Volume 11 Number 22 1983 Nucleic Acids Research

# The structure of rat 28S ribosomal ribonucleic acid inferred from the sequence of nucleotides in a gene

Yuen-Ling Chan, Jose Olvera and Ira G.Wool

Department of Biochemistry, University of Chicago, Chicago, IL 60637, USA

Received 8 August 1983; Revised and Accepted 11 October 1983

## ABSTRACT

The nucleotide sequence of a rat 28S rRNA gene was determined. The 28S rRNA encoded in the gene contains 4718 nucleotides and the molecular weight estimated from the sequence is $1.53 \times 10^6$. The guanine and cytosine content is 67%. The sequence of rat 28S rRNA diverges appreciably from that of Saccharomyces carlsbergensis 26S rRNA (about 50% identity), but more closely approximates that of Xenopus laevis 28S rRNA (about 75% identity). Rat 28S rRNA is larger than the analogous nucleic acids from yeast (3393 nucleotides) and X. laevis (4110 nucleotides) ribosomes. The additional bases are inserted in specific regions and tend to be rich in guanine and cytosine. 5.8S rRNA can interact with 28S rRNA by extensive hydrogen bonding at two sites near the 5' end of the latter.

## INTRODUCTION

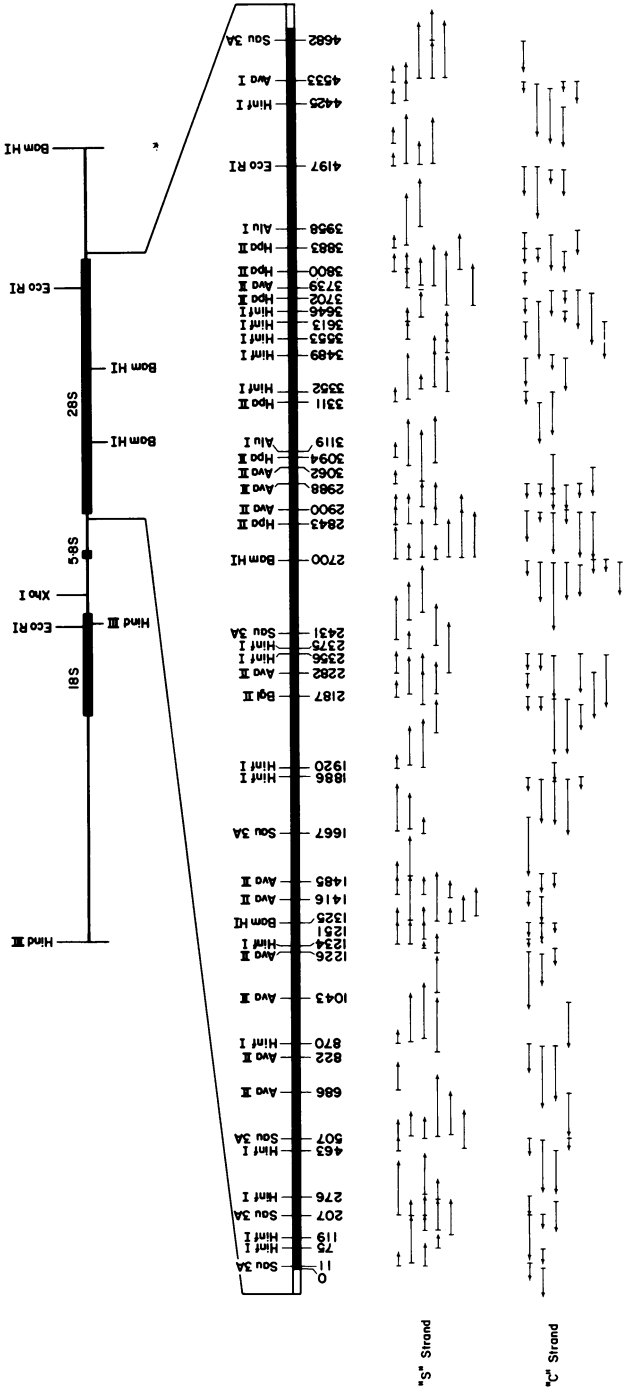
Solution of the structure of ribosomes requires information on the chemistry of the constituents. Eukaryotic ribosomes contain four molecules of RNA: 18S rRNA in the 40S subparticle, and 5S, 5.8S, and 28S RNAs in the 60S subparticle. The covalent structure of 5S (1), 5.8S (2,3), and 18S (4) RNAs from rat liver ribosomes have been determined. We have now established the structure of rat 28S rRNA from the sequence of a gene. For that purpose, we cloned a rat 28S rRNA gene carried on a phage (Charon 4A) in plasmid vectors and determined the sequence of the DNA.

## EXPERIMENTAL PROCEDURES

DNA from the λ phage Charon 4A in which rat ribosomal RNA genes had been cloned (5) was kindly provided by T.D. Sargent and J. Bonner of California Institute of Technology. The DNA from the recombinant plasmid p2.7 was provided by L.I. Rothblum of Baylor College of Medicine. The procedures used for the preparation of restriction endonuclease fragments and for the determination of the sequence of nucleotides in DNA have been described or cited before (4).

