Volume 12 Number 11 1984 Nucleic Acids Research

Primary structure of rabbit 18S ribosomal RNA determined by direct RNA sequence analysis

John F.Connaughton, Asha Rairkar, Raymond E.Lockard and Ajit Kumar

Department of Biochemistry, George Washington University, School of Medicine and Health Sciences, Washington, DC 20037, USA

Received 17 February 1984; Revised and Accepted 8 May 1984

ABSTRACT

The primary structure of rabbit 18S ribosomal RNA was determined by nucleotide sequence analysis of the RNA directly. The rabbit rRNA was specifically cleaved with T_1 ribonuclease, as well as with E. coli RNase H using a Pst 1 DNA linker to generate a specific set of overlapping fragments spanning the entire length of the molecule. Both intact and fragmented 18S rRNA were end-labeled with [^{32}P], base-specifically cleaved enzymatically and chemically and nucleotide sequences determined from long polyacrylamide sequencing gels run in formamide. This approach permitted the detection of both cistron heterogeneities and modified bases. Specific nucleotide sequences within E. coli 16S rRNA previously implicated in polyribosome function, tRNA binding, and subunit association are also conserved within the rabbit 18S rRNA. This conservation suggests the likelihood that these regions have similar functions within the eukaryotic 40S subunit.

INTRODUCTION

Ribosomal RNA is becoming increasingly important in our current understanding of the mechanism of action of ribosomes. Numerous studies have provided evidence in prokaryotes for a direct participation of 16S rRNA in translation. For example, base pairing interactions between the 3'-terminus of 16S rRNA and mRNA 5'-noncoding regions are important in discriminating mRNA initiation sites during translation (1,2). Direct 16S rRNA:RNA interactions are also implicated in both tRNA binding (3,4) and ribosomal subunit association (5,6). Similar interactions are likely to exist in eukaryotes. An intimate association between mRNA and 18S rRNA can be detected by psoralen cross-linking (7). Furthermore, E. coli N-acetyl-tRNA^{Val} can be cross-linked to yeast 18S rRNA when occupying the P-site of 40S subunits (8). Specific interactions between 18S rRNA and the α and γ subunits of eukaryotic initiation factor 2(eIF-2), as well as the 66,000 dalton subunit of eIF-3 also have been determined using the heterobifunctional protein:RNA cross-linker diepoxybutane (9,10). Such studies lend strong support for a direct involvement of the 18S ribosomal RNA during translation.

To understand the functional role of eukaryotic 18S rRNA within the 40S subunit, its primary structure must first be determined. Recently, the complete nucleotide sequences for yeast (11) frog (12), and rat (13,14) 18S rRNA have been deduced by DNA sequence analysis of the cloned ribosomal genes. These studies have provided the basis for a phylogenetic comparison between 18S rRNA and 16S rRNA (14-16). Even though secondary structure models have been proposed for 18S rRNA using such comparisons (14-16), analysis of the RNA directly will ultimately be required to refine current structure predictions. To this end, we have determined the primary structure of rabbit reticulocyte 18S rRNA by direct RNA sequence analysis. This approach has permitted detection of both cistron heterogeneities and modified bases. In addition, the use of both T_1 ribonuclease and E. coli RNase H to generate specific overlapping fragments for 18S rRNA, will permit direct structure analysis of the entire molecule by chemical modification (17).

METHODS

Fragmentation of Rabbit 18S rRNA with Ribonuclease T_1

Purification of 18S rRNA from rabbit reticulocytes was performed as previously described (18,19). Fragments of 18S ribosomal RNA were prepared by digestion with 2.5×10^{-5} units of T_1 RNase/μg RNA in structure buffer (20 mM Tris-HCl, pH 7.5, 300 mM KCl, 20 mM $MgCl_2$) for 15 min at 4°C. The reaction was terminated by incubating for 25 min with 300 μg/ml of autodigested proteinase K at 25°C. Following phenol extraction and ethanol precipitation, the T_1 RNase generated fragments were 5'-[^{32}P]-end-labeled with [γ-^{32}P]-ATP and T_4 polynucleotide kinase and fractionated on a 80 cm long 3.5% polyacrylamide gel run in 7M urea. The 5'-[^{32}P]-end-labeled fragments #1,#2 and #3 were excised from the gel (Fig. 2A) and sequenced enzymatically (18,21) and chemically (22) as previously detailed (18,21).

Fragmentation of Rabbit 18S rRNA with E. coli Ribonuclease H

Purified 18S rRNA was first preblocked at its 3'-terminus with nonradioactive pCp to enrich for only [^{32}P]-end-labeled RNase H digestion fragments. To 2.5 A_{260} units (100 pmol) of 18S rRNA in ligase buffer (50 mM Hepes-KOH, pH 8.3, 10 mM $MgCl_2$, 3.3 mM DTT, 10 μM ATP, 10% DMSO, 15% glycerol) a ten-fold molar excess of nonradioactive pCp was added with five units of T_4 RNA ligase, and incubated for 48 hr at 10°C (23). After incubation the reaction mixture was made 10 mM NH_4HCO_3, pH 8.0 and the nonradioactive pCp removed by Sephadex G-75 spin column chromatography (24). Following ethanol precipitation, 0.4 A_{260} units (16 pmol) of preblocked 18S rRNA was dissolved

in 40 µl of 1 mM EDTA to which 100 - fold molar excess of Pst 1 linker DNA (Collaborative Research, Inc.) was added. The mixture was heated to 90°C for 1 min, quick-cooled to 0°C, and made 50 mm Tris-HCl, pH 7.9, 4 mM $MgCl_2$, and 1 mM DTT. Following hybridization for 3 min at 50°C and for 30 min at 32°C, 21 units of E. coli RNase H was added and the reaction mixture was incubated at 32°C for 30 min. The RNase H generated fragments were precipitated with ethanol, labeled at their 3'-termini with [^{32}P]Cp and T_4 RNA ligase, and fractionated on an 80cm long 3.5% polyacrylamide gel run in 7M urea. The 3'-[^{32}P]-end-labeled fragments A,B,C and D (Fig. 2B) were excised from the gel and their nucleotide sequences determined enzymatically and chemically (18,22).

