Volume 11 Number 10 1983 Nucleic Acids Research

Identification and sequence of the initiation site for rat 45S ribosomal RNA synthesis

Christina A.Harrington* and Dona M.Chikaraishi

Division of Molecular and Cellular Biology, National Jewish Hospital and Research Center/National Asthma Center, Denver, CO 80206, and *Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309, USA

Received 25 October 1982; Revised 6 April 1983; Accepted 20 April 1983

ABSTRACT

The transcription initiation site for rat 45S precursor ribosomal RNA synthesis was determined by nuclease protection mapping with two single-strand endonucleases, S1 and mung bean, and one single-strand exonuclease, ExoVII. These experiments were performed with end-labeled ribosomal DNA from double-stranded pBR322 recombinants and from single-stranded M13 recombinants. Results from experiments using both kinds of DNA and all three enzymes showed that the 5' end of 45S RNA mapped to a unique site 125 bases upstream from the Hind III site in the ribosomal DNA gene. The DNA surrounding this site (designated +1) was sequenced from -281 to +641. The entire sequence of this region shows extensive homology to the comparable region of mouse. This includes three stretches of T residues in the non-coding strand between +300 and +630. Two sets of direct repeats adjacent to these T-rich regions are observed. Comparison of the mouse and human ribosomal DNA transcription initiation sites with the rat sequence reported in this paper demonstrates a conserved sequence at +2 to +16, CTGACACGCTGTCCT. This suggests that this region may be important for the initiation of transcription on mammalian ribosomal DNAs.

INTRODUCTION

Eukaryotic transcription initiation sites and the chromosomal features determining promoter activity at these sites have been a major focus of recent research on transcriptional control mechanisms. Both *in vivo* and *in vitro* studies of promoter activities for hnRNA synthesis have shown that several DNA sequences upstream from the initiation site are important for recognition and positioning of polymerase II molecules (1,2,3,4,5). In contrast to polymerase II, polymerase III recognition and positioning appear to involve sequences internal to the 5S and tRNA genes (for review see 6 and 7). The nature of the controlling elements for polymerase I, the activity responsible for synthesis of eukaryotic ribosomal RNA (rRNA), has not yet been identified.

Initiation sites for ribosomal RNA synthesis in several organisms have been determined, primarily by S1 nuclease protection mapping, and the DNA regions surrounding these sites sequenced: *Drosophila* (8), *Saccharomyces* (9),

© IRL Press Limited, Oxford, England.

Xenopus (10,11), mouse (12,13), and human (14,15). Comparison of the non-mammalian initiation site sequences with each other and with the mammalian sequences shows little conserved homology. Mammalian species (mouse, rat, and human), however, do share significant sequence homology around the initiation site, as discussed by Financsek et al. (14) and in this paper. Comparison among three *Xenopus* species showed some homology upstream from the initiation site and an exact homology at -9 to +4 (11).

The present work extends the information available on polymerase I initiation sites to include that of rat. Transcription mapping using the single-strand nucleases of *Aspergillus* (S1) and mung bean, and the single-strand exonuclease, Exo VII, has located the initiation site for ribosomal precursor RNA synthesis at approximately 4.5 kb upstream from the 18S rRNA gene. The DNA sequence of this area from -281 to +641 (with +1 being the initiation site for 45S rRNA synthesis) has been determined. The sequence shows extensive homology to the DNA surrounding the initiation site of mouse, and some homology to that of human. There are several interesting sequence features found in this region, most of which are also seen in mouse.

Recently, Financsek et al. (16) reported the DNA sequence for the rat 45S initiation site from -166 to +127. It differs from the sequence described in this work by two base substitutions and four single base insertions.

MATERIALS AND METHODS

Materials

Restriction enzymes, polynucleotide kinase and Exo VII were purchased from BRL. Bacterial alkaline phosphatase and mung bean nuclease were obtained from P-L Biochemicals. S1 nuclease was purchased from P-L Biochemicals and Miles Laboratories. DNase I was purchased from Worthington and purified of RNAse activity by the method of Maxwell et al. (17). Large fragment DNA polymerase I used in dideoxy-sequencing reactions was supplied by BRL and NEN. Protease type XI (proteinase K) was purchased from Sigma.

Isotope for the dideoxy-sequencing reactions was purchased either from Schwartz-Mann ([α-^{32}P] NTP, 2000 Ci/mmole) or NEN ([α-^{32}P] NTP, 600-800 Ci/mmole). The 26 bp DNA primer for the reaction was obtained from BRL.

Identification of rDNA recombinants

A rat (Sprague-Dawley) genomic library, consisting of DNA generated by partial digestion with Eco RI, in lambda Charon 4A (gift of Drs. R.B. Wallace, T.D. Sargent and J. Bonner) was screened in plaque hybridizations (18) with ^{32}P-labeled 18S and 28S rRNA. Selected recombinants were analyzed by complete

digestion with Eco RI, fractionation on 1% agarose gels, transfer to nitrocellulose and hybridization to ^{32}P-labeled 18S RNA (19).

Subcloning of rDNA fragments into pBR322, pBR325 and M13

A 10.9 kb Eco RI fragment of rDNA was cloned in pBR322 and propagated in *E. coli* HB101. This plasmid was cleaved with Hind III and Eco RI to generate two smaller rDNA fragments, 5.0 kb and 5.9 kb (see Fig. 1). These were cloned in pBR322 using the Hind III/Eco RI sites of the vector. A 0.51 kb Hind III/Xho I fragment adjacent to the putative initiation site was cloned in pBR325 using Eco RI linkers (BRL) that had been ligated to the isolated fragment.