## RESULTS AND DISCUSSION

The Determination of the Sequence of 28S rDNA. A restriction map of the region of the Charon 4A recombinant λ phage containing the rDNA genes had been constructed (4). In


BamHI
EcoRI
BamHI
28S
BamHI
5.8S
XhoI
EcoRI
HindIII
18S
HindIII
Sau 3A
AvaI
HinfI
EcoRI
AluI
HpaII
HpaII
HpaII
AvaI
HinfI
HinfI
HinfI
HinfI
HinfI
HpaII
HinfI
AluI
HpaII
HpaII
AvaI
AvaI
BamHI
HpaII
AvaII
HinfI
HinfI
Sau 3A
AvaII
BglII
AvaII
HinfI
HinfI
Sau 3A
AvaII
AvaII
BamHI
AvaII
HinfI
AvaII
AvaII
HinfI
Sau 3A
HinfI
HinfI
HinfI
Sau 3A
HaeIII
AvaII
Sau 3A
4682
4533
4425
4197
3958
3883
3800
3739
3702
3646
3613
3553
3489
3362
3311
3119
3094
3062
2988
2900
2843
2700
2431
2375
2356
2282
2187
1920
1886
1667
1485
1416
1325
1251
1234
1226
1043
870
822
686
507
463
276
207
119
75
11
0
"S" Strand
"c" Strand

order to amplify the 28S rDNA and to facilitate purification, the gene was subcloned in pACYC184 (6) at the EcoRI sites to form pRRR228 (4). The EcoRI fragment of 6.4 kb contains most of the 28S rRNA gene (it lacks only the 3' end) and two internal transcribed spacers. The 3' end of a 28S rRNA gene was contained in an EcoRI-BamHI fragment of 2.7 kb (7) which was subcloned in pBR322 to form the recombinant plasmid pBR 288 (originally designated p2.7) (Fig. 1).

A restriction map of pRRR228 revealed that a combination of the enzymes BamHI, EcoRI, and XhoI would release six fragments, three of which contained portions of the 28S rRNA gene. The three fragments had an approximate length in kbs of 2.9 (XhoI-BamHI), 1.4 (BamHI-BamHI), and 1.5 (BamHI-EcoRI) (Fig. 1). An analysis of the cleavage pattern generated from the 1.4 and 1.5 kb fragments with restriction endonucleases identified enzymes suitable for production of subfragments convenient for a determination of the nucleotide sequence and for formation of overlapping oligonucleotides (Fig. 1). The 2.9 kb XhoI-BamHI fragment which contains the 5' end of the 28S rRNA gene, the 5.8S rRNA gene, and most of the internal transcribed spacer was digested with Sau3A. Four of the subfragments were shown by Southern hybridization (8) to contain portions of the 28S rRNA gene. The 3' end of the 28S rRNA gene was isolated from pBR288 using BamHI, EcoRI, and AvaI; the presence of the 3' end of 28S rRNA in the restriction enzyme digests was confirmed by Southern hybridization (8) to the α-sarcin RNA fragment(9).

The sequence of the 28S rRNA gene was determined by the Maxam and Gilbert procedure (12, 13). The analysis included a determination of the sequence of almost all

Figure 1. Restriction endonuclease map of a rat 28S rRNA gene and a diagram of the overlapping fragments whose sequence of nucleotides was determined. The upper portion of the figure depicts the restriction endonuclease sites in a rat 28S rRNA gene that were used to generate fragments (lower portion of the figure) for the determination of the sequence of nucleotides. The numbers indicate the first nucleotide in the restriction endonuclease recognition sequence counting from the 5' end of the 28S rRNA gene. Each arrow designates a nucleotide sequence that was determined. The "S" strand is synonymous with the RNA, the "C" strand is complementary. Plasmid DNA was isolated by the clear lysate method (10, 11). The procedures used for the digestion of DNA with restriction enzymes, for the isolation of DNA fragments, and for the labeling of the fragments at the 5' end with [$\gamma$-$^{32}$P]ATP using T4 polynucleotide kinase were cited earlier (4). Double-stranded DNA fragments were separated by heating for 2 min at 90 °C in 50% DMSO, 50 mM Tris/HCl, pH 8.3, 50 mM borate, 2 mM EDTA. The presence of portions of the 28S rRNA gene in the fragments was established by Southern hybridization (8) to $^{32}$P-labeled 28S rRNA prior to undertaking the sequence determination. The sequence of nucleotides in the DNA fragments was determined by the Maxam and Gilbert method (12, 13) using thin (0.5 mm) gels of 8, 10, or 20% polyacrylamide (13, 14). There are several short regions of ambiguity: 669 to 675; 777 to 785; 1281 to 1285; 2003 to 2009; 2033 to 2040; and 2058 to 2064. In each case the number of nucleotides differed when read in the two directions due to compression in the sequencing ladder in one of the two gels. We report the sequence with the larger number of nucleotides

of both strands and extensive overlaps of portions of the sequences in separate fragments (Fig. 1). The order of the fragments was approximated from the restriction endonuclease map, and by reference to the sequence of the 5' (3) and 3' ends (15) of rat 28S rRNA.

The sequence of the nucleotides in 28S rRNA was inferred from the sequence of the gene (Fig. 2). Rat 28S rRNA has 4718 nucleotides. The molecular weight is $1.53 \times 10^6$ (based on the assumption of a molecular weight of 325 for a nucleotide) somewhat less than the $1.7 \times 10^6$ (5230 nucleotides) estimated from physicochemical measurements (16, 17). The guanine and cytosine content of rat 28S rRNA is 67%.