RESULTS

The primary structure of rabbit 18S ribosomal RNA was determined by nucleotide sequence analysis of the RNA directly. Sequence analysis of the entire rRNA molecule was accomplished by in vitro [^{32}P]-end-labeling of both intact and fragmented 18S rRNA, followed by base-specific cleavage using enzymes and chemicals (18,22). Initial sequence data was obtained with intact 18S rRNA which was labeled in vitro at either its 5'-terminus with [γ-^{32}P]-ATP and T_4 polynucleotide kinase or its 3'-end with [^{32}P]Cp and T_4 RNA ligase (23). After base-specific enzymatic and chemical digestion (18-22), the resulting digests were electrophoretically fractionated by size in adjacent lanes of 140 cm long polyacrylamide sequencing gels run in 90% formamide (18). The use of long formamide sequencing gels allows determination of nucleotide sequences of 300-400 residues from a [^{32}P]-end-labeled terminus (18). The nucleotide sequences of the 5' proximal 400 residues and the 3'

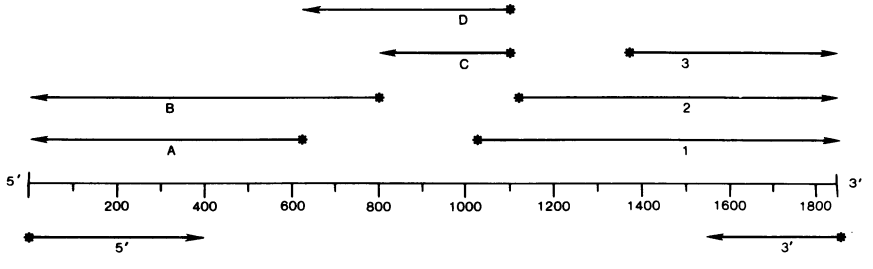


FIGURE 1. Map of the (*) [^{32}P]-end-labeled fragments used for nucleotide sequence analysis of rabbit 18S ribosomal RNA.

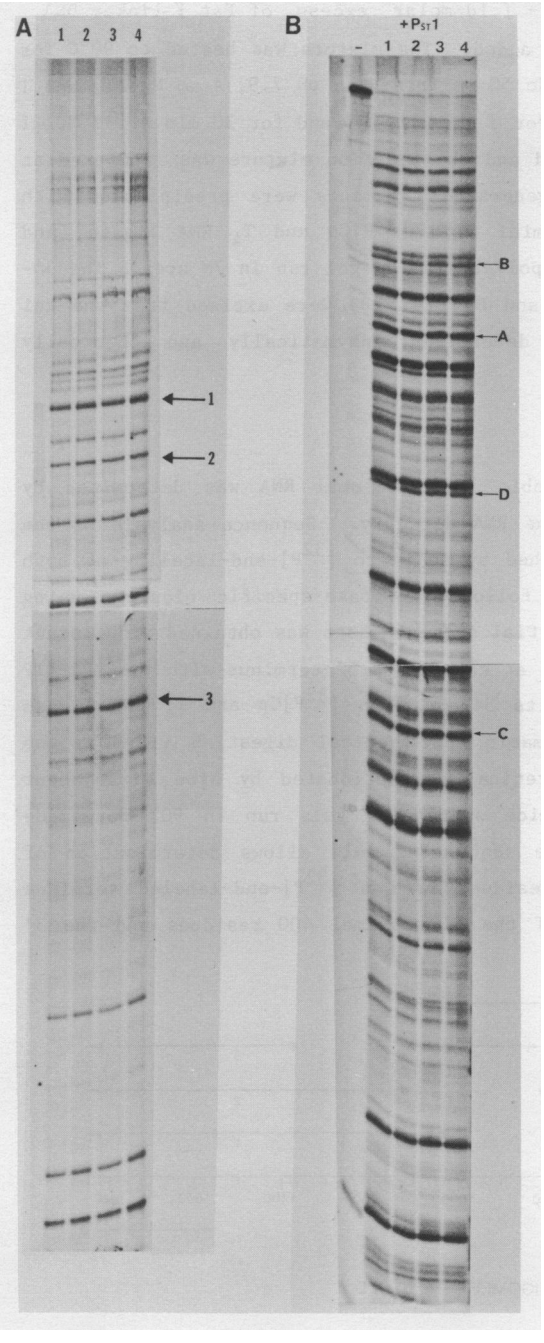


FIGURE 2A. Autoradiogram of the 5'-[^{32}P]-end-labeled T_1 RNase digestion fragments fractionated on a 3.5% polyacrylamide gel in 7M urea (80 cm long x 20 cm wide x 0.15 cm thick). 2B. Autoradiogram of the 3'-[^{32}P]-end labeled RNase H digestion fragments fractionated on a 3.5% polyacrylamide gel in 7M urea (80 cm long x 20 cm wide x 0.15 cm thick). (-) minus Pst 1 DNA linker, (+) plus Pst 1 DNA linker.