Restriction fragments of rDNA in pBR322 and pBR325 were ligated to M13mp7 (BRL), M13mp8 (gift of Mr. Sidney Shinedling) or M13mp9 (gift of Dr. John Sadler) RF-DNA and used to transform competent *E. coli* JM103 cells (20). Phage containing an insert were identified by the absence of β-galactosidase induction. The 0.51 kb Hind III/Xho I fragment was cloned in M13mp7 in both orientations using the Eco RI linker sites. A 0.41 kb Bam HI/Hind III fragment was cloned in M13mp8 at the Bam HI/Hind III sites and its reverse orientation was obtained by using M13mp9.

M13 transformations and screening were carried out according to the protocol described by J. Messing in the BRL M13mp7 cloning/'Dideoxy' Sequencing Manual (1980).

DNA Preparation

Plasmid DNA for sequencing and transcription mapping was isolated from cleared lysates by cesium chloride-ethidium bromide centrifugation. Restriction fragments for chemical sequencing were obtained by separating end-labeled DNA on low-melting point agarose gels and isolating individual bands.

Single-strand M13 DNA for enzymatic sequencing was prepared according to the following procedure. Phage suspensions from which JM103 cells had been removed by centrifugation were treated with 3.5% polyethylene glycol (PEG)-6000 and 0.5 M NaCl on ice for 30 min. The precipitated phage were pelleted by centrifugation at 10,000 xg for 15 min. Pellets were gently resuspended in 1/3 X TES (TES = 30 mM Tris, pH 8.0, 5 mM EDTA, 50 mM NaCl) at room temperature and centrifuged to remove remaining cells. The PEG precipitation was repeated and the pellet resuspended in TES plus 0.1% SDS and then incubated at 55°C for 10 min. to lyse the phage particles. The suspension was then treated with proteinase K (15 μg/ml) for 10 min. at 37°C, phenol/chloroform extracted twice, chloroform extracted once, and

ethanol precipitated. The yield of purified M13 was from 1-3 μg/ml of culture fluid.

DNA end-labeling

pEH5.0 DNA was cleaved at the single Hind III site, treated with alkaline phosphatase, phenol extracted, and 5' end-labeled with [γ-^{32}P] ATP (9000 Ci/mmole) and polynucleotide kinase (21). The DNA, labeled at 1-5 x 10^6 dpm/pmole of ends, was then cut with Sal I, resulting in two labeled fragments, a 290 bp fragment containing rDNA sequences and a 621 bp Hind III/Sal I fragment from pBR322. Labeling of the 0.51 kb Hind III/Xho I fragment was accomplished in a similar manner by cleaving pHX.51 at the Xho I site, performing the kination reaction as described, and then cutting with Eco RI.

Single-stranded M13 DNA containing rDNA inserts was end-labeled by cleaving purified phage DNA with Alu I at room temperature and performing the kination reaction as described above. The labeled products consisted of multiple DNA fragments corresponding to Alu I digestion products of the phage DNA and the rDNA insert.

Restriction mapping

Restriction mapping was performed by standard procedures and is discussed in detail elsewhere (Chikaraishi et al., submitted).

RNA preparation

RNA was prepared by modification of the method of Penman (22). For 45S RNA preparation, detergent prepared nuclei, isolated from RT4-D1 rat glial cells (23), were digested with DNase I (10 μg/ml) in the presence of 1 mM aurin tricarboxylic acid (ATA), at 37° for 5-10 min. and then centrifuged at 13,000g for 10 min. to pellet nucleoli. Nucleoli were suspended in RSB (10mM NaCl, 10mM Tris, pH 7.4, and 1.5mM $MgCl_2$), extracted with phenol/ chloroform and fractionated on linear sucrose gradients. The 45S RNA fraction appeared as a shoulder on the 32S RNA peak. This material was collected, ethanol precipitated, resuspended in 20mM Tris, pH 7.5, and stored at -20°C.

Total nucleolar RNA was prepared from Sprague-Dawley rat liver cells. Nucleoli were prepared as above, phenol/chloroform extracted and ethanol precipitated. Pellets were resuspended in 10mM Tris, pH 7.5, 10mM $MgCl_2$, 10 mM and digested again with DNase I (5 μg/ml) in the presence of 1mM ATA, at 37° for 20 min. This was followed by digestion with proteinase K (5 μg/ml) plus 0.1% SDS for 10 min. at 37°C. The nucleolar RNA was then extracted with phenol/chloroform and separated from small DNA fragments by chromatography on Sephadex G-50. The excluded peak was ethanol precipitated, resuspended in 20mM Tris, pH 7.5, and stored at -20°C.

Nuclease protection mapping

5' end-labeled double-strand DNA (15-30 ng) was hybridized with 2.8 μg of 45S RNA plus 3 μg carrier *E. coli* tRNA in 10 μl 80% formamide, 0.4M NaCl, 0.04M PIPES (piperazine - N,N'-bis [2-ethanesulfonic acid]), 0.1mM EDTA at 50°C overnight after boiling for 5 min. (24). The solution containing the hybrids was diluted 30-60 fold with S1 nuclease buffer (0.28 M NaCl, 0.03M Na acetate, pH 4.5, 4.5mM $ZnCl_2$, 20 μg/ml denatured calf thymus DNA) or mung bean nuclease buffer (50mM NaCl, 30mM Na acetate, pH 5.0, 2mM $ZnCl_2$, 20 μg/ml denatured calf thymus DNA) or exonuclease VII buffer (50mM KPO_4, pH 7.6, 10 mM EDTA, 1mM β-mercaptoethanol, 1 μg/ml denatured calf thymus DNA). In experiments using S1, buffers were adjusted to 50mM NaF to inhibit phosphatase activity. S1 nuclease digestions were carried out at room temperature or 37°C for 1 hr. at S1 concentrations of 200 U/ml. Mung bean nuclease digestions were incubated at 37°C for 1 hr. with either 20 U/ml or 80 U/ml enzyme activity. Exo VII digestions were also performed at 37° for 1 hr. at enzyme concentrations of 1.7 U/ml. Mung bean and S1 nuclease reactions were terminated with the addition of 1/6 volume of 2.5M NH_4 acetate, 50mM EDTA. Digestions were extracted with phenol/chloroform and ethanol precipitated after addition of tRNA to 100 μg/ml. Pellets were resuspended in 80-90% formamide, 5mM EDTA, 0.1% bromophenol blue, 0.1% xylene cyanol, boiled for 3 min. and electrophoresed on 6% polyacrylamide-8M urea gels.