Ribosomal nucleic acids are methylated. We have not identified the nucleotides in rat 28S rRNA that are methylated. However, the pattern of methylated bases in oligonucleotide digests of 28S rRNAs from human (HeLa), hamster (BHK/C13), and mouse (L) cells could hardly be distinguished, i.e. the identity is 99% (18). Moreover, the methylated nucleotides in 28S rRNAs from human (HeLa cells) and from X. laevis have at least 95% identity (18). Twenty-nine of the 30 methylation sites in yeast 26S rRNA (19) are conserved in the rat nucleic acid. Thus, the methylated bases in rat 28S rRNA are likely to be the same as those in other large eukaryotic rRNAs.

Comparison of the Sequence of Rat 28S rRNA with other Large Eukaryotic rRNAs. The sequence of nucleotides in yeast 26S rRNA (Saccharomyces carlsbergensis (19) and Saccharomyces cerevisiae (20)), and in X. laevis 28S rRNA(21) have been determined and can be compared with the structure of rat 28S rRNA. Yeast 26S rRNA has 3393 nucleotides whereas rat 28S rRNA has 4718 (Table I). When the sequences of the rat and yeast rRNAs were aligned there were 1946 identities (the same nucleotide at the same position). Thus, the percentage identity is 41 for rat and 57 for yeast; the latter is higher because it has a smaller number of nucleotides (Table I). In general, regions of close identity are interrupted by domains that are quite different (Fig. 3A). We have sought to identify the variable regions as well as sites of insertion of the additional nucleotides (Table II). There are three regions of rat 28S rRNA where large insertions have occurred. They are between positions 430 and 1206 (570 nucleotides), 1977 and 2180 (118 nucleotides), and 2745 and 3328 (440 nucleotides) (Table II). These three inserts account for 1128 or 85% of the 1325 additional nucleotides in rat 28S rRNA. The next largest inserts are of 43 nucleotides in the region 1244 to 1409, and of 53 nucleotides at 3698 to 3753 in the rat sequence (Table II). The regions of variability and of insertions in rat 28S rRNA are rich in guanine and cytosine (Table II). Indeed, the insertions account, in large part, for the difference between the content of these two nucleotides in rat (67%) and yeast (48%; ref. 19).

There is considerably more identity in the structure of rat and X. laevis 28S rRNAs than with yeast 26S rRNA. Rat 28S rRNA has 4718 nucleotides, X. laevis 28S rRNA has 4110. When the sequences are aligned there are 3248 identities, or 69% for rat and 79%

| | | | | | |
|---|---|---|---|---|---|
| pCGCGACCUCA | GAUCAGACGU | GGCGACCCGC | UGAAUUUAAG | CAUAUUAGUC | 50 |
| AGCGGAGGAA | AAGAAACUAA | CCAGGAUUCC | CUCAGUAACG | GCGAGUGAAC | 100 |
| AGGGAAGAGC | CCAGCGCCGA | AUCCCCGCCG | CGCGCCGCGG | CGCGGGAAAU | 150 |
| GUGGCGUACG | GAAGACCCAC | UCCCCGGCGC | CGCUCGUGGG | GGGCCCAAGU | 200 |
| CCUUCUGAUC | GAGGCCCAGC | CCGUGGACGG | UGUGAGGCCG | GUAGCGGCCC | 250 |
| CGGCGCGCCG | GGCCGGGUCU | UCCCGGAGUC | GGGUUGCUUG | GGAAUGCAGC | 300 |
| CCAAAGCGGG | UGGUAAACUC | CAUCUAAGGC | UAAAUACCGG | CACGAGACCG | 350 |
| AUAGCCAACA | AGUACCGUAA | GGGAAAGUUG | AAAAGAACUU | UGAAGAGAGA | 400 |
| GUUCAAGAGG | GCGUGAAACC | GUUAAGAGGU | AAACGGGUGG | GGUCCGCGCA | 450 |
| GUCCGCCCGG | AGGAUUCAAC | CCGGCGGCGC | GCGCCGGCCG | GCCGGUGGUC | 500 |
| CCGGCGGAUC | UUUCCCGCUC | CCCGUUCCUC | CCGACCCCUC | CACCCGCGCG | 550 |
| UCUCUCUCCC | CCCUCCCCGC | GUCCCGCCGU | CGCCGUCCCC | GCUCCUCCCU | 600 |
| CCGGGGGGGU | GUCGGCGGGC | GCUCCGGCGG | CGGGCGCGGG | GUGUGGUGGG | 650 |
| GGCGCGCGGG | CGGGGCCGGG | GGUGGGGUCG | GCGGGGGACC | GCCCCCGGUC | 700 |
| GGCGACCGGC | CGCCGCCGGG | CGCACUUCCA | CCGUGGCGGU | GCGCCGCGAC | 750 |
| CGGCUCCGGG | ACGGCUGGGA | AGGCCCGGCG | GGGAAGGUGG | CUCGGGGGGG | 800 |
| GCGGCGUCAC | CCGUGGGCGC | CGGACCACCC | CGCCCCGAGU | GUUACAGCCC | 850 |
| CCCGGCAGCA | GCGCUCGCCG | AAUCCCGGGG | CCGAGGGAGC | CGGAUACCCG | 900 |
| UCGCCGCGCU | CUCCCCCCGG | CCUCUCCCCU | CCCGCCCCUC | CCCGUGGGGU | 950 |
| GGCGGAAAGG | GGGGCGGUCG | CGGGGGCCGG | GCCGCCCCUC | CCACGGCGCG | 1000 |
| ACCGCUCUCC | CACCCCCCGU | CGCCUCCGUC | GUCCCUCUCG | GGGGUCCGGG | 1050 |
| GGCCCGGGGG | GCGGGGCGGA | CUGUCCCCAG | UGCGCCCCGG | GCGUCGUCGC | 1100 |
| GCCGUCGGGC | CCGGGGGGGC | CGUCGUCACG | CGCUCUCCCU | CCCCUUCUCG | 1150 |
| GGGUGGGGGG | GAGCGAAGCC | GAGCGCACGG | GGUCGGCGGC | GAUGUCGGCU | 1200 |
| ACCCACCCGA | CCCGUCUUGA | AACACGGACC | AAGGAGUCUA | ACGCGUGCGC | 1250 |
| GAGUCAGGGG | CUCGUCCGAA | AGCCGCCGUG | GCGCAAUGAA | GGUGAAGGGC | 1300 |
| CCCGUUCCCG | GGGGCCCCGA | GGUGGGAUCC | CGAGGCCUCU | CCAGUCCGCC | 1350 |
| GAGGGCGCAC | CACCGGCCCG | UCUCGCCCGC | CGCGCCGGGG | AGGUGGAGCA | 1400 |
| CGAGCGUACG | CGUUAGGACC | CGAAAGAUGG | UGAACUAUGC | UUGGGCAGGG | 1450 |
| CGAAGCAGAG | GAAACUCUGG | UGGAGGUCCG | UAGCGGUCCU | GACGUGCAAA | 1500 |
| UCGGUCGUCC | GACUUGGGUA | UAGGGGCGAA | AGACUAAUCG | AACCAUCUAG | 1550 |
| UAGCUGGUUC | CCUCCGAAGU | UUCCCUCAGG | AUAGCUGGCG | CUCUCGCAAC | 1600 |

