

Lessons from an Evolving rRNA: 16S and 23S rRNA Structures from a Comparative Perspective

ROBIN R. GUTELL,^{1*} NIELS LARSEN,² AND CARL R. WOESE²

MCD Biology, University of Colorado, Boulder, Colorado 80309-0347,¹ and Department of Microbiology, University of Illinois, Urbana, Illinois 61801²

INTRODUCTION	10
Comparative Sequence Analysis	11
Data Base	12
rRNA Higher-Order Structure	12
PRINCIPLES OF ORGANIZATION OF rRNA	15
Secondary Interactions: Unusual Structures	15
Noncanonical pairings in secondary structure	15
(i) G:U pairs	15
(ii) G:A pairs	16
(iii) Other noncanonical pairs	16
Unilaterally bulged residues	18
Bilaterally bulged residues	18
Terminal extensions of helices	18
Interesting secondary-structure motifs	19
Tertiary Interactions: Lone Pairs and Unusual Structural Elements	19
Lone pairs	19
A probable triple covariance in the small-subunit rRNA	20
Tetraloops	20
Pseudoknots	21
Coaxial helices	21
Parallel pairs	23
SUMMARY AND CONCLUSIONS	23
ACKNOWLEDGMENTS	24
REFERENCES	24

INTRODUCTION

Macromolecular structures reflect specific design principles; they can be described in terms of various structural motifs. However, outside of their capacity to form double-stranded helices, very little was understood until recently about the architectural principles underlying nucleic acid structure. rRNAs are certainly among the most interesting of macromolecules; they are the core of the ribosome, holding the key to the mechanism of translation, a process without which proteins could not even have begun to evolve. The number of rRNA sequences now known is sufficiently large that comparative analysis can be used effectively to deduce some of the basic design principles underlying rRNA structure and hence, perhaps, all RNA structure. These insights, in turn, will guide more detailed experimental approaches to nucleic acid structure.

Some of the major questions that guide our inquiry into rRNA structure are as follows. (i) Do biologically active RNAs adopt one (static) functional structure, or do they undergo functional conformational transitions (allosterism); and if so, do these conformational transitions involve major structural rearrangements, e.g., alternative double helical structures, and/or merely local perturbations of structure, e.g., the making

or breaking of individual base pairs or the coaxial stacking or unstacking of helices, etc.? (ii) To what extent are the primary determinants in RNA folding thermodynamic as opposed to kinetic? (iii) Do RNA molecules require ancillary molecules, e.g., ribosomal proteins, to effect or facilitate their folding? (iv) What, if any, is the common set of principles that establish all RNA structure? To what extent, for example, are the structural motifs characteristic of tRNAs (which are known in atomic detail) found in rRNAs, and vice versa? (v) All of these are facets of the more general question, what kind of machine is the ribosome; what are its molecular mechanics?

Although experimental methods such as nuclear magnetic resonance and thermodynamics (influenced in part by comparative analyses) have now extended our knowledge of RNA structure (reviewed in reference 9), much remains to be understood. Our ability to predict secondary structure, based in part on the study of a limited number of experimentally tested structures, has improved within the past few years (34) but still remains in a primitive state. The prediction of tertiary associations has yet to be seriously approached.

The purpose of this paper is to review what has been learned concerning the principles that underlie rRNA (in some cases, perhaps, all RNA) architecture from comparative analysis of the extensive rRNA sequence collections and how these insights, which are purely formal relationships, impinge on more direct approaches to the problem.

The higher-order structural models presented here incorporate all published and previously unpublished interactions that satisfy our comparative structure criteria. Detailed compara-

* Corresponding author. Mailing address: MCD Biology, Campus Box 347, University of Colorado, Boulder, CO 80309-0347. Phone: (303) 492-8595. Fax: (303) 492-7744. Electronic mail address: rgutell@boulder.colorado.edu.

tive evidence for these newer interactions, along with the minor adjustments they have caused in the secondary structure, will be discussed elsewhere (19a).

Comparative Sequence Analysis

One of the more important discoveries of the molecular era in biology is the simple principle that an enormous number of molecular sequences can correspond to the same three-dimensional structure and the same molecular function. The evolutionary process explores to differing extents variations on a given structure-function theme and presents us with those that satisfy the conditions necessary for survival. (This is not to imply that evolution explores the entire phase space of possible molecular sequences corresponding to a given function, but it does explore particular regions of that space extensively.) In other words, this process has performed an endless number of experiments on classes of structures that are consistent with particular functions; we observe the successful outcomes of those experiments and learn from them by noting their similarities and differences.

All rRNAs appear to be identical (or very nearly so) in function, for all are involved in the production of proteins. The overall three-dimensional rRNA structure that corresponds to this function shows only minor—but in some senses highly significant—variation. However, within this nearly constant overall structure, molecular sequences in most regions of the molecule are continually evolving, engaged in an unobtrusive game of molecular musical chairs, and undergoing change at the level of its primary structure while maintaining homologous secondary and tertiary structure, which never alters molecular function. It is this enormous variety of selectively neutral, or very nearly neutral, changes that has allowed comparative analysis to be applied so effectively to these molecules and given the biologist so much information regarding the structures of the rRNAs and other RNAs. (More recently, comparative analysis has been involved in the derivation of higher-order structures for a variety of different RNA molecules [see reference 22 and references therein].)

This comparative approach is based on the concept of positional covariance. Two positions covary when nucleotide substitutions at one column (position) in a sequence alignment are correlated with a similar pattern of substitutions at another position. Initially, this method was used to identify secondary-structure helices, but, as we will see, a number of unexpected structural motifs are now emerging.

The first use of comparative sequence analysis was to establish the so-called cloverleaf configuration of tRNA (32, 42, 54, 71), and later examination of a more extensive collection of sequences yielded covariation evidence for a few higher-order structural elements of tRNA (41). However, the bulk of the secondary and tertiary base pair interactions in tRNA structure were demonstrated by X-ray crystallography (36). Even so, all secondary-structure base pairs and most, if not all, tertiary base-base interactions can now be inferred, after the fact, by comparative analysis of a sufficiently large and comprehensive alignment of tRNA sequences (see references 26 and 50 and references therein). With rRNAs, which are more than an order of magnitude larger than tRNAs, direct demonstration of structure by X-ray crystallography is proving more difficult (although considerable progress has recently been made [70]). In these cases, comparative analysis has been the primary instrument for inferring higher-order structure, with the result that rather detailed structures for 5S, 16S, and 23S rRNAs are now known (4, 17, 18, 27, 48, 58, 66, 68, 72). These inferences have not been confined only to secondary

structure; various tertiary base-base interactions have been detected as the data sets have increased in size (19a, 25, 28–30, 38, 40, 63, 65).

Several aspects of comparative analysis as applied to nucleic acids should be stressed in that they are not commonly appreciated. Comparative analysis provides a very refined test of homology. In the world of sequence alignments, the frequency and pattern of change at any given nucleotide in a molecule are characteristics of its position and locale. This measure provides a most sensitive and accurate test of homology for each molecular structural element within different groups of organisms, because it in effect reflects most or all contacts made by a given base. Unless the frequency and pattern of nucleotide substitutions are comparable for the positions (within an alignment) that make up the element in any two groups of organisms, one cannot claim complete homology for the corresponding structure in those groups. Used in this way, comparative analysis will ultimately become part of detailed characterizations of higher-order structure. In addition, the variation that characterizes two interacting residues shows more than the fact of their interaction; the pattern of variation provides clues to the nature of that interaction. For example, covariation confined to the canonical (and G:U-type) pairing constraints almost certainly indicates a normal (Watson-Crick) type of base pairing—especially when all the possibilities are observed—whereas other patterns (see below) imply that the interaction is probably not of the Watson-Crick type.

The real advantage of a comparative approach to nucleic acid structure lies not in its use in isolation but in its use in conjunction with more direct experimental approaches to structure. Comparative analysis provides a very powerful way to identify, by their constancy, functionally important elements in a molecular structure. It can also identify which compositional variants of a given structure (e.g., a tetraloop [see below]) are functionally optimal and hence worthy of detailed structural determination. It can identify (functionally) false structures that have been detected or inferred by the more direct approaches (e.g., the denatured form of 5S rRNA, once thought to have functional significance [2]). It can, in principle, also detect alternative, functional configurations.

The central assumption underlying covariation analysis is that positions in a sequence alignment whose compositions covary have a structural relationship to one another, i.e., are in physical contact. Although in principle this is not necessarily true, in practice, when put to the experimental test, it almost always is. The covariation analysis used in inferring the structure of large rRNAs has to this point been a simple one. Initially, only positions that change composition in a fairly strict one-to-one correspondence (i.e., a change at one position is matched or compensated for by a change at its pairing position) were taken as having a structural relationship. Among these, to begin with, only those showing canonical patterns of variation (A:U ↔ C:G, etc.) were taken into consideration. This simple approach permitted identification of the major secondary structural elements.

What constitutes significant covariation evidence for a base-base interaction is a context-dependent matter. One could with considerable confidence, as was done initially, infer the existence of an entire double-helical element on the basis of covariation involving only a few of the base pairs therein, provided that no significant counterevidence existed, i.e., variation of one position without variation of its putative partner, or the occurrence of noncanonical juxtapositions, except for those of the G:U type. The large data set that now exists provides overwhelming comparative justification for the vast

majority of pairs in each helical element in 16S rRNA. A good example of this is the compound helical structure located between positions 588 and 651 in 16S rRNA (which binds ribosomal protein S8 [reviewed in references 5, 16, and 49] (see Fig. 1). Of the 24 (canonical or G:U-type) pairs shown in the structure, more than 20 phylogenetically independent (canonical or G:U) replacements occur for 13 of them; for an additional 6, 8 or more such replacements occur and the evidence we use here is taken from a (eu)bacterial alignment only (i.e., one that contains no organellar rRNAs). The remaining five pairs (588:651, 597:643, 605–606:633–632, and 617:623) are all located terminally in their respective helical segments, and each shows little or no compositional variation among (eu)bacteria. For one of them, however, U605:G633, the members of the archaea provide a convincing number of covariations with no counterevidence. (G606:U632 is problematic in that a small amount of covariation is accompanied by significant variation of one position in the absence of variation in the other; this pair is not considered proven.) For the remaining three, however, evidence is marginal. Positions 617 (G) and 623 (C) show three independent examples of covariation, which, however, are not all canonical; this is marginal proof of interaction. However, G597:C643 is completely invariant; and G588:C651 covaries in one phylogenetically independent case but also has an isolated example of noncovariance. In any case, we take these three pairs to be valid for other reasons: they are all of the G:C type, typical of terminal pairs in helices (as are their slow rates of change). The general rule we have used in the figures presented herein is that paired contiguous extensions that are unproven because of compositional invariance can be added to proven helical segments provided that no significant counterevidence exists. In these cases, however, the paired bases in question are shown juxtaposed but not joined by a connecting line (the symbol of properly demonstrated canonical pairs).

Once the basic secondary structures of the rRNAs became evident, the larger data sets could be used to refine them and to detect the less apparent tertiary interactions. Tertiary interactions are relatively difficult to detect in that they usually involve only a few bases whose compositions seldom vary and whose interactions do not necessarily involve canonical pairing or a regular antiparallel orientation. With larger data sets it becomes possible to search for interactions for which the covariation is far less strict than the telltale one-to-one correspondence, and one can also begin to consider idiosyncratic structures, i.e., those showing different forms in different groups of organisms. In parallel, more powerful and sensitive correlation methods, under development, will permit further refinements in rRNA structural detail (19a, 26).

