

# Evidence for several higher order structural elements in ribosomal RNA

(16S ribosomal RNA/tertiary structure/comparative analysis/mycoplasma/mitochondria)

C. R. WOESE\* AND R. R. GUTELL\*†

\*Department of Microbiology, University of Illinois, 131 Burrill Hall, Urbana, IL 61801; and †Cangene Corp., 3403 American Drive, Mississauga, ON L4V 1T4, Canada

Contributed by C. R. Woese, January 30, 1989

**ABSTRACT** Comparative analysis of small subunit ribosomal RNA sequences suggests the existence of two new higher order interactions: (i) a double-helical structure involving positions 505–507 and 524–526 (*Escherichia coli* numbering) and (ii) an interaction between the region of position 130 and the helix located approximately between positions 180 and 195. In the first of these, one of the strands of the helix exists in the bulge loop, and the other strand exists in the terminal loop of a previously recognized compound helix involving positions 500–545. Therefore, the new structure formally represents a pseudoknot. In the second, the insertion/deletion of a nucleotide in the vicinity of position 130 correlates with the length of the helix in the 180–195 region, the latter having a 3-base-pair stalk when the base in question is deleted and a stalk of  $\approx 10$  pairs when it is inserted.

The secondary structures of the small and large subunit rRNAs represent one of the great triumphs of comparative analysis (1–4). Now the question is whether comparative analysis will prove as useful in elucidating the three-dimensional (“tertiary”) folding of rRNA and the overall structure of the ribosome. Undoubtedly it will not. Tertiary structure is less simply defined (not merely a matter of simple 1:1 correspondences), is less extensive (does not involve relatively long contiguous stretches of base pairs), and tends to involve nucleotides whose compositions change seldom if at all. The tertiary-structure problem requires more direct experimental approaches; and one can see it beginning to emerge as these are successfully applied (5–8).

This is not to say, however, that comparative analysis will not make a significant contribution to the tertiary folding problem. Olsen has shown that tRNA sequence comparisons reveal the molecule’s secondary structure, deriving the base pairing rules in the process, and most of its tertiary interactions as well when the highly variable class of mitochondrial tRNAs are added to the set of sequences (9). Comparative analysis has also already detected a few tertiary interactions in rRNAs—e.g., the pseudoknot involving the two 16S rRNA helices at positions (9–13)·(21–25), and (17–19)·(916–918) (3) and the tertiary interaction involving positions 570 and 866 (10). It is clear that most of the direct experimental approaches now in use will not be able to define tertiary interactions in precise detail so that, to the extent it is successful, comparative analysis (together with its conceptual experimental equivalent, mutation second-site reversion studies) will help to refine the structures determined by these other methods and to test the validity of suggested interactions.

The present communication deals with two tertiary elements in the small subunit rRNA that can be detected through sequence comparisons: a pseudoknot (or alternative pairing) in the 500–545 region and a possible long-range interaction

involving a single nucleotide in the vicinity of position 130 and the helix in the 180–195 region.

## The (505–507)·(524–526) Interaction

In the current model of 16S rRNA structure, there exist only six contiguous stretches of 10 or more nucleotides in which none of the bases are involved in a recognized secondary or “tertiary” structural interaction (4). All are regions in which composition is highly conserved. The longest of these “unpaired” stretches occurs in the terminal loop of the helix shown in Fig. 1, the entire structure covering positions 500–545 (*Escherichia coli* numbering).

This compound helix has a stalk of 12 base pairs, interrupted by a bulge loop (in eubacteria) of six nucleotides in the middle of the (upstream side of the) stalk. It is capped by a loop of 16 nucleotides (assuming the terminal G·U pair shown in Fig. 1 to be valid). In archaeobacteria and eukaryotes, the structure has this exact same overall number of pairs and the same terminal loop size. However, the bulge loop is one nucleotide larger (total of seven) and ostensibly begins one base pair higher in the stalk (following the sixth, not the fifth, pair of the helix) (3, 4).

The composition of both loops is highly conserved in eubacteria. That of the stalk following the bulge loop (the upper helix in Fig. 1) varies, although infrequently so and only within narrowly prescribed limits. For the stalk preceding the bulge loop (the lower helix in Fig. 1), variation occurs somewhat more frequently and in a less constrained manner (refs. 3 and 4; unpublished analysis). [Some mitochondrial sequences introduce minor idiosyncrasy into the bulge loop or the stalk immediately after it, or both (4). These sequences also show more variety in the composition of the entire structure than do rRNAs of free living organisms.] Except for the sequences in Table 1, >98% percent of eubacterial 16S rRNAs conform to the general sequence shown in Fig. 1 in the two loops.