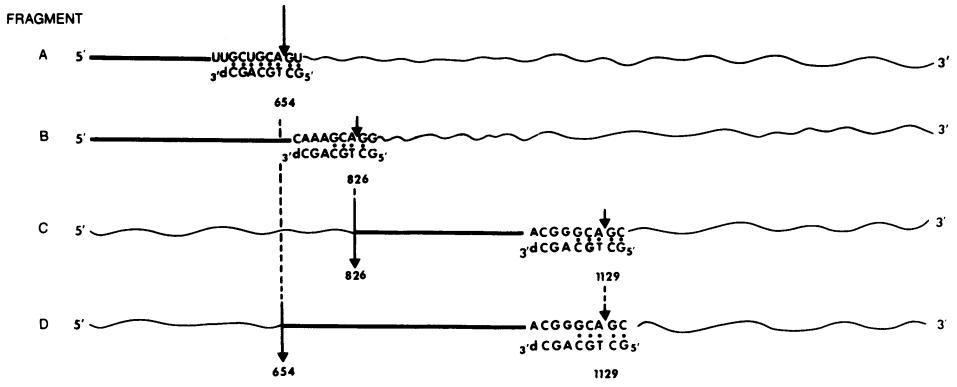
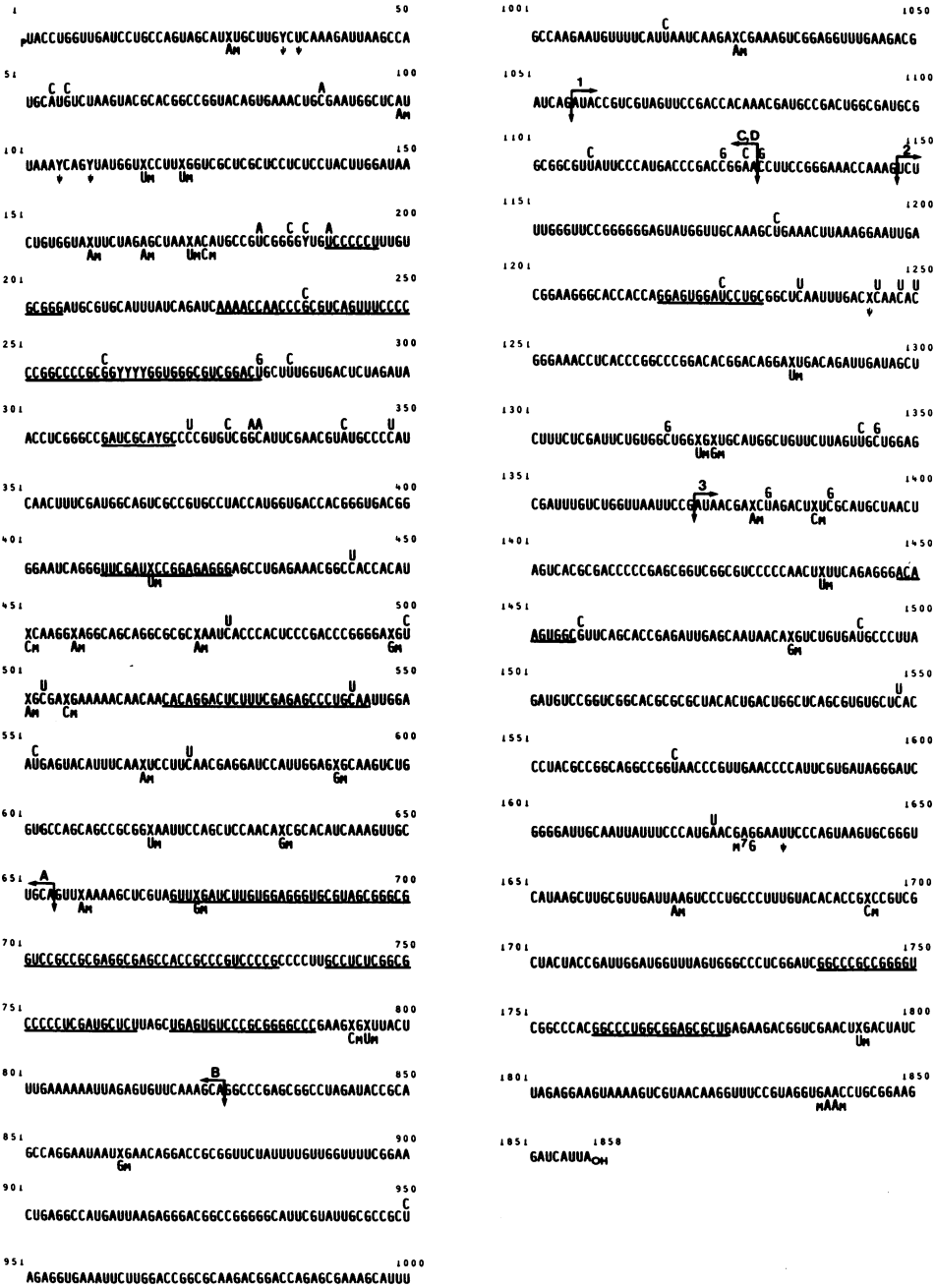


FIGURE 3. Digestion of rabbit 18S ribosomal RNA with RNase H using the Pst 1 DNA oligomer.

distal 300 nucleotides were determined in this manner (Fig. 1) as previously reported (18).

To extend the sequence analysis to internal regions of the RNA, large fragments were generated by partial digestion with T_1 RNase in high salt buffer. These fragments were labeled at their 5'-termini with [γ-^{32}P]-ATP and T_4 polynucleotide kinase and fractionated on a 80 cm long preparative 3.5% polyacrylamide gel (Fig. 2A). Partial digestion with T_1 RNase reproducibly generates several major fragments. Fragments #1, #2 and #3 (Fig. 2A) were overlapping and of sufficient purity to definitively extend the sequence of the 3' domain by an additional 504 nucleotides (Fig. 1).

Since the region encompassing nucleotides 400-1020 of the rabbit 18S rRNA molecule was relatively resistant to cleavage with T_1 RNase, a different approach was used to generate unique fragments in this region. Analysis of this area in yeast (12) and Xenopus laevis (13) 18S rDNA revealed several conserved restriction sites. DNA oligomers (commercially available linkers) corresponding to these restriction sites were used together with E. coli RNase H to generate site-specific cleavage of 18S rRNA. Hae III, Sma 1 and Pst 1 DNA oligomers were separately examined. A specific DNA oligomer was first heat denatured and hybridized to the rRNA. Site-directed cleavage of the 18S rRNA was then performed with E. coli RNase H (24). Only the Pst 1 DNA oligomer produced fragments with a unique 3'-terminus, as determined by end-group analysis of the resulting 3'-[^{32}P]-end labeled digestion fragments (21). Fragment A (Fig. 2B) resulted from cleavage of the rRNA at the region corresponding to the actual Pst 1 restriction site (Fig. 3). Fragments B, C and D



(Fig. 2B), however, are unique and were generated by partial hybridization of the Pst 1 DNA oligomer to the RNA (Fig. 3). These four fragments were of high enough specific activity and purity to complete the nucleotide sequence of the 18S rRNA (Fig. 1).