Data Base

Since the first complete 16S rRNA sequence was determined in 1978 (7, 8) the number of complete (or nearly complete) 16S-like rRNA sequences has reached over 2,200 (compilation as of August 1993 [21, 39]). The first complete 23S rRNA sequence was determined in 1980 (6), and complete (or nearly complete) sequences of this genre now number greater than 250 (for compilations, see references 24 and 39). The number and phylogenetic distribution of these sequences are given in Table 1. The organisms represented in the table cover all major taxa as well as a wide selection of organelles (mitochondria and chloroplasts). Extensive primary-structure alignments (phylogenetically ordered) and secondary-structure representations

TABLE 1. Approximate number and phylogenetic distribution of publicly available (or soon to be) rRNA sequences

Organism type	No. of sequences of:	
	16S rRNA	23S rRNA
<i>Archaea</i>	100	15
(<i>Eu</i>) <i>Bacteria</i>	1,500	70
<i>Eucarya</i>	500	40
Organelles		
Mitochondria	100	70
Plastid	40	60
Total	2,240	255

of many 16S and 23S rRNA sequences are publicly available (21, 24, 39, 46).

We will not discuss the process of aligning or the actual alignments in any detail here. Suffice it to say that aligning is done manually. It is an iterative process that begins with juxtaposition of regions of extensive primary structural similarity and is then refined by invoking higher-order structural constraints and so on. Alignment is a global process, although regions of idiosyncrasy have to be locally defined. The alignment process in principle can pit sequence homology against (convergent) structural similarity (analogy). In practice, situations of actual conflict between the two appear rare and, if they really exist, are confined to rRNAs of relatively closely related species. The reason for this apparent lack of conflict would appear to be that sequence homology cannot be preserved over extensive periods in the absence of overlying higher-order structural and functional homology. However, it is clear from any phylogenetically ordered alignment that higher-order structural constraints have given rise to a number of instances of local sequence convergence, e.g., the composition of tetraloops (see below).

rRNA Higher-Order Structure

The current versions of the *Escherichia coli* small- and large-subunit rRNA higher-order structure models (Fig. 1 and 2, respectively) are the result of 10 years of comparative analysis. The evidence for the few tertiary interactions shown therein, but not yet published, will be formally presented elsewhere (19a). All canonical pairs indicated by a connecting line in the figures are considered proven (highly likely) on the basis of covariation evidence, as are all G:U pairs indicated by dots and all G:A pairs indicated by open circles. (Those juxtaposed but not connected in this manner are possible but unproven, almost always because of compositional invariance.) In the vast majority of cases the canonical pairs indicated as proven show multiple phylogenetically independent examples of covariation, with no or relatively few counterexamples, in which variation in the one position is unaccompanied by variation in the other (G:C \leftrightarrow G:U variation excepted). Covariation within phylogenetically restricted groups (i.e., those in which overall sequence similarity is relatively high) is considered more significant than covariation involving phylogenetically isolated (distant) sequences.

Although the overall structures of both small- and large-subunit rRNAs are very similar within each of the three phylogenetic domains (the term "domain" here refers to a new phylogenetic taxon, which includes the three primary lines of descent, *Archaea*, *Bacteria*, and *Eucarya* [67]), there are characteristic differences in both small- and large-subunit rRNA

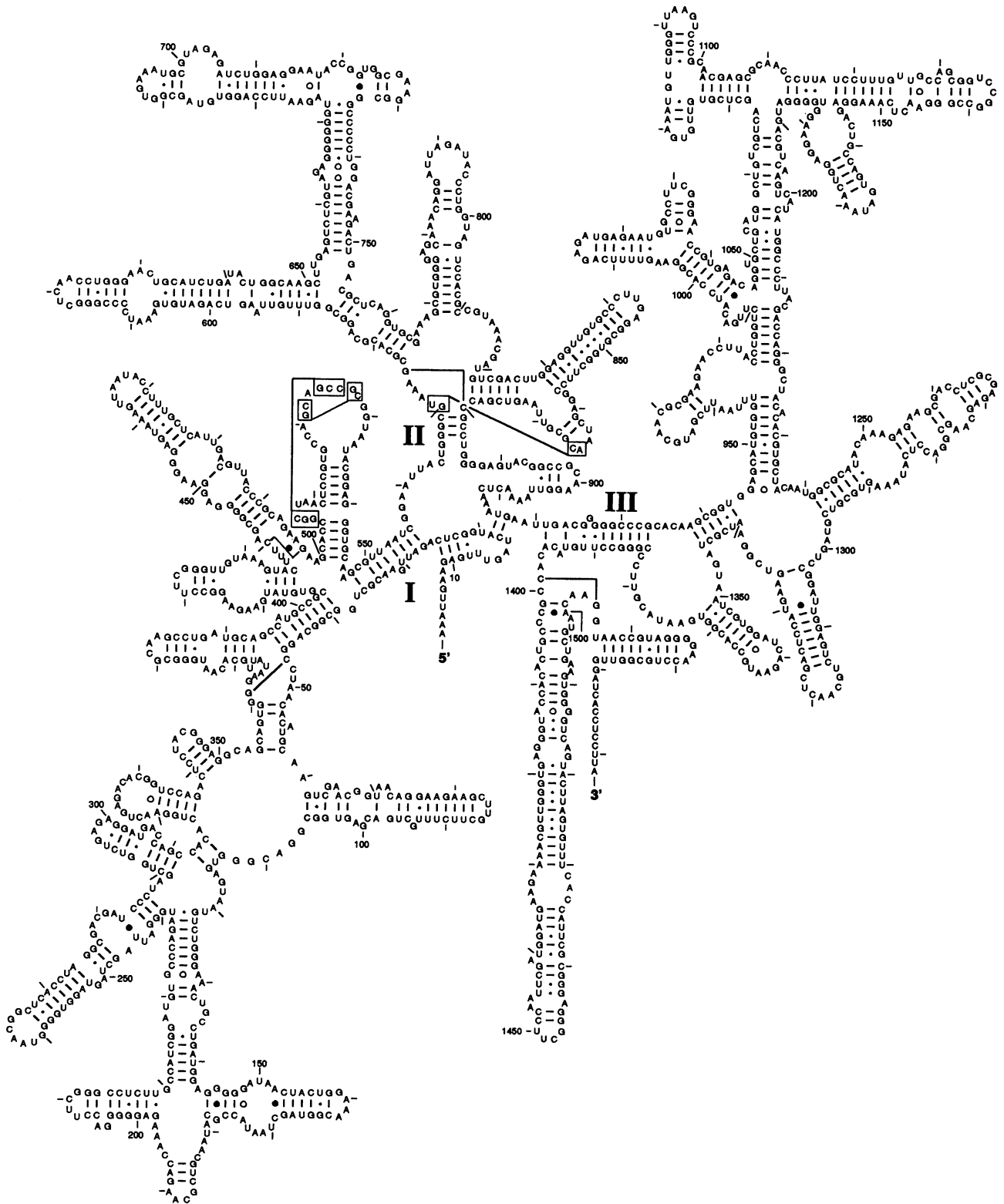


FIG. 1. Higher-order structure model for *E. coli* 16S rRNA. Canonical (C:G, G:C, etc.) base pairs are connected by lines, G:U pairs are connected by dots, A:G-type pairs are connected by open circles, and other noncanonical pairings (see text) are connected by solid circles. "Tertiary" interactions are connected by thicker (and longer) solid lines. Every 10th position is marked with a tick mark, and every 50th is numbered. Primary structure was determined by Brosius et al. (7, 8).

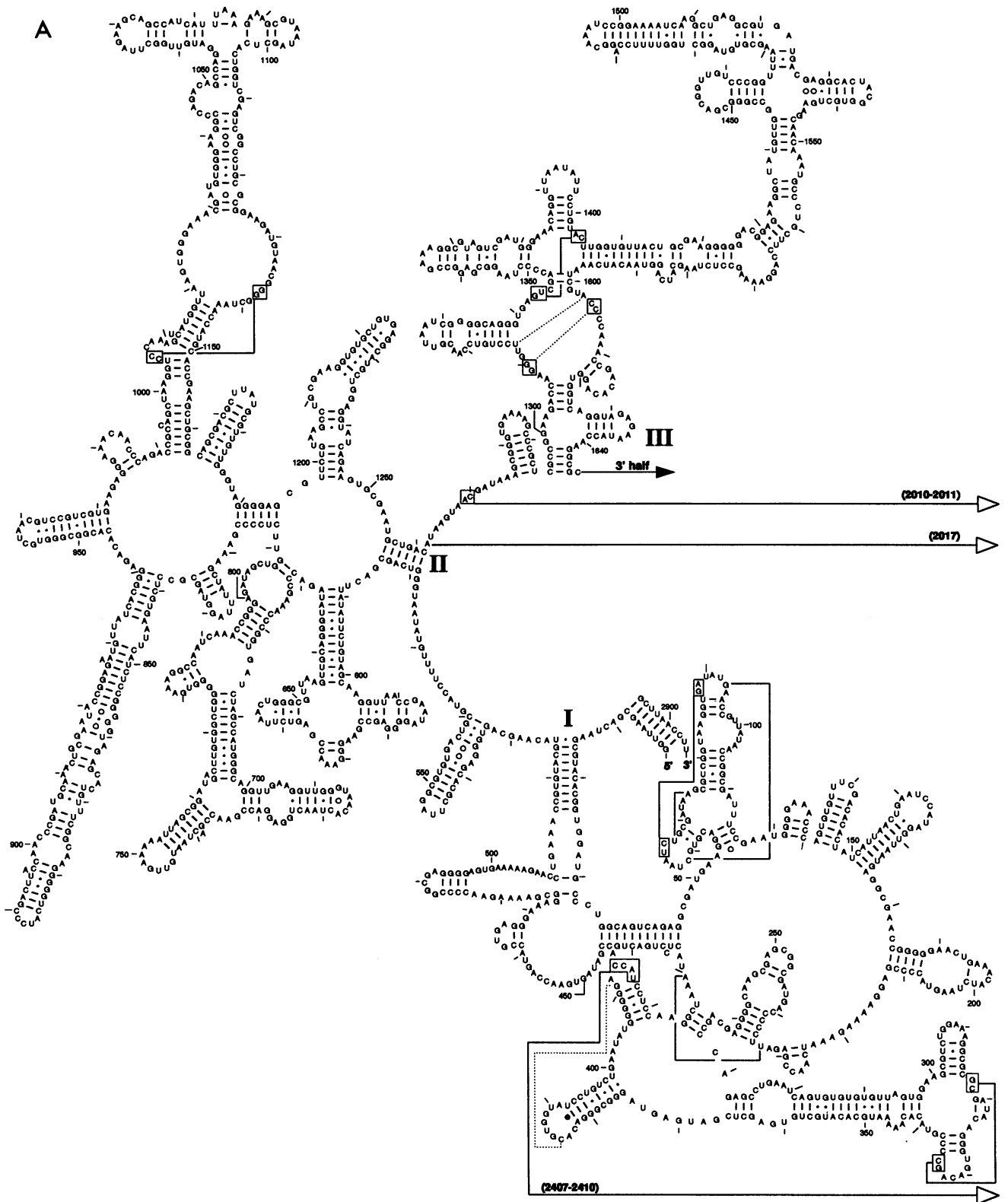


FIG. 2. Higher-order structure model for *E. coli* 23S rRNA. (A) The 5' half. (B) The 3' half. Secondary and tertiary interactions and sequence numbering follow the convention used for Fig. 1. Primary structure was determined by Brosius et al. (6).

