J Mol Evol (1985) 21:259–269

Journal of Molecular Evolution

Comparison of the Nucleotide Sequence of Soybean 18S rRNA with the Sequences of Other Small-Subunit rRNAs

Virginia K. Eckenrode[1]*, Jonathan Arnold[2], and Richard B. Meagher[2]

[1] Department of Biochemistry, University of Georgia, Athens, Georgia 30602, USA
[2] Department of Genetics, University of Georgia, Athens, Georgia 30602, USA

Summary. We present the sequence of the nuclear-encoded ribosomal small-subunit RNA from soybean. The soybean 18S rRNA sequence of 1807 nucleotides (nt) is contained in a gene family of approximately 800 closely related members per haploid genome. This sequence is compared with the ribosomal small-subunit RNAs of maize (1805 nt), yeast (1789 nt), *Xenopus* (1825 nt), rat (1869 nt), and *Escherichia coli* (1541 nt). Significant sequence homology is observed among the eukaryotic small-subunit rRNAs examined, and some sequence homology is observed between eukaryotic and prokaryotic small-subunit rRNAs. Conserved regions are found to be interspersed among highly diverged sequences. The significance of these comparisons is evaluated using computer simulation of a random sequence model. A tentative model of the secondary structure of soybean 18S rRNA is presented and discussed in the context of the functions of the various conserved regions within the sequence. On the basis of this model, the short base-paired sequences defining the four structural and functional domains of all 18S rRNAs are seen to be well conserved. The potential roles of other conserved soybean 18S rRNA sequences in protein synthesis are discussed.

Key words: Eukaryotic rRNA — Small-ribosomal-subunit rRNA — Secondary structure — Computer simulation

Offprint requests to: R.B. Meagher
* *Current address: Department of Genetics, North Carolina State University, Raleigh, North Carolina 27695, USA*

Introduction

Ribosomes from all prokaryotic, archaebacterial, and eukaryotic sources are composed of small and large subunits. These complexes of RNA and protein molecules have conserved overall structures and perform similar biological functions during protein synthesis (Wool 1980; Liljas 1982; Lake 1983). The ribosomal complex has been studied as an example of RNA–RNA and RNA–protein interactions. These interactions are reflected in a conserved structure for the ribosomal RNAs (rRNAs) (Noller and Woese 1981).

The rRNA molecule associated with the small ribosomal subunit from any source is referred to as the small-subunit rRNA. Small-subunit rRNAs are often grouped according to source and size: 18S rRNAs from eukaryotic cytoplasms, 16S rRNAs from prokaryotic sources, 12S rRNAs from animal mitochondria, etc. Although their lengths vary approximately twofold, from 954 nucleotides (nt) [12S from human mitochondria (Eperon et al. 1980)] to 1962 nt [18.5S from maize mitochondria (Chao et al. 1984)], small-subunit rRNAs contain certain structures that can be identified as shared by all. All the nuclear-encoded 18S rRNAs have lengths of close to 1800 nt and share significant nucleotide sequence homology. However, the eukaryotic 18S and prokaryotic 16S small-subunit rRNAs differ in length by about 300 nt and share little overall nucleotide sequence homology.

Relatively little is known about plant cytoplasmic ribosomes and rRNAs. The sequence of the small rRNA from maize, a monocotyledon, has recently been reported (Messing et al. 1984). To extend struc-

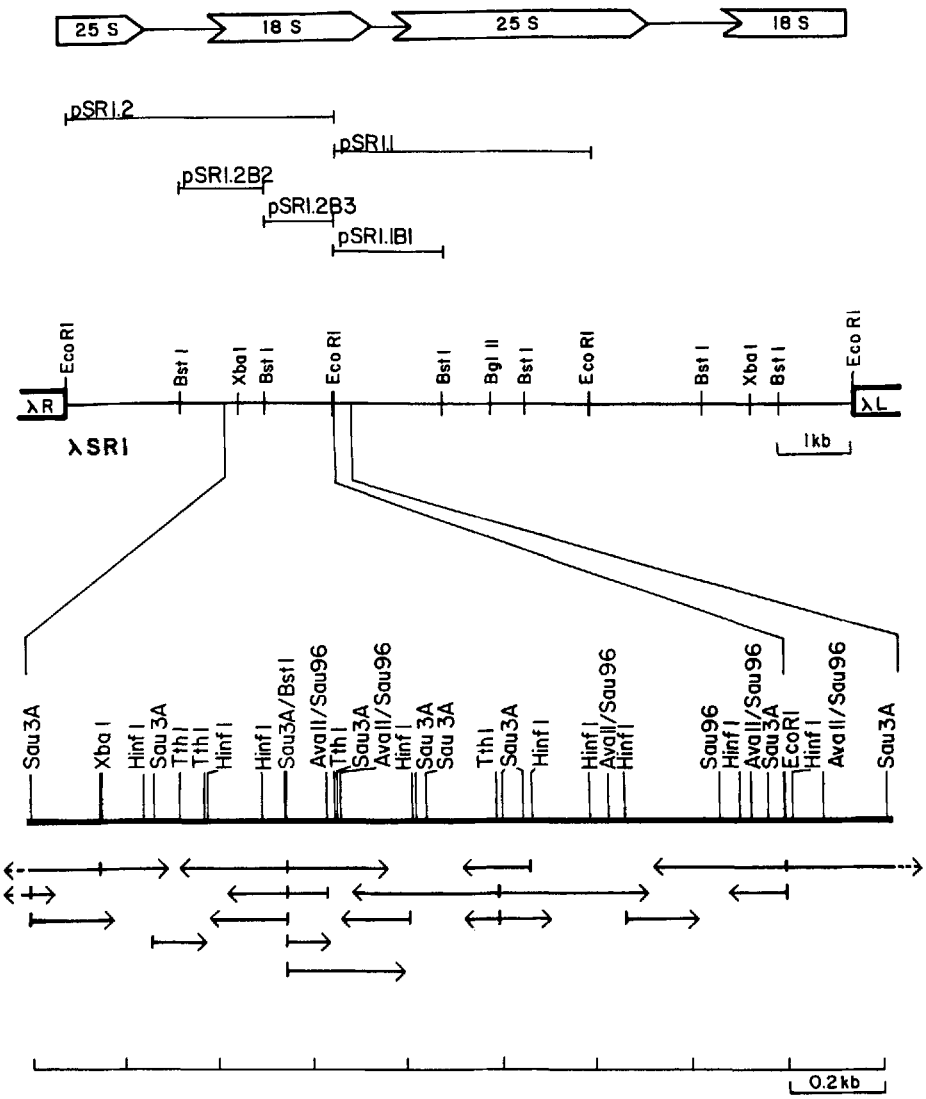


Fig. 1. Strategy for sequencing the gene encoding soybean 18S rRNA. The restriction endonuclease map of λSR1, a lambda clone containing one and one-half rDNA repeat units from soybean, is shown in the center of the figure. The approximate coding regions for 18S and 25S rRNAs are shown at the top of the figure. These coding regions were localized to within 200 bases by a combination of Southern and Northern analyses (Eckenrode 1983). Plasmid subclones used for Maxam and Gilbert (1980) sequence analysis of the region encoding 18S rRNA are indicated above the map of λSR1. An expanded restriction endonuclease map for the 18S rDNA region, derived from its nucleotide sequence, is shown below the map of λSR1. The sequencing strategy for the 18S rDNA is indicated below the expanded map

tural and evolutionary comparisons further, we have determined the nucleotide sequence of the 18S rRNA gene from soybean, a dicotyledon. This sequence is compared here with other nuclear-encoded eukaryotic small-subunit rRNAs, as well as with *Escherichia coli* 16S rRNA.

