The Nucleotide Sequence of a Rat 18 S Ribosomal Ribonucleic Acid Gene and a Proposal for the Secondary Structure of 18 S Ribosomal Ribonucleic Acid

Yuen-Ling Chan†, Robin Gutell‡, Harry F. Noller§, and Ira G. Wool¶
From the †Department of Biochemistry, The University of Chicago, Chicago, Illinois 60637 and the ¶Thimmann Laboratories, University of California at Santa Cruz, Santa Cruz, California 95064

The nucleotide sequence of a rat 18 S rRNA gene was determined. The 18 S rRNA encoded in the gene contains 1874 nucleotides, and the molecular weight estimated from the sequence is 6.09 × 10^6. The sequences of rat and *Xenopus laevis* 18 S rRNAs are very similar; the only differences of consequence are the insertions between nucleotides 197 and 206 and between 258 and 279 of the rat nucleic acid of sequences rich in guanine and cytosine. A proposal is presented for the secondary structure of rat 18 S rRNA based on a comparison with the sequences of 17 other small ribosomal subunit RNAs. While the primary sequences of rat and eubacterial RNAs are different, the deduced secondary structures are remarkably similar.

Knowledge of the chemistry of the constituents is required for an analysis of the structure of ribosomes. Information on the structure is, in turn, necessary to relate the molecular organization of ribosomes to their function in protein synthesis. Eukaryotic (rat liver) ribosomes contain four molecules of RNA: 18 S RNA in the 40 S subparticle and 5, 5.8, and 28 S RNAs in the 60 S subparticle. The covalent structures of 5 S (1) and 5.8 S (2, 3) RNAs from rat liver ribosomes and from the ribosomes of a number of other eukaryotic species (4) have been determined. There is less information on the structure of the large eukaryotic rRNAs. The sequences of 18 S rRNAs from *Saccharomyces cerevisiae* (5) and *Xenopus laevis* (6) and of 26 S rRNAs from two species of yeast, *S. cerevisiae* (7) and *Saccharomyces carlsbergensis* (8), have been inferred from the structure of the genes. The sequence of the large RNA(s) from the ribosomes of mammalian species had not been determined. The information is important since a great deal is known, perhaps more than for any other eukaryotic species, of the chemistry of the other molecular constituents of mammalian ribosomes (9) and of the biochemistry and regulation of protein synthesis in mammalian cells (10, 11). Moreover, the most important data for deriving the secondary structure of ribonucleic acids comes from a comparison of the sequence of the molecule in different species. For all these reasons, we have determined the covalent structure of a gene for rat 18 S RNA. We cloned a rat 18 S RNA gene carried on a plasmid (Charon 4A) in plasmid vectors and determined the sequence of the DNA.

Eukaryotic ribosomal RNA genes occur in clusters of tandemly repeated transcription units. The mammalian primary transcript from ribosomal DNA has a sedimentation coefficient of 45 and contains 18, 5.8, and 28 S rRNAs as well as transcribed spacers which are removed during processing (12).

EXPERIMENTAL PROCEDURES

Materials—DNA from the λ phage Charon 4A in which the rat ribosomal RNA genes had been cloned (13) was kindly provided by T. D. Sargent and J. Bosner of California Institute of Technology. Restriction endonucleases and phage T4 DNA ligase were obtained from Bethesda Research Laboratories or New England Biolabs. [γ-^32P]ATP (approximately 3 × 10^6 Ci/mmol) was from Amersham-Searle, and calf intestinal alkaline phosphatase and phage T4 polynucleotide kinase were from Boehringer Mannheim. Digestion of DNA with restriction enzymes was carried out as recommended by the supplier.

Preparation of Plasmids Containing rDNA—The phage Charon 4A, containing rDNA sequences derived from an Eco RI digest of the rat genome (13), was grown and purified as recommended (14). The presence of rDNA was confirmed by hybridization of rat liver 18, 5.8, and 28 S rRNAs to the recombinant phage DNA. The phage was dialyzed overnight against 10 mM Tris/HCl, pH 8.0, 50 mM EDTA, heated at 70 °C for 30 min, dialyzed exhaustively against 10 mM Tris/HCl, pH 8.0, 5 mM NaCl, 1 mM EDTA, and stored at 4 °C.