Fig. 4 presents the primary structure of rabbit 18S ribosomal RNA. The entire sequence contains 1,858 nucleotides. Nucleotides designated as X depict phosphodiester linkages resistant to both enzymatic and chemical cleavage and likely represent nucleotides with a 2'-O-methylated ribose. Y denotes a possible modified pyrimidine due to relatively weak enzymatic and chemical cleavages. In most instances, either 2'-O-methylated nucleotides or pseudouridines have been identified at both the X and Y positions, respectively. These were determined from previously catalogued T_1 RNase digestion fragments in Novikoff hepatoma (25), HeLa cell (26), and Xenopus laevis (27) 18S rRNA. Forty-two cistron heterogeneities have also been mapped as indicated by simultaneous cleavages with different base-specific enzymes and chemicals (Fig. 4). Most of the heterogeneities are clustered in specific areas of the molecule and likely reflect transcription from all or a portion of the two-hundred ribosomal genes present within the mammalian genome (28). During the sequence analysis numerous tracts within both the intact and fragmented rRNA molecule proved to be relatively resistant to cleavage with the base-specific enzymes, and could only be sequenced by chemical means (22). These regions underlined in Fig. 4 are G+C rich and likely reveal areas within mammalian 18S rRNA that are extensively base paired.

DISCUSSION

A phylogenetic alignment of the rabbit 18S rRNA sequence with corresponding rRNA sequences from yeast (11), Xenopus laevis (12), and rat (13) is shown in Fig. 5. The sequences have been arranged to give maximum homology and nucleotides conserved in 4 out of 4 species are indicated by an "'" subscript. A comparison of the rabbit sequence with those of yeast, frog, and rat shows a 67%, 91%, and 95% homology, respectively. The most striking differences occur within the region extending from #244-290 of the rabbit sequence (Fig. 5). Examination of this highly mutable region reveals numerous

FIGURE 4. Primary structure of rabbit 18S ribosomal RNA. X denotes a phosphodiester linkage resistant to chemical and enzymatic cleavage, and Y denotes a possible modified pyrimidine. Known modified bases at these positions within other 18S rRNAs are indicated parenthetically below. Two nucleotides within the same position indicate cistron heterogeneities. Regions underlined represent nucleotides resistant only to base-specific enzymatic cleavage and which are likely base paired.

