconberg Court, London W1V 5FG, U.K.

initiation, processing and termination. So far, however, no analogous comparison has been made for ribosomal genes which are transcribed by RNA polymerase I. This is mainly due to the fact that very little sequence data on the transcription initiation site and 5' flanking regions of rDNA are available (5-11). In this paper we present the sequence of 1140 nucleotides from the initiation region of mouse rDNA. Comparison of this nucleotide sequence with the presumptive promoter region from other eukaryotic ribosomal genes showed that there are no obvious common structural features in the initiation site of RNA polymerase I genes.

MATERIALS AND METHODS

rDNA clones and fragment preparation

The cloning and characterization of the large EcoRI fragment of mouse rDNA has been described before (3,4). For the present studies we have subcloned the fragment SalI-B which contains the initiation site by inserting the purified fragment into the SalI site of pBR322. Restriction enzyme digestions were routinely carried out in the assay conditions recommended by the suppliers. DNA fragments were either labelled at the 5'end with γ-^{32}P-ATP and polynucleotide kinase or labeled at the 3' end with the Klenow fragment of DNA polymerase I (8). The fragment purification, strand separation and sequencing technique was according to Maxam and Gilbert (12).

Isolation of 45S pre-rRNA and in vitro capping

45S pre-rRNA was purified from isolated nucleoli as described previously (1). The nucleolar RNA was treated with RNase-free DNase and run on a 15-30 % sucrose gradient in 10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 10 mM EDTA, 0.1 % sodium dodecylsulfate for 15 hours at 26.000 rpm in a SW41 rotor. Pooled fractions from the gradient were ethanol precipitated and dissolved in 50 mM Tris-HCl, pH 7.5, 0.1 mM EDTA at 1 mg/ml. Capping reactions were usually performed as described by Moss (13) in a volume of 100 µl containing 10 µg of RNA, 50 mM Tris-HCl, pH 7.5, 1 mM DTT, 2 mM $MgCl_2$, 2 mM GTP, 0.2 mM ATP, 50 µCi of S-adenosyl-(^{3}H-methyl)methionine (33 Ci/mMol) and 5 µl of enzyme extract. After 90 min. at 37°C the samples were TCA precipitated and washed on

glass fibre filters. Since the crude enzyme fraction used also transferred a significant amount of cap structures on RNA molecules with monophosphate termini a background level obtained by assaying 5 µg of 28S RNA in the reaction was subtracted from each value.

Hybridization, S1 nuclease digestion

DNA fragments to be hybridized with RNA were labeled with polynucleotide kinase and strand separated. The coding strand was mixed with about 1 µg of purified 45S rRNA in 50 µl of 80% formamide, 0.4 M NaCl, 40 mM MOPS pH 6.7 and 1mM EDTA. After 10 min. at $60^{o}C$ the mixture was hybridized at $49^{o}C$ overnight. The hybrid was diluted 10 fold into S1 buffer (50 mM Na-acetate, pH 4.5, 200 mM NaCl, 1 mM $ZnCl_2$) and treated for one hour at $45^{o}C$ with either 10^5 or $5x10^5$ units of S1 nuclease (Miles). After extraction with phenol and chloroform the samples were ethanol precipitated and lyophilized. They were analyzed in a 8% sequencing gel in parallel with the sequence ladder of the coding strand.

RESULTS AND DISCUSSION

Recently we have described the structure of the recombinant bacteriophage λMr974 that contains a large EcoRI rDNA fragment from mouse including parts of the non-transcribed spacer and about one third of the transcribed gene region. From previous electron microscopic studies on the secondary structure of mouse 45S pre-rRNA and determinations of the molecular weights of processing intermediates the length of the external transcribed spacer has been estimated to be 4.4 kb (14, 15). This places the 5'-terminal coding sequence for the 45S precursor at the left end of the fragment SalI-B of the cloned rDNA (Fig. 1).