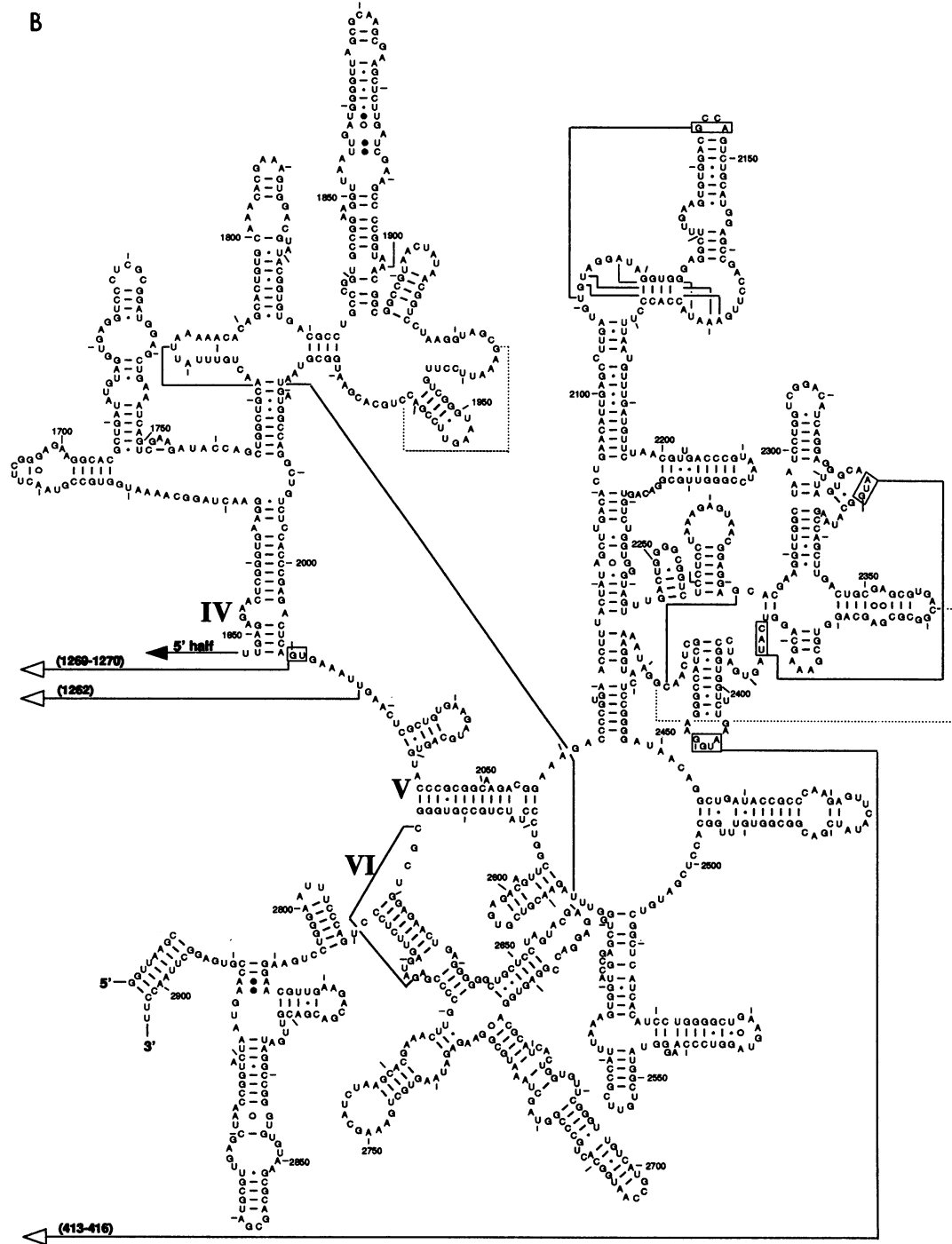


FIG. 2—Continued.

structures associated with each domain. (At a more refined level, especially in terms of a sensitive measure such as frequency and pattern of replacement, one can even see lesser differences in the different versions of rRNA within the different subgroups of *Bacteria*, *Archaea*, or *Eucarya*.) In that our purpose here is not phylogenetic comparison for its own sake, we will focus mainly on the higher-order structures of the (eu)bacterial small- and large-subunit rRNAs, using *E. coli* as the standard.

PRINCIPLES OF ORGANIZATION OF rRNA

Secondary Interactions: Unusual Structures

Noncanonical pairings in secondary structure. This section deals with some of the more interesting noncanonical pairings, bulges, and other features found in otherwise normal double-helical segments of rRNA secondary structure.

(i) **G:U pairs.** The G:U pair appears to play an incidental role in tRNA structure. Although G:U and U:G pairs can be

seen at a variety of positions in tRNAs, they are usually represented by canonical pairs in homologous tRNAs from even close relatives of the organism in question. In other words, G:U pairs are not characteristic, fixed features of tRNA structure. (An exception to this generalization was first pointed out by Madison et al. [42], when it was suggested that a G:U base pair in the tRNA acceptor stem could be important: "Is it possible that the activating enzymes could extract enough information from the G:U pair and other features of this double-stranded region for it to act as a recognition site?") Subsequently, the G:U base pair at positions 3:70 in alanine-tRNA was shown to be a primary recognition signal for its cognate synthetase [33, 43].)

The incidental type of G:U pair obviously occurs in rRNAs as well. In rRNAs G:U pairs are frequently characteristic (usually fixed) features of particular helical elements. Figure 3 shows the characteristic G:U pairs for the small-subunit rRNA (not all of which are considered proven). It is evident from patterns of covariation that the G:U pairs in 16S rRNA are of several different kinds. The first type we would call the invariant G:U pairs (represented by I's in Fig. 3); i.e., their composition is essentially constant within the (*Eu*)Bacteria (and often in all three domains). It must be realized, however, that G:U pairs of the I type in some or all cases may merely be examples of the types described below in which the sequence is highly conserved. The second type we would call a dominant (D-type) G:U pair, represented by D's in Fig. 3. These do vary in composition, interchanging with canonical pairs, but G:U occurs at significant levels in almost all groups across the spectrum of bacterial 16S rRNAs and is the predominant pair in many of them. Some D-type pairs show numerous instances of what appear to be direct U:G \leftrightarrow G:U replacements. A typical example of the D-type G:U pair is 157:164 (Fig. 1 and 3): in about 55% of (eu)bacterial 16S rRNAs this pair is either U:G (the predominant composition) or G:U, and at least 15 to 20 phylogenetically independent, and closely related, examples can be seen in which the U:G has been directly replaced by G:U (or vice versa). (Among the archaea all examples of this pair have the composition G:U rather than U:G.)

The remaining subclass of G:U pairs in 16S rRNA (the N, or nontypical type) tends to change composition only infrequently, but when change occurs, the replacement is almost always U:G \leftrightarrow C:A. Three clean examples of N-type G:U pairs in the small-subunit rRNA are 249:275, 1074:1083, and 1086:1099. The first has a constant U:G composition in 98% of (eu)bacteria, the exceptional 2% being made up of six phylogenetically independent occurrences of its C:A alternative. (The archaea show one example of C:A amid an otherwise constant U:G background.) The 1074:1083 N-type pair has a constant G:U composition among bacteria but switches from a G:U composition in the crenarchaeota to C:A in the euryarchaeota. The 1086:1099 pair is again U:G in 95% of (eu)bacteria but C:A in almost all of the remaining 5% (all confined to one particular subgroup of the flavobacterial phylum). This replacement pattern for N-type G:U pairs suggests that the interaction in question may not have a normal pairing geometry; structurally homologous U:G and C:A pairs can be formed when the two bases are in opposite orientation (syn- versus anti-) to their sugar moieties. Structurally isomorphic U:G and C:A pairs can also be formed by protonating the adenine. Physical nuclear magnetic resonance studies have shown this to be the case in a simple 16-mer, where the C:A "pair" is embedded in an otherwise normal helix (52). There would appear to be some hybrid G:U pairs as well, for example ones ostensibly of the N type primarily but having some canonical variants as well.

Enlarging the structure data set beyond the prokaryotes reveals a few cases in which pairs defined as C type in the bacteria become N type. The most notable here is 1512:1523 in the small-subunit rRNA, whose otherwise constant U:G composition becomes C:A in *Plasmodium fragile* (but remains U:G in the other *Plasmodium* species). Although examples of I-, D-, and N-type G:U pairs also exist in the large-subunit rRNA, we defer a discussion of these for another paper.

Perhaps the most spectacular use of G:U type pairs is seen in the 16S rRNA helix 829-840:857-846 (Fig. 1). This structure, which usually comprises 10 to 12 pairs (Fig. 1), almost always contains five or more G:U-type pairs in its central section. The composition of this helix is quite variable, so that G:U \leftrightarrow U:G replacements occur frequently and the exact position of the G:U pairs varies.

(ii) **G:A pairs.** Juxtaposition of the two purines G and A is another fairly common feature in the helices of the two major rRNAs. These occur both terminally and in the interior of helices, either as isolated pairs or in small contiguous groups. Comparative evidence can be adduced for a number of the putative G:A pairs, some being replaced by canonical pairs, while others show strict A:G \leftrightarrow G:A covariation. A well-documented example of a terminal A:G pair is 1357:1365 in the 16S rRNA (66). In all (eu)bacteria this is the closing pair for a loop of seven bases. Proof of pairing comes from two types of comparative evidence: (i) the fact that the corresponding positions form canonical pairs (of various compositions) only in all archaea and almost all eucarya, and (ii) the fact that the dominant form of this pair in bacteria (A:G composition) is replaced at least a dozen phylogenetically independent times by its opposite (G:A), this being the only replacement observed. G:A pairs at different positions can show different patterns of replacement, suggesting that not all G:A pairs have the same geometry. One such pattern is an A:G replaced mainly by G:A or A:A; sometimes only by A:A (meaning, of course, that the pairing remains unproven).

Two notable examples of contiguous multiple G:A pairs within helices are at 16S rRNA positions 1417-1418:1482-1483 and the somewhat more complex central section of the compound helix, i.e., 663-665:741-742. Chemical modification studies suggest that the former two pairs may not be in a standard Watson-Crick orientation (45). In the archaea the use of G:A pairs in the 665 region is pronounced; the region contains four of them in most cases. An even more pronounced use of G:A pairs by the archaea can be seen in the 1484 to 1505 region of the large subunit rRNA, where most species show four or five of them in succession, surrounded by comparatively proven canonical pairs.

(iii) **Other noncanonical pairs.** The remaining noncanonical pairing possibilities, G:G, A:A, and Y:Y, have also been seen and documented within various helices; however, they are relatively rare. In the small-subunit rRNA the pairing G:G occurs at positions 145:177 in all β and γ and some α and δ purple bacteria and is frequent in other major bacterial groups as well. In the latter groups, its alternatives are exclusively the various canonical pairs.

What appears to be a different type of G:G pair occurs in the 16S rRNA helix 61-82:87-106, which is highly variable in composition, length, and overall structure. It occurs in the interior of the helix, in which the helical versions are not completely homologous to the *E. coli* version; thus, its position cannot be given in terms of the numbering of Fig. 1. In this case the alternative to G:G is solely A:A. This particular pair will be mentioned again below, in the context of its surrounding structure.