| | | | | | |
|---|---|---|---|---|---|
| GCGUUCGCUC | GACAACCCGC | AGUUUUAUCC | GGUAAAGCGA | AUGAUUAGAG | 1650 |
| GUCUUGGGGC | CGAAACGAUC | UCAACCUAUU | CUCAAACUUU | AAAUGGGUAA | 1700 |
| GAAGCCCGGC | UCGCUGGCGU | GGAGCCGGGC | GUGGAUGCGA | GUGCCUAGUG | 1750 |
| GGCCACUUUU | GGUAAGCAGA | ACUGGCGCUG | CGGGAUGAAC | CGAACGCCGG | 1800 |
| GUUAAGGCGC | CCGAUGCCGA | CGCUCAUCAG | ACCCCAGAAA | AGGUGUUGGU | 1850 |
| UGAUAUAGAC | AGCAGGACGG | UGGCCAUGGA | AGUCGGAAUC | CGCUAGGAGU | 1900 |
| GUGUAACAAC | UCACCUGCCG | AAUCAACUAG | CCCUGAAAAU | GGAUGGCGCU | 1950 |
| GGAGCGUCGG | GCCCAUACCC | GGCCGUCGCC | GGCAGUCGGA | ACGGGACGGG | 2000 |
| AGCGGCCGCG | GGCGCGCGAC | CCCGGGGGCC | GGGCGGCGUC | GGCUUCGGCC | 2050 |
| GGCCGCCGCC | CGUCCACCCC | CGGGGCUCCC | CCCGCGGCGU | CGGGCCCCGC | 2100 |
| GGAGCCUACG | CCGCGACGAG | UAGGAGGGCC | GCUGCGGUCA | GCCUUGAAGC | 2150 |
| CUAGGGCGCG | GGCCCGGGUG | GAGCCGCCGC | AGGUGCAGAU | CUUGGUGGUA | 2200 |
| GUAGCAAAUA | UUCAAACGAG | AACUUUGAAG | GCCGAAGUGG | AGAAGGGUUC | 2250 |
| CAUGUGAACA | GCAGUUGAAC | AUGGGUCAGU | CGGUCCUGAG | AGAUGGGCGA | 2300 |
| GUGCCGUUCC | GAAGGGACGG | GCGAUGGCCU | CCGUUGCCCU | CAGCCGAUCG | 2350 |
| AAAGGGAGUC | GGGUUCAGAU | CCCCGAAUCC | GGAGUGGCGG | AGAUGGGCGC | 2400 |
| CGCGAGGCGU | CCAGUGCCGG | UAACGCGACC | GAUCCCGGAG | AAGCCGGCGG | 2450 |
| GAGCCCGGGG | AGAGUUCUCU | UUUCUUUGUG | AAGGGCAGGG | CGCCUUGGAA | 2500 |
| UGGGUUCGCC | CCGAGAGAGG | GGCCCGUGCC | UUGGAAAGCG | UCGCAGUUCC | 2550 |
| GGCGGCGUCC | GGUGAGCUCU | CGCUGGCCCU | UGAAAAUCCG | GGGGAGAGGG | 2600 |
| UGUAAAUCUC | GCGCCGGGCC | GUACCCAUAU | CCGCAGCAGG | UCUCCAAGGU | 2650 |
| GAACGCCUCU | GGCAUGUUGG | AACAAUGUAG | GUAAGGGAAG | CGGCAAGCCG | 2700 |
| GAUCCGUAAC | UUCGGGAUAA | GGAUUGGCUC | UAAGGGCUGG | GUCGGUCGGG | 2750 |
| CUGGGGCGCG | AAGGCGGGGC | UGGGCGCGCG | CCGCGGCUGG | ACGAGGCGCC | 2800 |
| GCCGCCCCCC | UCCCACGUCC | GGGGAGACCC | CCUCCUUUCC | GCCCGGGCCC | 2850 |
| GCCCUCCCCU | CUCCCCGCGG | GGCCCCGCCG | UCCCCCGCGU | CGUCGCCGUG | 2900 |
| GUCCCCUCCU | CCCUCCCUUC | UUCCCCGUCC | GCGGGGGGGA | CGGGGCGGGU | 2950 |
| GCGGGGGGGC | GCGCGCGCGC | GCGGCCCAGG | GGCGGCGGGU | CCAACCCCGC | 3000 |
| GCGGGCCGGA | GCGGGGGGAA | CCCGCGGGCC | CCCGGUGGGG | GGGGGCCCGG | 3050 |
| ACACCCGGGG | GGGACCGGCG | GCGGCGGCGA | CUCUGGACGC | GAGCCGGGCC | 3100 |
| CUUCCCGUGG | AUCGCCCCAG | CUGCGGCGGG | CGUCGCGGCC | GCUCCCGGGG | 3150 |
| AGCCCGGCGG | GUCGCCCGGC | GGGGUUUUCC | UCCGGCCUCG | UCCUCCCCCU | 3200 |

