MICROBIOLOGY REVIEWS, Dec. 1983, p. 621–669
0146-0749/83/040621-49$02.00/0

Vol. 47, No. 4

Detailed Analysis of the Higher-Order Structure of 16S-Like Ribosomal Ribonucleic Acids

CARL R. WOESE,[1]* ROBIN GUTELL,[2] RAMESH GUPTA,[1] AND HARRY F. NOLLER[2]

Department of Genetics and Development, University of Illinois, Urbana, Illinois 61801,[1] and Thimann Laboratories, University of California at Santa Cruz, Santa Cruz, California 95064[2]

INTRODUCTION

Translation physically links the genotype and the phenotype and thereby defines them. The process is a necessary precondition to the evolution of all macromolecular structure. The size of enzymes, their complexity, and their specificity reflect an underlying accuracy in the translation process. Is it not true, therefore, that translation embodies the essence of the cell?

Translation occurs in the framework of the ribosome, an enormous and complex molecular aggregate that comprises two ribonucleoprotein subunits (12). The smaller of these (in bacteria) contains about 20 separate protein components, positioned around a large ribonucleic acid (RNA) some 1,500 nucleotides in length. The larger subunit comprises an RNA of about twice that size (2,900 nucleotides), in addition to a much smaller 5S RNA (120 nucleotides) and about 30 distinct proteins (26, 40).

For almost three decades, biologists have sought to understand the molecular mechanics of translation, but with little beyond descriptive success. In approaching translation, we have tended to focus on the protein components as the elements that define ribosome function—the ribosomal RNAs (rRNAs) were seen as basically structural. With the advent of nucleic acid sequencing technology, however, has come an interest in the possible functional roles for rRNAs. If we can transform their linear molecular sequence into a detailed and dynamic three-dimensional form, we will almost certainly understand the all-important translation process.

The earliest speculations concerning rRNA secondary structure were based upon a partial (and, it turns out, incorrect) sequence from one organism (20) and therefore need not be considered here. Studies of transfer RNA (tRNA) and of 5S RNA (25) have forcefully demonstrated that the only reliable way to determine secondary structure presently available (outside of X-ray crystallography) is through comparative analysis of primary structure. A comparative approach to the 16S rRNA sequence, based upon the *Escherichia coli* sequence (9, 11), the nearly complete 16S rRNA sequence from *Bacillus brevis*, and T_1 ribonuclease (RNAse) oligonucleotide catalogs from over 150 other bacteria, gave us a first look at the true secondary structure of the molecule (51, 98). The subsequent publication of 16S RNA sequences from *Zea mays* chloroplasts (69) and mammalian mitochondria (1, 21) permitted some refinement of the original structure and provided further comparative proof for a number of helical elements (54).

There now exist in the literature a number of proposed secondary structures for 16S (and 23S) rRNA (6–8, 30, 54, 79, 98, 105). All agree to a first approximation (for all have used to some extent a comparative approach). However, there are differences in detail among them. Since it is virtually impossible for interested biologists to assess the relative merits of the various models—indeed, it even requires considerable effort to intercompare them—and since new sequences add complexity as well as information to the picture, it is of considerable value to review the topic of the constraints in rRNA sequence. The purpose of the present review are to bring up-to-date and discuss in detail the status of 16S rRNA secondary structure and to present an overview on this rapidly developing field.

The first part of the review concerns the evidence for the individual helical elements. In each case, extensive comparative evidence supporting the double-helical stalk will be given, and the structure will be further described in terms of the susceptibility of various residues therein to chemical modification and so on. Also, the entire locale will, whenever possible, be characterized in phylogenetic terms (conservation of sequence, patterns of variability, etc.).

The data base used herein contains the complete sequences for the 16S-like rRNAs from *E. coli* (9), *Z. mays* (69) and tobacco chloroplasts (86), and various mitochondria (1, 21, 39, 42, 74, 89; J. J. Seilhammer, G. M Olsen, and D. J. Cummings, unpublished data); the unpublished partial sequences from *B. brevis* (C. R. Woese and H. F. Noller, unpublished data) and *Bacillus stearothermophilus* (R. Gupta et al., unpublished data); the 16S rRNA sequence from the archaebacterium *Halobacterium volcanii* (33); the 18S rRNA sequences from *Saccharomyces cerevisiae* (64); *Xenopus laevis* (65), and *Dictyostelium discoideum* (R. McCarroll, G. J. Olsen, Y. D. Stahl, C. R. Woese, and M. L. Sogin, Biochemistry, in press) and the T_1 RNase catalogs for 16S rRNAs of over 200 organisms (mostly eubacteria) (24 and references cited therein; C. Woese et al., unpublished data). Also, the susceptibility of various residues to chemical modification has been measured by T_1 and pancreatic RNase cataloging assays, using bisulfite modification of C's (*E. coli* and *B. brevis* 16S rRNAs), glyoxal substitution of G's (*E. coli* 16S rRNA), *m*-chloroperbenzoic acid modification of A's (*E. coli* 16S rRNA) (C. R. Woese and L. J. Magrum, unpublished data), and kethoxal substitution of G's (in *E. coli* active [50] and inactive [36] 30S subunits and certain ribonucleoprotein fragments), all reagents that are known to respect secondary structure. Also used to some extent are the relative sensitivity of residues to enzymatic cleavage by T_1 and pancreatic nucleases (which would detect un-

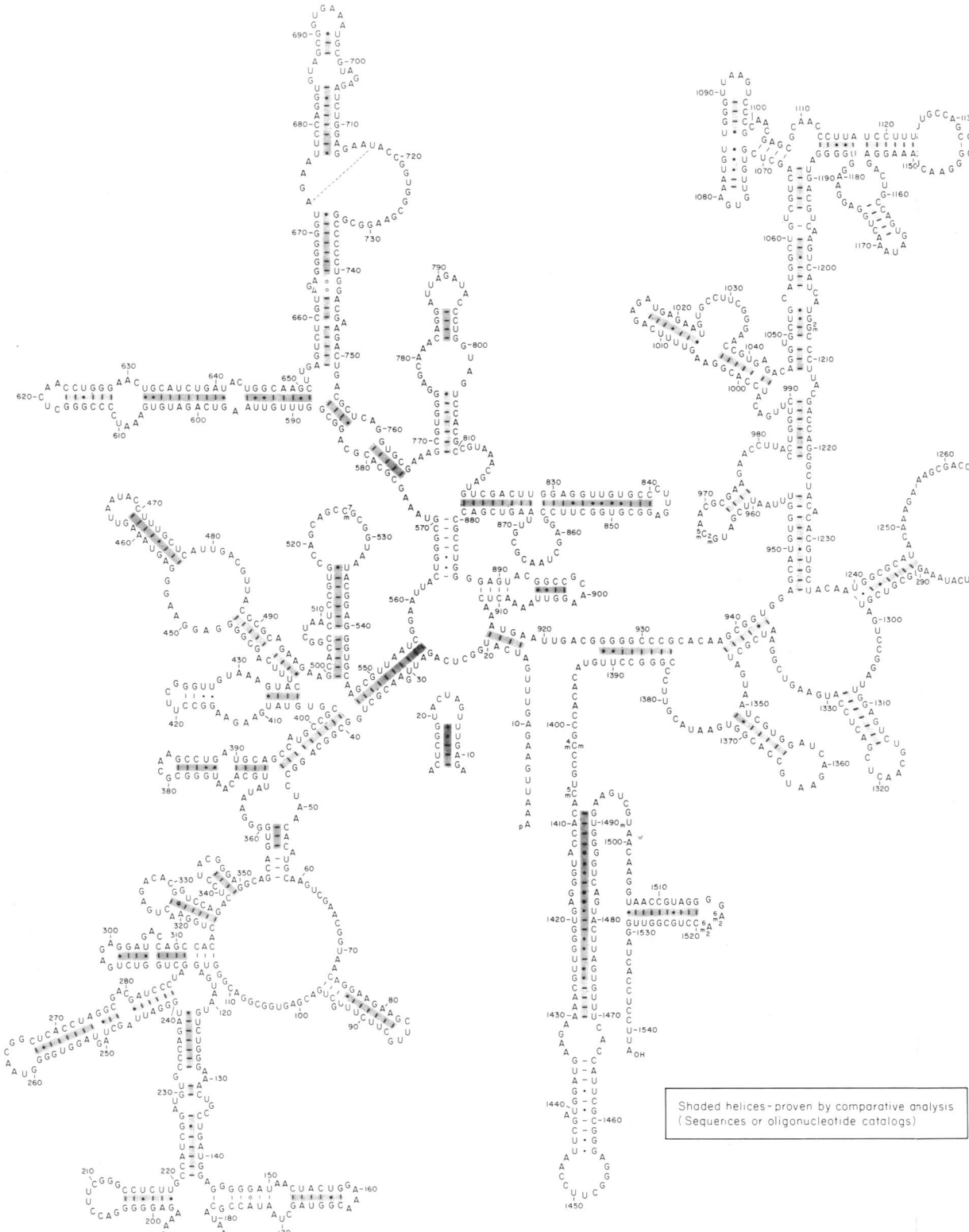


FIG. 1. Secondary structure of typical 16S-like rRNA, from *E. coli*. Numbering is according to that in reference 9. Canonical pairs are connected by lines, G · U pairs by dots, and suspected A-G pairs by open circles. Helices considered to be proven by comparative criteria are shaded.

structured residues) or by cobra venom nuclease (which attacks residues in double-helical conformation) acting on the 30S subunit (90, 91).

PART 1. DETAILED CONSIDERATION OF THE VARIOUS HELICAL ELEMENTS

In discussing 16S rRNA structure, we will consider a double-helical element as definitely existing only when it is supported by sufficient comparative data. Specifically, a putative helix is considered to exist when (i) it can be formed in at least two different 16S rRNAs (involving homologous segments in the molecule in all cases), and (ii) a canonical base pairing covariance can be demonstrated (i.e., a Watson-Crick pairing in one case being replaced by a different Watson-Crick pairing in some other case) for at least two pairs in the helix. Although it is not a necessary condition, bases in helical array should be relatively unreactive to chemical reagents that respect secondary structure and to nuclease attack, or sensitive to cobra venom endonuclease.

Figure 1 is an overview of eubacterial 16S rRNA secondary structure as it is now known, to which the reader can refer in the detailed discussion that follows. The molecule readily structures into several major domains, each comprising a number of helical elements and subdomains. The 5′ domain is defined by the helix linking positions 30 and 550, the central domain by that linking 565 and 885, the 3′ major domain by that linking 930 to 1390, whereas the 3′ minor domain comprises the sequence beyond the 1390 region. These domains and their subdomains are also to some extent defined in terms of certain of the ribosomal proteins, which stabilize various parts of the structure and so protect them from nuclease attack. For example, protein S4 in this way protects most of the 16S rRNA domain enclosed by the helix linking positions 30 and 550 (18, 19, 68, 102, 103) (Fig. 1). That this truly reflects structural organization of the RNA is evident from the fact that under carefully controlled conditions, the same domain is protected in the absence of protein (102, 103).

In the detailed discussion of the various helical regions, the numbering of residues will be that of the *E. coli* 16S rRNA sequence (i.e., that used for the *rrnB* operon from strain K) (8, 9).

In the analysis that follows, we will be using the following terms and symbols. (i) An apex loop is the (short) stretch of sequences that connects one chain of a double helix with the other, e.g., the anticodon loop in tRNA. (ii) A bulge loop comprises one or more (nonpaired) residues that protrude from one of the chains of an otherwise simple double-stranded helix. (iii) An interior loop can be considered adjacent bulge loops in opposite strands of a double helix. (iv) An inner helix is one whose sequence lies entirely within the loop defined by an outer helix. (v) An irregular helix is one that, in addition to canonical pairs and (a few) G · U pairs, contains single residue bulge loops, noncanonical "pairs," or excessive numbers of G · U pairs. (In the text, canonical pairs and G · U pairs are denoted by a dot [·], whereas noncanonical pairs are denoted by a hyphen [e.g., A-G].) (vi) A catalog is the set of oligonucleotides produced by complete digestion of a (16S ribosomal) RNA with RNase T_1. (vii) The three primary lines of descent (24, 97) are referred to as eubacteria (e.g., *E. coli*, *Bacillus* species, etc.), archaebacteria (e.g., methanogens, extreme halophiles), and eucaryotes. (viii) Post-transcriptionally modified nucleotides are designated by a superscript asterisk, e.g., Ȧ, when not otherwise identified. (ix) Residues

TABLE 1. Helix 9-13/21-25[a]

Organism/ organelle	Sequence
	10 20
E. coli	..... G A A [G A G U U] U G A U C• A U [G G C° U C•] A .....
H. volcanii	pA U [U C C G G U] U G A U C C U [G C C G G A] .....
X. laevis	pU A C [C U G G U] U G A U C C U [G C C A G] U .....
Human mitochondria	pA A U [A G G U U] U G G U C C U [A G C C U] U .....

[a] Secondary structural element(s) for region are boxed. (Breaks in boxes indicate either non-canonical pairs or a bulged base; see, for example, Table 3.) Symbols are as follows: superscript *, nucleotide is modified; superscript °, nucleotide is relatively resistant to chemical modification in free 16S rRNA; superscript •, nucleotide is relatively sensitive to chemical modification in free 16S rRNA (or in the case of a few G residues in the 30S subunit). Numbers indicate the position in the sequence. Chemical modification data involving C's (bisulfite), A's (*m*-chloroperbenzoic acid), and G's (glyoxal) are from the unpublished study of C. Woese and L. Magrum. References 10, 13, 35, 36, and 50 cover the interaction of G's in the 30S subunit with kethoxal. (The information in this footnote applies to Tables 1 through 38.)

TABLE 2. Helix 17-20/915-918

Organism/organelle	Sequence
Sequence	15 25 910 920
E. coli	U U G A U C A U G G C U C ..... C U C A A A U G A A U U G
Yeast mitochondria	A U G A U G U U G G U U C ..... C U C A A A A C A A U A G
Human mitochondria	U U G G G U C C U A G C C U ..... C U C A A A G G G A C C U G
Aspergillus mitochondria	G A G U U U G A U G A U G G C U C ..... C U G A A A U C A U U A G G A C C G
Catalog	
0.83 of eubacteria[a]	G A U C C U G ..... C U C A A A G G G A A U U G
0.13 of eubacteria	G A U C A U G ..... C U C A A A A U G A A U U G
0.027 of eubacteria	G A U U C U G ..... C U C A A A G A A A A U U G
0.006 of eubacteria	G A U C U U G ..... C U C A A A A A G A A U U G
1.00 of eucaryotes and 0.90 of archaebacteria	G A U C C U G ..... C U U A A A G G G A A U U G

[a] Protection data from *B. brevis*.

sensitive (resistant) to chemical modification are indicated by a closed (open) circle superscript or subscript. (x) The four normal nucleotides have their usual designations, whereas R, Y, and N refer respectively to purine, pyrimidine, and unspecified nucleotides.

Helix 9–13/21–25 (Table 1)

The first helix in the molecule (helix 9–13/21–25; Table 1) starts two to three residues from the 5′ end in archaebacteria and eucaryotes. It has the familiar form of the anticodon and TΨC arms in tRNA, i.e., a stalk of five to six pairs enclosing a loop of seven residues. Sequence in the structure is nearly constant for the three known eubacterial examples, and the occurrence of the T_1 oligonucleotides $AUC^{\bullet}_{\circ}UG_{21}$ and $CUCAG_{27}$ in almost all eubacterial catalogs strengthens this claim. In eucaryotes and archaebacteria, the sequence in the helix is different, but in each kingdom it appears to be nearly constant, as judged from complete sequences and catalog information; i.e., ...CYG_{11} covers all archaebacteria and eucaryotes so far characterized, as does $GAUCCUG_{21}$ (except for one archaebacterial example). Although the eubacterial examples contain five pairs, the archaebacterial examples and one eucaryotic example appear to extend to a sixth pair (distal to the loop).

Helix 17–20/915–918 (Table 2)

One of the strands of helix 17–20/915–918 (Table 2) exists within the loop of the previous helix, a situation somewhat analogous to the codon-anticodon interaction within the anticodon loop. This is the only documented example of such a structure in the rRNAs thus far. Whether this and the previous helix coexist in the ribosome is not known. Both sides of the helix are covered by T_1 oligonucleotides, and its four canonical pairs can be formed in 99% of catalogs. Variation in sequence in the helix occurs but rarely, and even so is highly constrained. (Three of the four eubacterial oligonucleotide examples shown in Table 2 are of multiple phylogenetic occurrence.) In human mitochondria, the helix may extend to 6 pairs, i.e., 15–20/915–920, and in *Aspergillus* mitochondria to 10. These examples suggest that this and the previous helix do not simultaneously exist in 16S rRNA.

In *B. brevis*, CC_{19} is refractory to bisulfite.

Helix 27–37/547–556 (Table 3)

Helix 27–37/547–556 (Table 3) is the first example of an irregular helix. The structure occurs in all three primary kingdoms, It is somewhat variable in sequence. Examples of bulged residues (e.g., G_{31} in eubacteria) are common, as

TABLE 3. Helix 27-37/547-556

Organism/ organelle	Sequence
	30 ... 550
E. coli	A G A U U G A A C G C U G ..... A A G C G U U A A U C G
Paramecium mitochondria	U G A U U U A A C G C U A ..... U A G C G U U A A U C G
Yeast mitochondria	A G A U U A A G C G C U A ..... G A G C G U U A A U C A
Human mitochondria	U U C U A — U U A G C U C ..... A A G U C A A U A G A A
Chloroplasts	A G G A U G A A C G C U G ..... A A G C G U U A U C C G
H. volcanii	A G G U C — A U U G C U A ..... A A G U G A U G A C C G
X. laevis	U A G C A — U A U G C U U ..... U A G C G U A U A U U A

are A-C "pairs" (e.g., mitochondria and some eucaryotes) and modified residues (in eucaryotes). The archaebacterial version of the helix is not irregular.

The structure is one of several protected from nuclease attack by ribosomal protein S4 (18, 19, 68, 103), and electron microscopic evidence suggests that this may be a major binding site for this protein (14). Chemical reactivity in and around the helix is consistent with this structure. Residues 557 and 558 are reactive with glyoxal, as is bulged G_{31} with kethoxal (50). However, residues 553 and 554 are relatively unreactive, as would be predicted. C_{556}, in the terminal base pair of the helix, reacts with bisulfite, which is consistent with the mechanism of bisulfite attack (3).

Eucaryotes modify A_{33} (2′ O-methyl [2′ OMe]); *X. laevis* modifies G_{548} (2′OMe) (65). U_{553} (2′OMe) is modified in the *Lemna* catalog (unpublished data), whereas *S. cerevisiae* modifies A_{557}. This density of modification is remarkable, as is its variability (among eucaryotes) and, indeed, its occurrence within a helix in the first place.

The eucaryotic version of the structure appears to be one pair longer than the bacterial versions, i.e., the pair at positions 26 and 557. However, an A-G "pair" at these positions is possible in *E. coli* and other organisms. (See discussion of A-G pairs below.)

The unpaired sequence beyond the helix on the 3′ side is quite constant. The generalized oligonucleotide $GA^{\Psi}_{A}UUAYUG_{566}$ covers 96% of the eubacterial catalogs.

Helix 39–47/394–403 (Table 4)

The eubacterial and archaebacterial structures of helix 39–47/394–403 (Table 4) are the same, with the exception of a bulged A_{397} residue present only in the eubacterial (and mitochondrial) examples. (Note that the archaebacterial version could also be made to conform strictly to its eubacterial counterpart, i.e., with a bulged A_{397}, without any loss of pairing.) Although the 5′ strand in eucaryotes seems analogous to the procaryotic examples, a complementary (3′) strand has not been located with certainty. The possibility shown in Table 4 is from a nonhomologous section of the sequence (covering position 520 in yeast 18S numbers). Therefore, the eucaryotic version of this structure cannot be

TABLE 4. Helix 39-47/394-403

Organism/organelle	Sequence
	40 ... 400
E. coli	G G C G G C A G G C C ..... G C C A U G C C G C G
Chloroplasts	G G C G G C A U G C U ..... G C A A U G C C G C G
Human mitochondria	C U U A G U A A G A U ..... G C U A U A C U A A C
Yeast mitochondria	A A A U A A G G A C A ..... G U U A C U U A U U A
H. volcanii	A U U G G G G U C C G ..... G G G A C C C C A A G
X. laevis	U G U C U C A A A G A ..... G A A U G A G U A C A

TABLE 5. Helix 52-58/354-359

Organism/organelle	Sequence
	50 60 360
E. coli	C C U A A C A C A U G C A A G ..... A G C A G U G G G
B. brevis	C C U A A U A C A U G C A A ..... A G C A G Y A G G
Human mitochondria	A U U — A C A C A U G C A A G ..... A G C A G U G A U
X. laevis	A U U A A G C C A U G C A C G ..... A G C A G G C G C
H. volcanii	A U U U A G C C A U G C U A G ..... A G C A G G C G C

considered proven. A possible (unproven) alternative is the pairing of (U)UGUC$_{41}$ with GACR(A)$_{400}$, producing a shortened version of the helix seen in procaryotes. Sequence is nearly constant among the eubacterial examples. Catalogs show it to be at least somewhat variable in the archaebacteria.

C's at position 47 and 48 are unreactive with bisulfite.