Most of the pyrimidine-pyrimidine pairs seen in rRNAs are

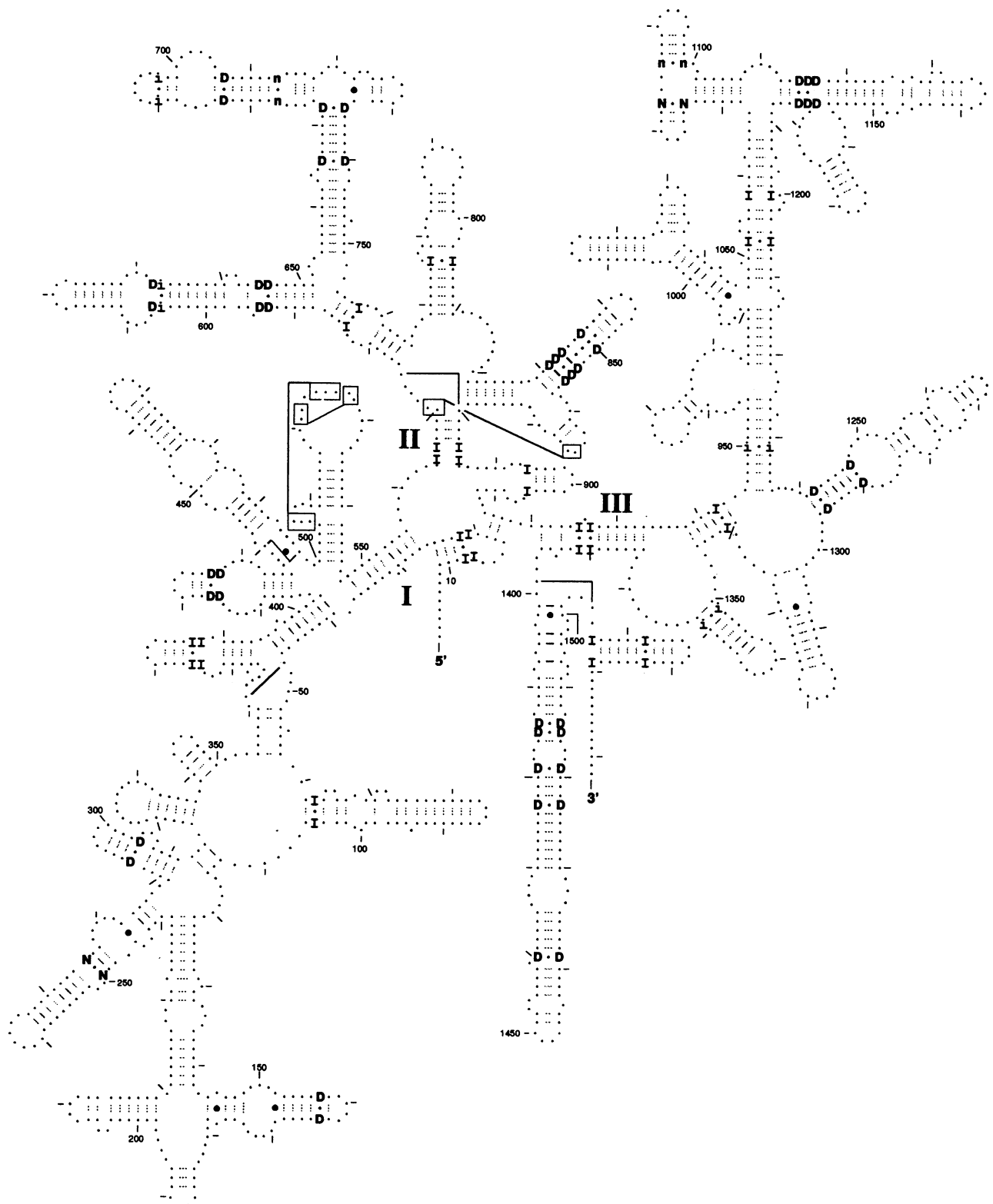


FIG. 3. Bacterial G:U base-pairing constraints mapped onto the *E. coli* 16S rRNA structure model (Fig. 1). G:U pairs of constant or invariant composition are represented by I's; those in which the G:U (or U:G) composition is dominant are represented by D's, and those showing a nontypical pattern of covariation are represented by N's. See text for details.

definitely not of the incidental type; they clearly play key roles in rRNA structure. Several will be encountered in the tertiary interactions discussed below. One that might be mentioned in the context of secondary structure is the U:U pair involving 16S rRNA positions 1307:1330, which immediately precedes a normal double-helical element (Fig. 1). Its predominant composition is U:U in all three domains. However, it occasionally transforms into C:C, its only known alternative (except in mitochondrial rRNAs); at least 15 phylogenetically independent examples of U:U \leftrightarrow C:C transitions have been recorded (28).

Unilaterally bulged residues. Single-base bulges from one side of a helix are common features of both small- and large-subunit rRNAs. Twelve of these are seen in the structure of Fig. 1. The vast majority of them are found in all other (eu)bacteria, and the majority of these are found in the other two domains as well. In principle, such residues could be protruding or, alternatively, intercalated, and it would seem that both situations occur. The bulged residue at position 31 would seem to be just that. It is found in all bacterial small-subunit rRNAs, but in the comparable helix in the archaea, which has the same overall length as its (eu)bacterial counterpart, no bulged nucleotide occurs. Residue 397, which appears to be bulged from the adjacent helix (in bacterial examples [Fig. 1]), is probably not so: the comparable base in the archaeal examples of this helix is always paired (64), suggesting intercalation of this residue in the bacterial cases.

One experimental study shows that a single nucleotide bulge loops out of the helix (reviewed in reference 9). Other experimental studies reveal that bulge nucleotides could bend the helix in which they are found (reviewed in reference 9). However, detailed experimental studies, taking into account a variety of different helix compositions containing a single nucleotide bulge, have yet to be done; therefore we cannot distinguish and discern the effects that bulge nucleotides have on their local structure.

Bilaterally bulged residues. Bilaterally bulged residues, i.e., internal loops, are extremely common in small- and large-subunit rRNAs. These are difficult to interpret and discuss, because one cannot say at this stage whether all are actually bulged; they could alternatively be noncanonically paired sections within proven helices. This alternative seems most likely when the opposing sides of the bulge have the same number of nucleotides and weak covariation (complex in pattern) occurs between opposing bases. An example of this situation is the structure located between small-subunit rRNA positions 1258 and 1277, which comprises two basal canonical pairs underlying a symmetrical three-base bilateral bulge, which is then overlaid by three additional canonical pairs that lead into the capping loop of the structure (Fig. 1). The potential 1260:1275 "pair," within the bilateral bulge, shows many examples of variation in the one position unaccompanied by variation in the other. However, in the flavobacterial phylum and in the archaea (Ribosomal Database Project [39]), the two positions do exhibit a strong covariation (though definitely not of a strict canonical sort). The second potential "pair," 1261:1274, shows no convincing covariation, except among the archaea. The final "pair," 1262:1273, shows considerable covariation, much of which is canonical, in a number of bacterial phyla and in the archaea (in which the pair has a strictly canonical form). However, many noncanonical juxtapositions are seen here among the bacteria, as are instances of variation at one position unaccompanied by variation at the other. This somewhat excruciating exercise is meant to point out that there may well still exist many atypical base-base interactions in rRNAs for which covariation rules are too

complex for them to be detected at the present levels of analysis and/or whose structures are not strictly homologous from one group of organisms to another.

Other examples of atypical pairings occur within positions 1852–1859:1883–1890 of the large-subunit rRNA (Fig. 2). This structure, found only in (eu)bacteria, contains four unusual pairings. Three of these, 1855:1887, 1856:1886, and 1859:1883, juxtapose U-U or U-C in *E. coli* and a small number of other (eu)bacteria, while the majority of (eu)bacterial sequences juxtapose canonical pairs, each with several compensatory changes. The fourth pair, 1858:1884, is another example of an A:G \leftrightarrow G:A interconversion (23) (see the G:A pair discussion above).

Terminal extensions of helices. As the data base of rRNA sequences grows, it is becoming increasingly apparent, as might be expected, that the bases immediately beyond the 5' and 3' (canonically paired) termini of established helical elements sometimes interact with one another, in loose, noncanonical, and hard-to-define ways. One type of extension was encountered above in the discussion of G:A base pairs at the ends of helices, but other noncanonical "pairings" appear to occur as well. The covariances seen between these "subterminal" bases are of a loose sort and sometimes are evident in certain groups of bacteria only, which might indicate that interaction occurs only in some groups of organisms. However, many of them are quite convincing.

One example of the foregoing is the extension of the helix whose terminal canonical pair is 769:810 of 16S rRNA. Covariation of positions 768 and 811 is evident among the bacteria, archaea, and mitochondria. It is far from pure (i.e., one to one) and basically noncanonical; among bacteria the pattern fluctuates predominantly between A:Y, the major form, and G:A, a minor variant. Interaction between these positions (768 and 811) is supported by the fact that both of them are protected against chemical modification in the 30S subunit; one of them, position 811, is protected in the isolated 16S rRNA as well (45).

A second example of subterminal "pairing" in the small-subunit rRNA involves positions 1257 and 1278, which would extend the complex 1258-to-1277 helix that contains a bilateral bulge, discussed above (Fig. 1). The covariance in this case is imprecise; however, there can be no doubt that it exists, because in the majority of cases the composition is either R:R or Y:Y, and these two motifs phylogenetically alternate frequently. In this case the subterminal pairing can even be extended back an additional "pair," 1256:1279, but the covariance is even looser in this case than in the previous one. We take the covariation as real, however, because within a rather wildly varying collection of compositions, dominated by Y:A, there occur some recognizable regularities. Almost without exception, the (phylogenetically rare) occurrence of G at position 1256 is accompanied by A at 1279 and, conversely, the (phylogenetically rare) occurrence of G at 1279 in turn is accompanied by A at 1256 (this pattern having arisen a significant number of independent times). The first of these subterminal pairs, 1257:1278, is protected against chemical modification in isolated 16S rRNA; however, the second is not (45). Finally, there is even some (marginal) comparative evidence to suggest that the compositions of the two pairs, 1257:1278 and 1256:1279, covary (unpublished analysis).

The last example here involves the subterminal "pair" 152:169, which underlies the helical stalk 153–158:163–168. Covariation of these two bases is strong and can be seen in all (adequately represented) major bacterial groups. It most frequently involves A:C \leftrightarrow G:U conversions but can involve several other combinations as well, none of which is canonical.

It should be recognized that with these unusual (subterminal) covariances one has to question the basic assumption of covariation analysis, i.e., that covarying bases physically interact. In some of these cases, perhaps the immediately preceding one, the bases covary in a way that ensures that they do not interact canonically; i.e., they may not physically interact at all.