In order to locate the initiation site for RNA polymerase I precisely at the nucleotide level one has to make sure that the 45S pre-rRNA represents a primary transcript. Since the processing takes place immediately after or even during transcription it is often difficult to say whether the largest pre-rRNA which can be detected in pulse labeling experiments is indeed the primary transcript of the ribosomal repeating unit. A rigorous proof that 45S pre-rRNA represents the primary transcription

Fig. 1. Structural organization of mouse rDNA. An expanded SalI-B fragment derived from pMr974 and some restriction endonuclease sites in the region of transcription initiation are shown at the bottom.

product would be the demonstration of a conserved di- or triphosphate residue on the 5' terminus. So far, polyphosphate termini of eukaryotic ribosomal precursor RNA have been reported for yeast, *Tetrahymena*, *Dictyostelium* and *Xenopus laevis* (16-19). In the case of mammalian pre-rRNA several investigators have failed to identify triphosphate residues at the 5' end of 45S RNA (20,21). In the present study we have used the capping enzyme from vaccinia virus that acts on 5' polyphosphate termini, adding a reverse GTP and two methyl groups to produce a cap 1 structure: ^{m7}GpppX^{m}pYpZp...

Figure 2 shows a sucrose gradient analysis of the distribution of di- or triphosphate bearing RNA molecules isolated from purified nucleoli. There are two classes of RNA molecules which could be capped *in vitro*, the 45S pre-rRNA and 24S RNA, the latter one representing the external transcribed spacer cleaved off the primary transcript during processing. The incorporation of (^{3}H)methyl groups was completely dependent on the presence of GTP in the assay (not shown) which indicates that the methylation is in cap structures and not at internal sites along the RNA chain. However, calculation of the capping efficiency of 45S pre-rRNA reveals that only 10-15 % of the 45S rRNA molecules

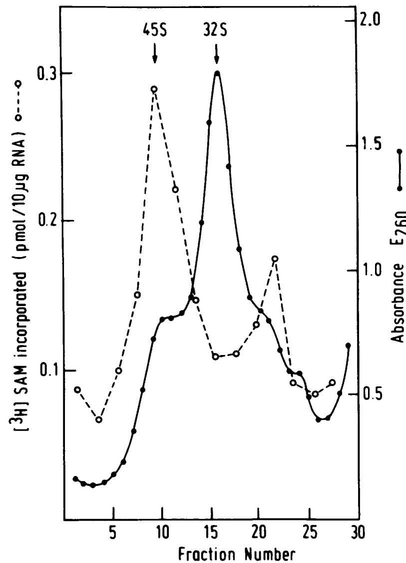

Fig. 2. Identification of 5' polyphosphate termini in nucleolar RNA.
Nucleolar RNA from Ehrlich ascites cells was run on a sucrose gradient. Pooled fractions from the gradient were tested for their ability to accept 5' terminal caps with S-adenosyl [^{3}H]methionine as label donor as described in Materials and Methods.

were able to accept caps. In Xenopus laevis, 25 % of nucleolar 40s rRNA could be capped in vitro (19). The low capping activity on mouse pre-rRNA can be explained by one of the following possibilities. (I) Either the secondary structure of 45S rRNA prevents efficient capping, (II) the polyphosphate termini have been hydrolyzed by phosphatase action, or (III) the primary rDNA transcript is very rapidly processed at the 5' end.