Experimental Procedures

Recombinant Clones. λSR1 is a recombinant lambda clone containing one and one-half rDNA repeat units from the soybean *Glycine max*, var. Wayne (Eckenrode 1983). The inserted DNA of λSR1 is composed of three EcoRI fragments: two identical 3.9-kb fragments and one 3.75-kb fragment (Fig. 1). Each of these two different EcoRI fragments was subcloned into the EcoRI site of pBR325 (Bolivar and Backman 1979) and relevant portions of these were further subcloned into the appropriate sites of pBR322 (Bolivar et al. 1977). Ligations were performed as described by Maniatis et al. (1982). The nomenclature of these subclones is presented in Fig. 1. Transformations of HB101 were performed as described by Kushner (1978). Colonies that tested as recombinant on the appropriate antibiotic plates were screened for the presence of inserted DNA in the vector plasmid (Meagher et al. 1977a). Inserted DNA fragments were identified on the basis of their sizes and restriction patterns. Plasmid DNA from spectinomycin- (pBR325) or chloramphenicol- (pBR322) amplified cells was purified as described by Meagher et al. (1977b).

Restriction enzymes were purchased from either Bethesda Research Labs (Gaithersburg, MD) or New England Biolabs (Waltham, MA) and used according to the manufacturers' specifications.

DNA Fragment Purification. All DNA fragments to be purified were separated by electrophoresis in 3-mm-thick nondenaturing 5% acrylamide gels (Maxam and Gilbert 1980). DNA fragments were electroeluted in a dialysis bag using the modified TEA buffer described by Carreira et al. (1980): 60 mM Trizma Base and 2 mM disodium ethylenediaminetetraacetate (EDTA), titrated to pH 8.3 with glacial acetic acid. The buffer containing the DNA sample (~3 ml) was concentrated tenfold by evaporating water from the sample in a Savant Speed-Vac. Particulate matter was removed by centrifuging the sample through a plug of siliconized glass wool in an Eppendorf tube with a hole in the bottom into a fresh Eppendorf tube (Maniatis et al. 1982). If the DNA sample had already been radioactively labeled, 15 μg phenol-extracted tRNA (Sigma, St. Louis, MO) was added to act as a carrier during ethanol precipitations. Three ethanol precipita-

tions were performed prior to sequencing (Maxam and Gilbert 1980).

Labeling the DNA. Recessed 3′ ends of restriction fragments were filled in using [α-^{32}P]dNTPs and the Klenow fragment of *E. coli* DNA polymerase I (both from New England Nuclear Corp., Boston, MA). The reaction conditions were as described by Shah et al. (1982), with the modification that the reaction mixtures contained no more than 10 pmol of fragment ends and approximately 12 pmol of the appropriate radioactive nucleotide (usually 50 μCi, with a specific activity of $\geq$3000 Ci/mmol).

DNA Sequence Determination. Purified DNA fragments were sequenced by the method of Maxam and Gilbert (1980), with the modification that the "G + A" reaction was performed in 70% formic acid for 5 min, as described by Krayev et al. (1980). Sequencing gels were run as described by Maxam and Gilbert (1980), except that high gel temperatures (>65°C) were necessary for obtaining reproducible DNA sequence patterns of this ribosomal gene region, most likely because of the high degree of potential secondary structure in rDNA. Because the rRNA transcripts from this gene probably have approximately 50% of their bases in a base-paired configuration (see Discussion), each DNA strand can also form intrastrand duplexes, which could cause aberrent sequencing patterns. Elevating the temperature probably denatured the intrastrand duplexes. Typical gels (43 cm long, 35 cm wide, 0.3 mm thick) were run at 150 W constant power.

Computer Analysis of the DNA Sequence Data. Sequencing data were stored and analyzed using the Stanford Gene Molgen Project with NIH SUMEX-AIM Facility (Stanford, CA). An Apple II Plus computer and a Zenith Z-19 terminal were used to interact with the Stanford computer via the TYMNET satellite communications system. Additional analyses were performed using the Intelligenetics system (Palo Alto, CA) and the same terminals. Files were transferred from the remote system to the Apple II Plus system using a BITS program (Software Sorcery, McLean, VA).

A program developed by Arnold et al. (J. Arnold, V.K. Eckenrode, K. Lemke, G.J. Phillips, and S.W. Schaeffer, manuscript submitted to Nucleic Acids Research) for a PDP 11/34A computer (Digital Equipment Corporation, Maynard, MA) was used for the final pairwise intergenic comparisons of the small-subunit rRNA nucleotide sequences. The program implements a dynamic programming algorithm described by Kruskal (1983) and originally proposed by Needleman and Wunsch (1970). This algorithm specifies a weight for each type of nucleotide mutation: transition, transversion, and insertion/deletion. A perfect match has no weight. A transition is defined as having a weight of 1. Other, less likely mutations are given higher weights. The optimal alignment of two sequences involves minimizing the sum of these weights; this minimum sum is defined as the evolutionary distance. Sankoff et al. (1976) suggested giving insertions/deletions weights of 2.25 and transversions weights of 1.75. These numbers were based on the results of multiway sequence comparisons among 5S rRNAs. We found that weighting transversions more heavily than transitions, as suggested by Brown and Clegg (1983), had little effect on our pairwise alignments. Therefore, transitions and transversions were both given weights of 1, as in the work of Erickson and Sellers (1983). For the sake of simplicity, insertions/deletions were weighted twice as heavily as transitions and transversions, i.e., were given a weight of 2. Mismatches were therefore favored over insertion/deletion events. All nucleotide sequence comparisons in Fig. 2 represent optimal pairwise alignments against the soybean 18S rRNA. Percentage homology between any two nucleotide sequences was calculated as the number of positions with the same nucleotide divided by 1919, the total number of positions needed to make the six-way match shown in Fig. 2.

To test the significance of the evolutionary distances, the distribution of evolutionary distances between pairs of random sequences was determined. Five hundred pairs of random sequences with lengths of 1819 nt and an average G + C composition of 51% (see Table 1) were generated using a multiplicative random-number generator with period $2^{31} - 1$ and multiplier 7^5 (Knuth 1981). The generator was tested as in Knuth (1981).

Generation of a Secondary-Structure Model. To facilitate the analysis of conserved structural regions and regions of significance in the soybean 18S rRNA sequence, we devised a potential secondary structure based on a secondary-structure model for yeast and *Xenopus* 18S rRNAs (Zwieb et al. 1981). Changes in the lengths and relative placements of duplex regions were made from the model for the yeast 18S rRNA to accommodate the soybean 18S rRNA sequences. The base pairing and structure rules of Erdmann et al. (1983) were followed for the determination of duplex regions.