The projected tertiary interaction involves a (canonical) pairing of positions 505–507 (in the bulge loop) with positions 524–526 (in the terminal loop) as indicated in Fig. 1. Two (phylogenetically independent) examples of convariation in this region have recently been detected in the eubacteria (among the mycoplasmas), and considerable support for the interaction exists in mitochondrial sequences. One particular subcluster of the *M. pneumoniae* subgroup of mycoplasmas shows guanosine → adenosine and cytidine → uridine replacements involving positions 506 and 525, respectively.‡ [Close relatives of these organisms retain the normal (ancestral) composition; from the Center for Prokaryote Genome Analysis (CPGA) rRNA sequence database, University of Illinois]. These same two replacements have occurred (independently) in fungal and *Drosophila* mitochondria (4, 13). *A. anaerobium*

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

‡*M. gallisepticum* 16S rRNA sequence, GenBank accession no. M22441.

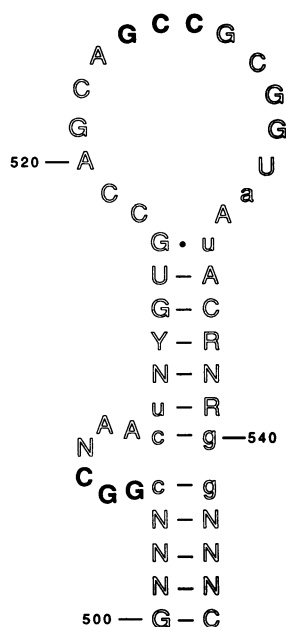


FIG. 1. Secondary structure of the 500–545 region in eubacterial small subunit rRNA. The *E. coli* numbering convention (3, 4) is used. Positions 505–507 and 524–526, which demonstrate covariation, are darkened. The sequence given is a consensus of a large and representative collection of eubacteria (CPGA rRNA sequence database, University of Illinois). Nucleotides in capital letters are essentially constant in composition; those in lowercase letters occur as shown in the vast majority of sequences. Y and R designate pyrimidine and purine nucleotides, respectively; N designates any nucleotide. Horizontal lines represent base pairs; in all cases nucleotides shown as paired are canonically so, except for the terminal G-U pair and rare G-U pairs elsewhere in the structure.

shows guanosine → uridine and cytidine → adenosine replacements for positions 505 and 526, respectively,<sup>8</sup> and the mitochondrion of *C. reinhardtii* (12) covaries positions 505 and 506 (G-G → C-C) with positions 525 and 526 (C-C → G-G) (see Table 1). Phylogenetically independent examples of the remaining covariation, position 507 and position 524 (cytidine → uridine and guanosine → adenosine respectively), are seen in the mammalian (4), sea urchin (14), and *D. virilis* mitochondria (13).

It is unlikely that the helix can be extended to include positions 508 and 523. Although variation is seen in these two positions (4), covariation of the two is the exception, not the rule. A number of replacements at position 508 unaccompanied by change in position 523 occur among eubacteria, which also yield one example of a change in the latter position unaccompanied by one in the former position as well (unpublished analysis). Their compositions also permit positions 504 and 527 to pair. However, a covariation suggesting this has never been observed.

There are no known counter examples among the eubacteria or mitochondria to the putative (505–507)–(524–526) canonical pairing, although several cases of single nucleotide replacements that would create (acceptable) G-U pairs are known (refs. 11 and 15; see Table 1).

The archaeobacteria add nothing to the picture developed from eubacterial and mitochondrial sequences. No variation in the composition of this structure is seen among the archaeobacteria, except for one case of base pairing between positions 507 and the 524 changing from C-G to U-G (16).

A slight amount of variation is seen among eukaryotic small subunit rRNAs. However, in this case the changes are

Table 1. Comparative evidence supporting a base-pairing relationship between positions 505–507 and 524–526 in small subunit rRNA

Source of small subunit rRNA	Positions	
	505–507	524–526
<b>Eubacteria</b>		
Ancestral version (3, 4)	GGC	GCC
<i>Mycoplasma gallisepticum</i> *	GAC	GUC
<i>Asteroleplasma anaerobium</i>	UGC	GCA
<i>Chlamydia psittaci</i> (11)	GGC	GCU
<b>Mitochondria</b>		
Plants and ciliates (4) <sup>†</sup> (n = 4)	GGC	GCC
<i>Chlamydomonas reinhardtii</i> (12)	CCC	GGG
Fungi (n = 2) (4)	GAC	GUC
<i>Drosophila yakuba</i> (13)	GAC	GUC
<i>Drosophila virilis</i> (13)	GAU	AUC
Mammals (n > 4) (4)	GGU	ACC
<i>Strongylocentrotus purpuratus</i> (14)	GGU	ACC
<i>Xenopus laevis</i> (15)	GGU	GCC

The variations from ancestral pattern are shown in bold type. \*Same sequence found in its immediate relatives, *Mycoplasma pneumoniae* and *Mycoplasma pirum* (C.R.W., unpublished data). <sup>†</sup>Compilation of a number of mitochondrial small subunit RNA sequences.

inconsistent with the proposed canonical pairing. In fact, the evidence (17–21) suggests a possible interaction between position 506 and position 524 (not 525). Given the fact, discussed above, that the structure of the bulge loop in Fig. 1 is not exactly the same in eukaryotes as in eubacteria (3, 4) and that eukaryotic sequences have provided “disproofs” for a number of structures in the small subunit rRNA that are actually well established (4), the present disproof is not considered compelling.