| | | | | | |
|---|---|---|---|---|---|
| UCCCCCUCCG | CGGGGUCGGG | GGUUCCCGGG | GUUCGGGGUU | CUCCUCCGCG | 3250 |
| CGGCGGUUCC | CCCGCCGGGU | GCGCCCCCCG | GGCGCGGUUU | CCCGCGGCCG | 3300 |
| CCCGCCUCGG | CCGGCGCCUA | GCAGCCGACU | UAGAACUGGU | GCGGACUAGG | 3350 |
| GGAAUCCGAC | UGUUUAAUUA | AAACAAAGCA | UCGCGAAGGC | CCGCGGCGGG | 3400 |
| UGUUGACGCG | AUGUGAUUUC | UGCCCAGUGC | UCUGAAUGUC | AAAGUGAAGA | 3450 |
| AAUUCAAUGA | AGCGCGGGUA | AACGGCGGGA | GUAACUAUGA | CUCUCUUAAG | 3500 |
| GUAGCCAAAU | GCCUCGUCAU | CUAAUUAGUG | ACGCGCAUGA | AUGGAUGAAC | 3550 |
| GAGAUUCCCA | CUGUCCCUAC | CUACUAUCCA | GCGAAACCAC | AGCCAAGGGA | 3600 |
| ACGGGCUUGG | CGGAAUCAGC | GGGGAAAGAA | GACCCUGUUG | AGCUUGACUC | 3650 |
| UAGUCUGGCA | CGGUGAAGAG | ACAUGAGAGG | UGUAGAAUAA | GUGGGAGGCC | 3700 |
| CCCGGCGCCC | CCCCGUUCCC | CGCGAGGGGU | CGGGGCGGGG | UCCGCCGGCC | 3750 |
| UCGCGGGCCG | CCGGUGAAAU | ACCACUACUC | UCAUCGUUUU | UUCACUGACC | 3800 |
| CGGUGAGGCG | GGGGGGCGAG | CCCCGAGGGG | CUCUCGCUUC | UGGCGCCGAA | 3850 |
| CGCGUCCGCG | CGCGCGGCGG | GCGCGACCCG | CUCCGGGGAC | AGUGCCAGGU | 3900 |
| GGGGAGUUUG | ACUGGGGCGG | UACACCUGUC | AAACGGUAAC | GCAGGUGUCC | 3950 |
| UAAGGCGAGC | UCAGGGAGGA | CAGAAACCUC | CCGUGGAGCA | GAAGGGCAAA | 4000 |
| AGCUCGCUUG | AUCUUGAUUU | UCAGUACGAA | UACAGACCGU | GAAAGCGGGG | 4050 |
| CCUCACGAUC | CUUCUGACCU | UUUGGGUUUU | AAGCAGGAGG | UGUCAGAAAA | 4100 |
| GUUACCACAG | GGAUAACUGG | CUUGUGGCGG | CCAAGCGUUC | AUAGCGACGU | 4150 |
| CGCUUUUUGA | UCCUUCGAUG | UCGGCUCUUC | CUAUCAUUGU | GAAGCAGAAU | 4200 |
| UCACCAAGCG | UUGGAUUGUU | CACCCACUAA | UAGGGAACGU | GAGCUGGGUU | 4250 |
| UAGACCGUCG | UGAGACAGGU | UAGUUUUACC | CUACUGAUGA | UGUGUUGUUG | 4300 |
| CCAUGGUAAU | CCUGCUCAGU | ACGAGAGGAA | CCGCAGGUUC | AGACAUUUGG | 4350 |
| UGUAUGUGCU | UGGCUGAGGA | GCCAAUGGGG | CGAAGCUACC | AUCUGUGGGA | 4400 |
| UUAUGACUGA | ACGCCUCUAA | GUCAGAAUCC | GCCCAAGCGG | AACGAUACGG | 4450 |
| CAGCGCCGAA | GGAGCCUCGG | UUGGCCCCGG | AUAGCCGGCU | CCCCGUCCGU | 4500 |
| CCCCGUCCGG | CGGGUCCCCG | CCUCGUCGCC | CCCCCGGGAA | ACGGGGUGCG | 4550 |
| GCCGGAAAGG | GGGCCGCCCU | CUCGCCCGUC | ACGCUUAACG | CACGUUCGUG | 4600 |
| UGGAACUUGG | CGCUAAACCA | UUCGUAGACG | ACCUGCUUCU | GGGUCGGGGU | 4650 |
| UUCGUACGUA | GCAGAGCAGC | UCCCUCGCUG | CGAUCUAUUG | AAAGUCAGCC | 4700 |
| CUCGACACAA | GGGUUUGU | | | | 4718 |