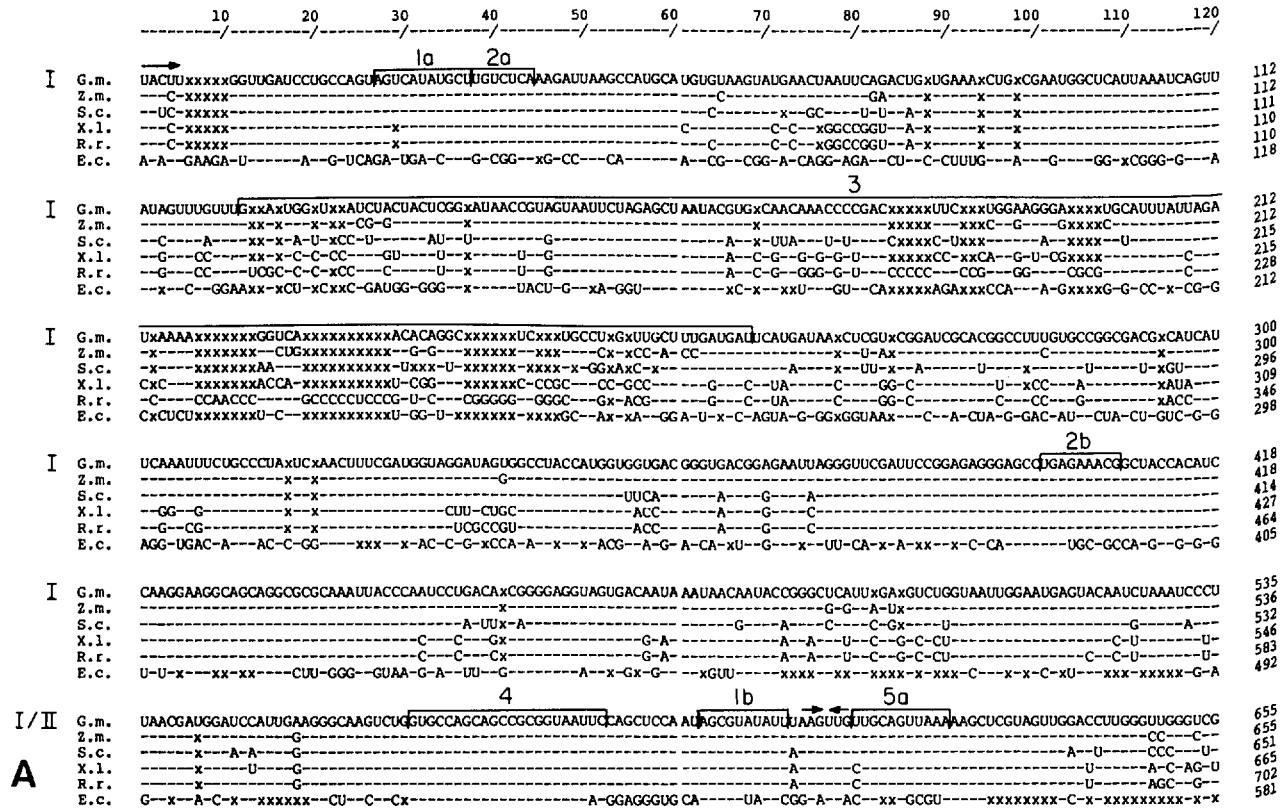
Analysis of Soybean Genomic DNA for Small-Subunit rRNA-Encoding Sequences. Soybean genomic DNA was analyzed (Southern 1975) for small-subunit rDNA sequences hybridizing to the small-subunit rDNA insert contained in pSR1.2B3. This 1.05-kb BstI–EcoRI fragment encodes nucleotides 544–1583 of the 18S rRNA gene of λSR1. Filters were prehybridized and hybridized in 50% formamide, 5× SSC, 5× Denhardt's solution (Denhardt 1966), at 56°C. Filters were washed three times in 0.2× SSC, 0.2% sodium dodecyl sulfate (SDS), at 56°C for 10 min each and exposed overnight to film.

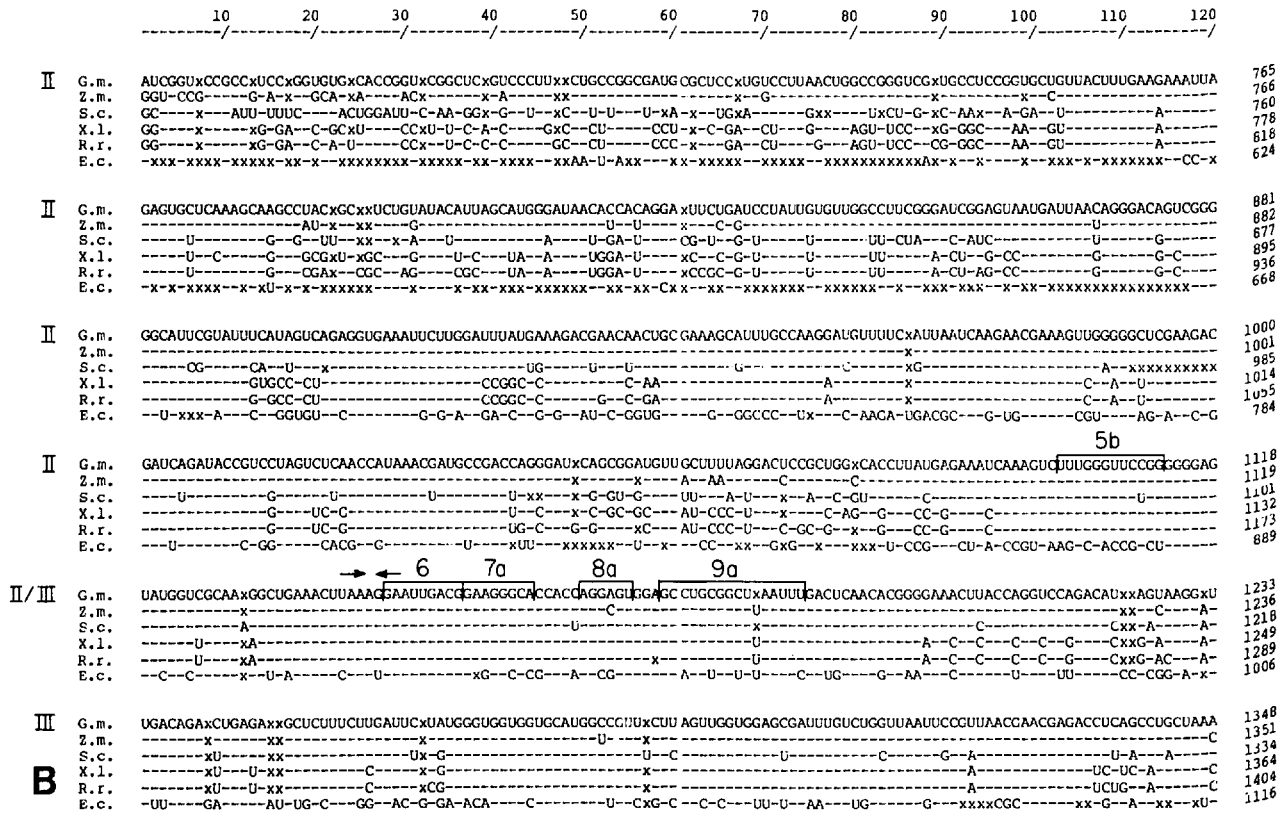
Results

Description of the Soybean 18S rRNA Sequence

The nucleotide sequence of a soybean 18S rRNA molecule, as inferred from the gene sequence contained in recombinant phage λSR1, is presented in full on the top line of Fig. 2. Also shown in Fig. 2 is a comparison of the sequence of the soybean 18S rRNA with the sequences of four other eukaryotic 18S rRNAs [maize (Messing et al. 1984), yeast (Rubstov et al. 1980), frog (Salim and Maden 1981), and rat (Torczynski et al. 1983)] and with the sequence of one prokaryotic 16S rRNA [*E. coli* (Noller and Woese 1981)]. Because the 5′ and 3′ ends of the 18S rRNAs from yeast (Krayev et al. 1980), frog (Salim and Maden 1981), and rat (Torczynski et al. 1983) have been experimentally determined and because of the perfect homology of the end nucleotides and the nearly perfect sequence homology of the first 70 and last 50 nt, the limits of the mature soybean 18S rRNA sequence were operationally defined by their agreement with those of these three other rDNA genes. The mature 18S rRNA from soybean is 1807 nt long. This is within the length range of the known eukaryotic 18S rRNAs, from 1789 [yeast (Krayev et al. 1980)] to 1869 [rat (Torczynski et al. 1983)] nt.