Helix 52–58/354–359 (Table 5)

Because of its near constancy of sequence, helix 52–58/354–359 (Table 5) is difficult to demonstrate by comparative evidence. The three pairs beyond bulged A$_{55}$ are not considered proven for this reason. Indeed the entire sequence from position 48 to 68 is rather conserved. For example, only one base replacement separates the *E. coli* and chloroplast versions, and 92% of the eubacterial catalogs are covered by the general sequence GCYU$^{U}_{A}$AYACAUG$_{57}$. Even the eucaryotic and archaebacterial versions are remarkably like their *E. coli* counterpart. The general formula AYUNA$^{G}_{C}$CCAUG$_{57}$ covers all sequences and catalogs for both kingdoms. The mitochondrial examples show considerable sequence variety in the region, and the distal three pairs in the helix are not formed in several such examples.

C$_{54}$ is unreactive with bisulfite. C$_{52}$ is reactive, as might be expected, since it is a terminal pair.

Helix 73–82/87–97 (Table 6)

Structure in region 60 to 110 is not completely clear and is to some extent idiosyncratic. *E. coli* and *Proteus vulgaris* have a helix, 73–82/87–97 (Table 6), which includes a bulged G$_{94}$, which appears proven except for the three pairs outside the bulge. Also, residues 71 to 81 are known to be unreactive with modifying reagents. G$_{86}$ and G$_{94}$ (in the apex loop and bulge loop, respectively) are reactive with kethoxal in active 30S subunits, however (50). Cobra venom RNase cuts after positions 72, 74, 78, 89, and 95 (90). Chloroplasts delete most of this structure, and the archaebacteria and eucaryotes appear to as well. However, archaebacteria may have a somewhat different, seven-pair helix in the area (65–71/98–104), which may have a counterpart in *D. discoideum*, but we consider these as yet unproven.

Helix 113–115/312–314

The small helix 113–115/312–314 occurs in two versions: GUG$_{115}$/CAC$_{314}$ in eubacteria and all of the mitochondria, and CUC$_{115}$/GAG$_{314}$ in archaebacteria and eucaryotes. The sequences are in entirely analogous positions in all cases, i.e., they are defined by surrounding homologous sequence or secondary structure or both. Conservation of sequence in the helix is further implied by the eubacterial catalogs, 92% of which contain an oligonucleotide of the form . . . *YCAC*AYUG$_{318}$. In yeast mitochondria, the pairing can be extended, i.e., *AAC*GUG$_{115}$/CAC*GUU*$_{317}$, creating two adjacent A · U pairs that ostensibly conflict with another helix (see Table 12); however, see discussion of coaxial helices in part 2 below.

TABLE 6. Helix 73-82/87-97

Organism/organelle	Sequence
	80
E. coli B	A A C A G G A A G C A G C U U G C U G C U U U G C U G
E. coli K-12	A A C A G G A A G A A G C U U G C U U C U U U G C U G
Proteus vulgaris	A A C A G G A G A A A G C U U G C U U U C U U G C U G

TABLE 7. Helices 122-128/233-239 and 136-142/221-227

Organism	Sequence
Sequence	130 220 230
E. coli	U A A U - G U C U G G G A A C U G C C U G A U G G A G . . . . U U G C C C A U G G A U U -
P. vulgaris	U A A U - G U A U G G G A U C U G - C C G A U A G A G . . . . G C G C U A U C G G A U G A A C C C A U A U -
H. volcanii	U A A C A C G U G G C C A A A C U A C C U A C A G A G . . . . G C G C U G U A G G A U G U - G G C U G C G G
D. discoideum	U A C A A C A G U G A U A A A C U A A U A G A C U U U C . . . . . G A C A A G U C U A C U G U - G U C A C U G -
B. brevis	U A A C A C G U A G G C A A C C U G C C U C U C A G A C . . . . . U C A C U G G A G A U G G G C C U G C N
Catalog	
Staphylococcus epidermidis	G A U A A C C U A C C U A A U A A G U C A C U U A U A G
Bacillus alvei	G U A A C C U G C C C A U A A G U C A C U U A U G
Bacillus polymyxa	G C A A C C U G C C C A C A A G U C A C U U G
Bacillus sphaericus	G A A C C U A C C C U A U A G G C U A U A G

Helices 122–128/233–239 and 136–142/221–227 (Table 7)

Helices 122–128/233–239 and 136–142/221–227 (Table 7) occur in all three kingdoms. However, in all eucaryotes except for *Dictyostelium*, the structures are irregular, especially the inner helix. (A slight irregularity is noted also in the two chloroplast examples of the outer helix, i.e., the pairing A_{126}-C_{235}.)

The interior loop defined by the two helices appears to be structured, as Table 7 shows; a canonical pairing covariance exists for positions 131 versus 231. Note also the $(A,G)_{129}$ versus $(G,A)_{232}$ covariance, suggesting additional structure. Sequence in the interior loop seems constrained, as does that preceding the helices (i.e., positions 116 to 121). However, within the helices themselves, some positions are readily variable. (Note that the catalog examples and one sequence are all taken from within the same [phylogenetically defined] genus.)

Many residues within each of the helices have been shown to be resistant to chemical modification, whereas the flanking sequences, at positions 119 or 120, 129 to 131, 132, and 134, are reactive with the modifying reagents.

In the mitochondria, only the fungal and protist examples clearly demonstrate both helices. For future reference, note that the archaebacterial sequence inserts an A residue after position 121 and a G after position 239.

Helices 144–147/175–178 and 153–158/163–168 (Table 8)

Helix 144–147/175–178 (Table 8) is not convincingly present in *E. coli* or most eucaryotes; however, it does form in *B. brevis*, chloroplasts, *Aspergillus* mitochondria, and the archaebacterium *H. volcanii*. Its extension by two pairs plus a G-A juxtaposition, i.e., GAU_{150}/AUA_{174}, is possible in all but chloroplasts, but remains unproven.

The second helix, 153–158/163–168 (Table 8), is quite variable in sequence except for the terminal C · G and G · C pairs; in the genus *Bacillus* alone at least six versions exist. Variation may be under (complex) constraint, however, for position 157 versus 164 is rarely a canonical pairing (it is U · G, G · U, or, in *D. discoideum*, A · C). Moreover, all eucaryotic sequences exhibit a $G \cdot G_{166}$ "pair." Sequence within the apex loop seems constant between eubacteria and archaebacteria, whereas most eucaryotes vary it somewhat and insert a residue therein.

All mitochondrial examples have an abbreviated form of this composit helix (and the mammalian versions delete it completely).

The resistance of residues 175 to 178 to chemi-

TABLE 8. Helices 144-147/175-178 and 153-158/163-168

Organism/organelle	Sequence
Sequence	150 170
E. coli	A G G G G A U A A A C U A C U G G A A A — C G G G U A G C U A A U A C C G C G A
Chloroplasts	A G G G G A A C A A C U A G G A A A — C G G G U U G C U A A U A C C C C G
H. volcanii	A G A A C G A U A A A C C U C G G G A A A — C U G A G G C U A A U A G U U C A
X. laevis	A C U U G G A U G A A C U G U G G U A A U U C U A G A G C U A A U A C A U G C
B. brevis	A C C G G G A U A A A C A U A G G G A A A — C U U A U G C U A A U A C C G G A
Catalog	
Lactobacillus plantarum	G A U A A A C A C C U G G A A A A — C A G N N G C U A A U A C C G
B. subtilis	G A U A A A C U C C G N G A A A A — C C G N N G C U A A U A C C G
B. alvei	G A U A A A C C C A C G G A A A A — C G N N N G C U A A U A C C G

cal modification is consistent with the first of these two helices. Residues in helix 153–158/163–168 are also protected against chemical modification, whereas those immediately flanking the helix, i.e., A's at positions 151 and 152, and C_{169}, are readily modified. The A residues in the apex loop are only moderately reactive (with *m*-chloroperbenzoic acid). Cobra venom nuclease cleaves after residues 155 and 156 of the inner helix (90).

The 180–220 Region (Table 9)

Structure in the 180 to 220 region (Table 9) is somewhat variable, as is the number of residues. The eucaryotic versions are the largest, all being over 100 residues in length. The version in *B. brevis* is also slightly larger (50 residues) than its *E. coli* counterpart. The two helices shown in Table 9 both have a reasonable amount of comparative data to support them. However, the first is not formed convincingly in *E. coli* (although the three-pair version indicated is possible), and the second is not formed convincingly in the archaebacterium *H. volcanii* (although, again, a three-pair version seems possible). *B. brevis* and all eucaryotes will form both, and the latter may have additional structure in the region not included in Table 9. The second helix is well defined by the sequence or structures (or both) surrounding it. In most cases, the helix is immediately preceded by a stretch of three A residues (or three purines in eucaryotes), and in all cases, the helix ends with position 219 (which can be defined not by this helix but by the following one [see Table 7] which begins at, and so defines, residue 221).

A number of positions in this region (184, 191, 193, 194, 214, 215, 217, 220) are known to be protected against chemical modification, whereas others (181, 182, 183, 196, 198, 204, 206, 207, 210) are reactive. The region is striking for its high ratio of purines to pyrimidines. *E. coli* contains a stretch of 11 contiguous purines, *B. brevis* 12, and chloroplasts 11 in approximately the same area (positions 190 to 205 in *E. coli*).

Helix 240–259/267–286 (Table 10)

Helix 240–259/267–286 (Table 10) has three sections separated by bulge loops. The two procaryotic versions resemble one another more than they do the eucaryotic version. The innermost helix (252–259/267–274) is highly constrained in sequence, as is the apex loop. The generalized oligonucleotide $CYYACC^{UA}_{AU}G_{275}$ covers at least 95% of eubacterial and the eucaryotic examples. Most archaebacterial examples are also covered by the same general sequence if a G_{273} possibility is included. The oligonucleotide $AUCCCUAG_{289}$, which would measure variability in the outermost helix, is not

TABLE 9. Region 180 to 220

Organism	Sequence
	180 200
E. coli	C A U A A C G U C G C - - - - - - - - - - - - - - - - - - A A G A C C A A A G A G G G G G A C C U U C G G G C C U C U U G
P. vulgaris	C A U G A C G U C U A - - - - - - - - - - - - - - - - - - C G G A C C A A A G C A G G G G U U C U U C G G A C C U U G C G
B. brevis	G A U A G G U U U U U G G A U U G - C G A U G A U C C G A A A A G A A A A G A U G G C - - - - U U C G - - G C U A U C A
H. volcanii	C A U A C G G G A G U C A U G C U G G A A U G C C G A C U C C C C G A A A A C G C - - - - - - - - - - - - - U C A G G C G
D. discoideum	A C A A G C G A U G G G U G A C U G G C A A C G G A A G C U C A G... A G A U C G A G G A - - - - U U U A U C U U C G A C

TABLE 10. Helix 240-259/267-286

Organism/ organelle	Sequence
	250 280
E. coli	G G G A U U A — G C U — — A G U A G G U G G G G U A A C G G C U C A C C U A G G C G A C G A U C C C U
Chloroplasts	C U G A U U A — G C U — — A G U U G G U G A G G C A A U A G C U U A C C A A G G C G A U G A U C A G U
H. volcanii	C C G A U U A — G G U — — A G A C G G U G G G G U A A C G G C C C A C C G U G C C G A U A A U C G G U
D. discoideum	C C C U A U C A A C U U U C G A U G G U A C G G U A U U G G C C U A C C A U G G U U G U A A C G G G U

conserved over any substantial phylogenetic distances, nor can relatives be recognized—implying a variable sequence.

The tested residues in the helices are not reactive to the modifying reagents, whereas four of the five tested residues in the apex loop are reactive, the exception is C_{264}. Cobra venom nuclease cuts after residue 272 in the innermost helix (90).

The outermost, but not the innermost, helix is retained in the RNA fragment protected from nuclease digestion by ribosomal protein S4 (18, 19, 68, 103). The latter helix (252–259/267–274) is protected, however, by the addition of protein S20 (18, 19, 102–104) and so appears to be (part of) the binding site for this protein; consistent with this is the observed photochemical cross-linking of S20 to this region of 16S RNA (18).

Helix 289–292/308–311 (Table 11)

The four-pair helix 289–292/308–311 shown boxed in Table 11 has ample comparative support. The region, however, has further, seemingly complex structure. All examples will form the helix of four pairs indicated by overlining (if an A-G pair is admitted in *Paramecium* mitochondria). The seven examples shown all have different sequences in this helix. However, another helix, which is an uninterrupted extension of the boxed helix (shown by underlining), also seems possible in four of the examples (*E. coli*, chloroplasts, *H. volcanii*, and *Paramecium* mitochondria). Note also that some of the bases that would be (exclusively) in this latter helix are relatively unreactive with the modifying reagents. Since this and the previous helix are mutually exclusive, we do not consider the comparative evidence sufficiently compelling to consider it (the smaller helix) proven. It is possible, of course, that these two helices each exist at different stages in the translation cycle, i.e., together they constitute a switch.

Sequence in the apex loop tends to be conserved (see *E. coli* versus *H. volcanii*), and the bases therein are strongly reactive with modifying reagents (unpublished experiments).

Helices 316–322/331–337 and 339–342/347–350 (Table 12)

Although its apex loop seems almost constant in sequence among the three kingdoms, there is some variation in sequence in helix 316–322/331–337 proper, enough to demonstrate its existence by sequence and oligonucleotide comparisons (Table 12). The eubacterial and eucaryotic versions of the structure each contain an A-G pair (at different positions), whereas their archaebacterial counterpart seems to add an eighth pair, $U_{323} \cdot A_{330}$. Residues flanking the helix and in the apex loop (seven) that can be tested are quite sensitive to chemical modification. The three or four tested C residues in the stalk were resistant. A cobra venom nuclease cut at position 337 is prevented upon 30S-50S subunit association (90).

Helix 339–342/347–350 and its apex loop are invariant in sequence among eubacteria and among archaebacteria, although the sequence in the helix proper is not the same in the two kingdoms. Their eucaryotic counterpart would contain an abnormal pairing A_{340}-A_{349}, except for *D. discoideum* (not shown), which has four canonical pairs. The C residues in the helix (*E. coli*) are surprisingly sensitive to bisulfite modification in the free RNA, and the area is sensitive to nuclease attack. Yet cobra venom cuts have been reported for the region, which suggests secondary structure (90).

The residues flanking and between the two helices tend to be highly conserved within a

TABLE 11. Helix 289-292/308-311

Organism/organelle	Sequence
	300
E. coli	U A G C U G — G U C U G A G A G G A U G A C C A G C
Chloroplasts	U A G C U G — G U C C G A G A G G A U G A U C A G C
H. volcanii	U A C G G G — U U G U G A G A G C A A G A G C C C G
Paramecium mitochondria	U A G C U G — A U U U G U G A G A A G A A U C A G C
Aspergillus mitochondria	U A G U C G U G A C U G A G A G G U C G A U C G A C
Yeast mitochondria	U A A U C G A U A A U G A A A G U U A G A A C G A U
X. laevis	A A C G G G — G A A U C A G G G U U C G A U U C C G

TABLE 12. Helices 316-322/331-337 and 339-342/347-350

Organism/organelle	Sequence
Sequence	
	320 340
E. coli	C A C A C U G G A A C U G A G A C A C G G U C C A G A C U C C U A C G G G A G
B. brevis	C A C A C U G G G A C U G A G A C A C G G C C C A G A C U C C U A C G G G A G
Yeast mitochondria	C A C G U U G A C U C U G A A A U A U - A G U C A A - U A U C U A U A A G A U A
H. volcanii	G A G A C G G A A U C U G A G A C A A G A U U C C G G C C C U A C G G G G C
X. laevis	G A G A G G G A G C C U G A G A A A C G G C U A C C A C A U C C A A G G A A G
Catalog	
Most eubacteria	C A C A C U G G A C U C C U A C G
Many eubacteria	C A C A U U G C C C A A A C U C C U A C G
Lactobacillus fermentum	C A C A A U G C C C A U A C U C C U A C G
Spiroplasma citri	C A C A U C G G A A C U C C U A C G
Deinococcus species	C A C A G U C C C A C U C C U A C G

Table 13. Helices 368-371/390-393 and 375-379/384-388

Organism/ organelle	Sequence
Sequence	
	370 390
E. coli	G A A U A U U G C A C A A U G G G C G C A A G C C U G A U G C A G C C A
B. brevis	G A A U U U U C C A C A A U G G A C G A A A G U C U G A U G G A G C A A
Human mitochondria	U A A C C U U U A G C A A U A A A C G A A A G U U U A A C U A A G C U A
H. volcanii	G A A A C C U U U A C A C U G C A C G C A A G U G C G A U A A G G G G A
X. laevis	G C A A A U U A C C C A C U C C C G A C G - C G G G G A G G U A G U G A
Catalog	
Various eubacteria	G A A U A U U G G A C A A U G G A U C C A G
Certain purple bacteria	G A A U C U U A G A C A A U G G A U C U A G

given kingdom, but vary to some extent among kingdoms.

Helices 368–371/390–393 and 375–379/384–388 (Table 13)

The composite structure of helices 368–371/390–393 and 375–379/384–388 (Table 13) is perhaps best considered a single irregular helix containing an interior loop. The bulged residues are largely conserved in sequence among the three kingdoms, whereas sequence in the capping loop is conserved between archaebacteria and eubacteria. Both helices vary somewhat in sequence, but variation seems constrained. Catalog data show that when G occurs at position 370, C always occurs at position 391 (in the oligonucleotide AUCCAG). In the same way, A_{369} covaries with U_{392}. Bulged C_{372} and the A's at 373 and 374 are reactive in chemical modification tests.

Although the pair between positions 367 and 394 will potentially form in all cases, G_{394} in eubacteria is taken to pair with C_{47} instead (Table 4). Archaebacteria and eucaryotes can form a sixth pair between positions 366 and 395, again overlapping the helix of Table 4 in the archaebacterial case. (See discussion of coaxial helices in part 2.)

The flanking sequence covering position 365 seems highly conserved among eubacteria; the T_1 oligonucleotide covering the region contains the general sequence GAAU$\overset{A}{\underset{Y}{}}UU_{368}$ in 97% of cases. The fungal, protist, and murine mitochondria, but not those of humans and cows, maintain the sequence as well. Its eucaryotic equivalent, GCAAAUU$_{368}$, seems universal in that kingdom, although the modification is not always present.

Helices 406–409/433–436 and 416–419/424–427 (Table 14)

Although both eubacteria and archaebacteria can form helices in the region 406–409/433–436 and 416–419/424–427 (Table 14), the two ver-

TABLE 14. Helices 406-409/433-436 and 416-419/424-427

Organism/ organelle	Sequence
	410 430
E. coli	G U G U A U G A A G A A G G C C U U C G G G U U G U A A A G U A C
B. brevis	G U G A A C G A U G A A G G U C U U C G G A U U G U A A A G U U C
Chloroplasts	G U G G A G G U G G A A G G C C U A C G G G U C G U C A A C U U C
H. volcanii	G U G C G A G G G C A U A U A - - - - - - - - - - G U C C U C G C

sions are not much alike, the former having the rather large interior loop, and the latter having a smaller structure, a single helix with no bulged residues. Comparative evidence for the eubacterial version is weak enough that only the outer of the two helices can be considered phylogenetically proven. Whether eucaryotes possess a true counterpart of these structures is uncertain (see discussion re Table 15).

Helices 437–446/488–497 and 455–462/470–477 (Table 15)

Again, the various kingdoms tend to structure the areas 437–446/488–497 and 455–462/470–477 somewhat differently. Even within the eubacteria there is some variation. For example, the chloroplast version deletes the inner helix entirely. However, note the sequence similarity in the outer helix between *H. volcanii* and chloroplasts. Although the archaebacterial and eubacterial versions of this structure seem to be truly homologous, as judged by sequence homology and position in the molecule (the position of AAG_{500} can be defined by the succeeding helix [Table 16]), their ostensible eucaryotic counterpart does not occupy a strictly homologous portion of the sequence, and may be idiosyncratic rather than homologous. (The eucaryotic helix in Table 15 precedes the 3′ strand of that of Table 4. This is the reverse of the order in procaryotes.) The eucaryotic structure is supported by some comparative evidence.

Cobra venom cuts have been reported for positions 461, 476, and 477 (90). A psoralen cross-link is reported between U_{458} and U_{473} (83, 88).

It is remarkable that human mitochondria specifically and precisely delete this whole region and the preceding one, i.e., 406 to 477 (replacing them by a stretch of four C residues), whereas fungal mitochondria precisely replace both by a large idiosyncratic structure of high A+U content.