Interesting secondary-structure motifs. In the small-subunit rRNA (Fig. 1), the backward extension of the helix located between positions 1308 and 1329 shows the sequence GGAU "pairing" with UGAA (forming in succession the pairs G:A, G:A, A:G, and U:U). Although the U:U pair is proven (see above), the three G:A pairs are not, but all four "pairs" are sandwiched between proven canonical pairs (Fig. 1). Very similar motifs are seen elsewhere in (some) small- and large-subunit rRNAs. Although the small-subunit rRNA of *E. coli* does not contain it, a clearly related structure occurs in the central section of (some versions of) the helix whose capping loop is located at position 83 in Fig. 1. The structure in question involves these five successive pairings (juxtapositions) G:A, R:R, G:A, A:G, and U:U, sandwiched between proven canonical pairs. (R:R covaries strictly between G:G and A:A and has been discussed above.) Another example at positions 25–29:511–515 of 23S rRNA contains five consecutive pairings, U:A, G:A, G:A, A:G, and U:U (Fig. 2), and is also flanked on both sides with standard canonical and comparatively proven pairings. While the first of these pairs, U:A, is variable in composition with no recognizable covariation, the remaining four are essentially invariant. Taken together, these three examples are identical in their last three pairings (G:A, A:G, and U:U). Physical and modeling studies could well offer some insight on this recurring motif.

A covariation of a very different sort, seen in the small-subunit rRNA, involves position 130 and the 180-to-195 structure (65). An extra nucleotide inserted after position 130 correlates with an increase in the length of the 184–186:191–193 helix from three to nine pairs. Numerous phylogenetically independent examples have been noted. The structural implications for this correlation are not readily apparent. The extra nucleotide inserted at position 130 could distort or change the angle of the underlying structure, thereby accommodating the longer version of the position 185 helix; alternatively, a more direct but unspecified physical connection between the two could occur.

Tertiary Interactions: Lone Pairs and Unusual Structural Elements

For want of a better term, "tertiary" is used here to describe all higher-order intramolecular interactions in rRNA that are too complex to be comfortably described as secondary structure (which we define as regular unknotted helical elements composed of canonical base pairs). The term mainly covers "lone" pairs, whether nearby or greatly separated in the secondary-structure representation, and pseudoknots (interactions involving nucleotides within the loops of helices). This section will consider some of the structural features in rRNA other than the standard helical elements. The tertiary interactions can be located in Fig. 1 and 2.

Lone pairs. Individual (isolated) pairs are not sufficiently stable in their own right to exist as such. They must somehow be supported, constrained, etc., by their surrounding structure, for example by stacking on adjacent bases. Nevertheless, comparative analysis does reveal a number of such lone pairs in rRNA. One example is the 245:283 pair in the small subunit, depicted in Fig. 1 as part of a complex coaxial helical structure. Variation in its composition is strictly confined to U:U ↔ C:C

in the bacteria and archaea, and a number of phylogenetically independent transitions between the two have been seen (28). Like the U1307:U1330 pair, discussed above, this one shows a U:U ↔ C:C covariation exclusively (among the bacteria and archaea). The same pattern of covariation, supported by a number of phylogenetically independent transitions (with only one exception to this rule), can be seen in the large-subunit rRNA, involving positions 1782 and 2586, positions very distant from each other in the secondary-structure representation (28). UV cross-linking studies lend physical support to this covariation (57). The U:U and C:C alternative compositions for this type of pair could be made structurally homologous, forming two hydrogen bonds, if the bases were in opposite (syn- versus anti-) orientation with regard to their sugar moieties. (Alternatively, both pairings can exist in homologous structures in the normal orientation if one of the cytosines is protonated, a structure suggested by nuclear magnetic resonance and thermodynamic studies [55]. We would note, however, that in this case the pair in question is sandwiched between canonical pairs, which is not the case in the three rRNA examples noted here.)

The lone pair involving positions 722 and 733 in the small-subunit rRNA has the composition G:G in Fig. 1. This composition alternates almost exclusively with A:A; this alternation has occurred almost 20 phylogenetically independent times (28). The same pattern of covariation characterizes the helix whose loop is located at position 83 of the small-subunit rRNA (mentioned above) and in 5S rRNA at positions 76:100 (*E. coli* numbering). Again, syn- versus anti-orientation of the two bases in the pair would make the two versions structurally equivalent in their interaction.

Recent in vitro genetic analysis has suggested that such noncanonical pairings can serve as protein recognition sites. Within the key RNA recognition site for the human immunodeficiency virus type 1 Rev protein lies a G-G juxtaposition. Protein recognition was significantly decreased when this pairing was changed to A-G but completely restored to "wild-type activity" upon changing the remaining G to A (3).

Other recognized lone pairs in the small-subunit rRNA are as follows. (i) The C:G composition of 47:361 varies only twice among the (eu)bacteria, to U:A in all spirochetes and G:A in all members of the planctomyces/chylmidia group; the composition is uniformly G:C among the archaea and all but three eukaryotes, where it becomes U:A. Protist mitochondria contain C:G in most cases, with two phylogenetically independent examples of U:A and one of G:U. Animal mitochondria show a number of canonical variants. The lone pair could potentially stack on the end of the immediately adjacent helix (Fig. 1). (ii) The variation at 438:496 among bacteria and archaea is confined almost exclusively to U:A ↔ G:G (more than 10 phylogenetically independent examples), strongly suggesting the pair to have an unusual geometry (28). (iii) The pair 450:483 exhibits about 10 phylogenetically independent examples of canonical variation (plus one of the G:U type) among the bacteria (28), which suggests that it has a normal geometry. This pair of bases is protected against chemical modification in isolated 16S rRNA, whereas three of the four bases immediately flanking it (position 449 excepted) are not protected (45). It is reasonable to assume that this pairing must be strengthened by some sort of stacking interaction with surrounding bases. Indeed, evidence for this exists: in many bacterial groups the underlying positions (positions 449 and 484) appear to covary in composition. (However, in most cases this "pairing" has an A:G or G:G, not a canonical composition.) Interestingly the composition of this underlying pair correlates with that of 450:483. The predominant composition of the latter is G:C,

found in over 90% of cases. It varies to Y:R (U:A or C:G) eight phylogenetically separate times. In seven of these instances the composition of underlying position 484 covaries to U, which is always then matched by either A or G at position 449. (iv) The G:C composition seen at 575:880 among the bacteria and archaea alters to A:U twice among the bacteria and three or more times among the mitochondria, with A:U being the characteristic composition among the eukaryotes. It would seem to be part of a rather complex interaction (Fig. 1; see also the discussion of 570:866 below). (v) Variation at 779:803 occurs seldom (five phylogenetically independent times in all) but is always canonical (including some transversion), and no counterevidence exists (28, 29).

The next and last set of pairings [pairs (vi) to (x)] associate the 1400 and 1500 regions of 16S rRNA, two of the most highly conserved sets of positions in all of biology, with variation found only among a few mitochondria. These two regions appear to be of critical functional significance (reviewed in reference 47), and thus the pairings in question represent the beginnings of our understanding of the structure of this important region. Three of the five interactions (1399:1504, 1401:1501, and 1405:1496) noted here were tentatively proposed in 1985 on the basis of a minimal amount of comparative evidence (27). Eight years later, support for these pairings has been bolstered with additional comparative (23) and experimental (10, 11, 13) data. More recently, two new correlations (1402:1500 and 1404:1497) have been identified that extend the antiparallel orientation of these interactions (23). (vi) C:G \leftrightarrow U:A interconversions at 1399:1504 are seen in several phylogenetically distant mitochondria; some U:G pairing occurs as well. Kinetoplast rRNAs appear to contain a U:U here, but their alignment is debatable in this region. (vii) The correlation at 1401:1501 is based on several examples of canonical variation (G:C \leftrightarrow A:U) in phylogenetically independent mitochondria. Kinetoplasts contain G:U. In vitro genetic analysis suggests a linkage between these two positions (10). (viii) Of this set of correlations, only the pair 1402:1500 is atypical, alternating between C:A and U:G. Three mitochondria, underlain with two separate phylogenetic events, contain a U:G pair; the dominant form is a C:A pair (also see the discussion of U-G pairs) (23). (ix) The C:G pair is found in at 1404:1497 all sequences (archaea, bacteria, eucarya, chloroplast, and mitochondrial) except in one mitochondrion, which contains a U:A pair (8a, 23). This one compensatory change does not prove this pairing; however, given its location next to the 1405:1496 pair, the canonical nature of the pairings, and recent experimental work (11), this proposed base pairing is highly suggestive. (x) The pair 1405:1496 is now based on three phylogenetically independent examples of canonical variation (G:C \leftrightarrow A:U) in the mitochondria, with no exceptions. Pairing is considered likely, especially in light of experimental analysis (10).

A probable triple covariance in the small-subunit rRNA. The three positions 440, 494, and 497 appear to covary, although in a manner less constrained than that observed for secondary-structure base pairings. Clear instances of positions 440 and 497 covarying can be seen among the α , β , and γ purple bacteria, the high-G+C-content gram-positive bacteria, and the fusobacteria (Ribosomal Data Base Project [39]); positions 440 and 494 covary in the cyanobacteria and one subsection of the order *Thermotogales*. All three covary within the myxobacteria, the *bdellovibrios*, and the archaea (the last of which shows one example of covariation of the pair 494:497 as well). Given that these pairs surround the adjacent (non-normal) 438:496 pair (see above), it is clear that the general structure in this locale is a complex one. One needs to consider

whether the triple covariation involving positions 440, 494, and 497 implies a base triple or functionally alternating binary interactions. Only position 440 of the three is protected against chemical modification in the isolated 16S rRNA (45). It should also be noted that the locale in question is immediately adjacent to what is arguably one of the most functionally important regions in the small-subunit rRNA, i.e., the 500-to-545 helix (47). It is conceivable that the complex of interactions (involving the putative triple, the unusual 438:496 pair, etc.) may be involved in a very sensitive positioning of this functionally important unit or even in some subtle cyclic conformational change therein.

Tetraloops. Perhaps the most prominent and easily recognizable structural element in rRNA is what has become known as the tetraloop, a loop of four nucleotides underlain by a double-stranded stalk of 2 bp or more. They account for the majority of all hairpin loops in rRNA. Their most interesting property is their sequence; of the 256 possible for a tetraloop, only an extremely limited subset are found in rRNAs. A strong correlation exists between the compositions of the first and last base of the sequence, the three predominant variants being U-G, C-G, and G-A (69). Their relationship has been elucidated by nuclear magnetic resonance solution structures for the two tetraloops C(UUCG)G and C(GMAA)G (where M = A or C) (31, 61), which indicate that the first and last bases of the tetraloop interact to form an atypical pair.

In addition, the interior two bases are far from randomly selected. In the small-subunit rRNA, the predominant compositions for each of the three major types are UUCG, CUUG, and GMAA. Moreover, the closing pair for the loop tends to correlate with the sequence of the loop: the UUCG type of loop strongly favors a C:G closing pair; CUUG favors a G:C closing pair, whereas the closing pair for the GCAA loop is less constrained but tends to be R:Y (69).

The sequence of the majority of tetraloops in the small-subunit rRNA (10 of the 18 seen in Fig. 1) changes extremely slowly over time and is highly constrained when it does so; in most cases the sequence (and its variations) conforms to one of the above three general types. However, three, perhaps four, of the tetraloops in the small-subunit rRNA change sequence with moderate to high frequency. Even in these cases, however, one sees an almost exclusive alternation among the three main types of sequence. For example, the tetraloop 83 to 86, the most variable of all, has the composition UUCG, CUUG, or GCAA in 93% of cases (69). Most of the very few alternatives to these are also tetraloops (often of related compositions), but a small fraction, about 3% of the total, are loops comprising three or five bases. The UUCG loop at this locale uses a C:G closing pair 91% of the time, the CUUG loop closes with a G:C pair in 95% of cases, and the GCAA loop closes with an R:Y (mainly A:U) pair 86% of the time.