DNA sequence of the transcription initiation region

In order to determine the nucleotide sequence of the transcription initiation site and surrounding regions we have subcloned the fragment SalI-B which is expected to contain the site at which initiation of transcription occurs. A detailed restriction map of the 1400 nucleotides between the SalI and the XhoI site has been constructed and selected overlapping fragments were sequenced according to Maxam and Gilbert (12). In Fig. 3

```
                   -600
GTCGACGCCAGTTGTTCCTTTGAGGTCCGGTTCTTTTCGTTATCCCCTCATTTTTGGGCCACCTC
 -550                                             -500
CCCAGTATGACTTCCTGTATTCTCTGTGGCCTGTCACTTTCCTCCCTGTCTCTTTTATGCTTGTG
                                    -450
ATCTTTTCTATCTGTTCCTAATGGAACTTGGAGATAGGTACTGACACGCTGTCCTTTCCCTATTA
                        -400
ACACTAAAGGACACTATAAAGAGACCCTTTCGATTTAAGGCTGTTTTGCTTGTCCAGCCTATTCT
      -350                                                  -300
TTTTACTGGCTTGGGTCTGTCGCGGTGCCTGAAGCTGTCCCCGAGCCACGCTTCCTGCTTTCCCG
                                             -250
GGCTTGCGCTTGCGTGTGCTTGCTGTGGCAGCTTGTGACAACTGGGCGCTGTGACTTTGCTGCG
                              -200
TGTCAGACGTTTTTCCCGATTTCCCCGAGGTGTCGTTGTCACACCTGTCCCGGTTGGAATGGTGG
            -150                                               -100
AGCCAGCTGTGGTTGAGGGCCACCTTATTTCGGCTCACTTTTTTTTTTTTTTTTTTCTCTTGGA
                                                -50
GTCCCGAACCTCCGCTCTTTTCTCTTCCCGGTCTTTCTTCCACATGCCTCCCGAGTGCATTTCTT
                                  1
TTTGTTTTTTTTCTTTTTTTTTTTTTTTTTTTGGGGAGGTGGAGAGTCCCGAGTACTTCACTCCT
                  50
GTCTGTGGTGTCCAAGTGTTCAGCCACGTGCCTCCCGAGTGCACTTTTTTTGTGGCAGTCGCTC
100                                                   150
GTTGTGTTCTCTTGTTCTGTGTCTGCCCGTATCAGTAACTGTCTTGCCCCGCGTGTAAGACATTC
                                      200
CTATCTCGCTTGTTTCTCCCGATTGCGCGTCGTTGCTCACTCTTAGATCGATGTGGTGCTCGCCG
                       250
GTTCTCTTCGGGCCAGGGCCTAAGCCGCGCTAGGCGAGGGACGGACATTCATGGCGAATGGCGGC
      300                                                    350
CGTTCTTCTCGTTCTGCCAGCGGGCCCCTCGTCTCTCCACCCCATCCGTCTGCCGGTGGTGTGTG
                                             400
GAAGGCAGGGGTGCGGCTCTCCGGCCCGAACGCTGCCCCGCGCGCACTTTTCTCAAGTGGTTCGC
                            450
GTGGTCCTTGTGGATGTGTGAGGCGCCCGGTTGTGCCCTCACGTGTTTCACTTTGGTCGTGTCTC
             500
GCTTGACCATGTTCCCAGAGTCGGTGGATGTGGCCGG
```

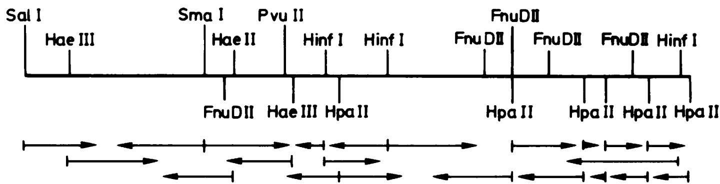

Fig. 3. Nucleotide sequence of the region surrounding the initiation site for pre-rRNA transcription.
The lower part of the diagram indicates the origin, direction and length of each sequence obtained. The nucleotide sequence of the non-coding DNA strand is shown. Numbers indicate nucleotide positions relative to the transcription initiation site. The underlined nucleotides represent sequences of homology.

the sequence arrangement at the boundary between spacer and transcribed sequences is shown, the region where control signals for initiation of transcription are likely to reside. Numbers indicate nucleotide positions relative to the transcription initiation site, the assignment of which will be justified below.

The most remarkable feature of this sequence is the presence of two clusters of thymidines in the non-coding strand of the non-transcribed spacer centered about 100 bp distant from each other. The longer stretch consists of 34 consecutive thymidines interspersed by only 3 other bases and is located just in front of the initiation point of 45S pre-rRNA synthesis. It is conceivable that the AT-rich blocks serve as binding sites for RNA polymerase I since the AT richness would destabilize the helix structure and thus facilitate RNA polymerase-mediated strand unwinding necessary for the initiation of transcription. A sequence of 25 nucleotides TTCTTCCACATGCCTCCCGAGTGCA is located in front of this long thymidine cluster (from position -63 to -39) which is repeated with only 3 bases mismatch at position 52-76. Whether this sequence exerts any relevant function in the expression of rDNA remains to be determined. In this context it should be noted that Urano et al. (11) recently have determined the nucleotide sequence of the ETS-NTS boundary of an independently derived mouse rDNA clone. Though those authors located the initiation site about 200 bp further downstream the published sequence data are almost identical to ours. This indicates that the rDNA clone we use is likely representative of most of the ribosomal genes from mouse.