Helix 500–517/534–545 (Table 16)

Helix 500–517/534–545 (Table 16) is a compound helix containing a bulge loop. The eubacteria (as evidenced by both sequences and oligonucleotide covariance) seem always to bulge six bases starting with residue 505. Archaebacteria and eucaryotes, on the other hand, seem to bulge seven bases starting with residue 506. The bulge loop in all cases [including catalogs] contains adjacent A residues—in *E. coli*, AA_{510}. Given this invariance, one wonders whether the placement of these A's in the loop is also invariant, i.e., all examples of the bulge loop begin at the same residue, 505, and contain six residues only. To accomplish this, the archaebacterial, eucaryotic, and some mitochondrial versions

TABLE 15. Helices 437-446/448-497 and 455-462/470-477

Organism/organelle	Sequence (440 … 460 … 480 … 500)
E. coli	U U U C A G C G G G G A G G A A G G G A G U A A A G U U A A U A C C U U U G C U C A U U G A C G U U A C C C G C A G A A G A A G
B. brevis	U G U U G U C A G G G A C G A A C A C G U C C G U U G G A A U A G G G C G G U A C C U U G A C G G U A C C U G A C G A G A A A G
Z. mays chloroplasts	U U U U C U C G G A G A A G A A A C - A A U G A C G G U A U C U G A G G A A U A A G
H. volcanii	U U U U C U C G A C C G - U A A G G - C G G U C G A G G A A U A A G
S. cerevisiae	C G A U A C A G G G C C C A U U C - G G G U C U U G U A A U U G

TABLE 16. Helix 500-517/534-545

Organism/organelle	Sequence
Sequence	500 520 540
E. coli	A A G C A C C - G G C U - A A C U C C G U G C C A G C A G C C G C G G U A A U A C G G A G G G U G C A
Chloroplasts	A A G C A U C - G G C U - A A C U C U G U G C C A G C A G C C G C G G U A A G A C A G A G G A U G C A
H. volcanii	A A G A G C U G G G C A A G A - C C G G U G C C A G C C G C C G C G G U A A U A C C G G C A G C U C A
X. laevis	U A U U G G A G G C A A G U - C U G G U G C C A G C A G C C G C G G U A A U U C C A G C U C C A A U
Catalog	
Some eubacteria	G C U - A A C U C U G G C C G C G G U A A U A C A G
Most mycoplasmas	G C U - A A C U A U G G C C G C G G U A A U A C A U A G
Myxococcus sp.	G C U - A A C U C U G U G G C C G C G G U A A U A C A G
Alternate Configuration for Bulge Loop	
H. volcanii	A A G A G C U G G G C A A - G A C C G G U G C C A G C C G C C G C G G U A A U A C C G G C A G C U C A

need to bulge an additional residue in the inner helix at position 510; Table 16. The mitochondrial examples all seem abnormal and idiosyncratic with respect to the bulge loop.

The entire 500 to 545 structure seems extremely important to ribosome function. It is present in all organisms and organelles, and sequence is highly conserved, being very nearly universal in the apex loop.

$GCCm^7GCG_{529}$ (m^7G is 7-methylguanine) is found in all but one of the eubacterial catalogs. The generalized sequence $GCUAACU^{A}_{C}YG_{515}$, which covers the bulge loop and part of the upper stalk, is found in 91% of the eubacterial catalogs, whereas $GUAAUACR_{537}$, which crosses the apex loop and stalk, is found in 93% of them. In all cases, a pairing relationship holds between the latter two oligonucleotides, e.g., CUAA CUAUG and UAAUA CAUAG (minor variants of both oligonucleotides found in mycoplasmas; Table 16).

The chemical reactivity of various residues in this region is consistent with the structures given. C_{507}, C_{525}, C_{526}, and C_{528} are highly reactive with bisulfite, as is G_{529} with glyoxal. Note that G_{530} is reactive with kethoxal in active 30S subunits and 70S ribosomes (13, 50), but becomes shielded in polysomes (10), underscoring the probable functional importance of this region. Residues 501, 503, 504, 511, 513, 514, 536, and 537, all in helical conformation, are not reactive with the modifying reagents.

The patterns of post-transcriptional modification in this region are notable. m^7G_{527} is found in all eubacteria and in two of the archaebacterial examples. Moreover, it is the first modification introduced into eubacterial 16S rRNA (73). $2'OMeU_{531}$ is found in a few eubacteria and in all eucaryotic examples. G_{506} is modified in *X. laevis* 18S rRNA (at least). One eubacterial example modifies G_{515} (Table 16).

A potential helix can form between positions 564 to 570 and 880 to 886. Sequence in the area is too conserved to provide convincing phylogenetic support for the structure. Moreover, mitochondrial, archaebacterial, and eucaryotic examples contain at least one mispair. Although the existence of this helix is therefore doubtful, it is mentioned here for the following reason. When prepared under the proper conditions, the 16S rRNA fragment protected from nuclease attack by protein S4 terminates at position 575 rather than at position 557 (19). In that case, the protected fragment also includes residues 819 to 858 and 870 to 898 (see Table 21). This indirect evidence is consistent with, if not suggestive of, the proposed helix. In addition, electron micrographs show that protein S4 sequesters the ends of a loop corresponding in approximate size and position to residues 570 through 880 (14).

Helices 576–580/761–765 and 584–587/754–757 (Table 17)

Helices 576–580/761–765 and 584–587/754–757 (Table 17) are seen in eubacteria and archaebacteria. Their counterpart in eucaryotes is uncertain. Table 17 shows two possibilities, the first of which is homologous with the procaryotic examples. The second, however, appears to be a better match, forming seven to eight contiguous canonical pairs. The second (647–653/752–759) is in effect an extension of the outer helix shown in Table 19. Sequence constancy in the eucaryotes prevents there being any comparative evidence to bear on the situation. This is the third example of eucaryotes having an ostensibly analogous helix in which one of the strands is from a nonhomologous area in the sequence (the

TABLE 17. Helices 576-580/761-765 and 584-587/754-757

Organism/organelle	Sequence
	580 760
E. coli	AAAG CGCAC GCA GGCG ... GA CGCU CAG GUGCG AAA
Chloroplasts	AAAG CGUCU GUA GGUG ... GA CACU GAG AGACG AAA
H. volcanii	AAAG CGUCC GUA GCCG ... GA CGGU GAG GGACG AAA
	580 760
	AAAAAG C UC G UA GUUG ... RUUAAU C A GAA C GAAA
Eucaryotes	650 760
	AUGAUUAA U/G AGG ... RUUAAUCAAGAACGAAA

other two being the helices shown in Tables 4 and 15).

Helix 588–617/623–651 (Table 18)

The compound structure 588–617/623–651 (Table 18) can be considered a single irregular helix with somewhat similar interior loops in the various cases. The outer part (588–606/633–651) seems to be the binding site for ribosomal protein S8 (67, 104). *E. coli* protein S8 will bind to the archaebacterial site in spite of the fact that the two sites are quite different in sequence and appear somewhat dissimilar in secondary structure (84).

The inner helix 612–617/623–628, which is not protected from nuclease attack by protein S8, appears to be regular in structure. Both its apex loop and (innermost) terminal base pair are highly conserved in sequence. A generalized oligonucleotide sequence $GCUYAACN_{624}$ fits 90% of the eubacterial catalogs and seems also to account for almost all archaebacterial examples as well. This helix provides a particularly good example of residues in stalks being protected against chemical modification (eight tested), whereas residues in accompanying loops and flanking sequences are not (at least six).

Eucaryotes have an idiosyncratic structure in this region; the 170 extra residues they contain account for much of the increased size of the eucaryotic 18S rRNA.

Helices 655–672/734–751 and 677–684/706–713 (Table 19)

The outer of the two helices 655–672/734–751 and 677–684/706–713 (Table 19) is, like the protein S8 binding site, an irregular helix of variable sequence. The position of the bulged residue seems to be phylogenetically variable. This helix too binds a ribosomal protein, S15 (48, 104). The helix is found in all three kingdoms. It is notable for the number of A-G pairings that seem to occur in or around its interior loop. The archaebacterial example exhibits four of these, the eucaryotic examples two or three. These can replace bona fide pairs in the *E. coli* version, which has two A-G juxtapositions itself.

As was the case for the structure in Table 18 (defined by ribosomal protein S8), the inner helix is not protected from nuclease attack by the ribosomal protein (48, 104) and is of rather conserved sequence. This is particularly true of the loop sequence between positions 690 and 700. The generalized sequence $GAAAU^{G}_{U698}$ can be located in 82% of eubacterial catalogs. $GAAAUC_{698}$ covers most archaebacterial examples, whereas eucaryotes are described by $GAAAUUCU_{700}$.

The apex loop defined by 677–684/706–713 contains further structure. The canonical covariance between positions 690 (A,G,C,U) and 697 (U,C,G,A) suggests the existence of the small helix of three pairs shown in Table 19.

Structure in the large asymmetric interior loop, i.e., positions 714 to 733, is uncertain. Its sequence is somewhat conserved, and some residues are mildly protected from chemical reagents. A convincing canonical covariance involves positions 673 versus 717, and in archaebacteria and eucaryotes (but not eubacteria), helices of three and five pairs, respectively, can cover this region (Table 19). (The larger helix in the eucaryotic example would compensate for a shortening of the underlying helix, in this case, i.e., 656–670/736–750.)

The possible functional importance of the apex loop is underscored by its being highly susceptible to chemical modification and by its containing two to three sites (positions 703, 705 [and probably 693]) reactive with kethoxal in active 30S ribosomal subunits (50). The first two (and reactive G_{674}) are protected in 70S ribosomes (13), whereas the last shows decreased reactivity in polysomes (10). This part of the structure is likely to be positioned, therefore, at or near the subunit interface.

Helices 769–775/804–810 and 783–786/796–799 (Table 20)

Both helices 769–775/804–810 and 783–786/796–799 (Table 20) are rather constant in sequence, but the apex loop, interior loop (except positions 776 to 778), and flanking sequences on both sides of the composite structure are even more constant. The sequences $GCRAACAG_{785}$ (87% of cases), $GAUUAG_{791}$ (99%), $GAUACCCUG_{799}$ (90%), $GUCYAYG_{809}$ (90%), and $^{G}_{C}UAAACG_{818}$ (96%) cover most of the eubacterial catalogs. $AUACCG_{797}$ accounts for six of the seven eucaryotic instances, and $AUACCCG_{798}$ accounts for 19 of 20 archaebacterial cases. Portions of the segment $GUCCACGCCGUAAACGAUG_{821}$ can be traced in over 98% of the catalogs—eubacteria, archaebacteria, and eucaryotes. The 11 residues in helical conformation tested are resistant to chemical modification, whereas those tested in the loops and flanking the helices are all reactive with modifying reagents.

U_{788} is post-transcriptionally modified (to pseudouridine) in some eubacterial and some eucaryotic groups. In fungi, C_{796} is modified (2'OMe).

The above facts, plus the facts (i) that the G residues at positions 791, 803, and 818 are reactive with kethoxal in active 30S ribosomal subunits (50) and (ii) that the entire area occurs practically unaltered in all mitochondria, suggest that this region is one of the more functionally important regions in the molecule. Shielding of

TABLE 18. Helix 588-617/623-651

Organism/organelle	Sequence
Sequence	
E. coli	G U U U G U U A A G U C A G A U G U G A A A U C C C C G G G C U C A A C C U G G G A A - C U G C A U C U G A U A C U G G C A A G C
Chloroplasts	G C U U U U C A A G U C C G C C G U C A A A U C C C A G G G C U C A A C C C U G G A C - A G G C G G U G G A A A C U A C C A A G C
H. volcanii	G C C A C G A A G G U U C A U C G G G A A A U C C G C C A G C U C A A C U G G C G G G C G U C C G G U G A A A A C C A C G U G G C
Catalog	
Rhodopseudomonas sphaeroides	G A A A U C C C A G G C U C A A C C C U G
Paracoccus denitrificans	G A A A U C C C G G C U C A A C C C C G
Bacillus insolitus	G A A A U C C C C A C G G C U C A A C C G

TABLE 19. Helices 655-672/734-751 and 677-684/706-713

Organism/organelle	Sequence
Helix 655-672/734-750	
E. coli	A G U C U C G U A G A G G G G G G U A G A A ..... G G C G G C C C C C C U G G A C G A A G A C U G
B. brevis	A G U G C A G A A G A G G A A A G C ..... G G C G G C U U U C U G G U C U G U A A C U G
H. volcanii	G G A C C G G A A G G C U C G A G G G G U A ..... A G C A C C U C G A G A A G A C G G A U C C G
D. discoideum	G G G A U G G A U G G G G G G U G U U C A U A ..... A G C A U U C A C C A A A U A C U U C C C C A
Helix 677-684/706-713	
E. coli	A G A A U U C C A G G U G U A G C G G U G A A A U G C G U A G A G G A U C U G G A G G A A U A C C G G U G G C G A A G
B. stearothermophilus	. . . G A A U U C C A C G U G U A G C G G U G A A A U G C G U A G A G G A U G U G G A G G A A C A C C A G U G G C G A A G
H. volcanii	G G U A C G U C C G G G G U A G G A G U G A A A U C C C G U A A U C C U G G A C G G A C C A C C G A U G G C G A A A
D. discoideum	U U C A U A U U G G U G G G C G A G A G G U G A A A U U C G U U G A C C C U A U C A A G A U G A A C U U C U G C G A A A
Paramecium mitochondria	U G A G C U U G U A A A C U A G U G A U A A A A U A C A G G G A G U U U A C A G G A A C A U U C A A - C G C G A A A

TABLE 20. Helices 769-775/804-810 and 783-786/796-799

Organism/organelle	Sequence
Sequence	770 790 810
E. coli	A A A G C G U G G G G A G C A A A C A G G A U U A G A U A C C C U G G U A G U C C A C G C C G U A A A C G A U G
Tobacco chloroplasts	A A A G C U A G G G G A G C G A A U G G G A U U A G A U A C C C C A G U A A G U C C U A G C C G U A A A C G A U G
H. volcanii	A A A G C U A G G G G U C U C G A A A C C G G G A U U A G A U A C C C G G G U A A G U C C U A G C U G U A A A C G A U G
X. laevis	A A A G U C G G A G G U U C G A A G A C G G A U C A G A U A C C C G U C G U A A G U U C C G A C C A U A A A C G A U G
Catalog	
Some mycoplasmas	G C A A A U A G G A U A C C C U A G
Acholeplasma sp.	G C A A A C A G G A U A C C C U G
Megasphaera sp.	G C A A A C G G A U A C C C C G
Deinococcus radiodurans	G C A A A C C G G A U A C C C G

the three kethoxal-reactive sites in 70S ribosomes (13) and discrimination of the sites by 50S subunits in modification-selection experiments would place this region at the subunit interface (35). This is also suggested by protection against cobra venom nuclease cleavage of position 773 in 70S ribosomes (90).

Helices 821–828/872–879 and 829–840/846–857 (Table 21)

Helices 821–828/872–879 and 829–840/846–857 (Table 21) could form a coaxial structure with a bulge loop. Sequence in the outer helix tends to be somewhat conserved, whereas that in the inner structure is highly variable. The latter is also extremely variable in overall size. By contrast, the bulge loop is of constant length in eubacteria and archaebacteria, and in eubacteria at least its sequence tends to be conserved as well.

Variability in sequence in the inner helix extends even to the species level. The structure is notable for its high density of G · U pairs, with both *E. coli* and *H. volcanii* showing four continguous such pairs. Mammalian mitochondria delete this helix. Its structure is uncertain in eucaryotes.

For the eubacteria, the generalized oligonucleotide GYUAACR$_{867}$ (in the bulge loop) accounts for 83% of the cases (GCUAACG$_{867}$ alone accounting for 75%). The residues in the bulge loop tested are reactive with the chemical modifying reagents, whereas those in the helices are not. Nevertheless, the bulge loop may contain structure. Among eubacteria, there exists a strong covariance (at least three phylogenetically independent examples) between position 862 (U versus C) and position 867 (A versus G). *H. volcanii* has G$_{862}$ and C$_{867}$.

G$_{844}$ is kethoxal reactive in 30S subunits (50). It is not significantly protected by 50S subunit association (13), in keeping with its nonconserved sequence. Cobra venom nuclease cleaves after position 840, and this too, is not prevented by 50S subunit association (90).

Helices 888–891/909–912 and 893–897/902–906 (Table 22)

For the helices 888–891/909–912 and 893–897/902–906 (Table 22), sequence in the outer helix tends to be highly conserved. As judged by the large T$_1$ oligonucleotide that covers the 3′ side of this helix, no variation exists within each of the three kingdoms. However, each kingdom is characterized by a unique pair involving positions 888 and 912, i.e., G · C in eubacteria, A · U in archaebacteria, and G · U in eucaryotes.

The inner helix has only four pairs (two of them G · U) in *E. coli*, but even among eubac-

TABLE 21. Helices 821-828/872-879 and 829-840/846-857

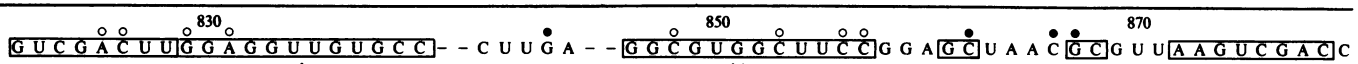
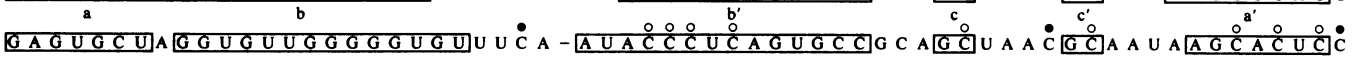
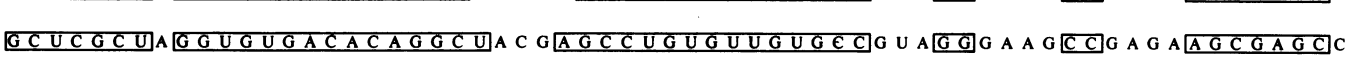
Organism	Sequence
E. coli	
B. brevis	
H. volcanii	

TABLE 22. Helices 888-891/909-912 and 893-897/902-906

Organism/organelle	Sequence
E. coli	GG[GAGU]AC[GGCC]GCAA[GGUU]AAA[ACUC]AAAUG
B. brevis	GG[GAGU]A[CGCUC]GCAA[GAGUG]AA[ACUC]AAAGG
Human mitochondria	AACACUAC[GAGC]CACA[GCUU]AAAACUCAAAGG
Yeast mitochondria	AA[GAGU]A[CGUUA]GCAA[UAAUG]AA[ACUC]AAAAC
H. volcanii	GG[AAGU]A[CGUCC]GCAA[GGAUG]AA[ACUU]AAAGG
X. laevis	GG[GAGU]A[UGGUU]GCAA[AGCUG]AA[ACUU]AAAGG

TABLE 23. Helix 926-933/1384-1391

Organism/organelle	Sequence
Sequence	930 1390
All eubacteria	A C G G G G G C C C G — C A C A A G ..... G U U C C C C G G G G C C U U G U A C
Mouse mitochondria	G C G G U A C U U U A — U A U C C A ..... A G C A A U G A A G U A C G C A C
H. volcanii	G C G G G G A G C A C U A C A A C ..... G U C C C C U G C U C C U U G C A C
X. laevis	A C G G A A G G G C A C C A C C A G ..... G U C C C C U G C C C C U U G U A C
Catalog	
0.99 of eubacteria	G C A C A A G ..... G U U C C C G N G G Y C U U G
Chlorobium vibrioforme	G C C C A C A C A A G ..... G U U C C C U G N G C C U U G

teria a fifth pair between positions 893 and 906 seems often to occur.

The interior loop separating the two helices seems to be of conserved sequence. The generalized sequence $RAAA_{909}$ accounts for all sequences and catalogs.

Its sequence conservation would suggest this structure to be one of the more functionally important ones in 16S-like rRNAs.

Helix 926–933/1384–1391 (Table 23)

Helix 926–933/1384–1391 (Table 23) is central to the structure if not the function of 16S-like rRNAs. It delimits and defines the 3′ major domain and occurs in all known examples of the 16S-like rRNA. Within a primary kingdom, the sequences in and around the helix (i.e., all four flanking sequences) are quite highly conserved, but variation from kingdom to kingdom, and in mitochondria, occurs. For example, the generalized sequences $GCACAAG_{939}$, $GCACCACCAG_{939}$, and $GCACYACAAC_{939}$ account, respectively, for all but 2 eubacterial catalogs, 6 of 7 eucaryotic catalogs, and at least 16 of 20 archaebacterial examples. Similarly, $GUUCCCG_{1385}$ accounts for 96% of eubacterial examples, whereas $UCCCUG_{1385}$ covers all eucaryotic and all but one archaebacterial example. Finally, $GUACACACCG_{1401}$ covers all eucaryotic and all but one eubacterial catalogs, whereas $GCACACACCG_{1401}$ is found in all archaebacteria tested. The only major exception to these constancies occurs in the mitochondria, particularly fungal mitochondria.

Cobra venom nuclease cuts within the helix after position 1389, a cleavage prevented by 70S particle formation (90).

A partial covariance involving G_{923} (usually A) and C_{1393} (usually U) is seen in archaebacteria and in some mitochondria, suggesting further structure on the 5′ side of the helix. Archaebacteria and (all but one of the) eucaryotes are distinguished from eubacteria by a pyrimidine insertion at about position 934.

Helix 938–943/1340–1345 (Table 24)

The structure of helix 938–943/1340–1345 (Table 24) seems conserved in sequence in eubacteria. The oligonucleotide $GAAUCG_{1343}$ is present in about 70% of eubacteria. (However, certain subgroups of eubacteria entirely lack it.) The pattern of chemical modification for the residues involved indicates that this structure may not exist in isolated 16S rRNA; G_{1343} is protected, and the preceding A's (1339 and 1340) are highly so, but the intervening C_{1342} is reactive.

G_{1338} is reactive with kethoxal in active 30S ribosomal subunits (50), a reactivity that remains in 70S ribosomes but is lost in polysomes

(10, 13). Some eucaryotes modify this residue (m^7G); some modify U_{1341}.