The average G + C content of the eukaryotic 18S


10 20 30 40 50 60 70 80 90 100 110 120
1a 2a
3
2b
4 1b 5a
I/II
A
G.m.
Z.m.
S.c.
X.l.
R.r.
E.c.


10 20 30 40 50 60 70 80 90 100 110 120
5b
6 7a 8a 9a
II/III
B
G.m.
Z.m.
S.c.
X.l.
R.r.
E.c.

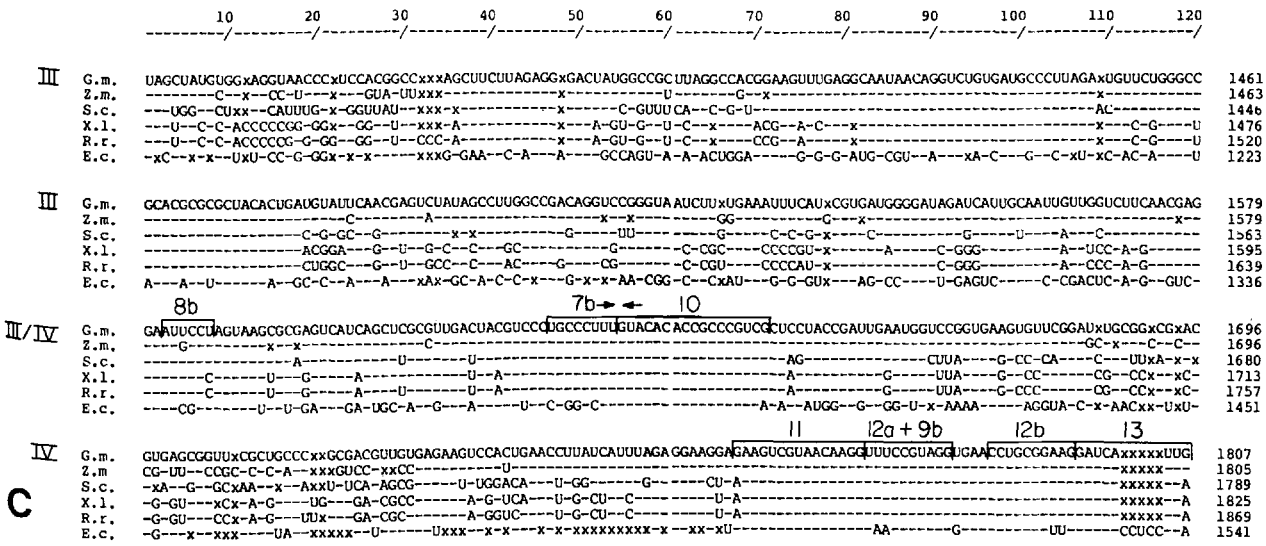


Fig. 2. Comparison of the soybean 18S rRNA nucleotide sequence with those of four eukaryotic 18S rRNAs and of one prokaryotic 16S rRNA. The nucleotide sequence of 18S rRNA from soybean (G.m.) was predicted from genomic DNA sequences. It is compared with the nucleotide sequences for the 18S rRNAs from maize (Z.m.), yeast (S.c.), *Xenopus leavis* (X.l.), and rat (R.r.), and the 16S rRNA from *E. coli* (E.c.). All nucleotide sequence comparisons were optimized for homology with the soybean sequence. A nucleotide in any of the other 18S rRNAs that is identical to that in the soybean sequence in that position is indicated by a dash. A nucleotide different from that in the soybean sequence is indicated by writing the letter symbol for the differing nucleotide in the appropriate position of the other 18S rRNA sequence. A deletion in any sequence is indicated by an "x" in the appropriate position. The four predicted structural domains of the soybean sequence are indicated by roman numerals on the left side of the figure and are roughly separated by opposing arrows. The total number of nucleotides presented for each sequence is tallied at the end of each line. Regions of particular interest are enclosed and numbered (see Discussion)

Table 1. Nucleotide compositions and lengths of various small-subunit rRNAs

Species[a]	Number of nucleotides				Length (nt)	G + C content
	A	U	G	C		
Soybean (G.m.)	451	469	491	396	1807	49%
Maize (Z.m.)	447	439	500	419	1805	51%
Yeast (S.c.)	475	509	458	347	1789	45%
Xenopus (X.l.)	432	411	516	466	1825	54%
Rat (R.r.)	421	408	543	497	1869	56%
E. coli (E.c.)	389	313	487	352	1541	54%
Eukaryotic average	445	447	502	425	1819	51%

[a] Abbreviations for organisms are as given in Fig. 2

rRNAs examined is 51% (Table 1), with a variation of 5% from this average. The G + C content of the mature 18S rRNA from soybean, 49%, is close to this average. Although the high degree of variation of G + C content among the eukaryotic 18S rRNAs has been noted previously (Salim and Maden 1981; Torczynski et al. 1983; Messing et al. 1984), the significance of this variation is unknown. Salim and Maden (1981) have noted, however, that this variation of G + C content is confined to the regions of low homology among the 18S rRNAs (see next section).

Sequence Comparisons among the Eukaryotic 18S rRNA Sequences

An examination of Fig. 2 reveals extensive nucleotide sequence homology among the eukaryotic 18S rRNAs. All of the eukaryotic 18S rRNAs examined share at least 74% nucleotide sequence homology with the soybean 18S rRNA (Table 2). Soybean 18S rRNA is more homologous (93.5%) in nucleotide sequence to maize 18S rRNA than to the 18S rRNAs from the organisms from other kingdoms. Similarly, even when they are optimized for alignment with the soybean sequence, *Xenopus* 18S rRNA is more homologous (91.6%) in nucleotide sequence to rat 18S rRNA than to the other sequences analyzed.

It is apparent from Fig. 2 that the five eukaryotic 18S rRNAs presented are collinear. In any pairwise comparison of the eukaryotic rRNAs, there are stretches of high sequence homology (greater than 90%) separated by stretches of low homology (less than 45%). For example, soybean nucleotides 1–65 and 81–124 are conserved in all eukaryotic 18S rRNAs. They are separated by 13 nucleotides that are nonhomologous among the various sequences. Other regions of shared high sequence homology include soybean nucleotides 135–170, 358–490, 505–647, 748–780, 860–895, 902–919, and 950–

Table 2. Percentage homologies (below the diagonal of null values) and evolutionary distances (above diagonal) between various small-subunit rRNAs

	G.m.	Z.m.	S.c.	X.l.	R.r.	E.c.
G.m.	0/100%	138	383	419	497	1126
Z.m.	93.5%	0/100%	412	416	496	1132
S.c.	82.8%	80.8%	0/100%	464	542	1117
X.l.	79.9%	79.4%	76.7%	0/100%	177	1154
R.r.	78.2%	77.3%	74.8%	91.6%	0/100%	1230
E.c.	53.3%	51.4%	47.8%	47.2%	44.4%	0/100%

All percentages were calculated by taking the total number of similarities between two sequences (using the alignment of Fig. 2) and dividing by 1919, the total number of positions necessary to accommodate all base insertions. In this way, all the pairwise comparisons were normalized to the same length. The upper half of the matrix gives the evolutionary distances based on the weighted values for transitions/transversions and insertions/deletions. An evolutionary tree for the rRNA family can be constructed by single linkage (Hartigan 1975) based on either percentage homology or evolutionary distance. The resulting trees are identical, and the common tree is superimposed on the matrix using boxes. The species abbreviations are as defined in the legend to Fig. 2

990. In all pairwise sequence comparisons, most of the regions of sequence divergence occur in the same relative positions. This pattern of 18S rRNA sequence conservation was first observed by Salim and Maden (1981) in a comparison between the *Xenopus* and yeast 18S rRNAs and was described by them as "extensive but interrupted" or "interspersed" homology. Our five-way comparison extends and confirms their observation.