Mapping of the 5' terminus of 45S pre-rRNA

Due to the high molecular weight of 45S pre-rRNA direct sequencing of its 5' terminus is difficult. Therefore we have used the S1 nuclease protection procedure (22) to locate the 5' end of the ribosomal precursor and thus the site of transcription initiation precisely within the sequenced DNA region. A 305 bp PvuII-FnuDII fragment (position -156 to +149) was labeled at the 5' end of the coding strand, hybridized to 45S pre-RNA and digested with the single strand-specific nuclease S1. The protected fragments were electrophoresed in a sequencing gel in parallel with the same 305 bp PvuII-FnuDII fragment cleaved by

base-specific reactions.As shown in Fig. 4 there is a major protected fragment 149 nucleotides long and a minor fragment which maps 64 bp upstream the FnuDII site. The relative intensity of the two protected fragments is 85 % and 15 %, respectively, as revealed by scanning of the autoradiogram.

There is one major difficulty in interpreting the S1 experiment. Both protected DNA fragments terminate just before a run of thymidines. It is known that the stability of A-U base pairs is considerably lower than that of A-T pairs (23). Under the stringent hybridization conditions used precursor molecules which extend into or over the T-cluster would probably not be detected. Therefore the possibility exists that the actual initation site is even further upstream than indicated in Fig. 3. We are currently doing oligonucleotide analysis of <u>in vitro</u> transcripts to determine the initiation point without ambiguity.

In the recent paper of Urano et al. (11) the presumptive point of transcription initiation has been located 202 bp further

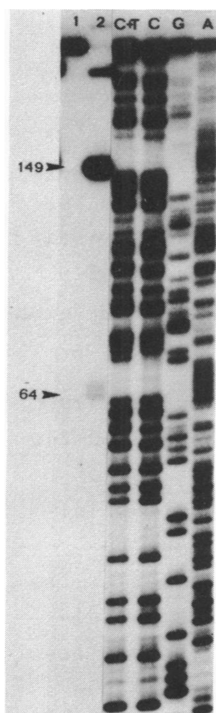

Fig. 4. Nuclease S1 mapping of 45S RNA.

The 305 bp PvuII-FnuDII (PF-305) fragment was labelled at the 5'end of the coding strand and hybridized to 45S pre-rRNA. The hybrids were digested with nuclease S1 and run in parallel with a sequence ladder of the same fragment (A, G, C, T).

1) PF-305 hybridized with the complementary DNA strand and treated with 10^5 units of nuclease S1.
2) PF-305 hybridized with 45S RNA and treated with 10^5 units of nuclease S1.

downstream. The S1 mapping procedure can identify the farthest upstream end of a transcript. However, downstream promoters cannot be distinguished from sites of RNA processing. Therefore, it is uncertain whether this site represents another promoter or a prefered site of 5'-terminal processing of the precursor. We find a significant amount of this 202 bp shorter transcript in S1 protection experiments, too (24), but since the amount of it varies with different 45S RNA batches we rather consider it as a processing product.

Furthermore it is still uncertain whether the weak site 64 bp upstream the FuvDII site (Fig. 4) could represent a second promoter 85 bp distant from the major initiation site. On the basis of the remarkable structure homology in the 5' flanking region of both sites (Fig. 5) we suggest that two promoters are located in front of the coding region of mouse rDNA. Whether both sites are differently used under different physiological conditions is currently being investigated.

Sequence comparison of ribosomal genes

A comparison of the primary structure of different transcription unitsand their flanking regions may reveal regulatory sequences involved in gene expression. This approach is based on the assumption that recognition sites for RNA polymerases or regulatory sequences should be conserved in evolution because of their importance to the living cell and the similarities of gene expression and its control in widely divergent organisms.

Though the information about the sequence of ribosomal genes is still very poor there seems to be no homology in the region of transcription initiation in different organisms. Se-

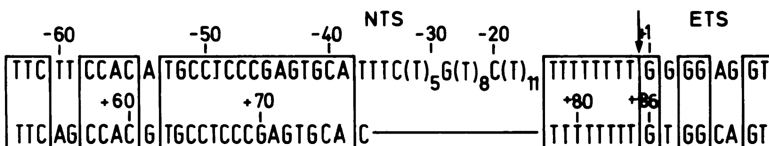

Fig. 5. Comparison of common structural features of two putative transcription initiation sites of mouse rDNA.