Helices 946–955/1225–1235 and 984–990/1215–1221 (Table 25)

Helices 946–955/1225–1235 and 984–990/1215–1221 (Table 25) seem to be a major structural feature of 16S rRNA. Sequence therein tends to be highly constrained if not fully conserved. For example, the generalized sequence . . . ACCNR$_{987}$ covers over 98% of the eubacterial catalogs, as does C^{UU}_{AA}CACRCG$_{1231}$. GCUACAAUG$_{1241}$ occurs in over 85% of the eubacterial and archaebacterial catalogs.

The two-pair extension of the first helix above bulged A_{1227} is well proven. The two oligonucleotides in question, i.e., GUUUAAUUCG$_{963}$ and GCUACACACG$_{1231}$, are very stable phylogenetically. In the few instances in which they vary, a canonical Watson-Crick covariance (U,C,A versus A,G,U) holds between positions 955 and 1225. The bulged A_{1227} appears to be a constant feature of all three kingdoms and is found in the mitochondria as well.

Chemical modification studies provide considerable support for the second helix, all consistent with the proposed structure. The eucaryotic examples of both helices, however, contain conserved mispairs.

Helix 960–963/972–975 (Table 26)

The small helix 960–963/972–975 (Table 26) defines one of the more interesting regions in the 16S rRNA. The region is universal and highly conserved in sequence. The generalized oligonucleotide NUUAAUUCG$_{963}$ covers all eubacterial examples. UUUAAUUG$_{962}$ covers all but one archaebacterial example, and eucaryotic catalogs all contain CUUAAUUUG. Sequence in the helix is so conserved that comparative evidence for it is minimal, although convincing. The generalized sequence in the loop for eubacteria is ANGCAACR$_{971}$ (m^2_2G or m^2G, and m^5C in almost all cases). Archaebacteria and eucaryotes conform to this also if G_{966} is replaced by a U. (This pyrimidine is uniquely hypermodified in both groups.) The same position is reactive with kethoxal in active *E. coli* 30S and 70S ribosomes, but becomes protected in polysomes (10, 13, 50). The remaining flanking sequence in almost all cases fits the general form RNAYCUYA$_{983}$.

In eucaryotes, the first three pairs in the helix are of the U · G type; in archaebacteria, the first two are so. The structure is supported by the little evidence that exists from chemical modification experiments, i.e., C_{962} and G_{963} are unreactive.

Helix 997–1012/1017–1044 (Table 27)

In eubacteria, helix 997–1012/1017–1044 (Table 27) is a composite helix with a pronounced bulge loop. The two-pair extension distal to the single bulged A_{1042} is supported by comparative evidence from oligonucleotide catalogs. However, this bulged base does not seem to be a constant feature in the eubacteria. The helix is highly variable in sequence, to the extent of species variation within the same genus. Yet flanking sequences are highly conserved. At least 90% of eubacterial examples can be accounted for by the sequence GACAU$_{997}$, whereas the general form N^{A}_{U}NACAG$_{1047}$ should account for a large fraction of the 3′ flanking sequences. G_{1015} in the apex loop is particularly sensitive to chemical modification (with glyoxal) in the free 16S rRNA. Cobra venom cuts have been reported for positions 999 to 1001, 1020, and 1021, none of which are prevented by 70S ribosome formation (90).

The apparent bulge loop (positions 1024 to 1036) in *E. coli* is of particular interest. It is enlarged in *B. stearothermophilus*, and a potential helix (c-c′ in Table 27) of seven pairs is evident. A smaller helix can be formed in the *B. brevis* case, and if a noncanonical pair is permitted, even the *E. coli* example can form a helix (underlined in Table 27). The pyrimidine stretch

TABLE 24. Helix 938-943/1340-1345

Organism/ organelle	Sequence
	940 ... 1340
All eubacterial sequences	A C A [A G C G G U] G G ..... G G A [A U C G C U] A G
Human mitochondria	C C C U [C U A G A] G G ..... G G A [U U U A G] C A G
Yeast mitochondria	U U A [A G C A G U] G G ..... A G A [A U U G C U] A G
H. volcanii	A C A [A C C G G A] G G ..... G G A [U U C G G U] A G
S. cerevisiae	A C C [A G G A G U] G G ..... G G A [A U U C C U] A G

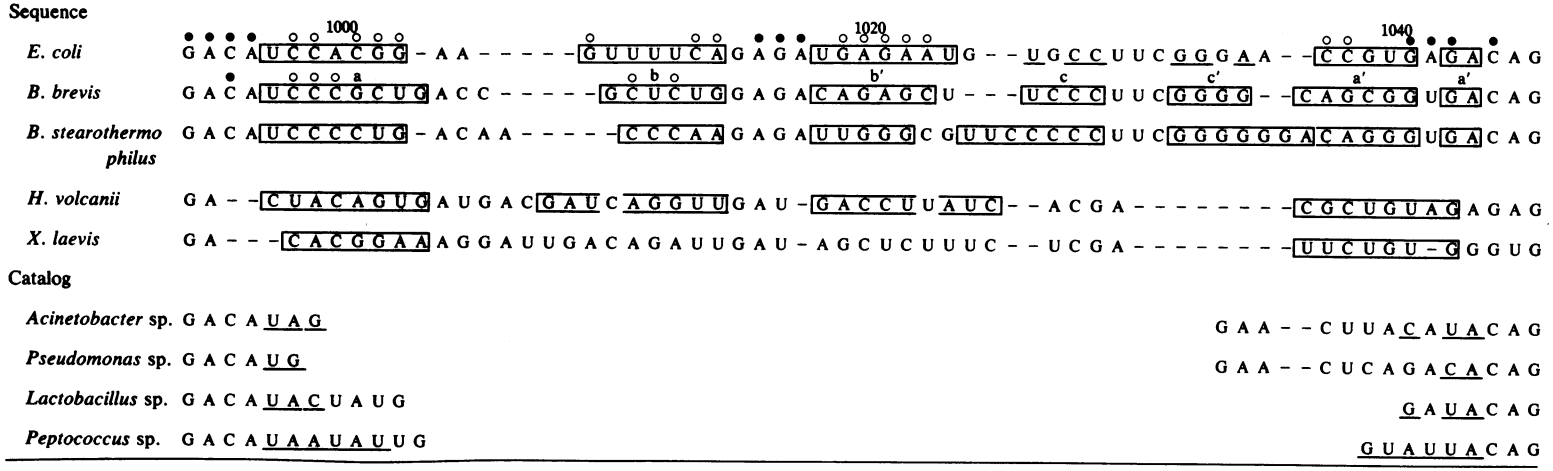
TABLE 25. Helices 946-955/1225-1235 and 984-990/1215-1224

Organism/organelle	Sequence
Sequence	950 990 1220 1240
E. coli	G A G C A U G U G G U . . A C C U G G U C U . . . C G A C C A G G G C U A C A C A C G U G C U A C A A U G
B. stearothermophilus	G A G C A U G U G G U . . A C C A G G U C U . . . U G A C C U G G G C U A C A C A C G U G C U A C A A U G
Human mitochondria	G A G C C U G U U C U . . A C C A C C U C U . . . U G A G G U G G – C A A G A A A U G G G C U A C A U U U
H. volcanii	G A G C C U G C G G U . . A C C A G C U C C . . . U G A G C U G G G C U A C A C G C G G G C U A C A A U G
X. laevis	G A G C C U G C G G C . . A C C C G G C C C . . . . . U G U C C G G G G C U G C A C G C G C U A C A C U G
Catalog	
Spirillum volutans	G A U U A U G G A . . . . . A C C U A C C C U . . . . . G C U U C A C A C G U A A U A C A A U G
Acholeplasma laidlawii	G A U C A U G G U . . . . . A C C A G G C U A C A A A C G G A U A C A A U G
Spiroplasma citri	G U . . . . . A C C A A G U A U C U U G G C U A C A C A C G G C U A C A A U G
Mycoplasma gallisepticum	G C . . . . . A C C U A G A C U . . . . . G U C U A G G C A A A C G G C U A C A A U G
Staphylococcus epidermidis	G U . . . . . A C C A A A U C U . . . . . G A U U U G G C U A C A C A C G G C U A C A A U G
Peptococcus glycinophilus	G U . . . . . A C C U A A A C U . . . . . G U U U A G G C U A C A C A C G G C U A C A A U G

TABLE 26. Helix 960-963/972-975

Organism/organelle	Sequence
	960 980
E. coli	U U U A A U U C G A U G C A A C G C G A A G A A C C U U A
Paramecium mitochondria	U U U A A U G C G A U A A U C C A C G U A A A A C C U U A
H. volcanii	U U U A A U U G G A C U C A A C G C C G G A C A U C U C A
X. laevis	C U U A A U U U G A C U C A A C A C G G G A A A C C U C A

TABLE 27. Helix 997-1012/1017-1044



capping this putative structure appears to be somewhat conserved; about 80% of the eubacterial oligonucleotide catalogs contain the sequence . . . CCUUCG. The structure is particularly interesting because (within the same genus, *Bacillus*) a pronounced difference exists between a thermophile and a mesophile. (Note the multiple adjacent G · C pairs in the case of the thermophile.) This is the most obvious example of the 16S rRNA apparently adjusting to a physical parameter of the organism's niche.

The overall 997–1012/1017–1044 structure is highly variable. The bulge loop is absent or abbreviated in the fungal mitochondrial example and in the other two kingdoms. Mammalian mitochondria seem to lack the structure completely. Although the eucaryotic sequences seem to show specific homology with the archaebacterial sequence, the helix b-b′ is not clearly evident in the former case. Overall length of the composite structure is not conserved. Moreover, the flanking sequence (position 995) in both eucaryotes (four residues) and archaebacteria (three residues) is shorter than in eubacteria (six residues).

Helix 1046–1067/1189–1211 (Table 28)

Helix 1046–1067/1189–1211 (Table 28) is a highly irregular structure in all organisms. However, sufficient canonical pairing covariance exists among its various examples that the structure proposed should be a reasonable approximation to the true situation. (The various examples of the helix are presented in Table 28 in a somewhat different format to facilitate comparisons of paired versus nonpaired positions.) The possible importance of this structure is indicated by its high degree of sequence conservation and by the fact that post-transcriptional modifications are occasionally introduced at a number of points. In the eubacteria, multiple scattered examples of post-transcriptional modification exist for residues U_{1194}, C_{1195}, A_{1201}, G_{1207}, U_{1211}, and one of four C's in the region 1207 to 1210. In eucaryotes, G_{1197} is modified in all catalogs but one; A_{1196} and U_{1210} are also occasionally modified. In eubacteria, the generalized sequence GUCAARUC$\overset{A}{C}$UCAUG$_{1206}$ covers 93% of the catalogs; $\overset{G}{C}$CCCU$_{1211}$ covers 98%.

Chemical reactivities also lend some support

Table 28. Helix 1046-1067/1189-1211[a]

Organism/organelle	Sequence
	1050 1060
E. coli	AGGUGCUG C AU GGCU GU CG UCA UCC CGGU UA CUGA C GCAGU 1210 1200 1190
B. stearothermophilus	G U C A
H. volcanii	A G C C G UG G U G AC
X. laevis	U G C U C U U C G A A U A C
Aspergillus mitochondria	A C C U U U U A A U
Paramecium mitochondria	A U GU U C C U C A

[a] The structure is composed in a paired form to facilitate comparison. All versions are compared with the *E. coli* form; when a residue is the same as in *E. coli*, it is replaced by a dash, and when unique, it is printed out. Vertical lines denote canonical pairs, internal dots G · U pairs.

TABLE 29. Helix 1068-1073/1102-1107

Organism/organelle	Sequence	
	1070	∘1105
eubacterial and archeabacterial sequences	GCUCGU	ACGAGC
Aspergillus mitochondria	GUUAAU	AUUAAC
Parmecium mitochondria	GUUCGU	ACGAAC

to this large irregular helix. C's at positions 1208 to 1210 are rather unreactive with bisulfite. (C_{1207} [occurring in *B. brevis*] is also unreactive.) G_{1198} is unreactive with glyoxal. However, C_{1195} (paired in Table 28) is reactive, whereas A_{1196} (not paired in Table 28) and A_{1197} are to some extent protected from the A reagent. One of the two C's at positions 1200 and 1203 is more reactive than the other; the structure would suggest the reactive one to be C_{1203}.

Interestingly, G_{1053} and G_{1068} become more reactive to kethoxal in 70S ribosomes (13), implying a 50S subunit-induced conformational change in the structure shown in Table 28.

Helix 1068–1073/1102–1107 (Table 29)

The sequence of helix 1068–1073/1102–1107 (Table 29) appears to be the same in all procaryotes. Two mitochondrial examples provide some comparative evidence for its existence. G_{1104} was found to be unreactive with glyoxal. However, the last two pairs of the six-pair structure shown in Table 29 are incompatible with the structure to be discussed next.

Helices 1072–1076/1081–1085 and 1086–1089/1096–1099 (Table 30)

The two contiguous helices 1072–1076/1081–1085 and 1086–1089/1096–1099 (Table 30) appear coaxial (see discussion in part 2) because their combined overall length (a′ plus b in Table 30) remains constant phylogenetically despite the fact that their individual lengths do not. The first helix is five pairs in length in eubacteria and archaebacteria. It has comparative evidence to support it within the eubacteria. (Note, however, that the archaebacterial example contains an A-C juxtaposition.) Because of sequence constancy, no comparative evidence for the second, four-pair helix is found within the eubacteria. However, the eubacterial version differs from the archaebacterial version by two transversions (and a U · G → C · G replacement). In eucaryotes, the first helix comprises six pairs (not five); the second three (not four), each with significant sequence variation from the corresponding *E. coli* versions. (*Paramecium* mitochondria, interestingly, can form the first helix with only three pairs, but the second with six.)

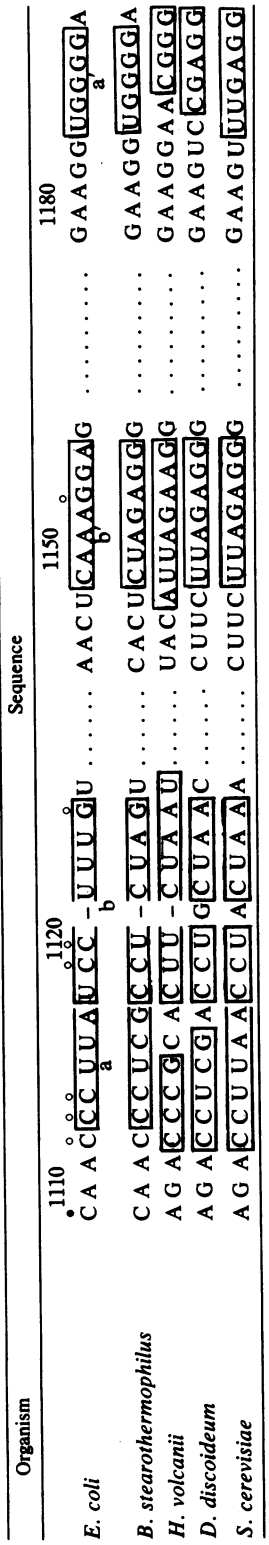
Sequence in the two loops $GYRA_{1080}$ and $UUAA^{G}_{U1095}$ tends to be universally conserved (except for some mitochondria). The residues tested in the b-b′ stalk, i.e., $C_{1096-1098}$ and G_{1099}, are at least moderately resistant to chemical modification in the free RNA, whereas G_{1094} (in the loop) is sensitive to kethoxal substitution in active 30S subunits (50).

The helices of Tables 29 and 30 partially overlap and so are to that extent mutually exclusive. If both exist (and evidence for those shown in Table 30 is strong), then they may somehow constitute alternate structures, forming at different stages in the translation cycle. (Indeed, they may be alternate coaxial structures; see discussion in part 2.) Such a dynamic switch is consistent with the observation that in the same general region of the molecule, several G residues,

TABLE 30. Helices 1072-1076/1081-1085 and 1086-1089/1096-1099

Organism/ organelle	Sequence
	1070 1080 1090 1100
E. coli	UC GUGUU GUGA AAUGU UGGG UUAAGU- CCCG CAACG (a, a′, b, b′)
Chloroplasts	UC GUGCC GUAA GGUGU UGGG UUAAGU- CUCG CAACG
Paramecium mitochondria	UCGU GUU UUGA AAU UUA G AA UUAAGUU UU A UAA ACG (a, a′, b, b′)
H. volcanii	UC GUA CC GUGA GG C GU CCUG UUAAGU- CAGG CAACG
X. laevis	U GGUGGA GCGA UUUGUC UGG UUAAUU- CCG AUAACG

TABLE 31. Helices 1113-1117/1183-1187 and 1118-1124/1149-1155



i.e., G_{1053} and G_{1068}, become more reactive toward kethoxal in 70S ribosomes (as opposed to active 30S subunits) (13, 35, 50). (Moreover, G_{1064} [in helical array] is not protected in 30S subunits, but is in 70S particles [13].)

Helices 1113–1117/1183–1187 and 1118–1124/1149–1155 (Table 31)

The outer helix of the adjacent pair 1113–1117/1183–1187 and 1118–1124/1149–1155 (Table 31) is rather conserved in sequence. The generalized sequence GCAACCC$^{A}_{Y}$YR$_{1117}$ accounts for all eubacterial catalogs. By contrast, the inner helix is variable in sequence, although constrained. Catalogs provide considerable additional comparative support for the inner helix. Covariances involving (paired) positions 1119 and 1154 (C · G, G · C), 1120 and 1153 (A · U, U · A, C · G), 1121 and 1152 (U · A, A · U), 1122 and 1151 (U · A, A · U), and 1123 and 1150 (U · A, A · U) are observed, which (together with some G · U pairs) cover almost all of the catalog examples.

C's at positions 1119 and 1120 are resistant to bisulfite. In *B. brevis*, C_{1149} is resistant, whereas C_{1147} (not in helix) is sensitive to the reagent. G_{1124} is resistant to glyoxal substitution.

The large apex loop (positions 1125 to 1148) defined by the inner helix appears to be structured, although only *E. coli* among the eubacteria shows a clear indication of this. It will form CGGU$_{1135}$/GCCG$_{1142}$. *H. volcanii* will form at least CAGC$_{1135}$/GCUG$_{1140}$. More extensive structure occurs in the eucaryotes, which extend the loop at position 1137.

Helix 1161–1165/1171–1175 (Table 32)

Helix 1161–1165/1171–1175 (Table 32) is a small helix that appears to exist in all eubacteria, archaebacteria, eucaryotes, and the fungal and protist mitochondria. Sequence in the apex loop and stalk seems constrained among eubacteria; the generalized oligonucleotide GAYRAAYYG$_{1174}$ accounts for 65% of the eubacterial catalogs.

G_{1166} (in the apex loop) is kethoxal reactive in 30S subunits but protected in 70S ribosomes (13). A cobra venom cut after position 1162 is not protected by 70S ribosome formation (90).

Helix 1241–1247/1290–1296 (Tables 33 and 34)

Helix 1241–1247/1290–1296 proper is found in all three kingdoms and most mitochondria (Tables 33 and 34). The large apex loop is of the same length (42 residues) in eubacteria and archaebacteria, but is of different sizes in the other examples. Sequence in the helix proper is variable among the eubacterial examples, and a number of catalogs can be used to strengthen the comparative case for the structure.

TABLE 32. Helix 1161-1165/1171-1175

Organism	Sequence
	1170
E. coli	A C U G C C A G U G A U A A A C U G G A G G A A
B. brevis	A C U G C C G U C G A C A A G A C G G A G G A A
H. volcanii	A C U G C C G C U G C U A A A G C G G A G G A A
D. discoideum	A C U A C C U G C C U C A A G C A G G C G G A A
X. laevis	A C A A G U G G C G U U C A G C C A C A C G A G

It is likely that the large apex loop contains additional structure. A number of residues therein are resistant to chemical modification, and a number of small, somewhat idiosyncratic helices can be formed in each case, e.g., 1263–1265/1270–1272; however, none are proven. Sequence in the loop, if not conserved, is interestingly constrained. Its most unusual feature is a quasi repeat of the sequence 1249 to 1266 at 1267 to 1285 (Table 34). Moreover, the data strongly suggest that this repeat is specific, i.e., "tuned", for each species (Table 34).

Flanking sequences tend to be conserved. The sequence $ACAAU_{1240}$ fits all eubacterial and all but one archaebacterial examples, whereas $ACA^{\hat{A}}_{C}U_{1240}$ covers the eucaryotic ones.

Helix 1308–1314/1323–1329 (Table 35)

The structure shown in Table 35 (helix 1308–1314/1323–1329) is found in all 16S-like rRNAs. Within the eubacteria, the sequence in the stalk appears constrained and that in the loop highly conserved; the oligonucleotide $GCAACUCG_{1323}$ is found in 88% of the eubacterial catalogs, whereas all eubacterial and archaebacterial cases are encompassed by the general sequence $G^{\hat{A}}_{C}AAYYCG_{1323}$. When tested, residues in the loop are very reactive with chemical modifying reagents in the free 16S rRNA, whereas residues tested in the stalk are protected.

Flanking sequences seem highly conserved.