The RNA genes were subcloned (15, 16) into the plasmid pACYC 184 (17). The Charon 4A recombinant DNA (1 µg) was digested with HindIII, and 0.25 µg of pACYC 184 DNA that had been hydrolyzed with the same restriction enzyme was added. After the HindIII was inactivated by heating at 65 °C for 15 min, the mixture was cooled, and sufficient buffer (66 mM Tris/HCl, pH 7.6, 7 mM MgCl2, 10 mM dithiothreitol, 0.4 mM ATP) was added to give a volume of 20 µl. T4 ligase was added and the sample was incubated at 22 °C for 2 h. Aliquots were used to transform calcium-treated *Escherichia coli* MC 1065. The cells were grown on Luria-Bertani agar plates (14) containing either tetracycline (10 µg/ml) or chloramphenicol (100 µg/ml) to detect transformants. The recombinant plasmid with most of the 18 S RNA coding sequence is referred to as pHeR 118. In a similar manner, an Eco RI insert in Charon 4A containing the 3' end of 18 S rDNA and most of the 28 S RNA coding region was cloned in pACYC 184. This plasmid is referred to as pRRR 228. The presence of the relevant inserts in the two plasmids was confirmed by colony hybridization (18, 19) to ^32P-labeled 18 and 28 S rRNAs. The plasmids were amplified using spectinomycin (17) and isolated (20).

DNA-RNA hybridization—The DNA was digested with restriction enzymes and separated by electrophoresis (0.5–2 V/cm) in agarose gels (0.8–1.0%) prepared in TBE buffer (50 mM Tris borate, pH 8.3, 40 mM NaCl, 1 mM EDTA) at 60 °C. Hybridization (18, 19) to ^32P-labeled 18 and 28 S rRNAs. The DNA fragments was in 50% formamide and 5 × SSC (1 × SSC: 0.15 M sodium citrate, 9.15 × NaCl incubated overnight at 37 °C. The filters were washed with 2 × SSC containing heat-treated pancreatic ribonuclease (0.001%) and then washed extensively with 2 × SSC. The filters were dried, and the autoradiographs were prepared at -70 °C using intensifying screens.
Preparation of Radioactive Nucleic Acids—The DNA or RNA was made radioactive by labeling at the 5' end with T4 polynucleotide kinase and [γ-32P]ATP after removal of phosphate by repeated addition of 0.1 unit of calf intestine alkaline phosphatase (23). The reaction mixture was boiled for 1 min before phenol extraction.

Isolation of DNA Fragments—DNA fragments resulting from digestion with restriction enzymes were separated by electrophoresis in either 1% agarose or 6-8% polyacrylamide gels. The gels were either stained with ethidium bromide or radioautographs were made. The bands containing DNA were cut out of the gel, and the nucleic acid was recovered from the slices by electrophoretic elution into dialysis bags (21). However, DNA prepared in this way contains contaminants that cause smearing on the gels used to determine the sequence. Thus, if the fragment was to be used directly for sequence determination, the DNA was recovered by a modification of the method recommended by Girvitz et al. (24) in which the orientation of the dialysis membrane was changed and the DNA was recovered by washing the filter paper and the dialysis membrane with 10 mM Tris/HC1, pH 7.5, 0.25 M NaCl, 1 mM EDTA.

The strands of DNA fragments were separated by heating at 90 °C for 2-3 min in TBE buffer containing 80% dimethyl sulfoxide, 0.05% bromphenol, 0.05% xylene cyanol. The sample was frozen immediately and thawed prior to electrophoresis which was as described above.

Restriction Endonuclease Digestion—Endonuclease sites were determined using the multiple enzymes (25) or the partial digestion methods (26). The fragments were separated by electrophoresis in 5-8% polyacrylamide gels.

Determination of the Sequence of DNA and RNA—The sequence of DNA was determined using the method of Maxam and Gilbert (23) with analysis on 8, 10, or 20% polyacrylamide gels (32 × 40 × 0.05 cm) (27). Radioautographs of the gels were prepared at -70 °C with or without intensifying screens. The sequence of the 5' end of 18 S rDNA was determined by both the enzymatic (28-30) and chemical (31) procedures.

RESULTS AND DISCUSSION

The Determination of the Sequence of 18 S rDNA—A preliminary restriction map of the region of the Charon 4A recombinant λ phage containing the rDNA genes was constructed (Fig. 1). In order to amplify the 18 S rDNA and to facilitate purification, the gene was subcloned in pACYC 184 at the HindIII and EcoRI sites. Plasmid pHRR 118 contains a 5.8-κb HindIII insert encompassing approximately 85% of the 18 S rRNA gene including the 5' terminus (Fig. 1). The remainder of the gene was contained in an EcoRI fragment of 6.4 kb that was cloned in pACYC 184 to form pRRR 228 (Fig. 1). The latter contains, in addition to the 5' end of the 18 S rRNA gene, the 5.8 S rRNA gene, most of the 28 S rRNA gene, and two internal transcribed spacers. It is important that the 3' end of the HindIII insert (pHRR 118) overlaps the 5' end of the EcoRI insert (pRRR 228).