It is conceivable that the interaction between positions 506 and 525 and positions 507 and 524 forms a pseudoknot within the previously defined (1, 3, 4) structure of the 500–545 region. However, it is tempting to view two helices that directly abut one another (that are not separated by intervening unpaired nucleotides on their shared common strand) as coaxial (3, 22), as occurs in tRNA (23)—which suggests a totally different interpretation of the above “tertiary” interaction in that both the (505–507)–(524–526) and (511–517)–(534–540) helices are potentially coaxial with the underlying [(500–505)–(541–545)] helix, but not simultaneously. The ribosome is almost certainly a molecular machine, which therefore must have moving parts. Alternate stacking of the two helices in question on the underlying helix might indeed play a role in such movement.

Its extremely conserved sequence and structure implies that the 500–545 region of the molecule is functionally important. The elegant protection studies of Noller and colleagues (5, 7, 24, 25) strongly point to this as well, as does the detection of a streptomycin-resistance mutation (adenosine → cytidine) at position 523 (26). The (505–507)–(524–526) interaction reported here provides another reason to focus on this region as we work toward the molecular basis of ribosome function.

#### A Covariance Between Position 130 and the Structure of the 180–195 Region

Among eubacteria there often occurs an insertion of a residue (almost always an adenosine) relative to the *E. coli* sequence either before or after position 130 (3, 4), the two alternatives being indistinguishable. Eubacteria also differ as to the size of the helix covering positions 184–193. In *E. coli* and certain other eubacteria, the helix in question has a stalk of three base pairs terminated by a loop of four bases. The remaining

<sup>8</sup>A. *anaerobium* 16S rRNA sequence, GenBank accession no. M22351.

Table 2. A correlation between a single base insertion in the vicinity of position 130 and the length of the helix covering positions 184–193

Phylum, subdivision, etc.	Insertion at position 130	Bases between positions 183 and 194, no.	Sequences, no.
Purple bacteria			
Subdivisions			
$\alpha, \beta, \gamma$	–	10	≈70
$\delta$	+	19–25	8
Exception: <i>Wolinella succinogenes</i>	+	10	1
Gram-positive bacteria	+	21–27	≈190
Cyanobacteria	–	10	1
Flavobacteria-bacteroides	+	25–28	12
Spirochetes	+	23–28	10
shrew str. 11616	–	10	1
Green sulfur bacteria	+	26	2
<i>Planctomyces</i> group			
<i>Planctomyces staleyii</i>	–	10	1
<i>Isocystis pallida</i>	+	25	1
Chlamydiae	+	25	2
<i>Thermus-Deinococcus</i> group	+	22	2
Green nonsulfur bacteria			
<i>Thermomicrobium roseum</i>	+	25	1
<i>Herpetosiphon aurantiacus</i>	+	16	1
Exception: <i>Chloroflexus aurantiacus</i>	+	10	1
<i>Thermotoga</i> group			
<i>Ferridobacterium nodosum</i>	–	10	1
<i>Thermotoga maritima</i>	+	24	1

Sequences used are in the CPGA rRNA sequence database, University of Illinois.

eubacteria show a much larger structure here, a somewhat irregular helix having a stalk of roughly 10 base pairs generally terminated by a loop of four nucleotides (3, 4). The insertion of a nucleotide in the vicinity of position 130 (relative to *E. coli*) correlates with the lengthening of the stalk of the 184–193 helix from 3 to about 10 pairs.

The known phylogenetically independent examples of the correlation are recorded in Table 2. In 5 of the 11 eubacterial phyla (27), all known sequences exhibit the insertion at position 130 and the extended form of the helix at 184–193. In one, only the short form of the 184–193 helix with no insertion at position 130 has so far been seen (28). In four phyla both types occur. There exist only two (independent) exceptions to this correlation, *W. succinogenes* (CPGA rRNA sequence database, University of Illinois) and *C. aurantiacus* (29), both of which exhibit the inserted nucleotide but the short form of the helix. [*H. aurantiacus* is atypical in having a helix of intermediate length (six “pairs,” some of which are noncanonical) accompanied by the insertion at position 130 (29)]. Given the amount and phylogenetically broad distribution of the supporting examples (27, 30), we do not consider the two exceptions reason to doubt the significance of the correlation.