**Figure 2.** The sequence of nucleotides in rat 28S rRNA inferred from the structure of a gene.

**TABLE I.** Comparison of the Structure of Rat, X. laevis and Yeast Large rRNAs

| | Nucleotides (no.) | Different Nucleotides (no.) | Similar Nucleotides (no.) | Identity (%)[a] |
|---|---|---|---|---|
| A. Rat | 4718 | 2772 | 1946 | 41 |
| Yeast | 3393 | 1447 | | 57 |
| B. Rat | 4718 | 1470 | 3248 | 69 |
| X. laevis | 4110 | 862 | | 79 |

[a]Because of the difference in the number of nucleotides in the large rRNAs in the three species the percent identity is conditioned by whether the comparison is between the larger and the smaller, or the smaller and larger.

for X. laevis (Table I). Once again, we have attempted to locate the variable sequences and to identify where the insertions are (Table III). There are five regions that account for 592 or 97% of the additional 608 nucleotides. They are at positions 492 to 725 (132 nucleotides), 891 to 1170 (101 nucleotides), 1984 to 2097 (102 nucleotides), 2807 to 3065 (167 nucleotides), and 3130 to 3292 (90 nucleotides) (Table III). The inserted sequences, again, are especially rich in guanine and cytosine (Table III). We take notice that there are two deletions in the rat sequence, in regions 1333 to 1352 (25 nucleotides) and 2404

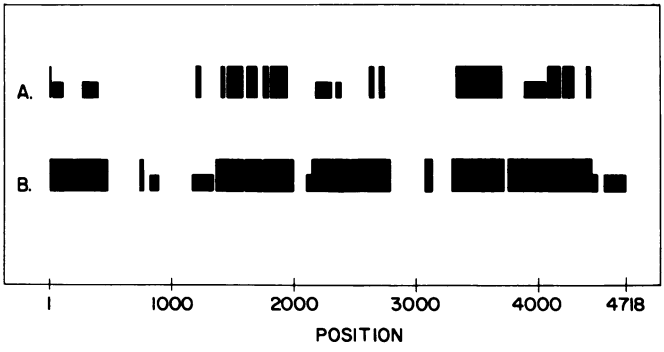



Figure 3. Regions of similarity in the structure of yeast 26S rRNA and X. laevis 28S rRNA compared to rat 28S rRNA. The sequences of the large rRNAs from S. carlsbergensis (19) and X. laevis (21) were compared to that of rat using the computer program of Queen and Korn (37) with the minimum number of matches set at 15 nucleotides. The tall blocks represent regions of yeast (S. carlsbergensis) 26S rRNA (A) or X. laevis 28S rRNA (B) with 85 to 100% identity with rat 28S rRNA; the short blocks are regions with 75 to 85% identity.

**TABLE II.** Comparison of Rat 28S rRNA and Yeast 26S rRNA: Variable Regions

| Rat | | | Yeast | | | |
|---|---|---|---|---|---|---|
| Position | Nucleotides (no.) | G + C (%) | Position | Nucleotides (no.) | G + C (%) | Nucleotides Inserted (no.) |
| 114-275 | 162 | 77 | 113-264 | 152 | 47 | 10 |
| 430-1206 | 777 | 81 | 424-630 | 207 | 48 | 570 |
| 1244-1409 | 166 | 75 | 667-789 | 123 | 49 | 43 |
| 1586-1619 | 34 | 71 | 968-979 | 12 | 42 | 22 |
| 1708-1754 | 47 | 77 | 1068-1114 | 47 | 47 | 0 |
| 1977-2180 | 204 | 84 | 1340-1425 | 86 | 60 | 118 |
| 2316-2343 | 28 | 71 | 1560-1578 | 19 | 47 | 9 |
| 2384-2437 | 54 | 74 | 1618-1649 | 32 | 38 | 22 |
| 2464-2609 | 146 | 62 | 1679-1818 | 140 | 51 | 6 |
| 2745-3328 | 584 | 84 | 1954-2097 | 144 | 59 | 440 |
| 3392-3400 | 9 | 100 | 2162-2170 | 9 | 33 | 0 |
| 3698-3753 | 56 | 89 | 2468-2470 | 3 | 33 | 53 |
| 3779-3887 | 109 | 76 | 2494-2588 | 95 | 47 | 14 |
| 4431-4718 | 288 | 66 | 3131-3393 | 263 | 46 | 25 |
| | | 79 | | | 51 | |