A restriction map of pHRR 118 revealed that a combination of the enzymes BamHI, HindIII, and XhoI would release two fragments containing portions of the 18 S rRNA gene (Fig. 1). The entire BamHI-HindIII fragment (1.07 kb) is located within the 18 S rRNA gene, whereas the XhoI-BamHI fragment (2.6 kb) contains the 5' end of the gene, the external transcribed spacer, and part of the nontranscribed spacer. An analysis of the cleavage pattern obtained with a series of restriction endonucleases identified enzymes suitable for generating fragments of a size convenient for a determination of the nucleotide sequence and for the production of overlapping sequences (Fig. 2). In a similar manner, the 3' end of the 18 S rRNA gene was isolated from pRRR 228 by digesting the plasmid with EcoRI and XhoI (Fig. 1). This fragment (0.6 kb) was digested with HpaII to generate subfragments suitable for a determination of the nucleotide sequence (Fig. 2). The presence of the 18 S RNA gene in a fragment was sometimes confirmed, prior to determination of the sequence, by Southern hybridization (22) to labeled 18 S RNA. Finally, a UNIX computer program was written to search the fragments whose sequence had been determined for additional restriction endonuclease sites to provide overlapping fragments.

The sequence of the 18 S rRNA gene was determined by the Maxam and Gilbert procedure (23). The analysis included a determination of the sequence of almost all of both strands and extensive overlaps of portions of the sequences in separate fragments (Fig. 2). The order of the fragments was approximated from the restriction endonuclease map, from comparison with the sequence of X. laevis 18 S rRNA (6), and by reference to the sequence of the 3' end of rat 18 S rRNA (32, 33).

We determined the sequence of the 5' end of 18 S rRNA by both the chemical (31) and enzymatic (28-30) methods (results not shown). This determination located the 5' end of the gene and confirmed the sequence obtained for the DNA. The sequence at the 3' end of 18 S rRNA had been done before (32, 33). The sequence of the nucleotides in 18 S RNA was inferred from the sequence of the gene (Fig. 3). Rat 18 S RNA has 1874 nucleotides. The molecular weight is approximately 6.09 × 10^6 (based on the assumption of a molecular weight of 325 for a nucleotide), somewhat less than the 7 × 10^6 (2154 nucleotides) estimated from physiochemical data (34, 35). The G + C content of rat 18 S rRNA (55.7%) is higher than that of X. laevis (53.8%; Ref. 6) and of S. cerevisiae (45%; Ref. 6).

Comparison of the Sequence of Rat 18 S rRNA with Other Eukaryotic 18 S rRNAs—When the sequences of rat and X. laevis 18 S rRNAs were aligned, there were 1752 identities (the same nucleotide at the same position) of 1822 possible matches, i.e. 96% identity. Of the nonidentities, 32 (or 26%) of the total of 122 nucleotides are in two regions, 197-206 and 258-279 (see below). It follows then that there are long stretches of the two rRNAs that are identical. Indeed, it is the great similarity in the two sequences that enabled us to use the structure of X. laevis 18 S rRNA to provisionally order the fragments of the rat gene. It is possible that some of the scattered single nucleotide differences are actually errors either in analysis or in the transcribing of the sequence. The similarity between the sequences of rat and yeast 18 S rRNA is not nearly so great. There is only 75% identity. While there are regions of extensive identity, they are interrupted by tracts

---

1 The abbreviation used is: kb, kilobases.
having little similarity (cf. the comparison of yeast and *X. laevis* 18 S rRNAs in Ref. 6). The regions of *X. laevis* 18 S rRNA that are different from the sequence in yeast tend to be rich in G and C. Nonetheless, the sites of nucleotide methylation are highly conserved (see below also). Finally, there is little similarity between the sequence of *E. coli* 16 S rRNA (38) and that of eukaryotic 18 S rRNA (*yeast, X. laevis, or rat*) except in scattered regions and near the 3' end (37).