For protection mapping with single-strand M13 recombinants, end-labeled DNA (40-125 ng) was hybridized with 0.5 μg of 45S RNA or 2.6 μg of total nucleolar RNA in 20 μl of 4 X SSPE (1 X SSPE = 0.18M NaCl, 10mM Na_2PO_4, pH 7.7, 1mM EDTA) at 63° overnight after boiling for 5 min. The hybridization samples were diluted 15-fold with S1 nuclease buffer and processed as described above.

Eco RII digestion products of SV-40 DNA were end-labeled with [γ-^{32}P] ATP and used as molecular weight markers.

DNA sequencing

DNA sequencing was performed by both the chemical cleavage method of Maxam and Gilbert (21) and the dideoxy-chain terminating method of Sanger et al. (25) as described by Smith (26). In the Sanger sequencing procedure, recombinants in the single-strand phage M13 were used as template. The 26 bp primer used in the synthesis reaction was homologous to the region immediately preceding the restriction sites used for cloning. Sequencing products were analyzed on 0.4mM thick 20%, 8% or 6% polyacrylamide-8M urea gels.

RESULTS

rDNA cloning and restriction mapping

A rat genomic DNA library in lambda was screened using plaque hybridization to ^{32}P-labeled 18S and 28S ribosomal RNA and three positive plaques were isolated for further analysis. These three lambda recombinants were digested to completion with Eco RI and analyzed by Southern blot hybridization to ^{32}P-labeled 18S rRNA. Autoradiography showed two of the lambda recombinants contained hybridizing bands at 10.9 kb and 6.7 kb. The third recombinant had two bands, at 6.7 kb and 4.6 kb.

Lambda recombinant 1A containing both the 10.9 kb and 6.7 kb Eco RI fragments was used for subcloning. The 10.9 kb fragment, which contains most of the 18S rRNA coding region and the putative initiation site for rRNA synthesis, was further subcloned using an internal Hind III site and the Eco RI ends. This resulted in pBR322 plasmids containing a 5.0 kb rDNA fragment, pEH5.0, and a 5.9kb rDNA fragment, pHE5.9. These recombinants, along with pEE10.9, were used extensively for restriction enzyme mapping and for subcloning into pBR325 and M13 vectors as described in Materials and Methods. A partial restriction map and the various subclones of the 10.9 kb rDNA fragment are shown in Figure 1.

The 10.9 kb fragment is highly representative of the rDNA repeat in rat genomic DNA (Chikaraishi et al., submitted). Various restriction enzyme

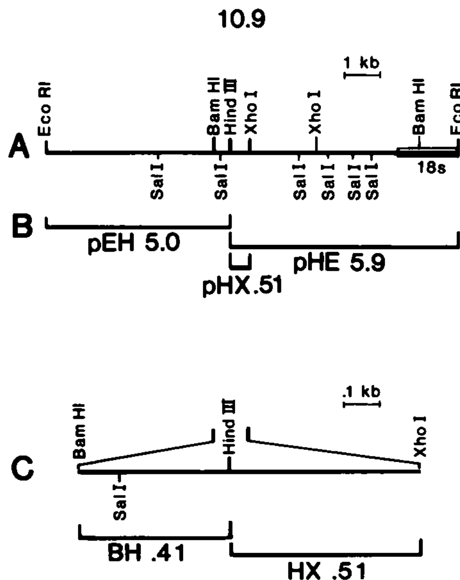

Figure 1 A. Map of the rat rDNA Eco RI fragment containing the initiation site for rRNA synthesis. B. rDNA fragments cloned in PBR322 and PBR325. C. rDNA fragments cloned in M13.

digests of genomic DNA were analyzed by Southern blot hybridization to labeled single-strand M13 rDNA recombinants. The results showed that the major hybridizing bands of the genomic DNA correspond to identical fragments in the 10.9 kb clone.

Determination of the initiation site for 45S transcription by nuclease protection mapping.

Initially, Southern blot hybridization of ^{32}P-labeled 45S pre-rRNA to restriction fragments of pEE10.9 localized the putative initiation site for ribosomal RNA transcription to a region approximately 5 kb upstream from the 18S gene. The 5' end of 45S RNA was precisely determined by nuclease protection mapping.

Plasmids containing the 5.0 kb Eco RI/Hind III fragment or the 0.51 kb Hind III/Xho I fragment were cleaved with Hind III or Xho I, respectively. The resultant linearized plasmid molecules were end-labeled as described in Materials and Methods and then digested with a second restriction enzyme: Sal I for pEH5.0 and Eco RI for pHX.51. The labeled DNA was hybridized to 45S RNA in 80% formamide as described. Following an overnight hybridization, the samples were treated with one of the following single-strand specific enzymes after dilution in the appropriate buffer: S1 nuclease, the single-strand endonuclease from mung bean, or the 5', 3'-exonuclease, Exo VII. After digesting the hybrids for one hr. to degrade all single-strand nucleic acid, the sizes of the protected DNA fragments were determined by electrophoresis on 1mm thick 6% polyacrylamide-8M urea gels or 0.4mm thick sequencing gels.