**TABLE III.** Comparison of Rat and X. laevis 28S rRNAs: Variable Regions

| Rat | | | X. laevis | | | |
|---|---|---|---|---|---|---|
| Position | Nucleotides (no.) | G + C (%) | Position | Nucleotides (no.) | G + C (%) | Nucleotides Inserted (no.) |
| 492-725 | 234 | 84 | 489-590 | 102 | 88 | 132 |
| 796-829 | 34 | 85 | 663-664 | 0 | | 34 |
| 891-1170 | 280 | 83 | 732-910 | 179 | 87 | 101 |
| 1302-1318 | 17 | 88 | 1043-1056 | 14 | 93 | 3 |
| 1333-1352 | 20 | 70 | 1071-1115 | 45 | 93 | -25 |
| 1594-1617 | 24 | 67 | 1359-1364 | 6 | 67 | 18 |
| 1984-2097 | 114 | 88 | 1739-1750 | 12 | 67 | 102 |
| 2404-2405 | 0 | | 2059-2094 | 36 | 94 | -36 |
| 2807-3065 | 259 | 86 | 2502-2593 | 92 | 86 | 167 |
| 3130-3292 | 163 | 82 | 2655-2727 | 73 | 86 | 90 |
| 3709-3756 | 48 | 88 | 3146-3160 | 15 | 67 | 33 |
| 4495-4554 | 60 | 83 | 3904-3956 | 53 | 81 | 7 |
| | | 85 | | | 87 | |

to 2405 (36 nucleotides). Thus, the increase in size of the large RNAs in the 60S ribosomal subunit is the result of the addition en bloc of skeins of guanine- and cytosine-rich oligonucleotides (Fig. 3B). Nonetheless, the total content of the two nucleotides is very similar in rat (67%) and X. laevis (65%; ref 21) 28S rRNAs.

We note in passing that there are two recognition sequences for the restriction endonuclease BamHI, at positions 1325 and 2700, in the rat 28S rRNA gene (3,7,15). The mouse 28S rRNA gene also has two BamHI sites (22, 23). However, the X. laevis 28S rDNA has only one (21, 24), and yeast none (19, 20). The loss of the second BamHI site in X. laevis 28S rDNA is the result of a guanine to adenine transition at postion 2398, changing the recognition sequence GGATCC to AGATCC.

Eukaryotic 80S ribosomes have increased in size during evolution from plants to animals (25, 26). The size of the 40S subunit has remained fairly constant so the increase in size is to be attributed to changes in the 60S subparticle and primarily to an increase in the number of nucleotides in the large RNA. The number of nucleotides in 18S rRNAs from rat (1874; ref. 4), X. laevis (1825; ref. 27), and yeast (1789; ref. 28) are similar, whereas the number in the large rRNAs progressively increase as one goes from yeast (3393; ref. 19, 20), to X. laevis (4110; ref. 21), to rat (4718). It is difficult to provide an explanation for this evolutionary change, or to rationalize the insertion sequences being rich in guanine and cytosine.

The 5.8S rRNA Contact Sites in 28S rRNA. In the large subunit of eukaryotic ribosomes 5.8S rRNA is non-covalently associated with 28S rRNA (29). The association between these RNA molecules is relatively stable, although the complex can be dissociated by agents that disrupt hydrogen bonds. Fragments containing the 5'-terminal 73 nucleotides and the 3'-terminal 83 nucleotides of 5.8S rRNA can associate independently with 28S rRNA (30). However, it is probable that only approximately the 3'-terminal 20 nucleotides and the 5'-terminal 20-30 nucleotides are involved in the interaction with 28S rRNA (31). While the exact location of the junction sites in 28S rRNA have not been established experimentally there is evidence that 5.8S rRNA interacts with a structural domain near the 5' end (23).

One can deduce what the identity of these contact sites might be by examining the proposal for the secondary structure of Escherichia coli 23S rRNA. The 5'-terminal region of E. coli 23S rRNA is structurally homologous to eukaryotic 5.8S rRNA (32, 33), indeed the basic difference between eukaryotes and prokaryotes in this regard is that in the former 5.8S rRNA has been processed out of the transcription unit in a manner that leads to formation of a separate molecule. It follows then, that one can examine the secondary structure of E. coli 23S rRNA (34), locate the region homologous to 5.8S rRNA, and take notice of where this region interacts with the remainder of the structure. The prediction arising from that exercise is that the 3' end of 5.8S rRNA

interacts with the immediate 5' end of 28S rRNA and that the 5' end of 5.8S rRNA is complementary to a region about 300 to 400 nucleotides from the 5' end of 28S rRNA (23). As a result of the determination of the sequence of yeast 26S rRNA, and elucidation of the possible secondary structure (19), the proposal could be tested. Two regions of extensive complementarity between the two nucleic acids were found (19) which are entirely consistant with the prediction. Moreover, an exactly analogous interaction has been suggested for mouse 5.8S and 28S rRNAs predicated on the entire sequence of the former and the 585 nucleotides at the 5' end of the latter (35).

Guided by the earlier work (19, 35) we have searched the sequences of rat 5.8S and 28S rRNAs for regions of complementarity. In conformity with the junction structures proposed for yeast and mouse rRNAs, there are two sites of possible extensive base pairing in rat 5.8S and 28S rRNAs. They encompass, in the first instance, 17 base pairs between the 3' end of 5.8S rRNA (positions 137 to 155) and the 5' end of 28S rRNA (positions 2 to 20); and in the second, 24 base pairs between the 5' end of 5.8S rRNA (positions 4 to 32) and a region of 28S rRNA centered about nucleotide 400. Thus, the results for rat are entirely compatible with the proposals for the contact sites in yeast (19) and mouse rRNAs (35). They do not lend credence to the submission that both ends of 5.8S rRNA interact with the 3' end of Neurospora crassa 25S rRNA (36), although the results do not entirely negate the possibility of some interaction of the 3' end of the large rRNA with either 5.8S rRNA or the 5' end of 28S rRNA.