The evolutionary pathway for which any one of these tetraloops (and their closing base pair) transforms to another is not readily apparent. There are cases in which it appears that this loop goes through a tri- and/or pentaloop intermediate (i.e., insertion-deletion events); however, other scenarios cannot be ruled out, nor should we assume a single evolutionary pathway for the maintenance of this loop constraint. In this context, it should also be pointed out that the constraints on tetraloops at different locations in the 16S and 23S rRNA vary in their tempo and mode.

There can be no doubt that these three loop compositions with their corresponding characteristic closing pairs have a strong selective advantage over any of their more than 4,000 alternatives. A question that remains to be answered is whether this selective advantage merely reflects structural and

energetic (1, 60) considerations confined to the tetraloop structure itself or whether selective advantage also derives from the interaction of these particular compositions with surrounding structures in the overall rRNA structural context, which it obviously does in a few cases (Fig. 1 and 2).

The fact that tetraloops at some locations in rRNA frequently vary in composition, whereas the composition varies not at all or very slowly and in a highly constrained way in others, strongly suggests that tetraloops play functionally different roles in overall rRNA structure. It seems entirely reasonable that some of them, such as the loop (position 83) discussed above, may be involved primarily in nucleating or controlling rRNA folding. And as we shall see, some are definitely involved in more-complex structures.

Pseudoknots. Pseudoknots are a recurring motif in rRNA. What is meant by pseudoknot in the present context is an interaction of the bases within a (simple) hairpin loop with bases external to the hairpin proper. Three such pseudoknots are shown in Fig. 1 for the small-subunit rRNA, and 15 or so are shown in Fig. 2 for the large subunit. The first of these, involving the loop starting at position 14, is within 20 nucleotides of the 5' terminus of the small-subunit rRNA, and for this reason could be either a pseudoknot or a true knot.

The pseudoknot structure involving the bulge loop and capping loop of the small-subunit rRNA structure between positions 500 and 545 is of particular interest because of the functional importance of this region of the molecule (47) and the high degree of constraint the structure would place on that region (if all helices were to form simultaneously). Experimental testing of the putative interaction of positions 505 to 507 with 524 to 526 lends strong support to the proposed structure (51).

The reader will note the pairing 521–522:527–528 (28) in 16S rRNA (Fig. 1). This interaction would constrain the pseudoknot sequence to a tetraloop. We consider this pairing very suggestive; its pairing is based on three phylogenetically independent examples of change in the mitochondria—*Chlamydomonas reinhardtii*, the two nematodes *Caenorhabditis elegans* and *Ascaris suum*, and the mushroom *Suillus sinuspaulianus* (8a; see reference 23 and references therein). In all three taxa, the C:G pair involving 522:527 changes to an A:U.

The third pseudoknot seen in the small-subunit rRNA, which involves the tetraloop at position 863 and the bases at positions 570 to 571, appears to be a complex and interesting structure (25). Note in Fig. 1 that the “external” bases in the pseudoknot (positions 570 to 571) are directly adjacent to the helix whose terminal base pair is 569:881. Adjacent position 880 (which must be spatially close to position 570) is also involved in a separate tertiary lone pair, with position 575. The two terminal nucleotides of this tetraloop, namely 863 and 866, tend to covary, although not in a strict one-to-one manner. When 866 is a purine, position 863 tends strongly to have a G composition; when 866 is a pyrimidine (predominantly C), position 863 tends strongly to have the composition U (23). Although in a sense, there is a possible triple interaction here (given the strict covariation between positions 866 and 570), it may be that variations in position 863 merely reflect those in position 866, the opposing terminal base in a tetraloop. However, the predominant pattern of covariation in this case, U:C ↔ G:R, is not that typical of a tetraloop (see above); a triple interaction should not be ruled out. (A similar situation will be encountered below, involving positions 2111, 2144, and 2147 in the large-subunit rRNA.)

The structure in this particular area of the small-subunit rRNA appears to be rather complicated. In addition to the pseudoknot in question, which lies above the helix 567–569:

881–883 (but cannot be coaxial with it without breaking the 861–862:868–867 helix), position 880 in the immediate vicinity covaries with position 575 (Fig. 1; see Fig. 11 in reference 56). Although not indicated in Fig. 1, there would appear to exist (on the basis of marginal evidence) covariation between positions 886 and 564, and two other “conflicting” marginal covariations, between positions 885 and 912 and between 887 and 910. If these weakly supported interactions are ultimately demonstrated, the geometry of the central section of the small-subunit rRNA (Fig. 1) will be complex indeed.

Between positions 53 and 117 of the large-subunit rRNA of bacteria and archaea is a tight structure involving several tertiary canonical pairings complexed within four helical segments (Fig. 2). These latter pairings (65–66:88–89 [40] and 61:93 [28]) associate hairpin loops 61–66 and 88–94. The antiparallel pairing between these two hairpin loops is quite possibly extended with pairings 62:92, 63:91, and 64:90, which all show some covariation, although not as pronounced as those diagrammed in Fig. 2. Also within this minidomain is a correlation between positions 67 and 74 (40) (Fig. 2). These tertiary pairings, in conjunction with the secondary structure, form a very tight and constrained structure. Computer modeling of this region demonstrates that all of these pairings can exist simultaneously. The two helices, 54–56:114–116 and 57–59:68–70, can be coaxially stacked and lie across the compound helix 76–87:95–110, with positions A61, U62, and A63 running through the major groove of the 57–59:68–70 helix (38a). This globular and stacked structure is very dense and in agreement with the lack of accessible nucleotides to chemical probes (15). This region has no homologous counterpart in the *Eucarya* large-subunit rRNA.

Within domain III of 23S rRNA, a pseudoknot helix is formed from two consecutive pairings, G1343:C1404 and U1344:A1403, bringing the bases of three helices into close proximity. Strong comparative evidence exists for both canonical pairs, suggesting the likelihood of this complex intercalation (28, 40). Recently, this proposed interaction was put to an experimental test (37). Various base pairs at positions 1343:1404 and 1344 to 1403 were evaluated by in vitro protein-binding studies, revealing that only transcripts with canonical pairs bound this protein at wild-type levels and thus providing convincing experimental evidence for the existence of this set of base pairings and their role in the binding of an important ribosomal protein.

The 23S rRNA pseudoknots discussed here represent only a small sampling of the total number so identified (Fig. 2). We leave it as an exercise for the future to explore in additional detail all of the large-subunit rRNA pairings of this type. However, before leaving this subject, it is worth noting that the vast majority of all proposed rRNA pseudoknot helices are short (e.g., 1, 2, or 3 bp in length), located in close proximity to the ends of secondary-structure helices, and with the possibility of orientating themselves onto more than one secondary structure helix. The recent experimental characterization of a simple pseudoknot structure revealed coaxial stacking of the stems (53), lending support to the idea that some if not all of these pseudoknot helices can be coaxially stacked onto more than one adjoining helix in a static or conformationally dynamic fashion (59; see below).

Coaxial helices. It is evident that coaxial helices will be important elements in rRNA architecture. However, at present one can make only the vaguest, speculative comments concerning which of the many helices are so arranged. The coaxially juxtaposed helices in the figures shown here should be taken merely as statements of faith or aesthetic preference on the parts of the authors. The three-dimensional structure of tRNA

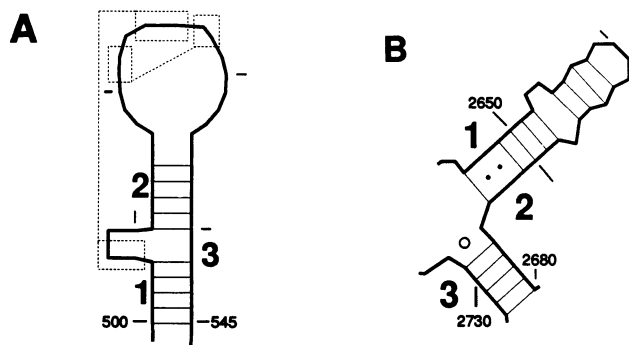


FIG. 4. Two examples of rRNA three-strand coaxial helices. (A) The 16S rRNA. The two helices 500–504:541–545 and 511–515:536–540 can stack upon one another, with strand 1 representing 500 to 504, strand 2 representing 511 to 515, and strand 3 representing 536 to 545. (B) The 23S rRNA. The two helices 2646–2652:2668–2674 and 2675–2680:2727–2732 can stack upon one another, with strand 1 representing 2646 to 2652, strand 2 representing 2668 to 2680, and strand 3 representing 2727 to 2732.

provides two examples of the simplest type of coaxial arrangement: in the arrangement of the stalks of the so-called common and amino acid acceptor arms and in that of anticodon and so-called dihydro-U arm stalks. The former is strictly coaxial: the two stalks directly abut, end to end. The other coaxial structure is only approximately so: the ends of the two adjacent canonically paired stalks are separated by an intervening A:G pair (formed at the terminus of the anticodon stalk; yeast-Phe), which gives a slight kink to their common helical axis at the point where the two helices join. (In many tRNAs, especially among the archaea [19], this A:G type of pair is replaced by a canonical one, suggesting that in some cases the two helices may be strictly coaxial.)

The structures seen in tRNA can be characterized as a “three-strand” coaxial helix, because three separate contiguous stretches in the molecule are involved in their formation. In principle, however, there exists another kind of coaxial helix, a “four-strand” coaxial helix, in which two disjoint double-helical units in an RNA become arranged coaxially because single-stranded extensions of each are complementary to one another. (This is somewhat analogous to restriction fragments joined by sticky ends.) All helices in rRNA that directly abut, or would do so by formation of an intermediate noncanonical pair (especially of the A:G type), should be considered potential (three-strand) coaxial helices, and evidence for their coaxiality should be sought. (To help illustrate three-strand coaxial helices, two examples are shown in Fig. 4. We will come back to these two examples later.)

Let us consider a few specific examples of helical elements having too few pairs to be stable in isolation. Some helices whose overall length appears to be merely two pairs, and certainly all those comprising a single pair, are prime candidates for coaxial stacking on adjacent longer helices. Although the *E. coli* small-subunit rRNA shows no lone canonical pairs, examples can be found in other prokaryotic small-subunit rRNAs and in the 23S rRNA (see below). The helix located between positions 198 and 219 in the small-subunit rRNA is variable in length, and in some cases comprises a stalk of only one pair, in both bacteria and archaea. Coaxial stacking on the adjacent helix (136–142:221–227), which requires the interpolation of an A:G pair (143:220) in *E. coli* (Fig. 1), seems a strong possibility. (Indeed, in a number of bacterial groups there exists a loose covariation between the positions in

question, i.e., 143 and 220. This covariation often involved U:G \leftrightarrow C:A interchanges, although other pairing patterns are seen as well.)

In the large-subunit rRNA (Fig. 2) three lone canonical pairs, 319:323, 1082:1086, and 1752:1756, that define loops of three bases have been identified; each could be stabilized by stacking upon an immediately adjacent helix. Three-dimensional modeling of one of them, the highly constrained L11 binding site (positions 1053 to 1106), shows that stacking 1082:1086 onto the 1057–1064:1074–1081 helix is compatible with and supports a structure that accommodates the known secondary pairings and experimental data (14). The model requires sharp turns for the bases in the loop, either between U1083 and A1084 or between A1084 and A1085, leaving these bases highly exposed. However, the bases are not accessible to chemical modification either in isolated RNA or in the 50S subunit (15). This may merely reflect the fact that in the three-dimensional model the “helix” in question (1082:1086) is buried in a cavity formed by the two flanking helices. In contrast, the terminal loops of the other two lone-pair helices, 319:323 and 1752:1756, are accessible to attack (15), suggesting that the function of these structural elements may be to cap the longer flanking helices, adjust the backbone of the RNA to maintain a tight and stable structure, or both.