Analysis of the protected fragments when the Hind III/Xho I DNA was used as a probe showed three bands (data not shown) that varied with S1 concentration and the temperature of the digestion. It was subsequently determined that these fragments contain long stretches of dA residues in the rDNA coding strand which encode stretches of rU residues in the 45S RNA. As discussed by Miller and Sollner-Webb (27), long rU/dA hybrids are not stable under most S1 digestion conditions. Therefore, our S1 digestions were probably generating internal cleavages at the sites of long rU/dA hybrids.

Transcription mapping with the Sal I/Hind III 0.29 kb DNA fragment, end-labeled at the Hind III site, showed one discrete band, approximately 125 bases upstream from the Hind III site (Fig. 2A). This site did not correspond to the location of any rU stretches in the RNA sequence and was seen with all three single strand specific nucleases. The protected fragment was a few bases longer when the digestion was performed with Exo VII. This was expected because Exo VII cutting generates a 2-3 base overhang (28).

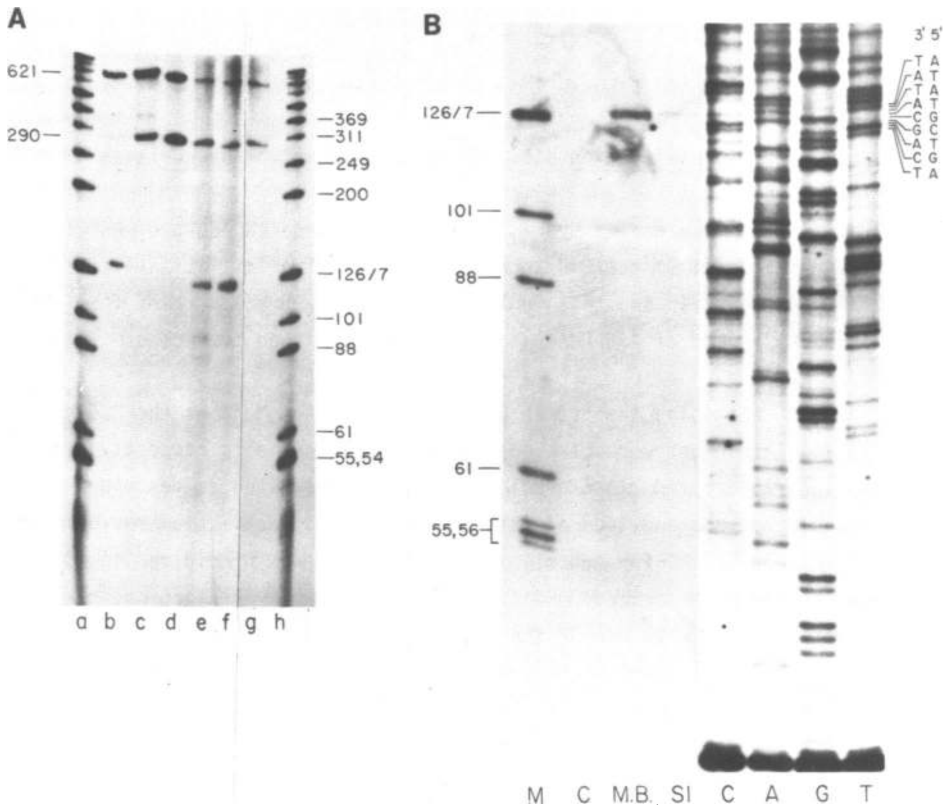

Figure 2 A. Nuclease protection mapping experiments in which double-strand DNA labeled at the Hind III site was denatured and hybridized to 45S RNA. Control hybridizations contained only tRNA. Lanes (a) and (h), SV40 DNA digested with EcoRII. Lane (b), hybrid digested with 1.7 U/ml Exo VII; (c) control digested with 1.7 U/ml Exo VII; (d) control, no enzyme digestion. Lane (e), hybrid digested with 80 U/ml mung bean nuclease; (f) hybrid digested with 20 U/ml mung bean nuclease; (g) control digested with 20 U/ml mung bean nuclease. DNA bands at 290 and 621 bases correspond to intact SalI/HindIII fragments. B. Determination of the 5' end of 45S RNA by nuclease protection mapping with rDNA labeled at the Hind III site. Lane (M), SV-40 DNA digested with Eco RII; (M.B.), 45S/rDNA hybrid digested with 80 U/ml mung bean nuclease; (S1) 45S/rDNA hybrid digested with 200 U/ml S1 nuclease. The protected material was electrophoresed on a .4 mm thick, 8% polyacrylamide-8M urea gel in parallel with a dideoxy-sequencing ladder of BH.41 digested with Hind III as described in the Results section. The two panels are different exposures of samples electrophoresed on the same gel as shown.

To determine whether the site identified above was the most upstream site protected by 45S RNA, pEH 5.0 DNA was labeled at the Sal I site 290 bases upstream from the Hind III site and used for protection mapping experiments.