ACKNOWLEDGMENTS

The expenses of the research were met by grants CA 19265 and GM 21769 from the National Institutes of Health. We are grateful to L. I. Rothblum of Baylor College of Medicine for providing DNA from the recombinant plasmid p2.7, and T.D. Sargent and J. Bonner of California Institute of Technology for DNA from the λ phage Charon 4A. We are indebted to S.A. Gerbi of Brown University for sharing with us the sequence of X. laevis 28S rRNA before its publication.

REFERENCES

1. Aoyama, K., Hidaka, S., Tanaka, T. and Ishikawa, K. (1982) J. Biochem 91, 363-367.
2. Nazar, R.N., Sitz, T.O. and Busch, H. (1975) J. Biol. Chem. 250, 8591-8597.
3. Subrahmanyam, C.S., Cassidy, B., Busch, H., Rothblum, L.I. (1982) Nucleic Acids Res. 10, 3667-3680.
4. Chan, Y.L., Gutell, R., Noller, H.F. and Wool, I.G. (1983) J. Biol. Chem., submitted.
5. Sargent, T.D., Wu, J.R., Sala-Trepat, J.M., Wallace, R.B., Reyes, A.A. and Bonner, J. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 3256-3260.
6. Chang, A.C.Y. and Cohen, S.N. (1978) J. Bacteriol. 134, 1141-1156.
7. Rothblum, L.I., Parker, D.L. and Cassidy, B. (1982) Gene 17, 75-77.
8. Southern, A.M. (1975) J. Mol. Biol. 98, 503-517.
9. Endo, Y. and Wool, I.G. (1982) J. Biol. Chem. 257, 9054-9060.
10. Clewell, D.B. and Helinski, D.R. (1969) Proc. Natl. Acad. Sci. U.S.A. 62, 1159-1166.

11. Kupersztoch, Y.M. and Helinski, D.R. (1973) Biochem. Biophys. Res. Comm. 54, 1451-1459.
12. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 560-564.
13. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
14. Sanger, F. and Coulson , A.R. (1978) FEBS Lett. 87, 107-110.
15 Chan, Y.-L., Endo, Y. and Wool, I.G. (1983) J. Biol. Chem., submitted.
16. Loening, U.E. (1968) J. Mol. Biol. 38, 355-365.
17. Weinberg, R.A. and Penman, S. (1970) J. Mol. Biol. 47, 169-178.
18. Shah, M., Khan, N., Salim, M. and Maden, B.E.H. (1978) Biochem. J. 169, 531-542.
19. Veldman, G.M., Klootwijk, J., de Regt, V.C.H.F., Planta, R.J., Branlant, C., Krol, A. and Ebel, J.P. (1981) Nucleic Acids Res. 9, 6935-6952.
20. Georgiev, O.I., Nikolaev, N., Hadjiolov, A.A., Skryabin, K.G., Zakharyev,V.M. and Bayev, A.A. (1981) Nucleic Acids Res. 9, 6953-6958.
21. Ware, V.C., Tague, B.W., Clark, C.G., Brand, R.C. and Gerbi, S.A. (1983) Nucleic Acids Res., submitted.
22. Kominami, R., Mishima, Y., Urano, Y., Sakai, M. and Muramatsu, M. (1982) Nucleic Acids Res. 10, 1963-1979.
23. Walker, T.A., Endo, Y., Wheat, W.H., Wool, I.G. and Pace, N.R. (1983) J. Biol. Chem. 258, 333-338.
24. Hall, L.M.C. and Maden, B.E.H. (1980) Nucleic Acids Res. 8, 5993-6005.
25. Cammarano, P., Romeo, A., Gentile, M., Felsani, A. and Gualerzi, C. (1972) Biochim. Biophys. Acta 281, 597-624.
26. Cammerano, P., Felsani, A., Gentile, M., Gualerzi, C., Romeo, A. and Wolf, G. (1972) Biochim. Biophys. Acta 281, 625-642.
27. Salim, M. and Maden, B.E.H. (1981) Nature 291, 205-208.
28. Rubtsov, P.M., Musakhanov, M.M., Zakharyev, V.M., Krayev, A.S., Skryabin, K.G. and Bayev, A.A. (1980) Nucleic Acids Res. 8, 5779-5793.
29. Pene, J.J., Knight, E. and Darnell, J.E. (1968) J. Mol. Biol. 33, 609-623.
30. Peters, M.A., Walker, T.A. and Pace, N.R. (1982) Biochemistry 21, 2329-2335.
31. Walker, T.A., Johnson, K.D., Olsen, G.J., Peters, M.A. and Pace, N.R. (1982) Biochemistry 21, 2320-2329.
32. Nazar, R.N. (1980) FEBS Lett. 119, 212-214.
33. Jacq, B. (1981) Nucleic Acids Res. 9, 2913-2932.
34. Noller, H.F., Kop, J., Wheaton, V., Brosius, J., Gutell, R.R., Kopylov, A.M., Dohme, F., Herr, W., Stahl, D.A., Gupta,R. and Woese, C. (1981) Nucleic Acids Res. 9, 6167-6189.
35. Michot, B., Bachellerie, J.P. and Raynal, F. (1982) Nucleic Acids Res. 10, 5273-5283.
36. Kelly, J.M. and Cox, R.A. (1981) Nucleic Acids Res. 9, 1111-1121.
37. Queen, C.L. and Korn, L.J. (1980) Methods Enzymol. 65, 595-609.