quences of the NTS/ETS boundary of rDNA from yeast (9), *Drosophila melanogaster* (10), three frog species (5-8) and mouse are available. A comparison of the sequences using a computer program revealed no obvious consensus sequence which is held in common between several eukaryotic ribosomal genes. However the organisms compared are evolutionarily widely distant species so that it cannot be ruled out that more closely related species share some sequence homologies at the 5' end of the ribosomal transcription unit. In fact in three *Xenopus* species analyzed so far, *X.laevis* (5-7), *X.borealis* and *X.clivii* (8) 13 nucleotides at the NTS-ETS boundary are identical pointing to a highly conserved and important function such as initiation of transcription. Recent experiments indicate that species specific factor(s) are required for the faithful transcription of mouse rDNA *in vitro* (24). Since the factor(s) seem to bind to DNA (I. Grummt, unpublished observation) it is suggested that specific protein(s) recognize a defined primary or secondary structure in the rDNA repeating unit. This specific DNA-protein interaction is a prerequisite for specific transcription initiation of rDNA by RNA polymerase I. This working hypothesis is now being investigated.

ACKNOWLEDGEMENTS

We are grateful to Dr. B. Moss for the gift of the capping enzyme and to M. Rahire for skilful technical assistance. This work was supported by grants to B.A. from the Swiss National Science foundation. I.G. was supported by a grant from the Stiftung Volkswagenwerk and the Deutsche Forschungsgemeinschaft.

+To whom correspondence should be addressed: Institut für Biochemie, Röntgenring 11, D-8700 Würzburg, GFR

REFERENCES

1. Grummt, I., Smith, V.A., and Grummt, F. (1976) Cell 7, 439-445.
2. Cox, R.F. (1976) Cell 7, 455-465.
3. Grummt, I., Soellner, C. and Scholz, I. (1979) Nuc. Acids Res. 6, 1351-1469.
4. Grummt, I. and Gross, H.J. (1980) Molec. Gen. Gen. 177, 223-229.
5. Boseley, P., Moss, T., Mächler, M., Portmann, R., and Birnstiel, M.L. (1979) Cell 17, 19-31.
6. Sollner-Webb, B. and Reeder, R.H. (1979) Cell 18, 485-494.
7. Moss, T., Boseley, P. and Birnstiel, M.L. (1980) Nucl. Acids Res. 8, 467-485.

8. Bach, R., Allet, B. and Crippa, M. (1981) Submitted for publication.
9. Valenzuela, P., Bell, G.I., Venegas, A., Sewell, E.T., Masiarz, F.R., Dennaro, L.J., Weinberg, F. and Rutter, W.J. (1977) J. Mol. Biol. 252, 8126-8131.
10. Long, E.O., Rebbert, M.L. and Dawid, J.B. (1981) Proc. Natl. Acad. Sci. USA 78, in press.
11. Urano Y., Kominami R., Mishima Y., and Muramatsu M. (198o) Nuc. Acids Res. 8, 6043-6058.
12. Maxam, A. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
13. Moss, B. (1977) Biochem. Biophys. Res. Commun. 74, 374-383.
14. Wellauer, P.K. and Dawid, I.B. (1975) Brookhaven Symp.Biol. 26, 214-233.
15. Dabeva, M.D., Dudor, K.P. and Hadjiolov, A.A. (1978) Biochem. J. 171, 367-374.
16. Nikolaev, N., Georgiev, O.J., Venkov, P.V. and Hadjiolov, A.A. (1979) J. Mol. Biol. 127, 297-308.
17. Niles, E. (1978) Biochemistry 17, 4839-4844.
18. Batts-Young, B. and Lodish, H. (1978) Proc. Natl. Acad. Sci. USA 75, 740-744.
19. Reeder, R.H., Sollner-Webb, B. and Wahn, H.L. (1977) Proc. Natl. Acad. Sci. USA 74, 5402-5406.
20. Choi, Y.C. and Busch, H. (1970) J. Biol. Chem. 245, 1954-1961.
21. Kominami, R. and Muramatsu, M. (1977) Nucleic Acids Res. 4, 229-240.
22. Berk, A. and Sharp, P. (1977) Cell 12, 721-732.
23. Martin, F.H. and Tinoco, J. (1980) Nuc. Acids Res. 8, 2295-2299.
24. Grummt, I. (1981) Proc. Natl. Acad. Sci. USA 78, in press.