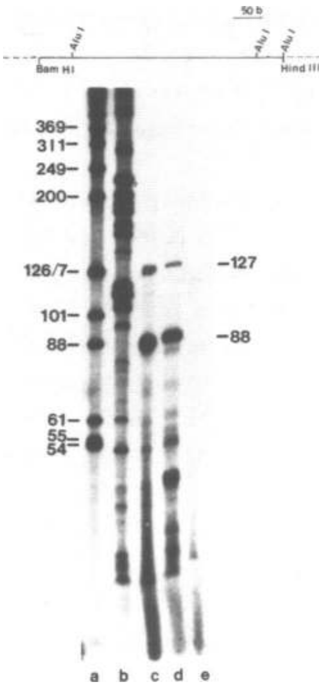

Figure 3. Nuclease protection mapping experiments in which the coding strand of BH.41 DNA, cut with Alu I and end-labeled, was hybridized to 45S RNA or total nucleolar RNA, as described in Materials and Methods. Lane (a), SV-40 DNA digested with Eco RII. Lane (b), control hybridization (no RNA), no nuclease digestion; (c) hybridization with 45S RNA, digested with S1 nuclease; (d) hybridization with total nucleolar RNA, digested with S1 nuclease, and (e) control hybridization (no RNA), digested with S1 nuclease.

No protection of this Sal I site by 45S RNA was seen (data not shown).

Protection mapping experiments were also performed with single-strand M13 recombinants. The coding strand of the rDNA fragment BH.41 was cut with Alu I, end-labeled, and hybridized to 45S RNA and to total nucleolar RNA at 63°C overnight and then digested with S1 nuclease. The protected fragments can be seen in Figure 3. Both RNAs strongly protect a band at 88 bases. This is the size of the band that is predicted by labeling at the Alu I site at +88 in the rDNA sequence (see Fig. 4). Incomplete cutting at this site produces a DNA probe 39 bases longer that is labeled at the Alu I site at +127. Protection of this DNA accounts for the less intense band at 127 bases. (In agreement with other studies (29) we are unable to achieve complete cutting of single-strand DNA with restriction enzymes). The smaller 39 base fragment that is produced by cuts at +88 and +127 is not well resolved in this system. Neither 45S RNA nor total rat liver nucleolar RNA produce any detectable protection of DNA sequences further than 125 bases upstream from the Hind III site. Protection experiments with nucleolar RNA produce some minor bands migrating faster than the 88 base fragment that are reduced or absent in

protection experiments with purified 45S RNA. These bands probably represent processing or degradation intermediates as they are also present in experiments in which the hybrid is digested with ExoVII (unpublished results). Other minor bands are probably due to sequence heterogeneity of the ribosomal DNA repeats.

The discrete and reproducible nature of the major protected band seen with DNA labeled at the Hind III site, as well as at the Alu I sites, and the lack of observable protection of sequences further upstream strongly suggest that 45S transcription begins approximately 125 bases upstream from the Hind III site. The exact location of this site was determined by electrophoresing 45S RNA-protected DNA labeled at the Hind III site alongside dideoxy-sequenced BH.41 DNA cut with Hind III (Fig. 2B). Since the only difference between the sequencing ladder and the protected fragment is the presence of a 3'-OH on the protected DNA, the 3' dideoxynucleotide corresponding to the 3' end of the protected DNA can be directly read from the sequencing ladder. This alignment allows us to pinpoint the G residue 125 bases upstream from the Hind III site in the noncoding strand as the first nucleotide (+1) of 45S rRNA.

Sequence of the DNA surrounding the 45S transcription initiation site

The DNA sequence of approximately 900 bases surrounding the putative 45S RNA initiation site has been determined (Fig. 4). This region extends from the Bam HI site 280 bases upstream from the initiation site to the Xho I site at +640. The sequences within 20-40 bases of the Sal I site, the Hind III site and the Xho I site were determined by the chemical cleavage method of Maxam and Gilbert. These sequences were confirmed and the sequence of the remaining region determined by the Sanger dideoxy-chain terminating procedure using rDNA recombinants in the single-strand phage M13. Both orientations of the Hind III/Xho I fragment and the Bam HI/Hind III fragment were sequenced.

The 5' end of the 45S RNA maps at the G residue labeled +1. The entire sequence of the rat rDNA reported here is very similar to that of the initiation site region of mouse rDNA (12,13). In Figure 4, the mouse sequence is aligned with the rat sequence to maximize homology. Identical bases are underlined. Approximately 80% of the bases from -169 to -1 are the same and there is a 70% homology from +1 to +200.

The rat rDNA sequence downstream from the initiation site contains several interesting features. The first is the presence of the aforementioned long stretches of T residues in the non-coding strand. There are three T-rich blocks: one from +352 to +407, a second from +502 to +545, and a third from +622 to +629. Only the stretch of 8 thymidines (+622 to +629) is