Helix 1350–1356/1366–1372 (Table 36)

The structure shown in Table 36 (helix 1350–1356/1366–1372) is again found in all 16S-like rRNAs. Oligonucleotide catalogs measure the loop with the T_1 RNAse pentamer $AUCAG_{1361}$, and the flanking sequences with $GUAAUCG_{1353}$ and $GAAUACG_{1379}$. (In that a significant fraction of the occurrences of a pentamer [AUCAG] may not represent the area in question, this oligonucleotide is only an approximate measure of the region; chloroplasts, indeed, are an example of such an exception.) GAUCAG is found in 90% of the eubacterial catalogs (and the two *Bacillus* sequences), suggesting substantial conservation in the loop among eubacteria. $GUAAUCG_{1353}$, present in 98% of the eubacterial catalogs and 95% of the archaebacterial catalogs (but no eucaryotic catalogs) and $GAAUACG_{1379}$, present in 88% of the eubacterial catalogs and 60% of the archaebacterial catalogs, argue for conservation of flanking sequences as well. (In eucaryotes, the equivalent $GA^{\hat{U}}_{U}AYG_{1379}$ accounts for all six catalogs and the sequences.)

Residues in the loop are reactive with chemical modifying reagents. In the free RNA, G_{1361} is the most reactive G residue encountered in the entire *E. coli* 16S rRNA; although this residue is also reactive in inactive (ion-depleted) 30S subunits, it is not so in active subunits (50).

Both archaebacterial and eucaryotic examples

TABLE 33. Helix 1241-1247/1290-1296

Organism	Sequence
	1240 1250 1290
E. coli	A U G G C G C A U A C A A A . . . . . A U A A A – G U G C G U C G U
B. stearothermophilus	A U G G G C G G U A C A A A . . . . . A A A A A – G C C G C U C U C
H. volcanii	A U G G U C G A G A C A A U . . . . . C U A A A – C U C G A U C G U
X. laevis	C U G A A C G G A U C A G C . . . . . C U G A A C C C C G U U C G U
D. discoideum	A U G U A G G A A A C A A A . . . . . U U G A A U U U C C U A C G U

TABLE 34. Helix 1241-1247/1290-1296[a]

Organism	Sequence	
	1250 1260	
E. coli	C A A A G A G — A A G C G — A C C U C G	
	x x x X x x X x x X x x x x x	(15)
	C G A — G A G C A A G C G G A C C U C A U A	
	1270 1280	
B. stearothermophilus	C A A A G G G — C U G C G — A A C C C G	
	x x x X x x x x x x x x	(12;2)
	C G A — G G G G G A G C G A A U C C C A A A	
B. brevis	C A A C G G G — A U G C U — A C C U C G	
	x x x x x x x x x x x x	(12;1,2)
	C G A — G G G G A C G C C A A U C U C U	
H. volcanii	C A A U G G G — U U G C U — A U C U C G	
	x x x x x x x x X x x x	(12;3,3,3)
	A A A — G A G A A C G C U A A U C U C C U A	
X. laevis	C A G C G U G — U G U C U — A C C C U G C G C C	
	x x x x X x x x x x x x x X	(12;6,6,7,6)
	C A G — G U G C G G G — U A A C C C G C U G	
S. cerevisiae	C A G C G A G — — — U C U A A C C U U G G C C G	
	x x x x x x x x x X X	(10;6,6,6,4,3)
	G A G — G U C U U G G — U A A U C U U G U G	
D. discoideum	C A A — — — A A A G G C U — — C C U G G U C C G	
	x x x x x x x x x X	(8;5,5,4,3,3,4)
	G A A — G G A U U G G G U A A U C A U U U G	

[a] Residues 1249 to 1266 are compared with 1267 to 1285 to show similarity of sequence in various organisms. The spaces introduced are constant for the bacterial examples. Lower-case x's denote identities; upper-case X's denote identities unique to a given organism at that position. The number of identities is the first number in parentheses after each row. The subsequent numbers are the number of identities each does not share with the row(s) above it.

TABLE 35. Helix 1308-1314/1323-1329

Organism/organelle	Sequence
	1310 1330
E. coli	C C G G A U U G G A G U C U G C A A C U C G A C U C C A U G A A G U C G G
B. brevis	U C G G A U U G U A G G C U G C A A C U C G C C U A C A U G A A G U C G G
Yeast mitochondria	A U G A A U U A U A A U C U G A A A U U C G A U U A U A U G A A A A A A G
H. volcanii	U C G G A U U G A G G G C U G A A A C U C G C C C U C A U G A A G C U G G
X. laevis	A G G G A U C G G G G A U U G C A A U U A U U U C C C A U G A A C G A G G

TABLE 36. Helix 1350-1356/1366-1372

Organism/organelle	Sequence[a]
	1350 1360 1370
E. coli	A G U A A U C G U G G *A* U C A G — A A U *G* C C A C G G U G A A
B. brevis	A G U A A U C G C G G A U C A G — C A U G C C G C G G U G A A
Chloroplasts	A G U A A U C G C C G *G* U C A G C C A U *A* C G G C G G U G A A
Human mitochondria	A G U A A A C U A A G A G U A G — A G U G C U U A G U U G A A
H. volcanii	A G U A A U C G C A U U U C A A — U A G A G U G C G G U G A A
X. laevis	A G U A A G U G C G G G U C A U — A A G C U C G C G U U G A U

[a] = , Define.

TABLE 37. Helices 1409-1430/1470-1491 and 1435-1445/1457-1466

Organism	Sequence
	1410 · 1420 · 1470 · 1490
E. coli	G U C A C A C C A U G G G A G U G G G U U G C A A A A G .... A C U U U G U G A U U C A U G A C U G G G G U G A A G
B. brevis	G U C A C A C C A C G G G A G U U U G C A A C A C C C G .... A A G G U G G G U A G A U G A C U G G G G U G A A G
H. volcanii	G U C A A A G C A C C C G A G U G A G G U C C G G A U G .... G A A U C U G G C U U C G C A A G G G G G C U U A A G
D. discoideum	G U C G C U C C U A C C G A U C G A A U G A U A C G G U .... A U C U - C A U U G U U U A G A G G A A G G A G A A G
	1430 · 1470
E. coli	A A G A A G U A G G U A G C U U A A C C U - U C G G G A G G G C G C U U A C C A C U U
B. brevis	C C G A A G U C G G U G A G G U A A C C G - C A A G G A G C C A G C C G C C G A A G G
H. volcanii	G A U G A - G G C C A C C A C - A C G G U G G U C - G A A U
X. laevis	A G U G A G G U C C U C G G A U C G G C C C C G C C G G G G U C G G C C A C G G C C C U G G C G G A G C G C C G A G A A G A C G A U C - A A A C U
	a · b · c · c′ · b′ · b′ · a

suggest that pairing extends further into the apex loop than may be apparent from the *E. coli* version. In all eubacterial and mitochondrial cases, there is an A-G juxtaposition on the inside. This becomes G-A in chloroplasts, suggesting an actual A-G pairing (Table 36).

Helices 1409–1430/1470–1491 and 1435–1445/1457–1466 (Table 37)

The two irregular helices 1409–1430/1470–1491 and 1435–1445/1457–1466 (Table 37) may best be viewed as a single composite structure with an interior loop. The outer helix is constrained if not conserved in sequence. The generalized sequence $UCA^{C}_{A}ACYAY^{G}_{C}{}_{1415}$ will account for almost all eubacterial examples. The structure is remarkable for the number of G-A juxtapositions it contains, several of which tend to be conserved phylogenetically. For example, A_{1413}-G_{1487} is found in both eubacteria and archaebacteria, while the two pairings GA_{1418}-GA_{1483} are found in eubacteria, eucaryotes, and some mitochondria. Interestingly, cobra venom nuclease has been shown to cleave between these two G-A pairs (in free 16S rRNA), strongly suggesting a genuine double-stranded arrangement (16a).

Archaebacteria possess an abbreviated inner helix, whereas eucaryotes expand the structure beyond that seen in eubacteria. In archaebacteria, the interior loop between the two helices is replaced by GA_{1432}-GA_{1469}, yet another example of G-A pairing in this region.

These two helices (and parts of the intervening sequence) are relatively resistant to RNases under certain conditions. The fragments so produced run as coherent units in polyacrylamide gel electrophoresis despite the fact that a section of the fragment has been removed in the vicinity of residue 1450 (i.e., the apex loop and parts of the inner helix) (J. Kop et al., manuscript in preparation). (In urea-containing gels, the unit then separates into two distinct pieces.) This fact indicates strong higher-order structure for the region. Residues in the helices are well protected against chemical modifying reagents as well. However, there are reports that residues in this structure are susceptible to various single-strand probes (45, 82).

The flanking sequences for the structure are very highly conserved (as seen above). $CCGCCCGUCR_{1408}$ and $AAGUCGUAACAAG_{1504}$ would seem to account for over 95% of all examples. Modified bases are frequent here, too. C_{1402} is (nearly) universally modified (2′OMe, and in eubacteria, a modification of the base as well [22]), whereas the base C_{1399} (or C_{1400}) is occasionally so, as are C_{1404}, C_{1407}, and C_{1409}. U_{1498} is modified in almost all eubacteria,

and A_{1500} is modified in all archaebacteria and almost all eucaryotes. A_{1499}, as well, is modified in one group of archaebacteria. Many residues in the flanking sequences are highly susceptible to chemical modification. G's at positions 1405 and 1497 are reactive with kethoxal in active 30S subunits (50) and are protected in 70S ribosomes, suggesting placement of this region at the subunit interface (13). The 5′ (wobble) base of the anticodon of tRNA can be photochemically cross-linked to C_{1400} (60). Colicin E3 inactivates 70S ribosomes by making a cleavage at the phosphodiester bond connecting residues 1493 and 1494 (70). In mitochondria resistance to the antibiotic paromomycin (which causes translation errors) involves a C → G transversion at position 1409, creating a mispair (42). These observations strongly implicate this region of 16S RNA directly in ribosome function, which is also suggested by its extreme sequence conservation.

The only 16S-like rRNAs to present significant abnormality (i.e., nonconservation) in the flanking regions are those of the fungal and protist mitochondria.

Helix 1506–1515/1520–1529 (Table 38)

Helix 1506–1515/1520–1529 (Table 38) is the only thoroughly studied helix in the 16S-like rRNAs (34). Nuclear magnetic resonance and temperature-jump (T-jump) studies on the isolated colicin E3 fragment (see above) have provided direct physical evidence for the helix as well as measurements of its thermodynamic stability (2, 100). Although sequence in the helix can vary somewhat, it is highly constrained. In the region from positions 1500 to 1534, 29 of the positions have identical residues when the *E. coli*, *H. volcanii*, and *X. laevis* sequences are compared. The adjacent dimethyladenosine (m^6_2A) residues (positions 1518 and 1519) occur in all known cases except (two) chloroplasts, in which only A_{1519} is dimethylated (C. Woese, unpublished data). In eubacteria, all known examples of this portion of the sequence are accounted for by the T_1 oligonucleotides A*A*CCUG, A*A*CG, A*A*AG, and A*A*G. Eucaryotic catalogs all possess A*A*CCUG, whereas the archaebacterial catalogs possess A*A*YCUG or (in one instance) A*A*C*CUG.

Several modified nucleotides occur in the helix proper. In eucaryotes, the sequence $UUUCCG_{1511}$ (Table 38) occurs as the T_1 oligonucleotide UUUCNG; and, as just mentioned, in one archaebacterium A*A*C*CUG$_{1523}$ occurs. (The sequence and electrophoretic data suggest the C modifications to be N^4-acetyl.)

Resistance to the antibiotic kasugamycin is associated with loss of methylation of the adjacent A residues (positions 1518 and 1519) (34),

TABLE 38. Helix 1506-1515/1520-1529

Organism/organelle	Sequence (1510 … 1530)
E. coli	G• U A° A° C° C° G° U A G G G* G A* A* C° C° U G C G G U U G• G• A U C• A C• C• U C• C• U U A_{OH}
B. brevis	G U A U C° C° G U A C° C° G G A* A* G G U G C G G A U G G A U C• A C• C• U C• C• U U U C• U_{OH}
Human mitochondria	G U A A G U G U A C U G G A A A G U G C A C U U G G A C G A A C ...
H. volcanii	G U A G C C G U A G G G G A* A* U C U G C G G C U G G A U C A C C U C C U N_{OH}
X. laevis	G U U U C C G U A G G U G A* A* C C U G C G G A A G G A U C A U U A_{OH}

suggesting in turn a functional role for these residues. (The loss of modification here plus the creation of the above-mentioned mispair at positions 1409 and 1491 in paromomycin resistance [42] suggest that antibiotic resistance often involves "detuning" of the 16S rRNA.)

G's at position 1505, 1516, and 1517 react with kethoxal in the active 30S subunit (50). G_{1505} and G_{1516} are protected in the 70S ribosome (13); G_{1517} becomes strongly protected in polysomes (10).

The sequence beyond position 1530 has been implicated in messenger RNA (mRNA) binding at the initiation of protein synthesis (17, 71, 72, 78); the sequence $CCUCC_{1539}$ is complementary to a class of sequences occurring 5′ to and near the initiation codon in most if not all bacterial mRNAs. Ninety-five percent of the eubacterial and all archaebacterial catalogs contain the sequence $GAUCACCUCC_{1539}$, whereas all eucaryotic examples contain GAUCAUU.

Comparison with Other Models

Experience with the much smaller 5S rRNA has shown that given one sequence, the number of secondary structures possible is large. For the 16S-like rRNAs, the number of possibilities is enormous: a moderately loose program (i.e., canonical pairs and G · U pairs, helix length at least four contiguous pairs) generates ca. 10,000 possible helices for any 16S rRNA sequence. Although (in retrospect) speculations on the existence of helices in 5S RNA served no useful purpose, comparable speculation concerning helices in 16S or 23S rRNAs would be nothing short of counterproductive. All discussion of a helical element must start from evidence strongly suggesting its existence. The most reliable evidence is in almost all cases generated from comparative sequence analysis. (Note, for example, that none of the original eubacterial 16S rRNA helices taken to be proven by comparative criteria were disproven by the archaebacterial sequence [33, 54, 98]; in fact, additional strong comparative support was provided for almost all of them.)

Admittedly, the comparative approach does not give direct (i.e., physical) evidence for a helix. In the case of tRNA, however, all examples of base pairing covariance correspond to true physical pairing. On the other hand, those methods that demonstrate helices directly (30, 83) can encounter artifacts, i.e., helices that do not exist in the functional condition. Ideally, comparative evidence for any helix should be supported by evidence of a direct nature.

It should also be recognized that a model for RNA structure should progress in stages, as was the case with 5S rRNA. The first-approximation model contains only those structures that are essentially canonical, strong, and extensive helices (25). Refinements (noncanonical pairs, tertiary interactions, coaxiality of helices, etc.) are introduced later (23, 43, 77, 81). To proceed otherwise would serve only to drown truth in a sea of speculation.

At present, a number of somewhat different proposals for 16S rRNA structure exist (7, 30, 54, 79, 98, 105). Since it is nearly impossible for interested scientists to distinguish among the several models, the tendency will be to disregard or distrust all models, at least temporarily, or to conclude that some approaches are not as revealing as they actually are.

We will not discuss the other models in detail here. However, Table 39 should help the interested reader to compare other models with the present one. The table lists those helices proposed by others that are not included in the present model and our reasons for not including them.

PART 2. BEYOND SECONDARY STRUCTURE

The comparative evidence presented in part 1 together with corroborating studies on chemical modification and so on have provided a good approximation to the secondary structure of the 16S RNA. Except for a few uncertainties, it would seem that all of the major (canonical base pairing) secondary structural elements in the molecule have been detected. The next step is to refine the various helices—define the structure at their termini, provide evidence concerning noncanonical pairs that would extend them, and begin to relate neighboring helices structurally to one another. Also, the molecule must be screened for general covariances, which, to be comprehensively done, will require more (complete) sequences than now exist. (However, the large amount of extant catalog information will permit a beginning in this direction.) Thus, reconstruction will progress through stages, using a variety of approaches, toward the ultimate goal of a fully folded and detailed three-dimensional structure that incorporates the interactions of the RNA with ribosomal proteins and with other parts of the translation apparatus and reveals its molecular mechanics.

What the present study makes clear is the subtlety of pairing (and, by implication, other interactions) in RNA and also the interconnectedness of the various elements in the molecule. Helices can be recognized and distinguished by their pairing characteristics: by (i) the degree of use of noncanonical pairings, (ii) the presence or absence of bulge loops and other irregularities, (iii) the degree of phylogenetic constancy of sequence and overall length, and (iv) the patterns of sequence and their phylogenetic varia-

tion. Also, many, if not most, of the helices can be precisely located merely by highly conserved sequences that flank them, by their apex loops, or by adjacent helices. (Particularly good examples are helices 769–775/804–810 [by preceding AAA and succeeding CGUAAA] and 1241–1247/1290–1296 [by preceding ACAAU and the A-rich purine stretch preceding position 1290]).

The comparative approach indicates far more than the mere existence of a secondary structural element; it ultimately provides the detailed rules for constructing the functional form of each helix. Such rules are a transformation of the detailed physical relationships of a helix and perhaps even reflection of its detailed energetics as well. (One might envision a future time when comparative sequencing provides energetic measurements too subtle for physical chemical measurements to determine.)

G · U, G-A, and Other Noncanonical Base Pairs

Although G · U pairs in RNA structure are well known, their frequency in some of the 16S rRNA helices is unusually high. As noted, helix 829–840/846–857 in both *E. coli* and *H. volcanii* contains four contiguous G · U pairs; in eucaryotes, three of the four pairs in helix 960–963/972–975 are G · U; in *E. coli* helix 1072–1076/1081–1085, three of five are G · U. Furthermore, the constancy of G · U at 942/1341 and 1512/1523 or the exclusive replacement of G · U by U · G at 157/164 suggests specific roles for some of them.

It is known from the tRNA crystal structure that G-A pairings can occur in the normal Watson-Crick configuration; one such occurs at the distal end of the $tRNA^{[phe]}$ anticodon stem (38, 63), between the anticodon and dihydro U stalks. The extent of their occurrence in RNA structure was not apparent, however, until the present comparative analysis. That these juxtapositions are structurally bona fide pairs is indicated by the facts (i) that (in one case) cobra venom nuclease is known to cleave between two of them, i.e., GA_{1418}-GA_{1483} (16a) and (ii) that they can replace known normal pairs in the interior of helices; the pairings at positions 321/332, 661–662/743–744, and 1425–1426/1474–1475 offer examples of this. There are, of course, many more examples of A-G juxtapositions at the termini of helices, which cannot be as certain to form bona fide pairs. The concentration of A-G pairs in helix 1409–1430/1470–1491 and the phylogenetic invariance of some of these is

TABLE 39. Helices not included in the present model[a]

Proposed helix or part thereof	Sequence	Reference(s)	Status
61-64/103-106	GUCG/UGGC	7, 30, 105	Ins prf, 1 Tn (Aspergillus mitochondria); ins dspr, 2 ncp
135-136/226-227	CC/GG	79	Disproven, >4 ncp (all pos)
143-148/214-219	AGGGGG/CCUCUU	7, 30, 79, 105	Disproven, >6 ncp (all pos)
198-202/206-210	GAGGG/CCUUC	7, 30, 105	Disproven, 1 ncp; alt proven
322-324/330-332	CUG/CGG	79	Disproven, 4 ncp
340-345/357-362	UCCUAC/GUGGGG	7, 30, 105	Disproven, >5 ncp; alt proven
410-413/419-422	GAAG/CUUC	7, 30, 105	Ins prf, 1 Tv; ins dsp, 2 ncp
686-687/701-702	UA/UA	7, 30, 105	Disproven, >5 ncp versus 1 Tv
954-957/976-979	GUUU/GAAC	7, 30, 105	Disproven, >5 ncp
1054-1061/1200-1206	CAUGGCUG/CAUCAUG	79	Disproven, >4 ncp
1061-1066/1187-1192	GUCGUC/GAUGAC	7, 30, 105	Disproven, >5 ncp
1112-1115/1184-1187	CCCU/GGGG	79	Unlikely, >3 ncp
1179-1182/1209-1212	AAGG/CCUU	54, 98	Disproven, >5 ncp
1187-1192/1198-1203	GAUGAC/GUCAUC	54, 98	Disproven, >4 ncp
1253-1255/1282-1284	GAG/CUC	7, 30, 105	Ins prf, 1 Tn
1263-1267/1272-1276	CUCGC/GCAAG	79	Disproven, 5 ncp
1265-1267/1276-1278	CGC/GCG	7, 54, 79, 98	Disproven, 4 ncp
1301-1305/1335-1339	UCCGG/UCGGA	7, 30, 105	Disproven, >4 ncp
1347-1349/1376-1378	GUA/UAC	7, 30, 105	Disproven, >3 ncp
1430-1432/1469-1471	AAG/CUU	7, 30, 105	Disproven, >3 ncp, all pos
1448-1449/1454-1455	CC/GG	7, 30, 105	Ins prf, 1 Tn
1503-1506/1534-1537	AGGU/ACCU	7, 30, 105	Unlikely[b]

[a] Those helices proposed in other models (7, 30, 54, 79, 98, 105) but not included herein are shown in the left-hand column, and the reason for their exclusion is given in the right-hand column. Abbreviations used: Tv, base pair transversion evidence; Tn, base pair transition evidence; ncp, noncanonical pair created; ins, insufficient; prf, proof; dspr, disproof; all pos, all positions covered; alt, alternate.

[b] Some mitochondrial 16S-like rRNAs appear to eliminate positions 1534 to 1537 without altering the sequence 1503 to 1506.

remarkable, as is the tight cluster of A-G pairs in the middle of the protein S15 binding site (helix 655–672/734–751), seen most dramatically in the archaebacterial example, which has four of them.