What differences there are in the sequences of rat and *X. laevis* 18 S rRNAs occur mainly in the 5' region; there are 62 nonidentities in the 400 nucleotides at the 5' terminus and only 18 in the 3' 400 nucleotides. The difference in the 5' region of rat and *X. laevis* 18 S rRNA is accounted for largely by 10 additional residues at position 197-206 and 22 additional residues at position 258-278 (Fig. 4). We presume the acquisition of the additional nucleotides is an evolutionary change since the two tracts do not occur in yeast or *X. laevis* 18 S rRNAs, but are found in the rabbit nucleic acid (Ref. 38; Fig. 4). The two insertions are particularly rich in G + C, 84% in rat and 72% in rabbit; indeed, they largely account for the slightly higher G + C content of rat compared to *X. laevis* 18 S rRNA. The functional significance, if indeed there is any, of the insertions is not known. However, variable nucleotide tracts are apparent when one compares *X. laevis* and yeast 18 S rRNAs, and they are G + C-rich also (6); indeed, it may be true for eukaryotes in general (39).

There is an oligonucleotide stretch (position 338-345) present in rat 18 S rRNA that is absent in rabbit. Since the sequence is present in *X. laevis* 18 S rRNA, an organism phylogenetically far more distant from rat than rabbit, it seems more likely to be an error in sequencing than an evolutionary change. Such errors are known to be more likely when one determines the sequence of the RNA, as was the case for rabbit 18 S rRNA (38), rather than the DNA. There are also some discrepancies in the 3'-terminal sequence determined directly from rat 18 S rRNA (33) and the sequence obtained here from the DNA (Fig. 5). Once again, we would ascribe that to the limitations inherent in the former procedure.

Ribosomal nucleic acids are methylated. We have not identified the nucleotides in rat 18 S rRNA that are methylated. However, the pattern of methylated bases in oligonucleotide digests from 18 S rRNAs from human (HeLa cells), hamster (BHK/c13 cells), and mouse (L-cells) could not be distinguished (40). Moreover, the methylated nucleotides in 18 S rRNA from humans (HeLa cells) and from *X. laevis* have at least 95% identity (40). All of the methylation sites in *X. laevis* 18 S rRNA are conserved in the rat nucleic acids. Thus, the methylated bases in rat 18 S rRNA are likely to be the same as those in HeLa cells.

**A Proposal for the Secondary Structure of Rat 18 S rRNA**

An analysis of eubacterial ribosomal RNAs has established that the most powerful and most reliable means available for deducing secondary structure is comparative sequence analysis (41-45). This method derives from recognition of compensating base changes in paired nucleotides in analogous helices of rRNAs from different organisms. It was originally assumed that, to a first approximation, the 16 S rRNAs from various eubacteria were likely to have the same secondary structure. Now that more than 8 eubacterial 16 S rRNAs are available, the assumption seems valid (44). However, before extending this approach to the analysis of eukaryotic 18 S rRNAs, which have 200-300 additional nucleotides, it would seem prudent to re-examine the assumption. A tentative secondary structure for rat 18 S rRNA was derived from comparative sequence analysis (Fig. 6) and compared with *E. coli* 16 S rRNA (see the schematic in Fig. 6). The additional nucleotides in rat 18 S rRNA are accounted for by three large insertions and moderate increases in the size of four stems.

The secondary structure of rat 18 S rRNA (Fig. 6) is based on the comparative analysis of 17 complete sequences of small ribosomal subunit RNAs, of which three are from eukaryotic species (yeast, *X. laevis*, and rat), and a RNase T1, T2, and T3, and up to four rat nucleic acid segments from human (HeLa cells), hamster (BHK/c13 cells), and mouse (L-cells) could not be distinguished (40). Moreover, the methylated nucleotides in 18 S rRNA from humans (HeLa cells) and from *X. laevis* have at least 95% identity (40). All of the methylation sites in *X. laevis* 18 S rRNA are conserved in the rat nucleic acids. Thus, the methylated bases in rat 18 S rRNA are likely to be the same as those in HeLa cells.

**Fig. 2. Restriction endonuclease map of a rat 18 S rRNA gene and a diagram of the overlapping fragments whose sequence of nucleotides were determined.** The upper portion depicts the restriction endonuclease sites in the rat 18 S rRNA gene that were used to generate fragments (lower portion) for the determination of the sequence of nucleotides. NTS, nontranscribed spacer, ETS, external transcribed spacer, ITS, internal transcribed spacer. The numbers indicate the first nucleotide in the restriction endonuclease recognition sequence counting from the 5' end of the 18 S rRNA gene. Each arrow designates a nucleotide sequence that was determined. "S" strand is synonymous with the RNA; "C" strand is complementary.
at position 680, which accounts for 175 additional nucleotides. The foldings of that insertion and of the 25-nucleotide insert at position 520 were not evident in terms of canonical base pairs; the sequence of additional eukaryotic rRNAs will probably be required to solve the secondary structure of these regions.