100 b

Bam HI Sal I Hind III XhoI

45S rRNA precursor

```
       -280        .         .      -250         .        . ▼       .
Rat:   GATCCTCCCCGGTCCCTGCCAGCGGGAGCATGGACTTCTGAGGCCGAGGGGAAGCTGTGTTCCAGGCTAT

         .      -200       .         .         .         .      -150
Rat:   CTACCATGGCCTCCTCGGTCTGTGAGCACTCAGGGTTCTAAGGTCGACCAGTTGTTCCTTTGCGGTCCGG
Mouse:                                            gtcgaccagttgttcctttgaggtccgg

         .         .         .         .       -100         .         .
Rat:   TTCTCTTTCTACATGGGGAC-CTCTTGGGGACACGTCACCGAACATGACTTCCAGACGTTCCGTGTGGCC
Mouse: ttct-tttcgttatggggtcattttggg-ccacctccacaggtatgacttccaggtattctctgtggcc

         .         .      -50          .         .         .         .
Rat:   TGTCATGTTTATCCCTGTGTCTTTTACACTTTTCATCTTTGCTATCTG-TCCTTATTGTACCTGGAGATA
Mouse: tgtcactttcctccctgtctcttttatgcttgtgatcttttctatctgttcc-tattggacctggagata

          +1=initiation site .         .         .       +50         .
Rat:   TATGCTGACACGCTGTCCTTTTGACTTCTTTTTGTCATTAAAGGACGTTGGAAGAGGCTTGCACC----A
Mouse: ggtactgacacgctgtcctttcccta-----ttaacactaaaggacactataagagaccctttcgattta

                .         .      ▼ .      +100         .         .   ▼ .
Rat:   AGGCTGTTT-GCTTGTCCAGCCCTAGCTCTTTTCTTCTGCGCATGGGCCTCTTCGA-TGCTTGAAGCTTA
Mouse: aggctgttttgct-gtccagcc-tat-tcttttta-ctggct-tgggtctgtcgcggtgcctgaagct--

                 .      +150       .         .         .         .    ▼
Rat:   GCGTCCCCCCATGAGTACGCGCTTCCTGCTTTCCCGTGCTTGC---TTGCCTGTGCTCTCTG-GGGCAGC
Mouse: --gtcccc----gagcca-cgcttcctgctttcccgggcttgctgcttgcgtgtgcttgctgtgggcagc

        +200         .         .         .         .         .         .
Rat:   TTTATGACAACCGTCCC------------GCGTGTCAGGCG---TTCCCGATTTCCCCGTGGTG--GTTG
Mouse: ttg-tgacaac--tgggcgctgtgactttgcgtgtcagacgttttcccgatttccccgaggtgtcgttg

                 .         .         .         .       +300         .
Rat:   TCGTCCGTTGCCGGTAGGAGTCGTTGGTGCCGAGTGCGAC-TGGA--AGGGTTTTCCCGTTTGGTGCTAG
Mouse: tcacatctcctgtcccggttggaatggtgg-gagccagctgtggttgagggccgccttattt--------

         .         .         .      +350         .         .         .
Rat:   TGACCCCCTGGCGTGCTCCTCTGCGGCCGACCGGTTTTTTTATTTGTTTTTTTTTTTTTTTTGTTTTTTTTG
Mouse: ----------------------cggctcac---------------ttttttttttttt---------

         .      +400         .         .         .         .       +450
Rat:   TTTTTTTTTTTTTTGTTTTTTGGAAGGAGTTCCCGAACCTCCGGTGCTTGGTGGTGTGTCC---CTTTCTT
Mouse: -----------------cctcttggagt-cccgaacctccgctctttctttcccggtctttctttctt

                 .         .         .         .      +500         .         .
Rat:   TCCTGCTGTGTGCCTCCCGAGTTGCACCTTTTCTCCTTCGAAGGGGATTTTATTTTTTTATTTTTATTTT
Mouse: -ccaca----tgcctcccgagt-gca---tttc-------------ttttgtttttttctttttttt

             .         .      +550         .         .         .
Rat:   TTTTTTATTTTTATTTTTTTTGAAGG-----AGTTCCCGAACCTCCGCTGCCCGTTGAGTCCCGTTCTTC
Mouse: tttttttttt-------ggggaggtggagag-tcccgagtacttcactcctgtctgcggtgtccaagt

         .      +600         .         .         .       +640
Rat:   CACG---CCACGTGCCTCCCGAGTGCAACGCTTCCTTTTTTTTCTCGCCCTCGAG
Mouse: gttcatcccacgtgcctcccgagtgca--------ttttttttgtggcagtcgct
```

Figure 4. DNA sequence of the region surrounding the start site of 45S RNA. The nucleotide sequence of the non-coding strand of rat rDNA is shown in capital letters. The sequence of mouse rDNA from Urano et al. (12) is shown in lower case letters. The two sequences are aligned to maximize homology; identical bases are underlined. Alu I sites are indicated by arrows above the sequence.

uninterrupted. The other two regions are either interspersed predominately with A residues (+502 to +545) or with G residues (+352 to +407). The T-rich regions are very homologous to T stretches at the same relative sites in the mouse sequence. The major difference in the mouse sequence is that the T stretch corresponding to the one in rat beginning at +352 is only 15 residues long.

Another apparent feature of this region is the presence of two sets of direct repeat sequences. The 21 bp sequence GAAGGAGTTCCCGAACCTCCG at +409 to +429 is repeated 137 bases downstream at +546 to +566. This sequence is located at the 3' end of the two large T stretches. The second direct repeat is a 16 bp sequence located at +464 to +480 and at +598 to +613. The repeated sequence is GTGCCTCCCGAGTGCA, but there is an extra T at +477. This sequence is located approximately 30 bases downstream from the repeats of the 21 bp sequence. A repeat identical to the 16 bp sequence is found in the same relative location in mouse rDNA (12,13).

DISCUSSION

In this paper we have mapped the 5' end of rat 45S precursor rRNA and reported the sequence of approximately 900 nucleotides surrounding this site. Three different single-strand specific nucleases were used in nuclease protection mapping experiments. This approach was taken in order to minimize the appearance of artifactual bands that have been described in experiments using S1 nuclease. In particular, it is known that the hybrid rU/dA is relatively unstable (30) and is sensitive to S1 nuclease (27). We found that rU/dA-rich sequences in 45S RNA/DNA hybrids were, indeed, sensitive to S1 nuclease under our digestion conditions and led to the appearance of variable numbers of protected DNA fragments on gels. Therefore, to reduce the problem of artifactual digestion due to the nature of an individual enzyme or the sequence in question, two different single-strand specific endonucleases were used, S1 and mung bean, and, of particular importance, one single-strand specific exonuclease, Exo VII. Exo VII should not cleave at any internal sequences even if poly rU/dA stretches were being melted under the digestion conditions.