In the case of $tRNA^{phe}$, the A-G pair is "propellered," i.e., it joins two helices whose axes are almost, but not quite, coaxial, one member of the A-G pair lying stacked along the anticodon axis, the other along the dihydro U axis (37, 38). Thus, A-G pairs could serve to effect slight bends in helices. (However, note that in the example cited, the G is methylated [2-Me], which itself might cause the twist [37].)

It is apparent that still other noncanonical pairings occur in proven helices. Mammalian mitochondrial examples of the various helices show a remarkable number of A-C juxtapositions. Dipyrimidine juxtapositions are a conserved feature of helices 946–955/1225–1235 and 984–990/1215–1221 in eucaryotes, as is the G-G juxtaposition in helix 153–158/163–168. The *E. coli* version of helix 144–147/175–178 would contain a G-G juxtaposition, too, and so on.

It is becoming increasingly obvious that noncanonical pairings play a significant role in RNA secondary structure. Although it seems as yet premature to attempt any quantitative analysis of these pairings, such will emerge ultimately from extensive comparative analyses of 16S and of 23S rRNAs.

Modified Residues

In all three primary kingdoms, the same regions of the 16S-like rRNA tend to be modified, although the eucaryotic versions contain, in addition, many other regions of modification, largely 2'OMe substitutions. For the most part, the regions of common modification are functionally important regions by three criteria: universality of structure, constancy of sequence, and the fact that the G residues therein are frequently kethoxal sensitive in active 30S subunits, but become protected in 70S couples or in polysomes. Examples of such regions are as follows: (i) the 520 to 535 region, in which m^7G_{527} occurs in all eubacteria and two archaebacteria, and U_{534} (probably 2'OMe) which occurs in some eubacteria and all eucaryotes; (ii) the loop at position 790, wherein eucaryotes, a few eubacteria, and several archaebacteria have Ψ_{788}; (iii) the loop at position 965, in which the base at position 966 (be it U or G) is variously modified in all cases, whereas C_{967} is modified (m^5C) in most and A_{968} in a few eubacteria; (iv) the 1400 to 1410 region, in which almost all organisms modify C_{1402}; a few eubacteria also modify (variously) position 1400 (or 1399), 1404, 1407, or 1409; one archaebacterial modification at position 1400 (or 1399) is also noted; and (v) the 1495 to 1505 region, in which almost all eubacteria modify U_{1498}, but almost all eucaryotes and all archaebacteria modify A_{1500} instead, and some archaebacteria also modify A_{1499}. Many of the regions that are modified only in the eucaryotes tend also for the above (three) reasons to be classified as important as well.

An evident characteristic of modifications in 16S-like rRNAs is their occurrence in clusters, in the sense that if one base is modified, other bases in the region tend to be modified in some other, if not the same, organism.

Many of the modifications are phylogenetically stable; if they are not universal, they tend to be kingdom specific, e.g., m^7G_{527} or m^3U_{1498} in eubacteria and $m_2^6A_{1500}$ in eucaryotes and archaebacteria. Others come and go phylogenetically, being characteristic of families, genera, or only species. It is possible that some of the modifying systems are subject to lateral, interspecific transfer. For example, the C_{m1409} is characteristic of gram-positive eubacteria, but unaccountably also turns up in one small subgroup (only) of the gram-negative purple bacteria (24, 29, 76).

Single Unpaired Residues Within a Helix

tRNAs do not contain unpaired, bulged single residues within helical regions. However, these are a notable feature of all rRNAs (including 5S RNA) (23, 30, 31, 53, 54, 58, 79, 81). In their zeal to create secondary structure, biologists too readily accept such bulged residues. Bulged residues can be reliably invoked only when the existence of the helical sections on either side of them is supported by comparative evidence. (A similar situation applies to noncanonical pairs and so on.) Many helices in Fig. 1 could be extended by bulging bases and the like. We have resisted this temptation in those cases where it cannot be rationalized by comparative evidence.

The *E. coli* 16S rRNA secondary structure as shown in Fig. 1 contains 10 bulged residues. If interior loops in which one side comprises a single residue are included, this number becomes 14. Of these, 10 are A residues, 2 are G residues, 1 is a C residue, and 1 is a U residue. A preference for bulged A residues can be seen in the other rRNAs also (23, 53, 77). Some bulged residues are phylogenetically stable in both position and composition, e.g., A_{389} and A_{1227}; others vary in both respects (or even disappear), e.g., G_{31}, A_{746}, and A_{1042}.

It is pointless to speculate on the function of bulged residues because we can say nothing about their physical relationship to the surrounding helices, and because they are merely one of the (more obvious) elements in tertiary structure, about which very little is known. In a number of cases, there is evidence to suggest

that so-called bulged residues are actually that. (i) Some, at least, are hyperreactive with reagents that do not react with stacked bases, e.g., A_{66} of the 5S rRNA (58), A_{1441} and G_{94} of 16S rRNA (H. F. Noller, unpublished data), and (ii) bulging is not always a phylogenetically constant feature, and when the bulged base is eliminated, no new pair is added to the helix, i.e., its length appears to remain the same. Examples here are G_{31} of 16S rRNA and N_{66} of 5S rRNA (23, 43, 58, 77).

Coaxial Helices

Helices immediately adjacent to one another (i.e., those between which no unpaired bases exist) could well be coaxial. This was first suggested for the common and CCA arms of tRNA (94) and proven by the tRNA crystal structure (38, 63). (As discussed above, the anticodon and dihydro U arms of $tRNA^{phe}$, with an intervening A-G pair, are also almost coaxial.) Determining which helices are coaxial would add major constraints to the developing rRNA structures, so the question of coaxiality, although highly speculative for now, is well worth consideration.

Although comparative evidence does not bear strongly on the question of coaxiality, it may be the only approach that presently provides any evidence concerning it whatsoever, short of X-ray diffraction. We would suggest that comparative evidence of the following type strongly implies coaxiality. If organism A has immediately adjacent helices of lengths m and n pairs, but in organism B, these helices are of lengths $(m + a)$ and $(n - a)$ pairs, i.e., the overall length of the combined structure is conserved, then the two helices are probably coaxial.

The most striking example of this rule involves the helix couple 1072–1076/1081–1085 and 1086–1089/1096–1099. In eubacteria, the first helix comprises five pairs, the second four pairs, but in eucaryotes, the first has six pairs and the second has three. Other candidates for coaxiality are the following. (i) Helices 9–13/21–25 and 27–37/547–556. In archaebacteria and eucaryotes, the two are immediately adjacent, for in archaebacteria, the first helix extends "forward" by one pair, while in eucaryotes, the second extends "backward" by one. In eubacteria, the two are separated by an A-G pair. (ii) Helices 9–13/21–25 and 17–20/915–918. As noted, this combined structure resembles somewhat the codon-anticodon helix perched on the anticodon stalk. (iii) Helices 39–47/394–403 and 368–371/390–393. In eubacteria, G_{394} seems a part of the first helix; in archaebacteria, it would belong to the second. (iv) Helices 113–115/312–314 and 289–295/305–311. (v) Helices 122–128/233–239 and 240–245/281–286. Strangely, the archaebacterial sequence contains two insertions (A after position 121 and G after position 239), which have the effect of inserting an A-G pair between the two helices. Their joint occurrence would seem to imply some relationship. (vi) Helices 584–587/754–757 and 588–595/644–651. (vii) Helices 673–675/715–717 and 666–672/734–740. In eucaryotes, the first helix is two pairs longer than in archaebacteria, i.e., five rather than three pairs, whereas the second is two pairs shorter. (The first of these is not convincingly present in eubacteria; see Table 19 and related discussion.) (viii) Helices 1113–1117/1183–1187 and 1118–1124/1149–1155.

Another characteristic of potentially coaxial helices seems to be that one of them can be lengthened by one or several pairs at the expense of the other, at the faces where they meet. Two examples of this are helices 122–128/233–239 and 240–242/284–286 in eucaryotes (which can also be formed as 124–128/233–237 and 238–242/284–288) and 112–115/312–315 and 316–322/331–337 in yeast mitochondria (whose alternatives are 110–115/312–317 and 318–322/331–335). Alternative formulations such as these could serve to strengthen the coaxial structure through an entropic contribution to its free energy.

If the contiguity criterion is relaxed to include those helices separated only by a single A-G pair, then many other helices become potentially coaxial, e.g., 316–322/331–337 with 339–342/347–350, or 136–142/221–227 with 144–147/175–178, or 27–37/547–556 with 39–47/394–403. As discussed above, the A-G pairs may be a means for introducing slight bends into coaxial structures.

By relaxing criteria this way, one can begin to envision large coaxial RNA "struts" that might crisscross the 30S subunit to give it shape and dynamic continuity. In the absence of more compelling evidence, however, speculation at this level is pointless.

Energetics of Helices (from a Biological Perspective)

Our understanding of the energetics of helices, the strength of interaction, is based on a set of empirical rules derived from simple model compounds (49, 66, 85). Experience with the 16S rRNA secondary structure teaches that (in their present form) such rules are of little value in deducing rRNA secondary structure. Many "energetically favored" helices have had to be rejected in constructing the present secondary structure because less-favored, mutually exclusive structures were consistent with the comparative evidence or because they were themselves inconsistent with the comparative evidence (or both). (What appears to be the strongest helix in the *E. coli* 16S rRNA sequence, i.e., 733–739/

925–931, ~−30 kcal, is phylogenetically disproven.) One does not discard the rules because of this, however. One merely applies them with less zeal, in a looser, more tentative fashion, and as a secondary, not primary, raison d'etre for double-stranded structure. The problem of the true energetics of helices in large molecules is a more complex one than we would or can now make of it.

In any case, it does seem that the various helices are not all energetically alike. There definitely seem to be strong and weak ones. The former may be less interesting than the latter. Note, for example, a small helix that occurs in all 16S-like rRNAs and so must be significant, i.e., 960–963/972–975. In eucaryotes, three of the four pairs are U · G. Similarly, in helix 894–897/902–905 of *E. coli*, two of the four pairs are G · U.

Ostensibly weak helices are known from 5S rRNA studies. The so-called tuned helix in 5S rRNA appears quite weak in certain marine gram-negative bacteria (25, 99). In fact, the apparent strength of this helix correlates with optimum growth temperature, being strongest in thermophilic bacteria (99). The only striking example of such a phenomenon in 16S rRNA was discussed above, i.e., the side bulge of the 997–1012/1017–1044 structure which is remarkably augmented in the thermophile *B. stearothermophilus*.

Weak helices may be transient structures in rRNA that open or otherwise deform at certain stages in the translation cycle. They are in effect energetically tuned.

Ribosomal Protein Binding Sites

Probably most of the 21 proteins that constitute the 30S subunit bind directly to the rRNA. However, only two of the binding sites, for proteins S8 and S15 (48, 67, 102), are well-enough localized to serve as models. S8 has been shown to bind to the 588–606/633–651 helix, and S15 to the 655–672/734–751 helix. Both are irregular helices; both are phylogenetically variable in sequence and in their irregularities. It is thus surprising that the *E. coli* version of both proteins will bind to a wide variety of 16S rRNAs (84). Zimmermann and co-workers have sequenced a number of such binding regions to determine the presumed constancies that underlie heterologous binding (84).

Other irregular, phylogenetically variable helices are then candidates for protein binding. The most obvious of these are 27–37/547–556, 821–840/846–879, 997–1012/1017–1044, and 1409–1445/1457–1491. The helix covering position 850 may bind protein S6 or S18 or both (61, 102).

Obviously, all protein-binding sites in 16S-like rRNAs need not be of the irregular, variable type. S20 protects the 252–259/267–274 region from nuclease cleavage, which protein S4 does not (19). This is a regular helix of highly constrained sequence in eubacteria.

Given the central role of protein S4 in ribosome assembly, structure, and function, its binding site is of especial interest. Nomura and co-workers have shown that S4 mRNA can be drawn in a configuration resembling helix 500–517/534–545 (56). This helix, plus the adjacent 27–37/547–556 structure (which delimits the domain protected by protein S4), are the reasonable binding site for this protein. The extent of modification in and around the latter helix in eucaryotes is remarkable. There is some group specificity in this and in the irregularities in the helix. In that alterations in protein S4 can affect the accuracy of translation (32), it is interesting to speculate that this helix, too, is involved somehow controlling accuracy.

Figure 2 summarizes what is known concerning the ribosomal protein binding sites on 16S-like rRNAs.

Overall Shape of 16S-like rRNA

It is a commentary on the state of our understanding of 16S rRNA structure that next to nothing can be said about the overall shape of the molecule or its positioning in the 30S ribosomal subunit. A combination of approaches (41, 47, 61, 80, 93) have provided a low-resolution picture of the position of the various ribosomal proteins on the subunit (as seen in the electron microscope). Some of these, in turn, position the regions of their binding sites in rRNA. In addition, the 3′ and 5′ ends of the 16S-like rRNA have been fixed in the 30S subunit (44, 46, 57, 71), as have those of two areas that involve modified bases, i.e., $m^6_2A_{1518/1519}$ and m^7G_{527} (59, 87; R. Gutell, unpublished data). Figure 3 summarizes what little is known here.

Ribosome Evolution and Essential Core 16S rRNA Structure

As defined by molecular phenotype, there exist three major groups or primary kingdoms of organisms—eubacteria, archaebacteria, and eucaryotes (97). Although the first two are procaryotes, outside of the common negative characteristic of not possessing certain eucaryotic cellular features, they resemble one another no more than either resembles the eucaryotes. There correspond to these three groupings three types of 16S-like rRNA, of characteristic length, sequence, and structure (9, 33, 65).

However, in addition to these three rRNA types, there exist mitochondrial 16S-like rRNAs more varied, more unique in structure, length, and sequence, than any of the three major organismal types (1, 21, 39, 42, 75, 89; J. J.

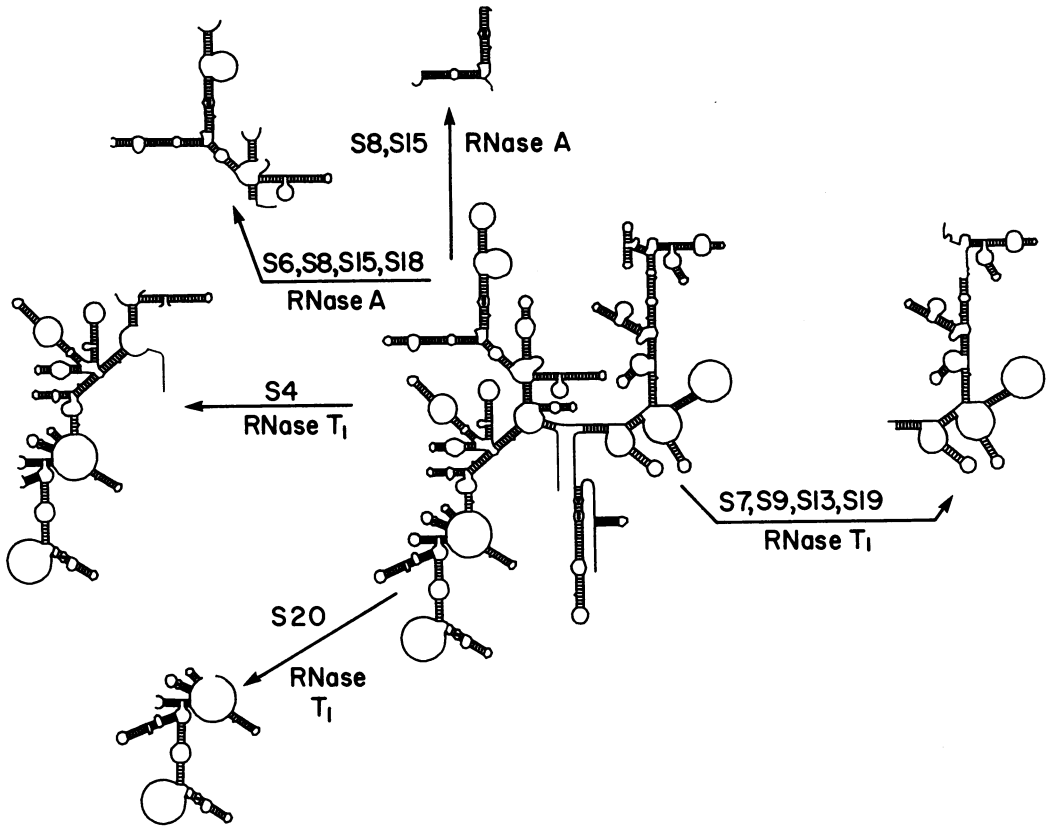


FIG. 2. RNA fragments found in ribonucleoprotein complexes after nuclease digestion of ribosomal protein-rRNA complexes (19, 48, 62, 102) or gently unfolded 30S ribosomal subunits (101). Note how the fragments correspond closely to predicted structural domains or subdomains.

Seilhamer et al., manuscript in preparation). Since mitochondria arose as endosymbionts, it has become popular to interpret their most unusual rRNAs as signifying a mitochondrial origin in simple entities too primitive to be procaryotes, or otherwise removed from the three groups of organisms. This is not the case, although we will not cast the argument here. Mitochondria have arisen from eubacterial ancestry, all probably from the same subgroup, i.e., the purple bacteria and relatives (4, 24, 29; C. R. Woese and G. E. Fox, unpublished calculations; M. Gray, personal communication). Although the mitochondrial rRNAs cannot be taken then to represent primitive stages in ribosome evolution, they are nevertheless of especial comparative value, for they appear to represent stripped-down if not streamlined versions of the translation apparatus, i.e., they tend to have retained the more essential features of the mechanism. (It must also be recognized that mitochondrial versions of the mechanism are probably cruder, less finely tuned, than are the normal versions thereof. At least, the mitochondrial rRNAs are less regular in structure than their normal counterparts, and so although of some value in defining the essence of the rRNA, they are of limited value for comparative purposes or in defining the fine points of rRNA structure and function.)

The three basic 16S-like rRNA types are contrasted in Fig. 4, an alignment of 16S-like rRNA sequences from one representative of each of the three primary kingdoms (*E. coli*, *H. volcanii*, and *D. discoideum*), and Fig. 5, a comparison of the three corresponding types of secondary structure together with the essential core of the molecule, i.e., the structure common to all versions thereof (mitochondrial type included). Of the three major types, the eucaryotic version is by far the most unique in both sequence and secondary structure. In those regions of the molecule of comparable secondary structure, the degree of sequence homology among the three is 63% (*E. coli-H. volcanii*), 56% (*H. volcanic-D. discoideum*), and 53% (*E. coli-D.*

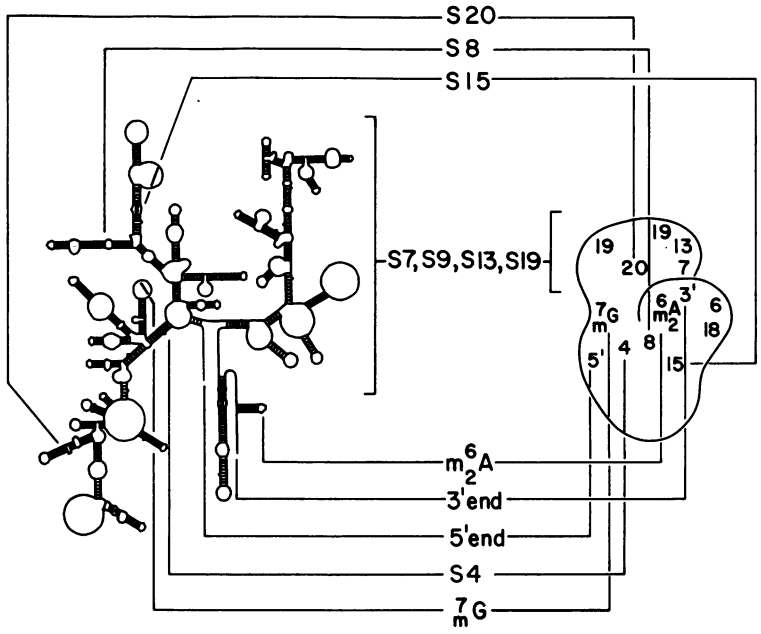


FIG. 3. Location of some structural features of 16S RNA in the electron microscope model for the 30S ribosomal subunit (41). Positions of ribosomal proteins (indicated by their respective numbers) are based on immunoelectron microscopy (41, 80) and neutron diffraction studies (47). Positions of the dimethyladenosines (m_2^6A), 7-methylguanine (m^7G), 5' terminus (5'), and 3' terminus (3') are also from immunoelectron microscopy (57, 59, 71, 80, 87).

discoideum). The unrooted phylogenetic tree for these three would then have the shortest branch for the archaebacterium, and the longest for the eucaryote (which configuration would be the same for any others of known eucaryotic or eubacterial sequences). Regardless of where the actual root of this tree occurs, the result means that the archaebacterial version of the 16S rRNA sequence is closer to the ancestral version common to all than at least one of the other two versions is.

The uniqueness of the eucaryote could reflect its having diverged from the common ancestral line before the remaining two groups split from one another, or alternatively, its having diverged further from an ancestral pattern common to all than have the others. We favor the latter alternative. It seems reasonable that the more primitive an rRNA, the more simple its structure. The eucaryotic version of the 16S-like rRNA is definitely more complex than those of its procaryotic counterparts. Helices are less regular (more noncanonical pairs appear); loops tend to be larger on average.