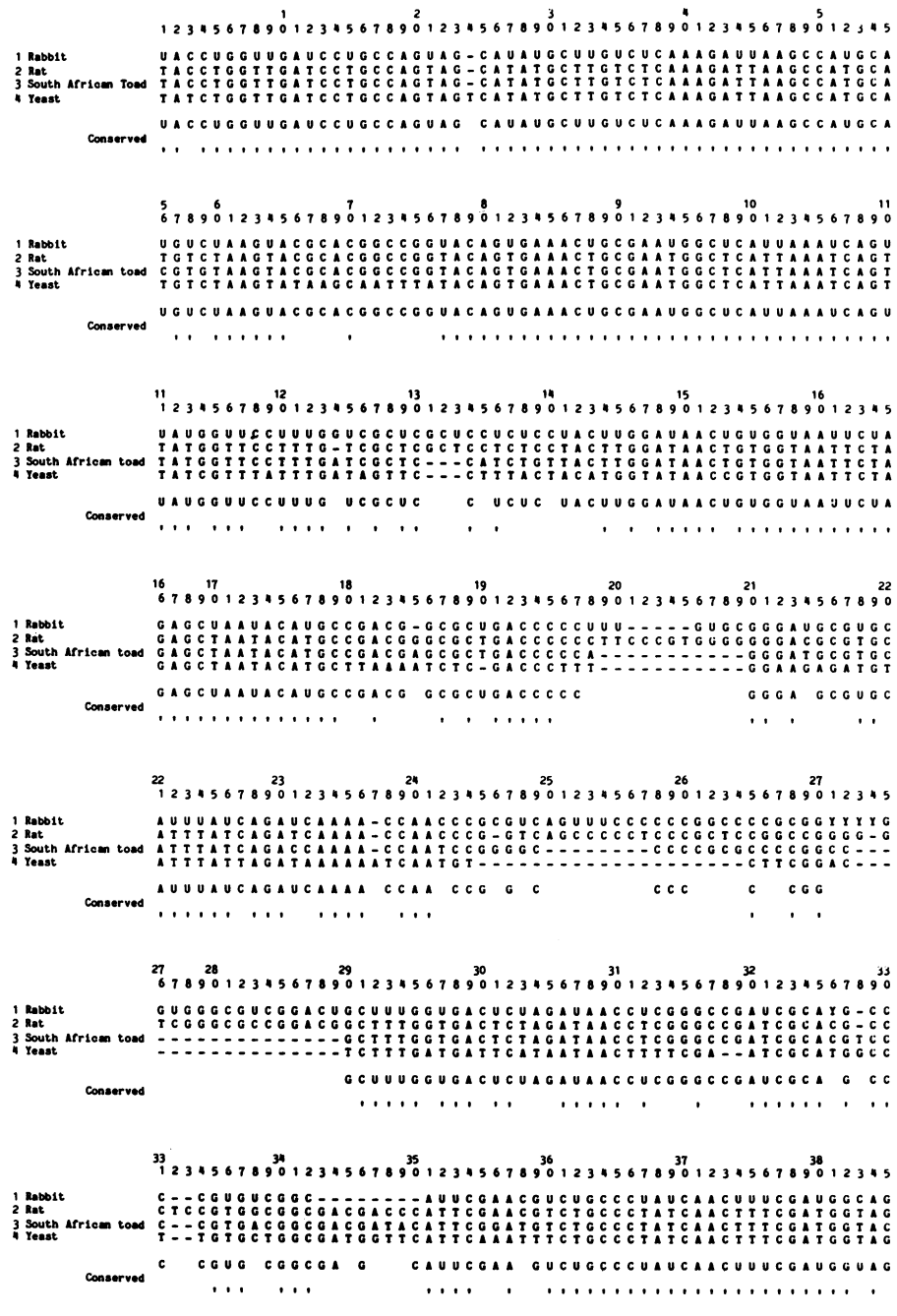
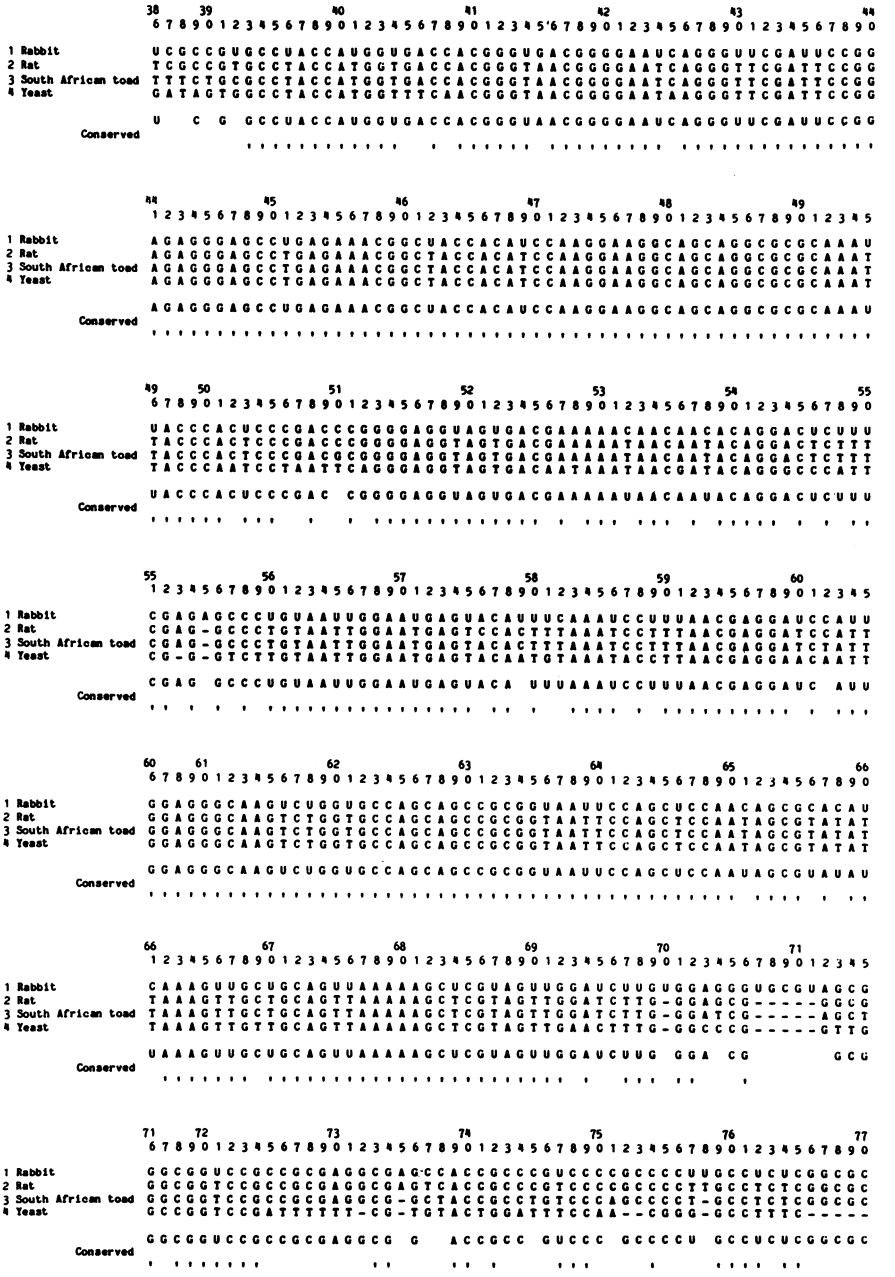
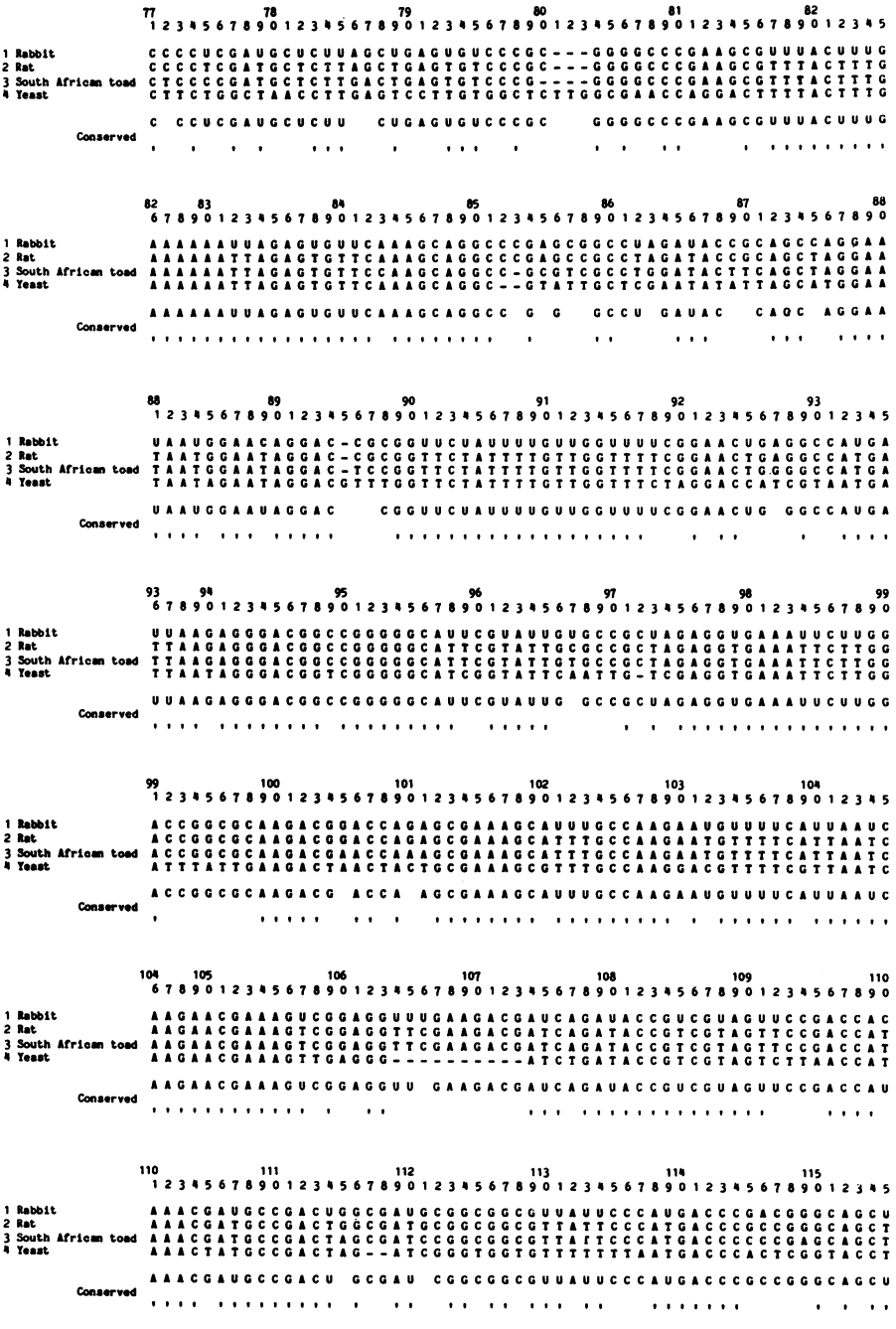