When protection experiments were done with 45S RNA hybridized to rDNA labeled at the Hind III site, all three enzymes produced the same size DNA fragment, approximately 125 bases in length. The 5' end of the precursor RNA was precisely determined by running the protected DNA fragment in parallel with a dideoxy-sequencing ladder of BH.41, digested with Hind III. The first

nucleotide of the 45S precursor RNA is G, 125 bases upstream from the Hind III site in the noncoding DNA strand.

Protection mapping experiments using double-strand DNA end-labeled at the Sal I site 290 bases upstream from the Hind III site showed no protection of this site. In addition, total rat liver nucleolar RNA, which would be expected to contain any potential larger rRNA precursors, was hybridized to end-labeled single-strand BH.41 DNA (coding strand). No detectable protection of sequences 5' to the G residue at +1 was observed.

Based on our protection experiments and the close homology between our site and the sites mapped by in vitro transcription on human (15) and mouse (31,32,33) rDNA recombinant plasmids, we believe that the site we have mapped at +1 with 45S pre-rRNA represents a true initiation site for rDNA transcription. This conclusion is in agreement with Financsek et al. (16) who identified and sequenced a transcription initiation site for rat 45S pre-rRNA almost identical to the one described in this work.

Financsek et al. (16) reported that approximately 12% of rat 45S RNA was capable of being capped with rat liver guanylyltransferase and thus had a 5' di- or triphosphate end indicating an initial transcription product. Using vaccinia capping enzyme, we have been unable to cap more than 0.1% of the 45S RNA as prepared by our procedures. There are several possible explanations for this inability to demonstrate a 5'-triphosphate end on all 45S RNA molecules. It has been shown both in vivo and in vitro that mouse precursor rRNA initiating at the +1 nucleotide is rapidly processed to a molecule approximately 650 bases smaller (27). Therefore, 45S RNA is actually a heterogeneous collection of processed and unprocessed molecules. It is also possible that some of the 45S molecules represent rDNA transcripts initiated upstream of the +1 position, then rapidly processed at the +1 site. Support for such an idea comes from studies on amphibian oocyte rDNA in which transcription is observed in the so-called "spacer" regions of some rDNA repeats that may be part of larger, rapidly processed pre-rRNA molecules (34,35,36). Despite these considerations, it seems probable that the site we have mapped at +1 is a major start site for precursor ribosomal RNA synthesis.

Comparison of the sequence reported here with that of the rDNA initiation site from -166 to +127 reported by Financsek et al. (16) shows only a few differences. Our sequence contains four additional bases in this region: a T at -21, a C at -12, and T residues at +19 and +24. Interestingly, the additional bases at -21 and -12 move the rat rDNA sequence into closer alignment with that of the mouse rDNA. The extra bases at +19 and +24 lie in

the loop of the potential hairpin structure discussed by Financsek et al. (14). There are also two base substitutions: a C at -86 in our sequence, instead of a T, and a T at +122, instead of a C.

Sequence comparison of the rat initiation site with that of the mouse initiation region (Fig. 4) shows extensive homology between the two rodent species. The spatial arrangement of this homology is interesting. The most frequent blocks of similar or identical sequences are seen upstream of the initiation site, up to and including the Sal I site at -165. The sequence surrounding the initiation site is identical in the two rodents from -14 to +18, except for G residues at -2 and -3 and an A residue at +1 in the mouse. Downstream from the initiation site there is also considerable homology, although not quite as extensive nor as spatially precise as that noted above. In particular, the three T-rich stretches in the noncoding strand are maintained in the same relative locations. The close similarity between the mouse and rat sequences is especially noteworthy in view of the fact that three *Xenopus* species show similar or less homology within 200 bases upstream of the rDNA initiation site (11).

When sequence comparisons are extended to other organisms, much less homology is seen. As shown by Financsek et al. (14), comparison of mouse, rat and human rDNA shows an exact homology at +2 to +16, the conserved sequence being: CTGACACGCTGTCCT. The homology in this region can be extended slightly further, from -9 to +18, to generate the following consensus sequence: GAGATATATGCTGACACGCTGTCCTTT. Other rDNAs, yeast (9), *Drosophila* (8), and *Xenopus* (10, 11), do not contain sequences at the initiation site resembling this consensus sequence. It is possible that this sequence is a characteristic limited to the mammalian rDNAs in which the ribosomal precursor RNAs are, in general, larger. It is important to note, however, that despite the close homology of this particular sequence among the mammalian rDNA initiation sites sequenced to date, *in vitro* transcription systems, utilizing S-100 extracts and plasmid DNAs, demonstrate polymerase I transcription to be species specific between mouse and human rDNA (37). This implies, therefore, that close homology at this consensus site is not sufficient for initiation of transcription in a heterologous system. Either regions outside this area or minor differences within this site have a major influence on the initiation of polymerase I transcription.

Upstream from the mammalian initiation site there are four sets of identical sequences that are found in approximately the same location in all three mammals. These are: 1) ATT at -19 in mouse and rat, -22 in human; 2)

ATCTTT at -38 in mouse and rat, -43 human; 3) TCCC at -61, mouse and rat, -63 human; and 4) TGGGG at -128 mouse and rat, -130 human. Recently, Grummt (38) has shown that deletion of sequences between -39 and -34 results in a 90% decrease in rDNA transcription in vitro. This suggests that the conserved ATCTTT sequence located at -38 to -33 in mouse and rat rDNA may be a functional element required for initiation of polymerase I transcription on rDNA.