The *D. discoideum* rRNA sequence represents the deepest known branching in the eucaryotic tree, branching from the common eucaryotic stem before the lines of animals, plants, fungi, and (at least some) protists branched from one another. (That this is a true deep branching, rather than a false deep branching which reflects a "fast clock" in the *D. discoideum* line of descent, is shown by the fact that the *D. discoideum* rRNA sequence is as close to the bacterial sequences as are any of the other eucaryotic rRNA sequences (R. McCarroll, G. J. Olsen, Y. D. Stahl, C. R. Woese, and M. L. Sogin, Biochemistry, in press). In its secondary structure, the *D. discoideum* rRNA is not as irregular as are the other eucaryotic rRNAs; if anything, its helices (and associated structure) are more like the bacterial versions than are those of the other eucaryotes. For example, (i) the form of helix 122-128/233-239, of helix 136-142/221-227 (Table 7), and of the interior loop in *D. discoideum* is precisely the bacterial form; both helices contain only canonical pairs (except for one G · U pair), and 9 of 11 residues in the interior loop are identical to their *H. volcanii* counterparts. In all other eucaryotes, neither of the helices can be constructed without one or more noncanonical pairs (indeed, the second helix is not convincingly present in these cases), and sequence in the putative interior loop does not resemble the bacterial versions. (ii) The apex loop at position 160 comprises four residues and

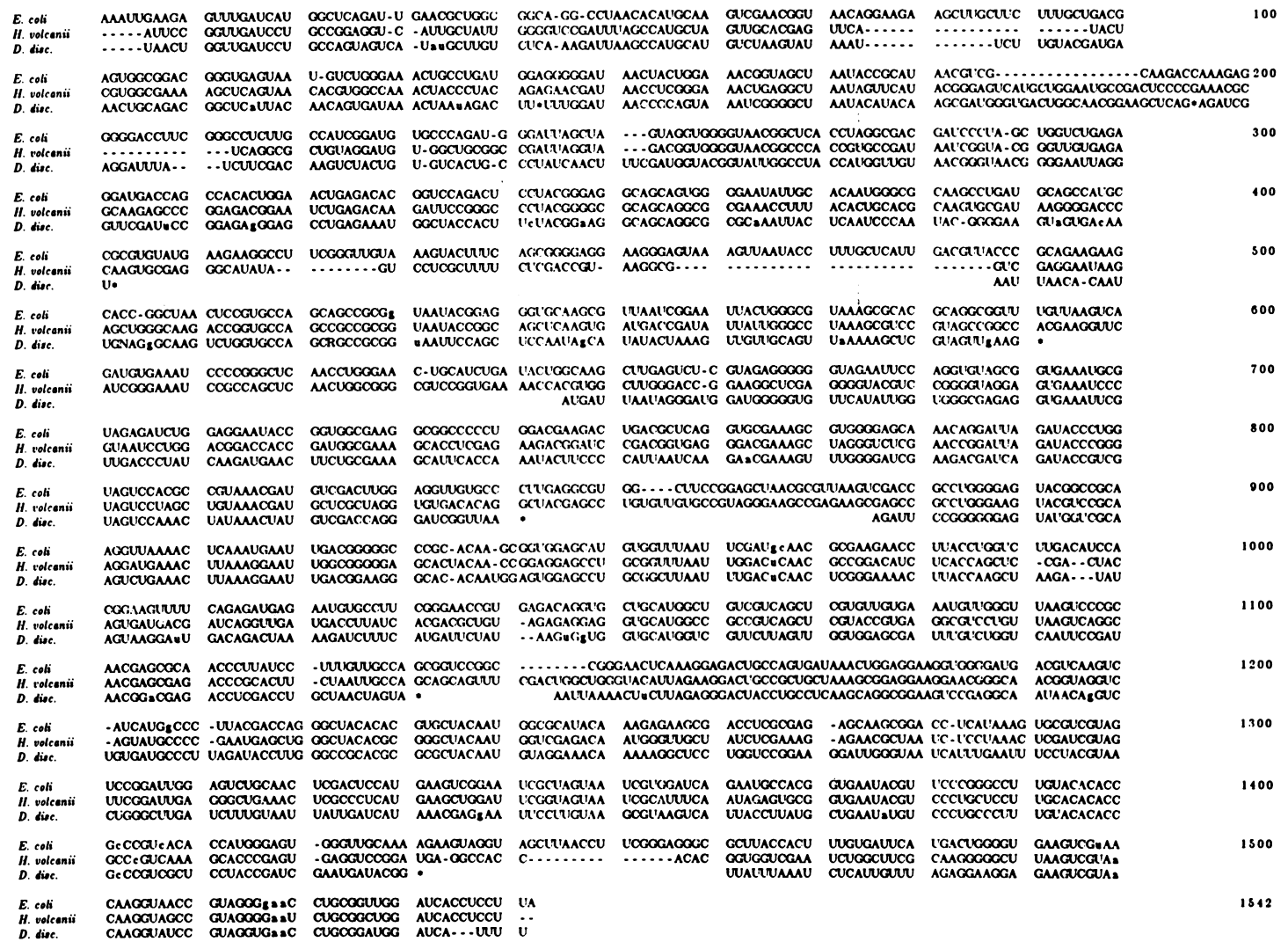


FIG. 4. Alignment of 16S-like RNA sequences from one representative of each of the three primary kingdoms, *E. coli* (8), *H. volcanii* (33), and *D. discoideum* (McCarroll et al., in press). The alignment is an empirical one based upon a "best fit" with regard to both primary and secondary structural homologies. An asterisk (*) denotes the start of (nonhomologous) portions of the eucaryotic sequences that are not reproduced.

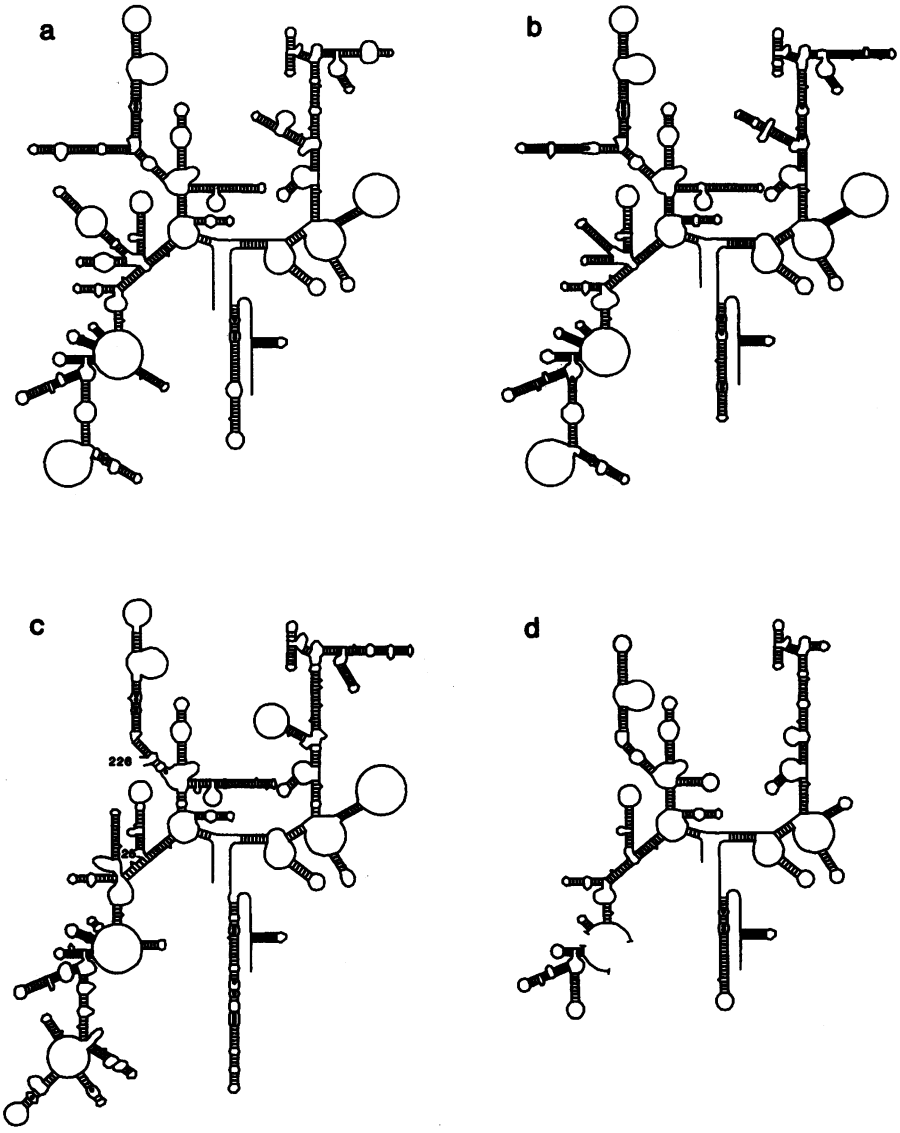


FIG. 5. Schematic comparison of secondary structure models for representative eubacterial (*E. coli*) (a), archaebacterial (*H. volcanii*) (b), eucaryotic (*S. cerevisiae*) (c), and a "minimal" small subunit (d) RNAs. The minimal structure contains only the structural features that are present in all 16S-like rRNAs thus far sequenced, including those from all three major phylogenetic lines as well as from chloroplasts and mitochondria. There is some question as to the existence of the structure corresponding to *E. coli* positions 1070 to 1100 in mammalian mitochondria. Two sections of *S. cerevisiae* 18S rRNA containing 25 and 226 nucleotides, after positions 520 and 634, respectively, are omitted from structure (c). Satisfactory base pairing schemes for these two sections that are compatible with the available evidence have not yet been determined.

a common composition, NAAA, in the procaryotic and *D. discoideum* examples, but five in the remaining eucaryotes, whose sequence is UAAUU. (iii) The small procaryotic helix 339-342/347-350 (Table 12) can be formed only in the *D. discoideum* sequence with canonical pairs and the procaryotic sequence in the loop; in the other eucaryotes, a noncanonical pair occurs in the stem and the loop sequence differs in two of the four residues from the procaryotic version. And (iv), *D. discoideum* is the only eucaryote that does not insert a residue in the vicinity of position 934 (which none of the 200-odd eubacterial examples do). Although few in number, these examples strengthen the case that the aboriginal eucaryotic 16S-like rRNA resembled

procaryotic 16S rRNAs more than its descendants now do, and so the eucaryotic uniqueness is derived, not aboriginal.

The essential core abstraction of 16S-like rRNA is surprisingly not all that much smaller than the procaryotic versions of the molecule. Although it is defined by structures common among the three kingdoms and mitochondria, approximately the same structure would emerge from an extensive characterization of eubacterial sequences (i.e., as the invariant portions of the molecule). It contains most of the sites reactive with kethoxal in active 30S subunits. Whether it is equivalent to a primitive version of the ribosomal RNA is a moot point.

Molecular Mechanics of Translation

There exists an enormous amount of descriptive information regarding ribosomes. Yet we have almost no idea as to the role(s) of ribosomes in the translation process. We do not even know which of the component macromolecules (or parts thereof) are the more important. Unfortunately, we often think we understand translation, for we can speak of tRNA "adapters" (15) that are "translocated" from an "A site" to a "P site" (92). This view of translation (i) is too imprecise to provide a genuine molecular understanding of the process; (ii) because it is strong dogma, inhibits other attempts at such an understanding; and (iii) is probably misleadingly wrong. tRNA, the adapter, is a passive, static entity that is processed by a ribosome whose "A and P sites" somehow define the underlying molecular mechanics of the process. Is any of this true? No evidence so far supports it (although much is consistent with this conceptually loose formulation). In trying to see molecular mechanism in rRNA structure (and the changes therein), the biologist may come to perceive this problem in a fresh, productive way.

On the simplest level, the mechanics of translation have three aspects: (i) a recognition of the codon (matching of codon to amino acid), (ii) a transfer of the peptide, and (iii) a movement of the mRNA relative to the peptide growing point. It is conceivable that the basis for all of these interactions resides in the tRNA molecule. (A specific proposal for such a molecular mechanism was made some time ago, but will not be discussed in detail here [95].) In any case, one can envision three general types of roles for the ribosome in translation. The ribosome could (i) define and provide the mechanical basis for mRNA movement (and peptidyl transfer), (ii) mechanically facilitate an mRNA movement (and peptidyl transfer) whose basic dynamics was inherent in the tRNA molecule, or (iii) play no direct mechanical role in translation (whose mechanism is solely defined by the tRNA molecule), but serve as some sort of damping system or thermal noise buffer, in the context of which tRNAs can function with greatly increased accuracy (96). The molecular mechanics of the ribosome would be rather different in each of the three cases.

A large body of experimental evidence attests to the fact that individual ribosomal proteins probably do not define ribosome functions. (For example, leaving one ribosomal protein out of a ribosome reconstitution rarely prevents ribosome function, provided the ribosomal particle can then form; it merely affects the quality of translation [55].) Recently, *E. coli* mutants have been isolated that lack various of the ribosomal proteins (16). However, it is easy to perturb the rRNA (in the ribosomal subunit) in ways that stop translation (5, 52, 70). Then, too, so-called nonenzymatic translation can occur in which an in vitro system can be constructed without elongation factors, without guanosine triphosphate, and so on (28). Although such a system translates slowly, it does so with at least as much accuracy as does a complete translation system (27). These facts suggest that ribosome function is in the main (if not solely) defined by rRNA and that the proteins serve only to facilitate this.

As a prelude to discussing possible general molecular mechanical movement in rRNA, it should be asked what evidence exists suggesting movement, deformation, etc., of the 16S rRNA (or ribosomal particles in general). Little does. Residues within a few helices are susceptible to chemical modification in free 16S rRNA, which, however, could be interpreted in other, trivial, ways than as evidence for functional change in structure. The most convincing evidence of this sort is that formation of the 70S ribosome renders certain G residues in 16S rRNA more reactive with kethoxal than they are in active 30S subunits. (Specifically, these are the G's at positions 1053 and 1068 in 16S rRNA [13].)

Another type of evidence suggesting movement would be the existence of two (or more) mutually exclusive helices both of which were proven by comparative evidence. Brimacombe (7) suggested three overlapping pairings for the region covering position 1060 in 16S rRNA. These are helices 386-400/1053-1068 (bulge U_{1062}), 35-45/1059-1069, and 1055-1066/1187-1202 (bulge 1193 to 1196). However, in other organisms (in which sequences are appreciably different than the *E. coli* sequence in these regions), these proposed structures cannot be formed without introducing a substantial number of mispairs, and it is absurd to think that only *E. coli* would possess "perfect" versions of the structures. Therefore, they do not reasonably exist. (For example, the *H. volcanii* and *E. coli* versions of the first proposed helix differ in

10 of the 30 positions, which produce eight mispairs; the one canonical pair replacement that does occur is not statistically significant. The chloroplast version of the structure differs by three base replacements, which introduce three mispairs. For the second proposed helix, comparisons reveal a similar degree of mispairing—four for *H. volcanii*, six for *X. laevis*, and five for *Aspergillus* mitochondria. The third helix agrees in part with that seen in Table 28; however, its upper portion (beyond the bulge loop) is in disagreement with the one presented here.)

Nevertheless, there does appear to be a little evidence for mutually exclusive helices (as discussed in connection with Tables 11, 29, and 30).

It is perhaps instructive at this point to give an extreme example of pairing that seems not really to exist, as a cautionary note to those who would play matching games with large rRNA (or other) sequences. The human mitochondrial 16S-like rRNA sequence (21) will form a perfect helix (684-697/711-724 in mitochondrial numbering) of 14 base pairs, only 2 of which are of the G · U type. Yet this structure is incompatible with another (554-563/707-717 [bulge A_{709}]), which has an exact counterpart in all other organisms, i.e., 946-955/1225-1235 (in *E. coli* numbers). The former helix is found only in human mitochondrial rRNA.

Three types of potential molecular movement in rRNAs and tRNAs are both sufficiently defined and general enough to be worth considering at this time. One is coiling and uncoiling of particular helices. Such a mechanism was suggested by the existence of energetically weak helices in all rRNAs, particularly the case in 5S rRNA in which the central, or tuned, helix seems to vary in strength according to the optimum growth temperature of the organism (99). However, the uncoiling of one helix would reasonably demand concomitant formation of some other structure, to minimize the overall change in energy.

A second mechanism would be the type suggested for the tRNA-mRNA interaction. (The anticodon-codon helix, of three pairs, can exist as a helical extension of either the 5′ or the 3′ strand of the underlying double-helical anticodon stem. A switching between the two conformations would then produce movement, in this case movement of mRNA through the translation mechanism [95].) As seen above, helix 17-20/915-918 is formed within the loop of seven residues defined by helix 9-13/21-25. It is also possible to form the helix 14-17/919-922 within this loop. The two helices that can be formed within this loop are mutually exclusive (if the loop exists in both cases) and are somewhat analogous to the two postulated conformation of anticodon-codon helix (95). Unfortunately, no comparative evidence exists for the second of the two possibilities, and mammalian mitochondria appear to offer, if not a disproof, a variation.

The third mechanism worth considering is movement generated by the stacking and unstacking of various helices. As seen above, a number of helices in the rRNAs may be coaxial. The breaking of coaxiality (with possible formation of alternate coaxial structures) would create significant deformation, movement, in rRNA. A case has been made for such a mechanism in 5S rRNA (77). In the present instance, the possibility for a coaxial "switch" exists for helix 27-37/547-556 either with helix 39-47/394-403 or with helix 500-517/534-545. (An A-G pair separates either couple in the *E. coli* sequence.) Another potential switch of this sort may be the protein S8 versus S15 binding sites with the helix that underlies both (584-587/754-757).

CONCLUSION

At the present state of our understanding, 16S rRNA is merely a collection of individual helices. It ultimately must come to be a three-dimensional structure that (probably) moves, interacts with other molecular species, and has a certain evolutionary history that is reflected in its function. In this development, comparative studies will play an important role, moreso because rRNAs are so large and complex in structure. Comparative evidence should prove of particular value in testing various hypothetical structures, e.g., 16S-23S interactions. Phylogenetic constancy of sequence is an indicator of (functionally) important areas in the molecule. And the permissible ways in which regions of the rRNA vary phylogenetically must measure some structural or energetic constraints (or both) basic to their functioning. What is not generally appreciated is that the translation apparatus is an evolutionary mechanism, and one cannot fully understand its workings without understanding its evolution. In this undertaking, the comparative approach will be essential.

ACKNOWLEDGMENTS

The (unpublished) experiments cited herein were supported by grants (to C.R.W.) from the National Aeronautics and Space Administration and (to H.F.N.) from the U.S. Public Health Service.

We are grateful to Norman Speedy Pace for discussions of this manuscript, to Tom Macke for his considerable help in preparing Fig. 4, and to K. Katze for sharing with us various unpublishable insights.

LITERATURE CITED

1. **Anderson, S., M. H. L. deBruijn, A. R. Coulson, I. C. Eperon, F. Sanger, and I. G. Young.** 1982. Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. J. Mol. Biol. 156:683–717.
2. **Baan, R. A., C. W. Hilbers, R. Van Charldorp, E. Van**

Leerdam, P. H. Van Knippenberg, and L. Bosch. 1977. High-resolution proton magnetic resonance study of the secondary structure of the 3'-terminal 49-nucleotide fragment of 16S rRNA from *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. 74:1028–1031.

3. Bhanot, O. S., and R. W. Chambers. 1977. Bisulfite-induced C → U transitions in yeast alanine tRNA. J. Biol. Chem. 252:2551–2559.
4. Bonen, L., R. S. Cunningham, M. W. Gray, and W. F. Doolittle. 1977. Wheat mitochondrial 18S ribosomal RNA: evidence for its prokaryotic nature. Nucleic Acids Res. 4:663–671.
5. Bowman, C. M., J. E. Dahlberg, T. Ikemura, J. Konisky, and M. Nomura. 1971. Specific inactivation of 16S ribosomal RNA induced by colicin E3 *in vivo*. Proc. Natl. Acad. Sci. U.S.A. 68:964–968.
6. Branlant, C., A. Krol, M. A. Machatt, J. Pouyet, J. Ebel, K. Edwards, and H. Kössel. 1981. Primary and secondary structures of *Escherichia coli* MRE 600 23S ribosomal RNA. Comparison with models of secondary structure for maize chloroplast 23S rRNA and for large portions of mouse and human 16S mitochondrial rRNAs. Nucleic Acids Res. 9:4303–4324.
7. Brimacombe, R. 1980. Secondary structure homology between *Escherichia coli* 16S and *Saccharomyces cerevisiae* 18S ribosomal RNA. Biochemistry Int. 1:162–171.
8. Brosius, J., T. J. Dull, D. D. Sleeter, and H. F. Noller. 1981. Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli*. J. Mol. Biol. 148:107–127.
9. Brosius, J., M. L. Palmer, J. P. Kennedy, and H. F. Noller. 1978. Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. 75:4801–4805.
10. Brow, D. A., and H. F. Noller. 1983. Protection of ribosomal RNA from kethoxal in polyribosomes: implications of specific sites in ribosome function. J. Mol. Biol. 163:27–46.
11. Carbon, P., C. Ehresmann, B. Ehresmann, and J. P. Ebel. 1979. The complete nucleotide sequence of the ribosomal 16S RNA from *Escherichia coli*. Experimental details and cistron heterogeneities. Eur. J. Biochem. 100:399–410.
12. Chambliss, G., G. R. Craven, J. Davies, K. Davis, L. Kahan, and M. Nomura (ed.). 1980. Ribosomes: structure, function, and genetics. University Park Press, Baltimore.
13. Chapman, N. M., and H. F. Noller. 1977. Protection of specific sites in 16S RNA from chemical modification by association of 30S and 50S ribosomes. J. Mol. Biol. 109:131–149.
14. Cole, M. D., M. Beer, T. Koller, W. A. Strycharz, and M. Nomura. 1978. Electron microscopic determination of the binding sites of ribosomal proteins S4 and S8 on 16S RNA. Proc. Natl. Acad. Sci. U.S.A. 75:270–274.
15. Crick, F. H. C. 1963. The recent excitement in the coding problem, p. 163–217. *In* J. N. Davidson and W. E. Cohn (ed.), Progress in nucleic acid research. Academic Press, Inc., New York.
16. Dabbs, E. R., R. Ehrlich, R. Hasenbank, B. H. Schroeter, M. Stöffler-Meilicke, and G. Stöffler. 1981. Mutants of *E. coli* lacking ribosomal protein L1. J. Mol. Biol. 149:553–578.