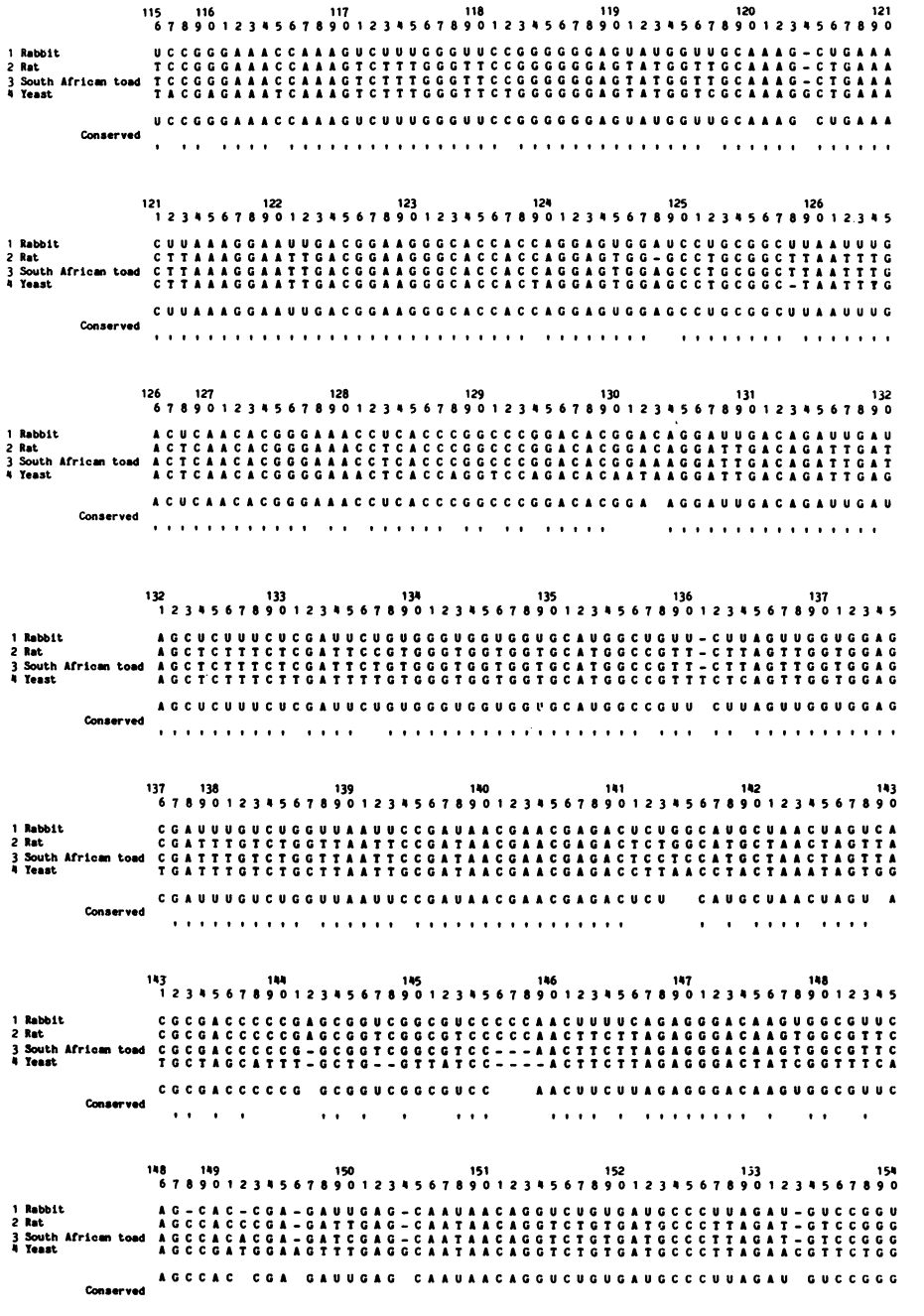
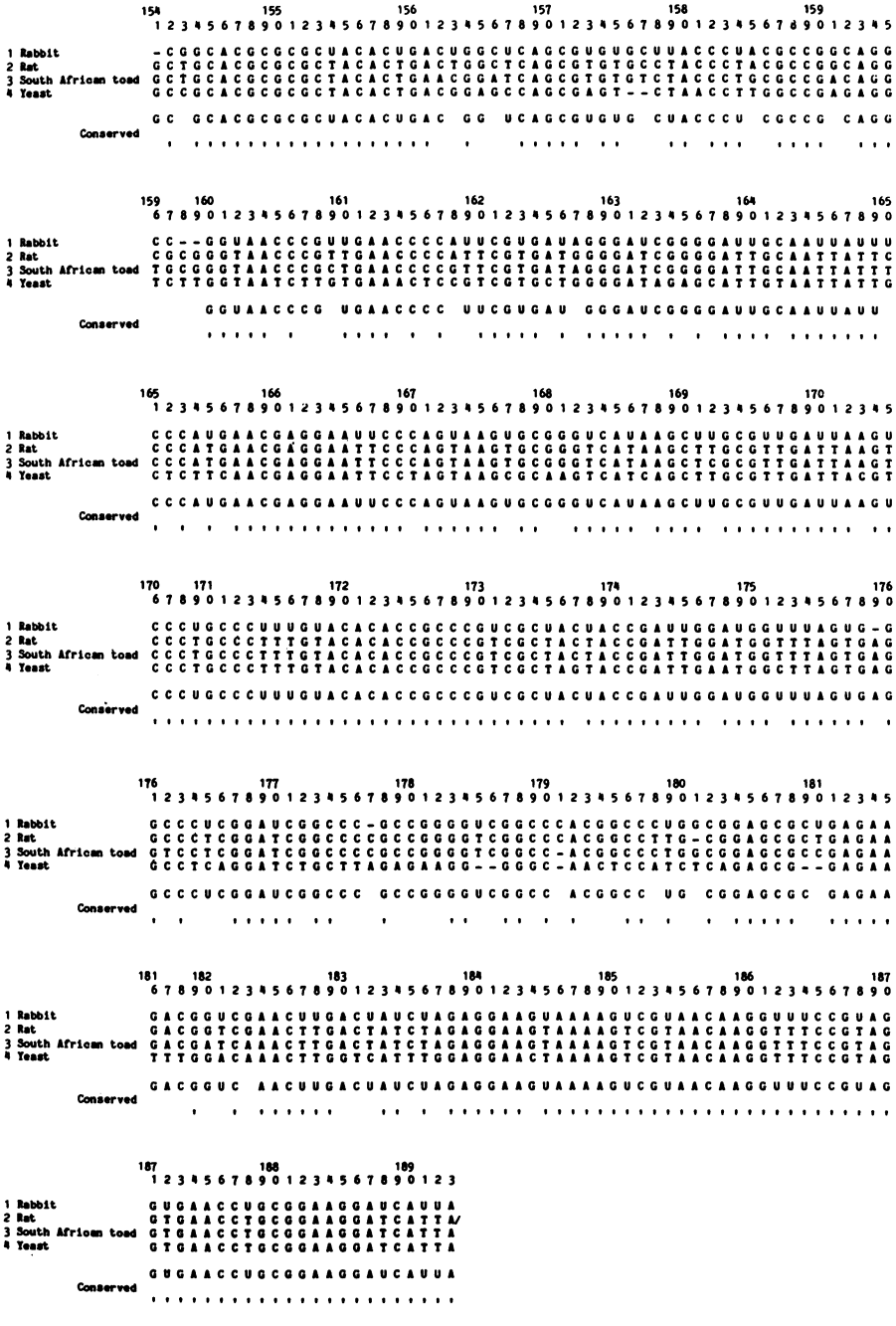