Most striking is the 93.5% homology between the soybean and maize 18S rRNA sequences, which is accounted for by the fact that there are only 129 nucleotide replacements between the two genes, which are scattered throughout the sequences. In contrast, both the rat and frog sequences are less than 80% homologous with the soybean sequence, while being 91.6% homologous with each other. In fact, the rat 18S rRNA is 62 nt longer than the soybean 18S rRNA molecule. Most of the additional nucleotides in the rat 18S rRNA occur in region 3, rat nucleotides 119–306 (Figs. 2 and 5), in which the nucleotide sequences also vary among the other four sequences. The differences in length among the non-rat rRNA molecules are much less dramatic and are scattered throughout the whole sequence.

Sequence Homology Between the Soybean and E. coli Small-Subunit rRNAs

Comparison of the soybean and *E. coli* sequences provides examples of the degrees and types of sequence homology between eukaryotic and prokaryotic small-subunit rRNAs. In contrast to the high degree of nucleotide sequence homology among the eukaryotic small-subunit rRNAs, the homology between the soybean 18S rRNA sequence and the *E. coli* 16S rRNA sequence is only 53%. This calculation was based on a total length of 1919 nucleotides, with insertions placed where necessary to maximize the homology. Because the total length used for the match was longer than the soybean sequence and far longer than the *E. coli* 16S sequence, this percentage is undoubtedly an overestimate of the actual homology. The percentage homology between two random sequences of approximate length 1800 is 50% using our matching program. Therefore, the actual homology between the soybean and *E. coli* small-subunit rRNAs is not very different from the homology between two random sequences. What is significant is that there are short stretches of high sequence homology between the soybean and *E. coli* small-subunit rRNAs. For example, there are five perfect matches of ten bases or more, represented by soybean nucleotides 565–584 (region 4), 1146–1155 (region 6), 1631–1649 (region 10), 1760–1769 (region 11), and 1785–1794 (region 12). As in the work of Brown and Clegg (1983, p. 118), the probability of finding a run of ten perfect matches between *E. coli* and soybean in 1807 bases is estimated to be $(266/1807)(1541/1807)^{10} = 0.030$, where the length difference between *E. coli* and soybean is 266 bases. The fact that all five matches occur in an order consistent with collinear alignment of the two sequences makes it highly improbable that these homologies arose by chance. It is also notable that these homologies occur in regions that are conserved among all the eukaryotic 18S rRNA sequences. The potential structural and functional significance of some of the conserved regions within the small-subunit rRNA sequences will be addressed in the Discussion.

Statistical Significance of Homologies

The evolutionary distance between each pair of eukaryotic 18S rRNAs was calculated using the pro-

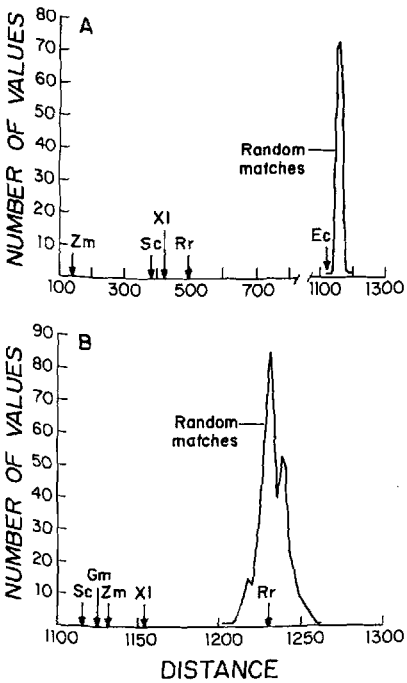


Fig. 3A, B. The empirical distribution of "evolutionary" distances between 500 pairs of random sequences. **A** The average base composition (51% G + C) of the random sequences and the length (1819 nt) of each random sequence are the averages given in Table 1 for eukaryotic 18S sequences. Arrows indicate the observed evolutionary distances between soybean and the various eukaryotic small-subunit rRNAs relative to these random comparisons. **B** In each of these comparisons, the base composition and length of one random sequence are based on those of soybean 18S rRNA (49% G + C, 1807 nt) and those of the other random sequence are based on those of *E. coli* 16S rRNA (54% G + C, 1541 nt). Arrows indicate observed distances between *E. coli* and the various eukaryotic small-subunit rRNAs on this second scale of random comparisons. Abbreviations for organisms are as in Fig. 2

gram developed by J. Arnold. The higher the evolutionary distance, the less the homology should be. By this criterion, the calculated data (Table 2) are in agreement with the homology percentages based on the optimal pairwise alignments of each sequence with the soybean sequence.

As in the studies of Sankoff and Cedergren (1973) and Elleman (1978), a Monte Carlo method was used to test the significances of the homologies among the sequences of the five eukaryotic 18S rRNAs and the one prokaryotic 16S rRNA shown in Fig. 2. Figure 3A shows the distribution of the evolutionary distances between 500 pairs of random sequences with a length of 1819 nt and an average G + C composition of 51%—the average length and G + C content for the eukaryotic 18S rRNAs compared (Table 1). Also shown is the position in this distribution of the evolutionary distance between soybean and each of the small-subunit rRNAs in Fig. 2. The fraction of simulated evolutionary distances falling below an observed evolutionary distance between a pair of real 18S rRNA sequences provides an estimate of the significance level of this observed distance. The observed distances between eukaryotic 18S rRNAs are significant at the 0.2% level (i.e., no random comparison gave a distance less than any observed distance between eukaryotic sequences). The evolutionary distance between the soybean and *E. coli* small-subunit rRNAs is also shown for comparison.

Figure 3B shows the distribution of evolutionary distances between 500 pairs of random sequences in which one member of each pair has the length and base composition of *E. coli* 16S rRNA (1541 nt, 54% G + C content) and the other has the length and base composition of soybean 18S rRNA (1807 nt, 49% G + C content). The mean evolutionary distance between random sequences is 1231, whereas the distance between *E. coli* and soybean sequences is 1100. None of the 500 simulated distances fall below this level. Therefore, despite the large size diffference between *E. coli* and soybean small-subunit rRNAs, this overall match is significant at the 0.2% level. Only the rat–*E. coli* small-subunit rRNA comparison fails to be significant by this criterion (i.e., the observed distances fall within the distribution of simulated distances). Note that both the rat and *E. coli* sequences were optimized for comparison with the soybean sequence before being compared with each other. This may account for the relatively high observed evolutionary distance.