ACKNOWLEDGEMENTS

We thank Barbara Masters and James Beeson for identifying and isolating lambda clone 1A. We are indebted to Lisa Buchanan for technical assistance and to Cathy Inouye and Lynn Van Meter for typing of the manuscript and Figure 4. We thank Drs. Elaine Lewis and Thomas Walker for helpful discussion and critical reading of the manuscript.

This work was supported by Grant GM 29302 from the National Institute of Health. C.A.H. was supported by NIH grant 5 T32 GM 07385-04 from 1/81 to 7/81.

REFERENCES

1. Grosschedl, R. and Birnstiel, M.L. (1980) Proc. Natl. Acad. Sci. USA 77, 1432-1436.
2. Benoist, C. and Chambon, P. (1981) Nature 290, 304-310.
3. Mathias, D.J. and Chambon, P.O. (1981) Nature 290, 310-315.
4. McKnight, S.L., Gaves, E.R., Kingsbury, R. and Axel, R. (1981) Cell 25, 385-398.
5. McKnight, S.L. and Kingsbury, R. (1982) Science 217, 316-324.
6. Korn, L.J. (1982) Nature 295, 101-105.
7. Hall, B.D., Clarkson, S.G. and Tocchini-Valentini, G. (1982) Cell 29, 3-5.
8. Long, E.O., Rebbert, M.L. and Dawid, I.B. (1981) Proc. Natl. Acad. Sci. USA 78, 1513-1517.
9. Bayev, A.A., Georgiev, E.I., Hadjiolov, A.A., Kermekckiev, M.B., Nikolaev, N., Skyrabin, K.G. and Zakharyev, V.M. (1980) Nucleic Acids Res. 8, 4919-4926.
10. Sollner-Webb, B. and Reeder, R.H. (1979) Cell 18, 485-499.
11. Bach, R., Allet, B. and Crippa, M. (1981) Nucleic Acids Res. 9, 5311-5330.
12. Urano, Y., Kominami, R., Mishima, Y. and Muramatsu, M. (1980) Nucleic Acids Res. 8, 6043-6058.
13. Bach, R., Grummt, I. and Allet, B. (1981) Nucleic Acids Res. 9, 1559-1569.
14. Financsek, I., Mizumoto, K., Mishima, Y. and Muramatsu, M. (1982) Proc. Natl. Acad. Sci. USA 79, 3092-3096.
15. Miesfeld, R. and Arnheim, N. (1982) Nucleic Acids Res. 10, 3933-3949.
16. Financsek, I., Mizumoto, K. and Muramatsu, M. (1982) Gene 18, 115-122.
17. Maxwell, I.H., Maxwell, F. and Hahn, W.E. (1977) Nucleic Acids Res. 4, 241-246.

18. Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstradiatis, A. (1978) Cell 15, 687-701.
19. Southern, E. (1975) J. Mol. Biol. 98, 503-517.
20. Messing, J., Crea, J. and Seeburg, P.H. (1981) Nucleic Acids Res. 9, 309-321.
21. Maxam, A.M. and Gilbert, W. (1980) in: Methods of Enzymology, Eds. Grossman, L. and Moldave, K., Vol. 65, pp. 499-560, Academic Press, New York.
22. Penman, S. (1969) in: Fundamental Techniques in Virology, Eds. Habel, K. and Salzman, N.P., pp. 35-48, Academic Press, New York.
23. Imada, M. and Sueoka, N. (1978) Devel. Biol. 66, 97-108.
24. Favaloro, J., Treisman, R. and Kamen, R. (1980) in: Methods of Enzymology, Eds. Grossman, L. and Moldave, K., Vol. 65, pp. 499-560, Academic Press, New York.
25. Sanger, F., Niklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
26. Smith, A.J.H. (1980) in: Methods in Enzymology, Eds. Grossman, L. and Moldave, K., Vol. 65, pp. 560-580, Academic Press, New York.
27. Miller, K.G. and Sollner-Webb, B. (1981) Cell 27 (Part 2), 165-174.
28. Sharp, P.A., Berk, A.J. and Berget, S.M. (1980) in: Methods of Enzymology, Eds. Grossman, L. and Moldave, K., Vol. 65, pp. 750-768.
29. Hofer, B., Ruhe, G., Koch, A. and Koster, H. (1982) Nucleic Acids Res. 9, 2763-2773.
30. Martin, F.H. and Tinoco, I. (1980) Nucleic Acids Res. 8, 2295-2305.
31. Grummt, I. (1981) Proc. Natl. Acad. Sci. USA 78, 727-731.
32. Grummt, I. (1981) Nucleic Acids Res. 9, 6093-6102.
33. Mishima, Y., Yamamoto, O., Kominami, R. and Muramatsu, M. (1981) Nucleic Acids Res. 9, 6773-6785.
34. Rungger, D., Achermann, H. and Crippa, M. (1979) Proc. Natl. Acad. Sci. USA 76, 3957-3961.
35. Franke, W.W., Scheer, U., Spring, H., Trendelenburg, M.F. and Krohne, G. (1976) Exp. Cell Res. 100, 233-244.
36. Moss, T. and Birnstiel, M.L. (1979) Nucleic Acids Res. 6, 3733-3743.
37. Grummt, I., Roth, E. and Paule, M.R. (1982) Nature 296, 173-174.
38. Grummt, I. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 6908-6911.
39. Rothblum, L. I., Reddy, R. and Cassidy, B. (1982) Nucleic Acids Res. 10, 7345-7362.

NOTE ADDED IN PROOF

After submission of this manuscript, Rothblum et al. (39) published the sequence of 1,100 nucleotides surrounding the transcription initiation site of rat. Comparison of the 900 bases of their sequence that overlap the one reported here shows 13 base differences, none of which are in the region from -110 to +20.