Comparative support for coaxial helices of the three-strand type would consist of instances in which the overall coaxial structure retains a constant length in two groups of organisms; in one group one helix is shorter than it is in the other group, while the other helix is compensatingly longer than in the other phylogenetic group (66). An example of such is known for the 500-to-545 region of 16S rRNA, involving the helices 500–504:541–545 and 511–515:536–540 (Fig. 1 and 4). Taken together, these two helices total 10 bp. In (*Eu*)*Bacteria* the two helices are 5 and 7 bp in length, while in the *Archaea* and *Eucarya* both are 6 bp in length (23, 64).

Another potential coaxial stack is possible at the base of the α -sarcin helix in the large-subunit rRNA, involving the two helices 2646–2652:2668–2674 and 2675–2680:2727–2732 (Fig. 2 and 4). The two have a combined length of 13 bp, but their individual lengths differ in three phylogenetic domains (23). In (*Eu*)*Bacteria* the two have respective lengths of 7 and 6 bp, whereas in *Archaea* and *Eucarya* the corresponding lengths are 8 and 5 bp. The *E. coli* sequence (Fig. 2) has an A:G pair “hinge” joining the two coaxial helices. It is of interest that this pair (2675:2732) covaries in (*Eu*)*Bacteria* between A:G and G:U.

Another indication of coaxiality might be when an abutting pair or pairs can be formed in more than one way, i.e., as an extension of either one of the helices. This would in effect amount to an entropic contribution to the energy of coaxial stacking (66). To illustrate these principles, consider the potential coaxial helix in the small-subunit rRNA formed from the two (adjacent) helices in the region between positions 315 and 351 (Fig. 1). These two helices could be made coaxial by forming either the pair 315:338 or the pair 338:351. In (*Eu*)*Bacteria* and *Eucarya* the 338:351 pair (only) is A:G, while in *Archaea* the 315:338 pair (only) is A:G. In the mitochondria several independent examples exist in which the presumed (bacterial) ancestral form of 338:351 (A:G) varies either to C:G or to U:A. In the latter case the possibility exists for two alternative joining pairs, A315:U338 or U338:A351.

In the case of four-strand coaxial structures the bases in the extensions of the component helices in the putative coaxial structure will covary. Although coaxial structures of this type are possible in the small-subunit rRNA, two potential cases exist in the large-subunit rRNA, involving proven tertiary

interactions. By means of the pseudoknot pairing 317–318:333–334, the lone canonical pairing 319:323 and the secondary-structure helix 325–327:335–337 could be brought into coaxial juxtaposition. Similarly, the pseudoknot pairing 1343–1344:1403–1404 brings three helices, 1345–1347:1599–1601, 1385–1389:1398–1402, and 1405–1415:1587–1597, together, creating the potential for their coaxial stacking. Other examples of coaxial stackings are possible in the immediate vicinity of other 23S rRNA pseudoknot structures.

Parallel pairs. Comparative analysis suggests two structural elements that may involve parallel pairing, one in the small-subunit rRNA and one in the large, near the so-called E site (44). In the small-subunit rRNA, the complex interaction in the 437-to-440 versus 494-to-498 region, discussed above, may well involve parallel pairing, if the well-established pairing 438:496 and the potential pair 440:497 occur simultaneously. (Remember, however, that 440:497 may well be a triple interaction, including position 494 as well [see above].)

There is now strong evidence for the two adjacent parallel pairs, 2112:2169 and 2113:2170, in the large subunit (28). A number of phylogenetically independent occurrences of G:A \leftrightarrow A:G are seen between position 2112 and 2169, whereas the U:A pair at 2113:2170 can be replaced by C:G and G:A pairings. The G2112:A2169 pairing may not involve N1 and N2 of G2112, because these positions are strongly reactive to kethoxal in isolated rRNA and 50S subunits (15). In addition to these two pairings in parallel, there is now some comparative evidence suggesting a canonical pair between positions A2117 and U2172, which, if substantiated, would extend this set of parallel interactions (Fig. 2). The 2112 and 2170 regions have been cross-linked (12), offering additional evidence for this set of unusual pairings.

The nucleotide immediately adjacent to the first of these parallel pairings, i.e., position 2111, covaries with the two terminal nucleotides of the tetraloop 2144 to 2147 (38) (Fig. 2); see discussion in the section on tetraloops. This latter covariation, 2111 with 2147, alternates between U:A and C:G (a number of phylogenetically independent times). It is conceivable that a triple-base interaction is involved in this case, in which a canonical pair forms between U2111 and A2147 and in which N3 and N2 of A2147 interact with G2144. Alternatively, the two implied pairings of A2147 (with U2111 and G2144) need not occur simultaneously. If the latter were true, this would be another example of transient interactions, forming only at particular stages in the translational cycle.

SUMMARY AND CONCLUSIONS

The 16S and 23S rRNA higher-order structures inferred from comparative analysis are now quite refined. The models presented here differ from their immediate predecessors only in minor detail. Thus, it is safe to assert that all of the standard secondary-structure elements in (prokaryotic) rRNAs have been identified, with approximately 90% of the individual base pairs in each molecule having independent comparative support, and that at least some of the tertiary interactions have been revealed. It is interesting to compare the rRNAs in this respect with tRNA, whose higher-order structure is known in detail from its crystal structure (36) (Table 2). It can be seen that rRNAs have as great a fraction of their sequence in established secondary-structure elements as does tRNA. However, the fact that the former show a much lower fraction of identified tertiary interactions and a greater fraction of unpaired nucleotides than the latter implies that many of the rRNA tertiary interactions remain to be located. (Alternatively, the ribosome might involve protein-rRNA rather than

TABLE 2. Degree of known structure in tRNA, 16S, and 23S rRNA

RNA type	Total no. of nucleotides	% of total nucleotides with:		
		Secondary-structure base pairings	Tertiary interactions ^a	No base pairs (unpaired)
tRNA	76	55	20	25
16S rRNA	1,542	60	3	37
23S rRNA	2,904	58	2	40

^a Including base triples.

intramolecular rRNA interactions to stabilize three-dimensional structure.)

Experimental studies on rRNA are consistent to a first approximation with the structures proposed here, confirming the basic assumption of comparative analysis, i.e., that bases whose compositions strictly covary are physically interacting. In the exhaustive study of Moazed et al. (45) on protection of the bases in the small-subunit rRNA against chemical modification, the vast majority of bases inferred to pair by covariation are found to be protected from chemical modification, both in isolated small-subunit rRNA and in the 30S subunit. The majority of the tertiary interactions are reflected in the chemical protection data as well (45). On the other hand, many of the bases not shown as paired in Fig. 1 are accessible to chemical attack (45). However, in this case a sizeable fraction of them are also protected against chemical modification (in the isolated rRNA), which suggests that considerable higher-order structure remains to be found (although all of it may not involve base-base interactions and so may not be detectable by comparative analysis).

The agreement between the higher-order structure of the small-subunit rRNA and protection against chemical modification is not perfect, however; some bases shown to covary canonically are accessible to chemical modification (45). For example, in both the small subunit and the isolated rRNA therefrom, position 66 is readily modified chemically; however, this position shows strong canonical covariation with position 103—there are over 30 phylogenetically independent examples among the bacteria, without exception. These discrepancies cannot be unequivocally interpreted. They could reflect the 16S rRNA or the 30S subunit not being in a functionally optimal state *in vitro*; they might imply that the bases inferred as paired by comparative analysis are not paired throughout the entire translation cycle. However, it is highly unlikely that discrepancies of this kind imply that these bases (at positions 66 and 103) do not pair at all. It is interesting that the protections that hold for the isolated rRNA do not agree with those shown by the corresponding intact subunit in a few cases, which could be interpreted to mean that the functional structure of rRNA, although primarily inherent in the rRNA itself, is secondarily determined by association with protein (45). A good example here may be the tertiary interaction discussed above, 722:733, a G:G \leftrightarrow A:A covariation. In the isolated small-subunit rRNA, position 722 is strongly modified by chemical reagents and the following position, 723, is protected. In the small subunit itself, however, position 722 is protected while position 723 is strongly modified (45). It would appear that the noncanonical lone pair interaction 722:733 forms only in the presence of ribosomal proteins. This is particularly interesting in view of the previously noted finding regarding the role of G:G (or alternatively A:A) pairs in the Rev protein recognition site on human immunodeficiency virus type 1 RNA (3).

It should be noted that although there is some agreement between the higher-order structures adduced by comparative analysis of the large rRNAs and those predicted from folding algorithms, primarily involving local structures (e.g., hairpins), overall agreement tends to be poor (34). One wonders whether such algorithms, especially those that do not take a great deal of empirical evidence into account, will ever be able to predict these structures with a reasonable (useful) degree of accuracy. In any case, folding algorithms that take comparative structure into account are very much needed.

Along these lines, we would like to make a more general point. Comparative analysis has played a very strong role in determining the structure of several RNA molecules (reviewed in reference 22), including a smaller, highly variable RNA, the RNA moiety of RNase P (35). In this case many of the predictions have been tested and a "minimal" functional form of the molecule has been predicted, genetically engineered, and shown to be functional (62). Examples such as these, together with the rRNA structures, clearly point to the role comparative analysis should be playing in experimental approaches to molecular structure. To this should be added the value of comparative analysis in detecting the modulo 3 variability spikes in genetic sequences (corresponding to the third codon position) that identify protein-coding genes.

At this junction we question what additional information about RNA structure can be inferred by using comparative methods. Our comparative rationale for RNA structure determination is based on the simple concept of a homologous structure for the RNA molecule under study. Our primary method for identifying this isomorphic structure relies on the search for compensatory base substitutions or positional covariance, which has revealed a secondary structure and the beginnings of its tertiary structure for the 16S and 23S rRNAs. This search has, to a first approximation, identified the structural elements in common with all sequences in their respective data sets. However, we have noted a few examples of structural features common only to members of certain phylogenetic groupings (noted herein [64]). (Another example involves the 1850 region of 23S rRNA, where the basic helical region is different between the three phylogenetic domains [20] and also different within the (*Eu*)*Bacteria*, where specific noncanonical pairings in *E. coli* and related purple bacteria are replaced with several examples of canonical pairings [23].) With the large and diverse 16S and 23S rRNA sequence collection now available, we can begin to systematically search for other minor structural elements common only to a subset of the entire 16S and 23S data bases. Such studies should result in additional refinement of 16S and 23S rRNA structure.

Further refinements in rRNA structure will also come from a more exhaustive and quantitative analysis of the 16S and 23S rRNA data sets. Newer quantitative correlation algorithms, under development, are more sensitive than previous methods and are beginning to identify helical base pairings constrained by surrounding base pairs and other nucleotides not considered to be directly involved in structural interactions (19a, 26). With the ever-increasing 16S and 23S rRNA sequence collections and the more powerful and dynamic correlation analysis algorithms, we should expect to find more structural constraints in these RNA molecules, many of which will not involve positions that covary in the strict one-to-one manner. Ultimately, comparative methods will go beyond the search for positions that covary in a simple one-to-one manner as the paradigm for homologous structure. Instead, these methods will utilize a growing appreciation for RNA conformations and a mapping between sequence and its secondary and tertiary structure to assist in the search for isomorphic structure.