Analysis of the Soybean Genome for 18S rRNA Sequences

It has been reported that the haploid soybean genome contains approximately 800 copies of the rDNA repeat (Friedrich et al. 1979). Does the 18S rRNA gene contained in the repeat unit isolated with the λSR1 recombinant phage and described in this paper represent a typical 18S rRNA gene in the soybean genome, or are there other, divergent DNA repeat units? Due to the conserved nature of the 18S rRNA gene itself, we felt that a subfragment from the gene could be used as a probe for all the 18S rRNA genes in the genome. Figure 4 presents the genomic hybridization data. The 1.05-kb BstI–EcoRI fragment used as a probe is an internal portion of the 18S rRNA gene. The lengths of the observed fragments are 3.9 kb for soybean genomic DNA digested with EcoRI and 2.5 kb for digestions with BstI. These two genomic bands are consistent with those observed in the λSR1 clone. No other genomic fragments encoding the soybean 18S rRNA gene are detected. On digestion with HindIII the genomic DNA fragments are longer than 23 kb. These data

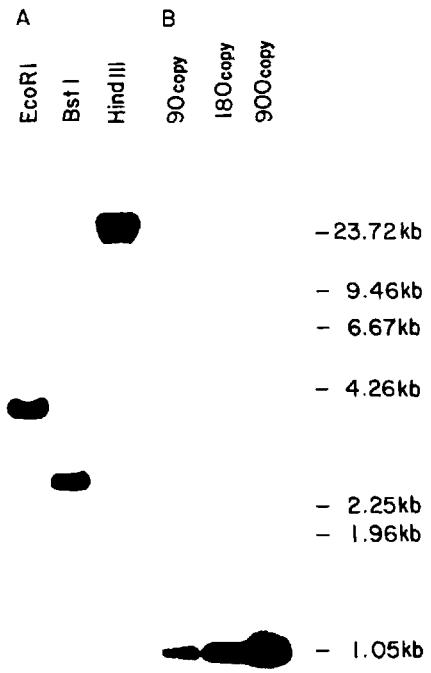


Fig. 4A, B. Analysis of soybean genomic DNA for 18S rRNA sequences. The **A** lanes contain 5 μg soybean DNA digested with EcoRI, BstI, and HindIII, respectively. The **B** lanes represent copy-number reconstructions and contain pSR1.2B3 plasmid DNA digested with BstI and EcoRI to yield the 1.05-kb insert fragment containing an internal portion of the soybean 18S rRNA gene described in this paper: 2 ng plasmid or 90 copies, 4 ng plasmid or 180 copies, and 20 ng or 900 copies, respectively. Based on a haploid genome size of 1.2×10^9 bp for soybean, each copy of the 18S gene would represent approximately 7 pg of the 5 μg of DNA loaded

are consistent with there being no observed HindIII sites in the typical 7.7-kb soybean rDNA repeat. Varsanyi-Breiner et al. (1979) and Jackson and Lark (1982) also observed no HindIII sites in soybean genomic rDNA, although Friedrich et al. (1979) reported the existence at one HindIII per soybean rDNA repeat unit.

It can be seen from this experiment that the 18S rRNA genes are present in about 500–800 copies per haploid genome, as reported by Friedrich et al. (1979). We conclude that the clone λSR1 examined in this paper is representative of the multigene family encoding soybean rRNA.

Discussion

Secondary Structure

The high degree of sequence conservation within the eukaryotic 18S rRNAs implies a high degree of similarity of potential secondary structure (Noller and Woese 1981; Stiegler et al. 1981; Zwieb et al. 1981). To provide a perspective on the types of sequence homology in small-subunit RNAs, we present in Fig. 5 a rough proposal for the secondary structure of the soybean 18S rRNA, constructed by using as a guide the proposed secondary structures for the yeast and *Xenopus* 18S rRNAs (Zwieb et al. 1981).

Despite the limited sequence homology between the eukaryotic 18S rRNA class and the prokaryotic 16S rRNA class [e.g., the 53% homology between the soybean 18S and the *E. coli* 16S rRNAs (Fig. 2)], there is a great deal of secondary-structure similarity between the small-subunit rRNAs of the two different classes (Küntzel and Köchel 1981; Stiegler et al. 1981; Zwieb et al. 1981).

Description of the Model

The model of the soybean 18S rRNA presented in Fig. 5 contains 53% of its nucleotides in base-paired configurations. This is comparable to the percentage proposed by Zwieb et al. (1981) for both the yeast and *Xenopus* 18S rRNA secondary structures. Furthermore, there are no extended perfect duplexes in the proposed secondary structure of soybean 18S rRNA. The longest perfect duplex is 14 bp long (nucleotides 117–130 pair with nucleotides 203–210). This, too, is comparable in length to the longest proposed perfect duplex of 13 bp in yeast 18S rRNA (Zwieb et al. 1981) and the longest proposed perfect duplex in the *E. coli* 16S rRNA of 12 bp (Noller and Woese 1981; Zwieb et al. 1981). Noller and Woese (1981) have suggested that this lack of extended perfect duplexes in small-subunit rRNAs provides them with increased structural flexibility when packaged in the small ribosomal subunit.

More significantly, examination of Fig. 5 reveals that the proposed secondary structure of soybean 18S rRNA can be geometrically divided into four structural domains, I–IV. These are comparable to the four structural domains in Stiegler et al.'s (1981) proposed generalized secondary structure for small-subunit rRNAs. These structural domains are believed to be related to the four functional domains determined by Herr et al. (1979) from biochemical data prior to the development of reliable secondary-structure models for *E. coli* 16S rRNA. Herr et al. (1979) suggested that domain I functions in the structural organization and assembly of the small ribosomal subunit, that domain II functions in the contact of the small ribosomal subunit with the large ribosomal subunit, that domain III lines the pocket created at the interaction with the large ribosomal subunit, and that domain IV functions in the interaction with the large ribosomal unit and plays a key role in the initiation of protein synthesis.

The four domains are separated from each other and stabilized by a core of five central duplex regions (regions 1, 2, 5, 7 and 8 in Figs. 2 and 5). Then