FIGURE 5.


1 Rabbit
2 Rat
3 South African toad
4 Yeast
Conserved






1 Rabbit
2 Rat
3 South African toad
4 Yeast
Conserved

G+C rich insertions into both the rabbit and rat sequence. These insertions may reside within a helix proposed by Brimacombe and co-workers between residues #222-245 of the yeast sequence (15), and may serve to extend and stabilize this duplex in higher eukaryotes.

Recently, it has been suggested that sequence data derived entirely from analysis of 18S rRNA directly is not as reliable as that obtained from the corresponding cloned rDNA (14). However, the appropriate use of both enzymes and chemicals to induce base-specific cleavage, in conjunction with sequence analysis of 5'- and 3'-^{32}P-end-labeled overlapping RNA fragments on highly denaturing formamide sequencing gels is general and can be used for sequence determination of other comparably large RNA molecules (18-22). Even though the 140 cm long formamide sequencing gels have the ability to resolve at least 150 nucleotides (18), the sequences were confirmed by expanding and resolving individual regions of both the intact and the fragmented rRNA on multiple polyacrylamide sequencing gels (10,15, and 20%) of various lengths run in both formamide and 8M urea. Nucleotide sequences were additionally verified by extending the sequence analysis of all RNase H and T_1 fragments within their overlapping regions. For example, 3'-^{32}P-end-labeled fragment A (#654) could be read to nucleotide #280, thereby overlapping and confirming a substantial portion of the sequence determined from intact 5'-^{32}P-end-labeled rRNA. Likewise, 3'-^{32}P-end-labeled fragment D(#1129) could be read to nucleotide (#809) verifying not only the sequences within fragment C and the beginning of B, but also, most of the sequence determined from the 5'-^{32}P-end-labeled T_1 RNase fragment 1. It is noteworthy that the nucleotide sequence for rat 18S rRNA determined in two different laboratories by sequence analysis of the ribosomal genes shows twenty-five base differences more than half of which are either insertions or deletions (13,14).

The extensive sequence homology among the eukaryotic 18S rRNAs clearly suggests a conservation of structure and function within the 40S subunit. However, unlike in prokaryotes there is little information regarding the function of specific nucleotide sequences. Nevertheless, many functionally important regions which have been identified in E. coli 16S rRNA are conserved in 18S rRNA. For example using kethoxal modification, specific guanine residues in 16S rRNA (G_{530}, G_{693}, G_{966}, G_{1388} and G_{1517}) implicated in polyribosome function (29) are also present within nucleotide sequences conserved

FIGURE 5. Comparison of the rabbit 18S rRNA nucleotide sequence with yeast, frog and rat 18S rRNA.

in rabbit 18S rRNA (C_{605}-C_{622}, U_{950} - U_{961}, U_{1228} - A_{1247}, G_{1671} - C_{1699}, G_{1838} - G_{1847} respectively; Fig. 4). Likewise, guanine residues (G_{703}, G_{791}, G_{803}, G_{1497}, G_{1505} and G_{1517}) implicated in the association of the 30S and 50S subunits (5,6) are present within sequences conserved in rabbit 18S rRNA (G_{972} - A_{979}, G_{1050} - C_{1060}, G_{1064} - C_{1074}, A_{1813} - A_{1821}, A_{1823} - U_{1827}, G_{1838} - G1847, respectively). Furthermore, the specific nucleotide in 16S rRNA, C_{1400}, which can be cross-linked to E. coli N-Acetyl-$tRNA^{Val}$ and N-Acetyl-$tRNA^{Ser}$ when occupying the ribosomal P site (3,4,8), lies within a sequence also present in rabbit 18S rRNA (C_{1679} - C_{1699}). The remarkable conservation of these functionally important regions of E. coli 16S rRNA in rabbit 18S rRNA, strongly suggests a similar role within the eukaryotic 40S subunit.