16a. Douthwaite, S., A. Christensen, and R. A. Garrett. 1983. Higher order structure in the 3'-minor domain of small subunit ribosomal RNAs from a Gram negative bacterium, a Gram positive bacterium, and a eukaryote. J. Mol. Biol. 169:249–279.

17. Dunn, J. J., E. Buzash-Pollert, and F. W. Studier. 1978. Mutations of bacteriophage T7 that affect initiation of synthesis of the gene 0.3 protein. Proc. Natl. Acad. Sci. U.S.A. 75:2741–2745.
18. Ehresmann, B., C. Backendorf, C. Ehresmann, and J. P. Ebel. 1977. Characterization of the regions from *E. coli* 16S RNA covalently linked to ribosomal proteins S4 and S20 after ultraviolet irradiation. FEBS Lett. 78:261–266.
19. Ehresmann, C., P. Stiegler, P. Carbon, E. Ungewickell, and R. A. Garrett. 1980. The topography of the 5' end of 16S RNA in the presence and absence of ribosomal proteins S4 and S20. Eur. J. Biochem. 103:439–446.
20. Ehresmann, C., P. Stiegler, P. Fellner, and J. Ebel. 1975. The determination of the primary structure of the 16S ribosomal RNA of *Escherichia coli*. III. Further studies. Biochimie 57:711–748.
21. Eperon, I. C., S. Anderson, and D. P. Nierlich. 1980. Distinctive sequence of human mitochondrial ribosomal RNA genes. Nature (London) 286:460–467.
22. Fellner, P., and F. Sanger. 1968. Sequence analysis of specific areas of the 16S and 23S ribosomal RNAs. Nature (London) 219:236–238.
23. Fox, G. E., K. R. Luehrsen, and C. R. Woese. 1982. Archaebacterial 5S ribosomal RNA. Zentralbl. Bakteriol. Parasitenkd. Infektionskr. Hyg. Abt. 1 Orig. Reihe C 3:330–345.
24. Fox, G. E., E. Stackebrandt, R. B. Hespell, J. Gibson, J. Maniloff, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. Tanner, L. J. Magrum, L. B. Zablen, R. Blakemore, R. Gupta, L. Bonen, B. J. Lewis, D. A. Stahl, K. R. Luehrsen, K. N. Chen, and C. R. Woese. 1980. The phylogeny of prokaryotes. Science 209:457–463.
25. Fox, G. E., and C. R. Woese. 1975. 5S RNA secondary structure. Nature (London) 256:505–507.
26. Garrett, R. A. 1979. The structure, assembly, and function of ribosomes. Int. Rev. Biochem. 25:121–177.
27. Gavrilova, L. P., O. E. Kostiashkina, V. E. Kotelianskv, N. M. Rutkevitch, and A. S. Spirin. 1976. Factor-free ("non-enzymic") and factor-dependent systems of translocation of poly-uridylic acid by *Escherichia coli* ribosomes. J. Mol. Biol. 101:537–552.
28. Gavrilova, L. P., and A. S. Spirin. 1971. Stimulation of "non-enzymic" translocation in ribosomes by p-chloromercuribenzoate. FEBS Lett. 17:324–326.
29. Gibson, J., E. Stackebrandt, L. B. Zablen, R. Gupta, and C. R. Woese. 1979. A phylogenetic analysis of the purple photosynthetic bacteria. Curr. Microbiol. 3:59–64.
30. Glotz, C., and R. Brimacombe. 1980. An experimentally-derived model for the secondary structure of the 16S ribosomal RNA from *Escherichia coli*. Nucleic Acids Res. 8:2377–2395.
31. Glotz, C., C. Zwieb, and R. Brimacombe. 1981. Secondary structure of the large ribosomal RNA from *Escherichia coli*, *Zea mays* chloroplast, and human mitochondrial ribosomes. Nucleic Acids Res. 9:3287–3306.
32. Gorini, L. 1971. Ribosomal discrimination of tRNAs. Nature (London) New Biol. 234:261–264.
33. Gupta, R., J. M. Lanter, and C. R. Woese. 1983. Sequence of the 16S ribosomal RNA from *Halobacterium volcanii*, an archaebacterium. Science 221:656–659.
34. Helser, T. L., J. E. Davies, and J. E. Dahlberg. 1972. Mechanism of Kasugamycin resistance in *Escherichia coli*. Nature (London) New Biol. 235:6–9.
35. Herr, W., N. M. Chapman, and H. F. Noller. 1979. Mechanism of ribosomal subunit association: discrimination of specific sites in 16S RNA essential for association activity. J. Mol. Biol. 130:433–449.
36. Hogan, J. J., and H. F. Noller. 1978. Altered topography of 16S RNA in the inactive form of *Escherichia coli* 30S ribosomal subunits. Biochemistry 17:587–593.
37. Holbrook, S. R., J. L. Sussman, R. W. Warrant, and S.-H. Kim. 1978. Crystal structure of yeast phenylalanine transfer RNA. II. Structural features and functional implications. J. Mol. Biol. 123:631–660.
38. Kim, S.-H., F. L. Suddath, G. J. Quigley, A. McPherson, J. L. Sussman, A. Wang, N. C. Seeman, and A. Rich. 1974. Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. Science 185:435–440.
39. Köchel, H. G., and H. Küntzel. 1981. Nucleotide sequence of the *Aspergillus nidulans* mitochondrial gene coding for the small ribosomal subunit RNA: homology to *E. coli* 16S rRNA. Nucleic Acids Res. 9:5689–5696.

40. **Kurland, C. G.** 1977. Structure and function of the bacterial ribosome. Annu. Rev. Biochem. **46**:173–200.
41. **Lake, J. A.** 1980. Ribosome structure and functional sites, p. 207–236. *In* G. Chambliss, G. R. Craven, J. Davies, K. Davis, L. Kahan, and M. Nomura (ed.), Ribosomes: structure, function, and genetics. University Park Press, Baltimore.
42. **Li, M., A. Tzagoloff, K. Underbrink-Lyon, and N. C. Martin.** 1982. Identification of the paromomycin-resistance mutation in the 15S rRNA gene of yeast mitochondria. J. Biol. Chem. **257**:5921–5928.
43. **Luehrsen, K. R., G. E. Fox, M. W. Kilpatrick, R. J. Walker, H. Doomday, G. Krupp, and H. J. Gross.** 1981. The nucleotide sequence of the 5S rRNA from the archaebacterium *Thermoplasma acidophilum*. Nucleic Acids Res. **9**:965–970.
44. **Lührmann, R., M. Stöffler-Meilicke, and G. Stöffler.** 1981. Localization of the 3′ end of 16S rRNA in *Escherichia coli* 30S ribosomal subunit by immunoelectron microscopy. Mol. Gen. Genet. **182**:369–376.
45. **Mankin, A. S., E. A. Skripkin, N. V. Chichkova, A. M. Kopylov, and A. A. Bogdanov.** 1981. An enzymatic approach for localization of oligodeoxyribonucleotide binding sites on RNA. Application to studying rRNA topography. FEBS Lett. **131**:253–256.
46. **Mochalova, L. V., I. N. Shatsky, A. A. Bogdanov, and V. D. Vasiliev.** 1982. Topography of RNA in the ribosome. Localization of the 16S RNA 5′ end by immune electron microscopy. J. Mol. Biol. **159**:637–650.
47. **Moore, P. B.** 1980. Scattering studies of the three-dimensional organization of the *E. coli* ribosome, p. 111–133. *In* G. Chambliss, G. R. Craven, J. Davies, K. Davis, L. Kahan, and M. Nomura (ed.), Ribosomes: structure, function, and genetics. University Park Press, Baltimore.
48. **Müller, R., R. A. Garrett, and H. F. Noller.** 1979. The structure of the RNA binding site of ribosomal proteins S8 and S15. J. Biol. Chem. **254**:3873–3878.
49. **Ninio, J.** 1979. Prediction of pairing schemes in RNA molecules. Loop contributions and energy of wobble and non-wobble pairs. Biochemie **61**:1133–1150.
50. **Noller, H. F.** 1974. Topography of 16S RNA in 30S ribosomal subunits. Nucleotide sequences and location of sites of reaction with kethoxal. Biochemistry **13**:4694–4703.
51. **Noller, H. F.** 1980. Structure and topography of ribosomal RNA, p. 3–22. *In* G. Chambliss, G. R. Craven, J. Davies, K. Davis, L. Kahan, and M. Nomura (ed.), Ribosomes: structure, function, and genetics. University Park Press, Baltimore.
52. **Noller, H. F., and J. B. Chaires.** 1972. Functional modification of 16S ribosomal RNA by kethoxal. Proc. Natl. Acad. Sci. U.S.A. **69**:3115–3118.
53. **Noller, H. F, J. Kop, V. Wheaton, J. Brosius, R. R. Gutell, A. M. Kopylov, F. Dohme, W. Herr, D. A. Stahl, R. Gupta, and C. R. Woese.** 1981. Secondary structure model for 23S ribosomal RNA. Nucleic Acids Res. **9**:6167–6189.
54. **Noller, H. F., and C. R. Woese.** 1981. Secondary structure of 16S ribosomal RNA. Science **212**:403–411.
55. **Nomura, M., S. Mizushima, M. Ozaki, P. Traub, and C. V. Lowry.** 1969. Structure and function of ribosomes and their molecular components. Cold Spring Harbor Symp. Quant. Biol. **34**:49–61.
56. **Nomura, M., J. L. Yates, D. Dean, and L. E. Post.** 1980. Feedback regulation of ribosomal protein gene expression in *Escherichia coli*: structural homology of ribosomal RNA and ribosomal protein mRNA. Proc. Natl. Acad. Sci. U.S.A. **77**:7084–7088.
57. **Olson, H. M., and D. G. Glitz.** 1979. Ribosome structure: localization of 3′ end of RNA in small subunit by immunoelectron microscopy. Proc. Natl. Acad. Sci. U.S.A. **76**:3769–3773.
58. **Peattie, D. A., S. Douthwaite, R. A. Garrett, and H. F. Noller.** 1981. A "bulged" double helix in a RNA-protein contact site. Proc. Natl. Acad. Sci. U.S.A. **78**:7331–7335.
59. **Politz, S. M., and D. G. Glitz.** 1977. Ribosome structure: localization of N6,N6-dimethyladenosine by electron microscopy of a ribosome-antibody complex. Proc. Natl. Acad. Sci. U.S.A. **74**:1468–1472.
60. **Prince, J. B., B. H. Taylor, D. L. Thurlow, J. Ofengand, and R. A. Zimmermann.** 1982. Covalent crosslinking of tRNA1val to 16S RNA at the ribosomal P site: identification of crosslinked residues. Proc. Natl. Acad. Sci. U.S.A. **79**:5450–5454.
61. **Ramakrishnan, V. R., S. Yabuki, I.-Y. Sillers, D. G. Schindler, D. M. Engelman, and P. B. Moore.** 1981. Positions of proteins S6, S11 and S15 in the 30S ribosomal subunit of *Escherichia coli*. J. Mol. Biol. **153**:739–760.
62. **Rinke, Y., A. Yuki, and R. Brimacombe.** 1976. Studies on the environment of protein S7 within the 30S subunit of *Escherichia coli* ribosomes. Eur. J. Biochem. **64**:77–89.
63. **Robertus, J. D., J. E. Ladner, J. T. Finch, D. Rhodes, R. D. Brown, B. F. C. Clark, and A. Klug.** 1974. Structure of yeast phenylalanine transfer RNA at 3 Å resolution. Nature (London) **250**:546–551.
64. **Rubtsov, P. M., M. M. Musakhanov, V. M. Zakharyev, A. S. Krayev, K. G. Skryabin, and A. A. Bayev.** 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.
65. **Salim, M., and E. H. Maden.** 1981. Nucleotide sequence of *Xenopus laevis* 18S ribosomal RNA inferred from gene sequence. Nature (London) **291**:205–208.
66. **Salser, W.** 1977. Globin mRNA sequences: analysis of base pairing and evolutionary implications. Cold Spring Harbor Symp. Quant. Biol. **42**:985–1002.
67. **Schaup, H. W., M. L. Sogin, C. G. Kurland, and C. R. Woese.** 1973. Localization of a binding site for ribosomal protein S8 within the 16*S* ribosomal ribonucleic acid of *Escherichia coli*. J. Bacteriol. **115**:82–87.
68. **Schaup, H. W., M. L. Sogin, and C. Woese.** 1971. Characterization of an RNA "binding site" for a specific ribosomal protein of *Escherichia coli*. Mol. Gen. Genet. **114**:1–8.
69. **Schwarz, Z., and H. Kössel.** 1980. The primary structure of 16S rDNA from *Zea mays* chloroplast is homologous to *E. coli* 16S rRNA. Nature (London) **283**:739–742.
70. **Senior, B. W., and I. B. Holland.** 1971. Effect of colicin E3 upon the 30S ribosomal subunit of *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **68**:959–963.
71. **Shatsky, I. N., L. V. Mochalova, M. S. Kojouharova, A. A. Bogdanov, and V. D. Vasiliev.** 1979. Localization of the 3′ end of *Escherichia coli* 16S RNA by electron microscopy of antibody-labelled subunits. J. Mol. Biol. **133**:501–515.
72. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. U.S.A. **71**:1342–1346.
73. **Sogin, M. L., B. Pace, N. R. Pace, and C. R. Woese.** 1971. Primary structural relationship of p16 to m16 ribosomal RNA in *E. coli*. Nature (London) **232**:48–49.
74. **Sor, F., and H. Fukuhara.** 1980. Sequence nucleotidique du gene de l'ARN ribosomique 15S mitochondrial de la levure. C.R. Acad. Sci. **291**:933–936.
75. **Sor, F., and H. Fukuhara.** 1982. Nucleotide sequence of the small ribosomal RNA gene from the mitochondria of *Saccharomyces cerevisiae*, p. 255–262. *In* P. Slonimski, P. Borst, and G. Attardi (ed.), Mitochondrial genes. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
76. **Stackebrandt, E., B. J. Lewis, and C. R. Woese.** 1980. The phylogenetic structure of the coryneform group of bacteria. Zentralbl. Bakteriol. Parasitenkd. Infektionskr. Hyg. Abt. 1 Orig. Reihe C **1**:137–149.
77. **Stahl, D. A., K. R. Luehrsen, C. R. Woese, and N. R. Pace.** 1981. An unusual 5S rRNA from *Sulfolobus acidocaldarius*, and its implications for a general 5S rRNA

structure. Nucleic Acids Res. 9:6129–6137.
78. **Steitz, J., and K. Jakes.** 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. U.S.A. 72:4734–4738.
79. **Stiegler, P., P. Carbon, J. P. Ebel, and C. Ehresmann.** 1981. A general secondary-structure model for procaryotic and eucaryotic RNAs of the small ribosomal subunits. Eur. J. Biochem. 120:487–495.
80. **Stöffler, G., R. Bald, B. Kastner, R. Lührmann, M. Stöffler-Meilicke, G. Tischendorf, and B. Tesche.** 1980. Structural organization of the *Escherichia coli* ribosome and localization of functional domains, p. 171–205. *In* G. Chambliss, G. R. Craven, J. Davies, K. Davis, L. Kahan, and M. Nomura (ed.), Ribosomes: structure; function, and genetics. University Park Press, Baltimore.
81. **Studnicka, G. M., F. A. Eiserling, and J. A. Lake.** 1981. A unique secondary folding pattern for 5S RNA corresponds to the lowest energy homologous secondary structure in 17 different prokaryotes. Nucleic Acids Res. 9:1885–1904.
82. **Teterina, N. L., A. M. Kopylov, and A. A. Bogdanov.** 1980. Segments of the 16S RNA located on the surface of the 30S subunits of *Escherichia coli* ribosomes. FEBS Lett. 116:265–268.
83. **Thompson, J. F., and J. E. Hearst.** 1983. Structure-function relations in *E. coli* 16S RNA. Cell 33:19–24.
84. **Thurlow, D. L., and R. A. Zimmermann.** 1982. Evolution of protein binding regions of archaebacterial, eubacterial and eukaryotic ribosomal RNAs, p. 347. *In* O. Kandler (ed.), Archaebacteria. Gustav Fischer, Verlag, Stuttgart.
85. **Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. 246:40–41.
86. **Tohdoh, N., and M. Sugiura.** 1982. The complete nucleotide sequence of a 16S ribosomal RNA gene from tobacco chloroplasts. Gene 17:213–218.
87. **Trempe, M. R., K. Ohgi, and D. G. Glitz.** 1982. Ribosome structure. Localization of 7-methylguanosine in the small subunits of *Escherichia coli* and chloroplast ribosomes by immunoelectron microscopy. J. Biol. Chem. 257:9822–9829.
88. **Turner, S., J. F. Thompson, J. E. Hearst, and H. F. Noller.** 1982. Identification of a site of psoralen crosslinking in *E. coli* 16S ribosomal RNA. Nucleic Acids Res. 10:2839–2849.
89. **Van Etten, R. A., M. W. Walberg, and D. A. Clayton.** 1980. Precise localization and nucleotide sequence of the mouse mitochondrial rRNA genes and three immediately adjacent noval tRNA genes. Cell 22:157–170.
90. **Vassilenko, S. K., P. Carbon, J. P. Ebel, and C. Ehresmann.** 1981. Topography of 16S RNA in 30S subunits and 70S ribosomes accessibility to cobra venom ribonuclease. J. Mol. Biol. 152:699–721.
91. **Vassilenko, S. K., and V. C. Ryte.** 1975. Isolation of highly purified ribonuclease from Central Asian cobra (*Naja oxiana*) venom. Biokhimya 40:578–583. (In Russian.)
92. **Watson, J. D.** 1964. The synthesis of proteins upon ribosomes. Bull. Soc. Chim. Biol. 46:1399–1425.
93. **Winkelmann, D. A., L. Kahan, and J. A. Lake.** 1982. Ribosomal protein S4 is an internal protein: localization by immunoelectron microscopy on protein-deficient subribosomal particles. Proc. Natl. Acad. Sci. U.S.A. 79:5184–5188.
94. **Woese, C. R.** 1967. The genetic code: the molecular basis of genetic expression. Harper & Row, Publishers, New York.
95. **Woese, C. R.** 1970. Molecular mechanics of translation: a reciprocating ratchet mechanism. Nature (London) 226:817–820.
96. **Woese, C. R.** 1972. The evolution of cellular tape reading processes and macromolecular complexity, p. 326–365. *In* Brookhaven Symposium.
97. **Woese, C. R., and G. E. Fox.** 1977. The phylogenetic structure of the procaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. U.S.A. 74:5088–5090.
98. **Woese, C. R., L. J. Magrum, R. Gupta, R. B. Siegel, D. A. Stahl, J. Kop, N. Crawford, J. Brosius, R. Gütell, J. J. Hogan, and H. F. Noller.** 1980. Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. Nucleic Acids Res. 8:2275–2293.
99. **Woese, C. R., C. D. Pribula, G. E. Fox, and L. B. Zablen.** 1975. The nucleotide sequence of the 5S ribosomal RNA from a photobacterium. J. Mol. Evol. 5:35–46.
100. **Yuan, R. C., J. A. Steitz, P. B. Moore, and D. M. Crothers.** 1979. The 3′ terminus of 16S rRNA: secondary structure and interactions with ribosomal protein S1. Nucleic Acids Res. 7:2399–2418.
101. **Yuki, A., and R. Brimacombe.** 1975. Nucleotide sequences of *Escherichia coli* 16S RNA associated with ribosomal proteins S7, S9, S10, S14 and S19. Eur. J. Biochem. 56:23–34.
102. **Zimmermann, R. A., G. A. Mackie, A. Muto, R. A. Garrett, E. Ungewickell, C. Ehresmann, P. Stiegler, J. P. Ebel, and P. Fellner.** 1975. Location and characteristics of ribosomal protein binding sites in the 16S RNA of *Escherichia coli*. Nucleic Acids Res. 2:279–302.
103. **Zimmermann, R. A., A. Muto, P. Fellner, C. Ehresmann, and C. Branlant.** 1972. Location of ribosomal protein binding sites on 16S ribosomal RNA. Proc. Natl. Acad. Sci. U.S.A. 69:1282–1286.
104. **Zimmermann, R. A., and K. Singh-Bergmann.** 1979. Binding sites for ribosomal proteins S8 and S15 in the 16S RNA of *Escherichia coli*. Biochim. Biophys. Acta 563:422–431.
105. **Zwieb, C., C. Glotz, and R. Brimacombe.** 1981. Secondary structure comparisons between small subunit ribosomal RNA molecules from six different species. Nucleic Acids Res. 9:3621–3640.