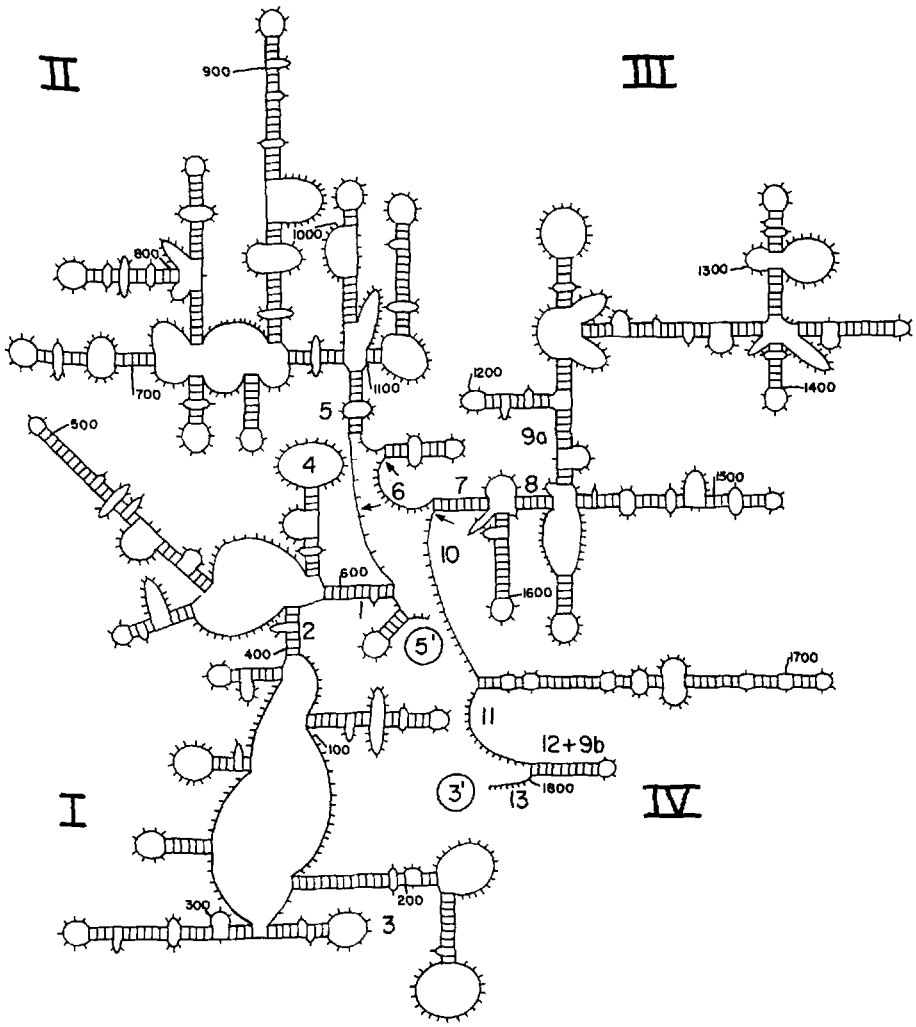


Fig. 5. Model of a potential secondary structure for soybean 18S rRNA. This model was based directly on the secondary structures proposed for the 18S rRNAs from yeast and *Xenopus* by Zwieb et al. (1981). Nucleotide positions are numbered in intervals of 100 nucleotides, and the 5′ and 3′ ends are indicated. The four domains of the molecule (Stiegler et al. 1981) are indicated by roman numerals: I (bases 1–611), II (bases 612–1144), III (bases 1145–1633), and IV (bases 1634–1807). Junctions between domains are indicated by arrows. Regions discussed in the text are indicated by Arabic numerals. The four domains and the numbered regions can also be correlated with Fig. 2

nucleotides that form the antiparallel strands of these central helices are also indicated in Fig. 2. It should be noted that the regions that form the comparable helices in the *E. coli* 16S rRNA secondary structure occur in comparable, if not identical, positions in Fig. 2, despite the limited sequence homology between the soybean 18S and *E. coli* 16S rRNAs. The antiparallel strands of each of these central duplexes are separated from each other by more than 300 bases of sequence. Duplex formation between such distant sequences has been defined as long-range interaction (Noller and Woese 1981) and is a hallmark of the proposed secondary structures for rRNAs.

Suggested Functions of Selected Conserved Regions of the Small-Subunit rRNA Molecule

In the eukaryotic 18S rRNAs, the sequences forming the central core of duplexes composed of nucleotides involved in long-range interactions (regions 1, 2, 7 and 8, in Figs. 2 and 5) all occur within more extensive regions of conserved sequence. The sequences forming these central duplexes have an average of greater than 90% sequence homology. It is likely that the conservation of these regions maintains the proposed long-range interactions.

It has been established that the rRNA molecule does indeed play active roles in protein synthesis (Noller and Woese 1981). These roles include mRNA recognition and binding, tRNA binding and codon–anticodon recognition, and subunit recognition (Noller and Woese 1981). Brimacombe (1982) has suggested that the small-subunit rRNA is also involved in elongation and translation. We will now discuss two of these functions in relation to the soybean 18S sequence and its proposed secondary structure.

mRNA Recognition and Binding. The 3′ end of small-subunit rRNAs has been implicated in mRNA recognition by the ribosome. In view of the high degree of sequence homology among the eukaryotic and *E. coli* small-subunit rRNAs (Fig. 2), it would

not be surprising if a significant function was performed by this region. Shine and Dalgarno (1974) proposed that a specific sequence (CCUCC) at the 3′ end of *E. coli* 16S rRNA is an mRNA recognition site [*E. coli* nucleotides 1534–1538, in region 13 (Figs. 2 and 5)]. Steitz and Jakes (1975) have shown that there is a sequence homologous to the Shine–Dalgarno sequence at the 5′ ends of the majority of *E. coli* mRNAs. However, the Shine–Dalgarno sequence is missing in all the eukaryotic 18S rRNAs we have examined (Fig. 2, between the soybean bases 1804 and 1805). The mechanism of mRNA recognition in eukaryotes may rely on another specific sequence at the 3′ end of 18S rRNAs. This theory has been tested (Kozak 1983) by comparing the sequence at the 3′ end of 18S rRNAs with that at the 5′ end of eukaryotic mRNAs. To date, no consistent conclusions have been drawn from these data. Noller (1980) has proposed that there may be switches between two alternate base-pairing schemes that permit the binding of specific molecules at specific stages during protein synthesis. Based on psoralen cross-linking studies, Thompson and Hearst (1983) have suggested that a specific switch between two alternate base-paired conformations, one consisting of region 12 paired as shown, and the other of region 9a paired with region 9b (Figs. 2 and 5), permits mRNA recognition by *E. coli* 16S rRNA. In view of the high degree of conservation between *E. coli* and the eukaryotic small-subunit rRNAs in these regions, both in sequence and in potential secondary structure, this interaction may also take place in eukaryotic 18S rRNAs.

tRNA Binding and Codon–Anticodon Recognition. Ofengand et al. (1982) have demonstrated that tRNA in the P site of a preinitiation complex including mRNA, tRNA, and the small ribosomal subunit covalently binds a nucleotide in a conserved region corresponding to region 10 (soybean nucleotide 1642). This happens with both *E. coli* and yeast small ribosomal subunits. In view of the high degree of nucleotide conservation of this region, it is possible that it plays the role in tRNA binding suggested by Ofengand et al. in all small-subunit rRNAs.

Summary

It can be seen from our data that soybean 18S rRNA is typical of small-subunit rRNAs. Considering the degree of sequence conservation, potential conservation of secondary structure, and conservation of function, it is likely that the major aspects of ribosomal function in higher plants are very similar to those in other eukaryotes. The data presented here showing partial conservation of sequence between soybean and *E. coli* small-subunit rRNAs and a high degree of potential secondary-structure homology suggest that many of the extensive genetic and biochemical studies on prokaryotic ribosomes are applicable to eukaryotic ribosomes. Further examination of plant rRNA structures should be helpful in this regard.

During the preparation and review of this manuscript the complete sequences of two eukaryotic small-subunit rRNAs were published [rabbit (Connaughton et al. 1984) and *Dictyostelium* (McCarroll et al. 1983)]. The rabbit 18S rRNA sequence is quite similar to the rat sequence we have examined. A detailed comparison of these new sequences to soybean has not yet been done.

Acknowledgments. We wish to acknowledge the continuing interest and encouragement of D.M. Shah, R.C. Hightower, R.T. Nagao, and J.L. Key. We would specifically like to thank R.C. Hightower for her help with the genomic blot and B. Rutledge for her assistance with the manuscript. This research was done in partial fulfillment of the requirements for a Ph.D. by V.K.E. It was supported in part by the Genetics Mechanisms section of USDA-SEA grant No. 83CRCR-1-1254 and grant No. S901-0410-8-0068-0, and by the U.S. Army Research Office under Training Grant No. DAA G29-83-G-0111.