Three regions of the secondary structure differ significantly between eukaryotes and eubacteria. The first is around positions 120 and 350, which in rat 18 S rRNA can form a series of loosely paired, interrupted helices, which contrast with the two stems separated by an internal loop posited for E. coli 16 S rRNA. The flanking structural and sequence similarities indicate the two eukaryotic loops are analogous to the 122–142/221–239 region in E. coli 18 S rRNA. The arrows designate a deleted nucleotide, and a space indicates that no nucleotide is present.

**Fig. 3. The sequence of nucleotides in rat 18 S rRNA.**

**Fig. 4. Comparison of the 5’-terminal region of rat, rabbit, and X. laevis 18 S rRNAs.** Differences from the sequence of rat 18 S rRNA are indicated by placement of the nucleotide outside of the box. A colon designates a deleted nucleotide, and Y a possible modified pyrimidine in the rabbit sequence.

**Fig. 5. A comparison of the 3’-terminal region of rat, rabbit, and X. laevis 18 S rRNAs.** The arrows indicate nucleotides that diverge from the sequence of the rat 18 S rRNA gene. A space indicates that no nucleotide was found at that position.
FIG. 6. A proposal for the secondary structure of rat 18 S rRNA. This structure is based on a general structure for eukaryotic 18 S rRNAs derived from comparative sequence analysis (44). Shaded helices are those considered to be proven on the basis of comparative sequence analysis, i.e. there are two or more compensating base changes (42). A drawing of the proposed secondary structure for E. coli 16 S rRNA (41-45) is in the lower right for comparison. The sequences designated a and b have not been assigned secondary structure and belong, as indicated, after positions 572 and 679, respectively. The dots after position 79 indicate where four nucleotides occur in eubacterial 16 S rRNA for which there are no counterparts in eukaryotic 18 S rRNA. The separate hairpin structure in the center shows an alternate base-paired structure for residues 3-21.
comparative sequence analysis (41–45). The second major difference involves a presumptive pairing of 564–570 and 880–886 in E. coli, which we relate to 663–669/1165–1171 in rat. There is not yet sufficient evidence from comparison of sequences to establish the existence of the stem in eubacteria owing to conservation of the primary structure. Evidence from psoralen cross-linking (46) and from ribosomal protein binding (47, 48) experiments is, however, consistent with a stem in that region of 16S rRNA. Here again, it may be the case that eukaryotic rRNA tends to be more irregular in its secondary structure of the primary structure. Evidence from conservation of the secondary structure of rat rRNA does not appear to be analogous to the E. coli secondary structure, which appears to be missing, and its location, although nearby, is displaced from that of the E. coli. It is of interest that eukaryotic 18S rRNAs do not bind E. coli ribosomal protein S8.

More provocative are the similarities in eukaryotic and eukaryotic rRNA sequences. The length of stems, the size of loops, and the sequence of unpaired regions are often identical or nearly so (Fig. 6). The unpaired sequences 617–632, 1698–1714, and 1830–1842 are nearly identical with their counterparts in E. coli and are enclosed by nearly identical secondary structural features. These three sequences have been directly implicated in the translational process. Kethoxal modification experiments have shown that a site in the first of these (617–632) is protected only when tRNA is bound to ribosomes (51), whereas the latter two sequences (1698–1714 and 1830–1842) contain nucleotides shielded by ribosomal subunit association (52). Finally, the wobble base of the anticodon of tRNA can be cross-linked to position 1706 (or its analogue) in eukaryotic and eukaryotic ribosomes (53, 54). Thus, it appears that the portions of the secondary structures of rRNAs that are highly conserved are likely to be important for function.

Acknowledgments—We are grateful to Jose Olvera for technical assistance, to Francis W. H. Leung for writing the computer programs, to C. Woes for valuable advice while we were formulating the secondary structure of rat 18S rRNA, and to Joann Kop, Joany Shah, M., Khan, N., Salim, M., and Maden, B. E. (1979) J. Mol. Biol. 110, 119–146.


The Structure of Rat 18S rRNA

The nucleotide sequence of a rat 18 S ribosomal ribonucleic acid gene and a proposal for the secondary structure of 18 S ribosomal ribonucleic acid.
Y L Chan, R Gutell, H F Noller and I G Wool


Access the most updated version of this article at http://www.jbc.org/content/259/1/224

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/259/1/224.full.html#ref-list-1