The correlation in this case is between an isolated nucleotide insertion/deletion and the overall size of a remote secondary structural element. Whether this actually speaks a physical contact, as opposed, for example, to some distortion produced by the insertion at position 130 that then permits the accommodation of the longer form of the helix (at 184–193) in the overall structure of the rRNA, is an open question. However, it would seem reasonable that, were such an accommodation required (in switching from the short to the long form of the helix), it could be accomplished in more than one very precise way. Therefore, some sort of physical contact between the elongated helix and the inserted nucleotide seems likely.

This work was supported by a grant to C.R.W. from the National Aeronautics and Space Administration (NSG-7044).

1. Woese, C. R., Magrum, L. J., Gupta, R., Siegel, R. B., Stahl, D. A., Kop, J., Crawford, N., Brosius, J., Gutell, R. R., Hogan, J. J. & Noller, H. F. (1980) *Nucleic Acids Res.* **8**, 2275–2293.
2. Noller, H. F., Kop, J., Wheaton, V., Brosius, J., Gutell, R. R., Kopylov, A. M., Dohme, F., Herr, W., Stahl, D. A., Gupta, R. & Woese, C. R. (1981) *Nucleic Acids Res.* **9**, 6167–6189.
3. Woese, C. R., Gutell, R. R., Gupta, R. & Noller, H. F. (1983) *Microbiol. Rev.* **47**, 621–669.
4. Gutell, R. R., Weiser, B., Woese, C. R. & Noller, H. F. (1985) *Prog. Nucleic Acids Res. Mol. Biol.* **32**, 155–216.
5. Moazed, D. & Noller, H. F. (1987) *Nature (London)* **327**, 389–394.
6. Brimacombe, R., Atmadja, J., Stiege, W. & Schüler, D. (1988) *J. Mol. Biol.* **199**, 115–136.
7. Stern, S., Powers, T., Changchien, L.-M. & Noller, H. F. (1988) *J. Mol. Biol.* **201**, 683–695.
8. Stern, S., Weiser, B. & Noller, H. F. (1988) *J. Mol. Biol.* **204**, 447–482.
9. Olsen, G. J. (1983) Dissertation (Univ. of Colorado, Denver).
10. Gutell, R. R., Noller, H. F. & Woese, C. R. (1986) *EMBO J.* **5**, 1111–1113.
11. Weisburg, W. G., Hatch, T. P. & Woese, C. R. (1986) *J. Bacteriol.* **167**, 570–574.
12. Boer, P. H. & Gray, M. W. (1988) *Cell* **55**, 399–411.
13. Clary, D. O. & Wolstenholme, D. R. (1987) *J. Mol. Evol.* **25**, 116–125.
14. Jacobs, H. T., Elliott, D. J., Math, V. B. & Farquharson, A. (1988) *J. Mol. Biol.* **202**, 185–217.
15. Roe, B., Ma, D.-P., Wilson, R. K. & Wong, J. F.-H. (1985) *J. Biol. Chem.* **260**, 9759–9774.
16. Leinfelder, W., Jarsch, M. & Böck, A. (1985) *Syst. Appl. Microbiol.* **6**, 164–170.
17. Sogin, M. L., Elwood, H. J. & Gunderson, J. H. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1383–1387.
18. Schnare, M. N., Collings, J. C. & Gray, M. W. (1986) *Curr. Genet.* **10**, 405–410.
19. Looker, D., Miller, L. A., Elwood, H. J., Stickel, S. & Sogin, M. L. (1988) *Nucleic Acids Res.* **16**, 7198.
20. Johansen, T., Johansen, S. & Haugli, F. B. (1988) *Curr. Genet.* **14**, 265–273.
21. Clark, C. G. & Cross, G. A. M. (1988) *Mol. Biol. Evol.* **5**, 512–518.
22. Woese, C. R. (1967) in *The Genetic Code* (Harper & Row, New York), p. 116.

23. Quigley, G. J. & Rich, A. (1976) *Science* **194**, 796–806.
24. Moazed, D. & Noller, H. F. (1986) *Cell* **47**, 985–994.
25. Stern, S., Wilson, R. C. & Noller, H. F. (1986) *J. Mol. Biol.* **192**, 101–110.
26. Melancon, P., Lemieux, C. & Brakier-Gingras, L. (1988) *Nucleic Acids Res.* **16**, 9631–9639.
27. Woese, C. R. (1987) *Bacteriol. Rev.* **51**, 221–271.
28. Tomioka, N. & Sugiura, M. (1983) *Mol. Gen. Genet.* **191**, 46–50.
29. Oyaizu, H. & Woese, C. R. (1985) *Syst. Appl. Microbiol.* **6**, 257–263.
30. Achenbach-Richter, L., Gupta, R., Stetter, K. O. & Woese, C. R. (1987) *Syst. Appl. Microbiol.* **9**, 34–39.