References

Bolivar F, Rodriguez RL, Greene PJ, Betlach MC, Heyneker HL, Boyer HW (1977) Construction and characterization of new cloning vehicles: II. A multipurpose cloning system. Gene 2:95–113

Bolivar F, Backman K (1979) Plasmids of *E. coli* as cloning vectors. Methods Enzymol 68:245–267

Brimacombe R (1982) The secondary structure of ribosomal RNA, and its organization within the ribosomal subunits. Biochem Soc Symp 47:49–60

Brown AHD, Clegg MT (1983) Analysis of variation in related DNA sequences. In: Weir BS (ed) Statistical analysis of DNA sequence data. Marcel Dekker, New York, pp 107–132

Carreira LH, Carlton BC, Bobbio SM, Nagao RT, Meagher RB (1980) Construction and application of a modified "Gene Machine": A circular concentrating preparative gel electrophoresis device employing discontinuous elution. Anal Biochem 106:455–468

Chao S, Sederoff RR, Levings CS III (1984) Nucleotide sequence and evolution of the 18S ribosomal RNA gene in maize mitochondria. Nucleic Acids Res 12:6629–6644

Connaughton JF, Rairkar A, Lockard RE, Kumar A (1984) Primary structure of rabbit 18S ribosomal RNA determined by direct RNA sequence analysis. Nucleic Acids Res 12:4731–4745

Denhardt D (1966) A membrane-filter technique for the detection of complementary DNA. Biochem Biophys Res Commun 23:641–646

Eckenrode VK (1983) Sequence analysis of an 18S rRNA gene and physical mapping of an rDNA repeat unit from soybean. Ph.D. dissertation, University of Georgia, Athens, Georgia

Elleman TC (1978) A method for detecting distant evolutionary relationships between protein or nucleic acid sequences in the presence of deletions and insertions. J Mol Evol 11:143–161

Eperon IC, Anderson S, Nierlich DP (1980) Distinctive sequence of human mitochondrial ribosomal RNA genes. Nature 286:460–467

Erdmann VA, Huysmens E, Vandenberghe A, De Wachter R

(1983) Collection of published 5S and 5.8S ribosomal RNA sequences. Nucleic Acids Res 11:r105–r133

Erickson BW, Sellers PH (1983) Recognition of patterns in genetics sequences. In: Sankoff D, Kruskal JB (eds) Time warps, string edits, and macromolecules. Addison-Wesley, Reading, Massachusetts, pp 55–91

Friedrich H, Hemleben V, Meagher RB, Key JL (1979) Purification and restriction endonuclease mapping of soybean 18S and 25S ribosomal RNA genes. Planta 146:467–473

Hartigan JA (1975) Clustering algorithms. John Wiley & Sons, New York, pp 191–215

Herr W, Chapman NM, Noller HF (1979) Mechanism of ribosomal subunit association: discrimination of specific sites in 16S RNA essential for association activity. J Mol Biol 130: 433–449

Jackson PJ, Lark KG (1982) Ribosomal RNA synthesis in soybean suspension cultures growing in different media. Plant Physiol 69:234–239

Knuth D (1981) The art of computer programming. In: Seminumerical algorithms, 2nd ed vol 2. Addison-Wesley, Reading, Massachusetts, pp 1–77

Kozak M (1983) Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol Rev 47: 1–45

Krayev AS, Kramerov DA, Skryabin KG, Ryskov AP, Bayev AA, Georgiev GP (1980) The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse foldback RNA. Nucleic Acids Res 8:1201–1215

Kruskal JB (1983) An overview of sequence comparison. In: Sankoff D, Kruskal JB (eds) Time warps, string edits, and macromolecules. Addison-Wesley, Reading, Massachusetts, pp 1–40

Küntzel H, Köchel HG (1981) Evolution of rRNA and origin of mitochondria. Nature 293:751–755

Kushner SR (1978) An improved method for transformation of *Escherichia coli* with colE1 derived plasmids. In: Boyer HW, Nicosia S (eds) Genetic engineering. Elsevier/North Holland, New York, pp 17–23

Lake JA (1983) Evolving ribosome structure: domains in archaebacteria, eubacteria, and eucaryotes. Cell 33:318–319

Liljas A (1982) Structural studies of ribosomes. Prog Biophys Mol Biol 40:161–228

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol 65: 499–560

McCarroll R, Olfen GJ, Stahl YD, Woese CR, Sogin ML (1983) Nucleotide sequence of the *Dictyostelium discoideum* small-subunit ribosomal ribonucleic acid inferred from the gene sequence: evolutionary implications. Biochemistry 22:5858–5868

Meagher RB, Tait RC, Betlach M, Boyer HW (1977a) Protein expression in *E. coli* minicells by recombinant plasmids. Cell 10:521–536

Meagher RB, Shepherd RJ, Boyer HW (1977b) The structure of cauliflower moisaic virus: I. A restriction endonuclease map of cauliflower mosaic virus DNA. Virology 80:362–375

Messing J, Carlson J, Hagen G, Rubenstein I, Oleson A (1984) Cloning and sequencing of the ribosomal RNA genes in maize: the 17S region. DNA 3:31–40

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino-acid sequence of two proteins. J Mol Biol 48:443–453

Noller HF (1980) Structure and topography of ribosomal RNA. In: Chambliss G, Craven GR, Davies J, Davis K, Kahan L, Nomura M (eds) Ribosomes: structure, function and genetics. University Park Press, Baltimore, Maryland, pp 3–22

Noller HF, Woese CR (1981) Secondary structure of 16S ribosomal RNA. Science 212:403–411

Ofengand J, Gornicki P, Chakraburtty K, Nurse K (1982) Functional conservation near the 3′ end of eukaryotic small subunit RNA: photochemical crosslinking of P site-bound acetylvalyl-tRNA to 18S RNA of yeast ribosomes. Proc Natl Acad Sci USA 79:2817–2821

Rubstov PM, Musakhonov MM, Zakharyev VM, Krayev AS, Skryabin KG, Bayev AA (1980) The structure of the yeast ribosomal RNA genes: I. The complete nucleotide sequence of the 18S ribosomal RNA gene from *Saccharomyces cerevisiae*. Nucleic Acids Res 8:5779–5794

Salim M, Maden BEH (1981) Nucleotide sequence of *Xenopus laevis* 18S ribosomal RNA inferred from the gene sequence. Nature 291:205–208

Sankoff D, Cedergren RJ (1973) A test for nucleotide sequence homology. J Mol Biol 77:159–164

Sankoff D, Cedergren RJ, Lapalme G (1976) Frequency of insertion–deletion, transversion, and transition in the evolution of 5S ribosomal RNA. J Mol Evol 7:133–149

Shah DM, Hightower RC, Meagher RB (1982) Complete nucleotide sequence of a soybean actin gene. Proc Natl Acad Sci USA 79:1022–1026

Shine J, Dalgarno L (1974) The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc Natl Acad Sci USA 71:1342–1346

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517

Steitz JA, Jakes K (1975) How ribosomes select initiator regions in mRNA: base pair formation between the 3′ terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*. Proc Natl Acad Sci USA 72:4734–4738

Stiegler P, Carbon P, Ebel J-P, Ehresmann C (1981) A general secondary-structure model for prokaryotic and eucaryotic RNAs of the small ribosomal subunits. Eur J Biochem 120: 487–495

Thompson JF, Hearst JE (1983) Structure–function relations in *E. coli* 16S RNA. Cell 33:19–24

Torczynski R, Bollon AP, Fuke M (1983) The complete nucleotide sequence of the rat 18S ribosomal RNA gene and its comparison with the respective yeast and frog genes. Nucleic Acids Res 11:4879–4890

Varsanyi-Breiner A, Gusella JF, Keys C, Honsman DE, Sullivan D, Brisson N, Verma DPS (1979) The organization of a nuclear DNA sequence from a higher plant: molecular cloning and characterization of soybean ribosomal DNA. Gene 7: 317–334

Wool IG (1980) The structure and function of eucaryotic ribosomes. In: Chambliss G, Craven GR, Davies J, Davis K, Kahan L, Nomura M (eds) Ribosomes: structure, function and genetics. University Park Press, Baltimore, Maryland, pp 797–824

Zwieb C, Glotz C, Brimacombe R (1981) Secondary structure comparisons between small subunit ribosomal RNA molecules from six different species. Nucleic Acids Res 9:3621–3640

Received July 23, 1984/Revised December 4, 1984