There is some evidence for conservation of structure, as well. The rabbit nucleotide sequences surrounding the guanine residues G_{1055} and G_{1371} (Fig. 4) cleaved by T_1 ribonuclease and generating fragments #1 and #3, respectively, are both conserved in other eukaryotic 18S rRNAs (Fig. 5), as well as, within E. coli 16S rRNA (30). These same G residues are reactive to kethoxal modification in E. coli 30S subunits and likely reside within single-stranded regions on the surface of the molecule (30). The conservation of both these sequences within rabbit 18S rRNA (G_{1050} - C_{1060} and G_{1361} - G_{1371}, respectively; Fig. 4) and their extreme accessibility to T_1 RNase, suggests a similar structure within the 40S subunit (14,15).

Finally, the methods described here, for generating specific overlapping fragments spanning the entire 18S rRNA molecule, will now permit direct secondary structure analysis using structure-specific chemical probes. Furthermore, mapping the topography of 18S rRNA within both the free 40S subunit, as well as, within a functionally engaged ribosome should reveal the regions likely involved in both subunit association and those interacting with mRNA, tRNA, and initiation factors. Such studies are currently in progress.

ACKNOWLEDGEMENTS

We wish to thank Dr. Robert Crouch (NIH) for the kind gift of E. coli RNase H. We also are indebted to Dr. D.G. George (National Biomedical Research Foundation, Georgetown University Medical Center) for his assistance in determining nucleotide sequence homologies by computer. We also wish to thank Mr. Madhaven Kutty for technical assistance, and Ms. Rachel Quynn and Jean Brereton for preparation of the manuscript. For Mr. John F. Connaughton, this work was in partial fulfillment of the Ph.D. requirements at The George Washington University. This work was supported by NIH grant GM 28985 to Raymond E. Lockard and NSF (PCM 820176) to Ajit Kumar.

REFERENCES
1. Shine, J. and Dalgarno, L. (1974). Proc. Nat'l. Acad. Sci. U.S.A. 71, 1342-1347.
2. Steitz, J.A. and Jakes, K. (1975). Proc. Nat'l. Acad. Sci. U.S.A. 72, 4734-4739.
3. Zimmermann, R.A., Gates, S.M., Schwartz, I., and Ofengand, J. (1979). Biochemistry 18, 4333-4336.
4. Prince, J.B., Taylor, B.H., Thurlow, D.L., Ofengand, J., and Zimmermann, R.A. (1982). Proc. Natl. Acad. Sci. U.S.A. 79, 5450-5454.
5. Herr, W., Chapman, N.M., and Noller, F.H. (1979). J. Mol. Biol. 130, 433-449.
6. Herr, W., Chapman, N.M., and Noller, H.F. (1979). J. Mol. Biol. 130, 421-432.
7. Nakashima, K., Darzynkiewicz, E., and Shatkin A.J. (1980). Nature 286, 226-230.
8. Ofengand, J., Gornicki, P., Chakraburtty, K., and Nurse, K. (1982). Proc Nat'l. Acad. Sci. U.S.A. 79, 2817-2882.
9. Westermann, P., Nygard, O., and Bielka, H. (1981). Nucleic Acids Res. 8, 3065-3071.
10. Nygard, O. and Westermann, P. (1982). Nucleic Acids Res. 10, 1327-1334.
11. Rubstov., P.M., Musakhanov, M.M., Zakharyev, V.M., Krayev, A.S., Skryabin, K.G., and Bayev, A.A. (1980). Nucleic Acids Res. 8, 5779-5794.
12. Salim, N. and Maden, B.E.H. (1981). Nature 291, 205-208.
13. Torczynski, R., Bollon, A.P., and Fuke, M. (1983). Nucleic Acids Res. 11, 4879-4890.
14. Chan, Y-L., Gutell, R., Noller, H.F., and Wool, I.G. (1984). J. Biol. Chem. 259, 224-230.
15. Stiegler, P., Carbon, P. Ebel, J.P., and Ehresmann, C. (1981). E.J. Biochem. 120, 487-495.
16. Zwieb, C., Glotz, C., and Brimacombe, R. (1981). Nucleic Acids Res. 9, 3621-3640.
17. Peattie, D.A. and Gilbert, W. (1980). Proc. Nat'l. Acad. Sci. U.S.A. 77, 4679-4683.
18. Lockard, R.E., Connaughton, J.F., and Kumar, A. (1982). Nucleic Acids Res. 10, 3445-3457.
19. Connaughton, J.F., Kumar, A., and Lockard, R.E. (1983). In Eukaryotic Gene Expression, Vol II, Edit. by A. Kumar, Plenum Press, 203-221.
20. Donis-Keller, H., Maxam, A.M., and Gilbert, W. (1977). Nucleic Acids Res. 4, 2527-2538.
21. Lockard, R.E., Alzner-DeWeerd, B., Heckman, J.E., MacGee, J., Tabor, M.W., and RajBhandary, U.L. (1978). Nucleic Acids Res. 5, 37-56.
22. Peattie, D.A. (1980). Proc. Nat'l. Acad. Sci., U.S.A. 76, 1760-1764.
23. England, T.E. and Uhlenbeck, O.C. (1978). Nature 275, 560-561.
24. Donis-Keller, H. (1979). Nucleic Acids Res. 7, 179-192.
25. Choi, Y.C. and Busch, H. (1978). Biochemistry 17, 2551-2560.
26. Maden, B.E.H. and Salim, M. (1974). J. Mol. Biol. 88, 133-164.
27. Maden, B.E.H. and Kahn, M.S.N. (1977). Biochem J. 167-211-221.
28. Long, E. and Dawid, I.B. (1980). Ann. Rev. Biochem. 49, 727-764.
29. Brow, D.A. and Noller, H.F. (1983). J. Mol. Biol. 163, 27-46.
30. Noller, H.F. and Woese, C.R. (1981). Science 212, 403-411.