The comparative method, although inferring base pairings, does not imply or require that all of these pairings occur simultaneously. Unfortunately, the manner in which these secondary-structure figures are drawn suggests just this, a static structure devoid of possible structural alternations. The essence of ribosomal function most probably involves dynamic movement in rRNA structure. Transforming these static 16S and 23S rRNA structures presented here into a functioning ribosome will entail experimental approaches. The ultimate goal of all analysis of the ribosome is to determine its structure and relate this to its function. Comparative analysis clearly cannot carry rRNA structure to this point. Only a combination of approaches can produce the picture of the functioning ribosome.

ACKNOWLEDGMENTS

R.R.G. is supported by the NIH (grant GM 48207), N.L. is supported by postdoctoral grant 11-8804 from the Danish Natural Science Research Council, and the work of C.R.W. on the ribosome is supported by a grant from NASA. R.R.G. and C.R.W. are both associates in the Evolutionary Biology Program of the Canadian Institute for Advanced Research.

The programming expertise of Bryn Weiser and Tom Macke is greatly appreciated in these endeavors. Finally, we also thank the W. M. Keck Foundation for their generous support of RNA science on the Boulder campus.

REFERENCES

1. Antao, V. P., S. Y. Lai, and I. Tinoco, Jr. 1991. A thermodynamic study of unusually stable RNA and DNA hairpins. *Nucleic Acids Res.* **19**:5901-5905.
2. Aubert, M., J. F. Scott, M. Reynier, and R. Monier. 1968. Rearrangement of the conformation of *Escherichia coli* 5S rRNA. *Proc. Natl. Acad. Sci. USA* **61**:292-299.
3. Bartel, D. P., M. L. Zapp, M. R. Green, and J. W. Szostak. 1991. HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell* **67**:529-536.
4. Branlant, C., A. Krol, M. A. Machatt, J. Pouyet, J. P. Ebel, K. Edwards, and H. Kossel. 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.
5. Brimacombe, R., B. Breuer, P. Mitchell, M. Osswald, J. Rinke-Appel, D. Schuler, and K. Stade. 1990. Three-dimensional structure and function of *Escherichia coli* 16S and 23S rRNA as studied by cross-linking techniques, p. 93-106. *In* W. E. Hill, A. Dahlberg, R. A. Garrett, P. B. Moore, D. Schlessinger, and J. A. Warner (ed.), *The ribosome: structure, function, and evolution*. American Society for Microbiology, Washington, D.C.
6. Brosius, J., T. Dull, and H. F. Noller. 1980. Complete nucleotide sequence of a 23S ribosomal RNA gene from *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **77**:201-204.
7. Brosius, J., T. 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.
8. Brosius, J., M. L. Palmer, P. J. Kennedy, and H. F. Noller. 1978. Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **75**:4801-4805.
- 8a. Bruns, T. Unpublished data.
9. Chastain, M., and I. Tinoco. 1991. Structural elements in RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **41**:131-177.
10. Cunningham, P. R., K. Nurse, A. Bakin, C. J. Weitzmann, M. Pflumm, and J. Ofengand. 1992. Interaction between the two conserved single-stranded regions at the decoding site of small subunit ribosomal RNA is essential for ribosome function. *Biochemistry* **31**:12012-12022.
11. Cunningham, P. R., K. Nurse, C. J. Weitzmann, and J. Ofengand. 1993. Functional effects of base changes which further define the decoding center of *Escherichia coli* 16S ribosomal RNA: mutation

- of C1404, G1405, C1496, G1497, and U1498. *Biochemistry* **32**: 7172-7180.
12. Doring, T., B. Gruer, and R. Brimacombe. 1991. The three-dimensional folding of ribosomal RNA: localization of a series of intra-RNA cross-links in 23S RNA induced by treatment of *Escherichia coli* 50S ribosomal subunits with bis-(2-chloroethyl)-methylamine. *Nucleic Acids Res.* **19**:3517-3524.
 13. Doring, T., B. Gruer, and R. Brimacombe. 1992. The topography of the 3'-terminal region of *Escherichia coli* 16S ribosomal RNA: an intra-RNA cross-linking study. *Nucleic Acids Res.* **20**:1593-1597.
 14. Egebjerg, J., S. R. Douthwaite, A. Liljas, and R. A. Garrett. 1990. Characterization of the binding sites of protein L11 and the L10(L12)₄ pentameric complex in the GTPase domain of 23S ribosomal RNA from *Escherichia coli*. *J. Mol. Biol.* **213**:275-288.
 15. Egebjerg, J., N. Larsen, and R. A. Garrett. 1990. Structural map of 23S rRNA, p. 168-179. In W. E. Hill, A. Dahlberg, R. A. Garrett, P. B. Moore, J. Schlessinger, and J. R. Warner (ed.), *The ribosome: structure, function and evolution*. American Society for Microbiology, Washington, D.C.
 16. Ehresmann, B., C. Ehresmann, P. Romby, M. Mougel, F. Baudin, E. Westhof, and J.-P. Ebel. 1990. Detailed structures of rRNAs: new approaches, p. 148-159. In W. E. Hill, A. Dahlberg, R. A. Garrett, P. B. Moore, D. Schlessinger, and J. R. Warner (ed.), *The ribosome: structure, function, and evolution*. American Society for Microbiology, Washington, D.C.
 17. Fox, G. E., and C. R. Woese. 1975. 5S RNA secondary structure. *Nature (London)* **256**:505-507.
 18. Glotz, C., C. Zwieb, and R. Brimacombe. 1981. Secondary structure of the large subunit ribosomal RNA from *Escherichia coli*, *Zea mays* chloroplast, and human and mouse mitochondrial ribosomes. *Nucleic Acids Res.* **9**:3287-3306.
 19. Gupta, R. 1984. Halobacterium volcanii tRNAs: identification of 41 tRNAs covering all amino acids, and the sequences of 33 class 1 tRNAs. *J. Biol. Chem.* **259**:9461-9471.
 - 19a. Gutell, R. Unpublished data.
 20. Gutell, R. R. 1992. Evolutionary characteristics of 16S and 23S rRNA structures, p. 243-309. In H. Hartman and K. Matsuno (ed.), *The origin and evolution of prokaryotic and eukaryotic cells*. World Scientific Publishing Co., River Edge, N.J.
 21. Gutell, R. R. 1993. Collection of small subunit (16S- and 16S-like) ribosomal RNA structures. *Nucleic Acids Res.* **21**:3051-3054. (Data base issue.)
 22. Gutell, R. R. 1993. Comparative studies of RNA: inferring higher-order structure from patterns of sequence variation. *Curr. Opin. Struct. Biol.* **3**:313-322.
 23. Gutell, R. R. 1993. The simplicity behind the elucidation of complex structure in ribosomal RNA, p. 477-488. In Nierhaus et al. (ed.), *The translational apparatus*. Plenum Publishing Corp., New York.
 24. Gutell, R. R., M. W. Gray, and M. N. Schnare. 1993. A compilation of large subunit (23S- and 23S-like) ribosomal RNA structures. *Nucleic Acids Res.* **21**:3055-3074. (Data base issue.)
 25. Gutell, R. R., H. F. Noller, and C. R. Woese. 1986. Higher order structure in ribosomal RNA. *EMBO J.* **5**:1111-1113.
 26. Gutell, R. R., A. Power, G. Z. Hertz, E. J. Putz, and G. D. Stormo. 1992. Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. *Nucleic Acids Res.* **20**:5785-5795.
 27. Gutell, R. R., B. Weiser, C. R. Woese, and H. F. Noller. 1985. Comparative anatomy of 16S-like ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **32**:155-216.
 28. Gutell, R. R., and C. R. Woese. 1990. Higher order structural elements in ribosomal RNAs: pseudo-knots and the use of non-canonical pairs. *Proc. Natl. Acad. Sci. USA* **87**:663-667.
 29. Haselman, T., D. G. Camp, and G. E. Fox. 1988. Phylogenetic evidence for tertiary interactions in 16S-like ribosomal RNA. *Nucleic Acids Res.* **17**:2215-2221.
 30. Haselman, T., R. R. Gutell, J. Jurka, and G. E. Fox. 1989. Additional Watson-Crick interactions suggest a structural core in large subunit rRNA. *J. Biomol. Struct. Dyn.* **7**:181-186.
 31. Heus, H. A., and A. Pardi. 1991. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **253**:191-194.
 32. Holley, R. W., J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick, and A. Zamir. 1965. Structure of a ribonucleic acid. *Science* **147**:1462-1465.
 33. Hou, Y.-M., and P. Schimmel. 1988. A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature (London)* **333**:140-145.
 34. Jaeger, J. A., D. H. Turner, and M. Zuker. 1989. Improved predictions of secondary structures for RNA. *Proc. Natl. Acad. Sci. USA* **86**:7706-7710.
 35. James, B. D., G. J. Olsen, J. Liu, and N. R. Pace. 1988. The secondary structure of ribonuclease P RNA, the catalytic element of a ribonucleoprotein enzyme. *Cell* **52**:19-26.
 36. Kim, S.-H. 1979. Crystal structure of yeast tRNA-phe and general structural features of other tRNAs, p. 83-100. In P. R. Schimmel, D. Soll, and J. N. Abelson (ed.), *Transfer RNA: structure, properties, and recognition*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 37. Kooi, E. A., C. A. Rutgers, A. Mulder, J. Van't Riet, J. Venema, and H. A. Raue. 1993. The phylogenetically conserved doublet tertiary interaction in domain III of the large subunit rRNA is crucial for ribosomal protein binding. *Proc. Natl. Acad. Sci. USA* **90**:213-216.
 38. Larsen, N. 1992. Higher order interactions in 23S rRNA. *Proc. Natl. Acad. Sci. USA* **89**:5044-5048.
 - 38a. Larsen, N. Unpublished data.
 39. Larsen, N., G. J. Olsen, B. L. Maidak, M. J. McCaughey, R. Overbeek, T. J. Macke, T. L. Marsh, and C. R. Woese. 1993. The ribosomal database project. *Nucleic Acids Res.* **21**:3021-3023. (Data base issue.)
 40. Leffers, H., J. Kjems, L. Ostergaard, N. Larsen, and R. A. Garrett. 1987. Evolutionary relationships amongst Archaeobacteria. A comparative study of 23 S ribosomal RNAs of a sulphur-dependent extreme thermophile, an extreme halophile and a thermophilic methanogen. *J. Mol. Biol.* **195**:43-61.
 41. Levitt, M. 1969. Detailed molecular model for transfer ribonucleic acid. *Nature (London)* **224**:759-763.
 42. Madison, J. T., G. A. Everett, and H. K. Kung. 1966. On the nucleotide sequence of yeast tyrosine transfer RNA. *Cold Spring Harbor Symp. Quant. Biol.* **31**:409-416.
 43. McClain, W. H., and K. Foss. 1988. Changing the identity of a tRNA by introducing a G-U wobble pair near the 3' acceptor end. *Science* **240**:793-796.
 44. Moazed, D., and H. F. Noller. 1989. Interaction of tRNA with 23S rRNA in the ribosomal A, P, and E sites. *Cell* **57**:585-597.
 45. Moazed, D., S. Stern, and H. F. Noller. 1986. Rapid chemical probing of conformation in 16S ribosomal RNA and 30S ribosomal subunits using primer extension. *J. Mol. Biol.* **187**:399-416.
 46. Neefs, J.-M., Y. Van de Peer, P. De Rijk, S. Chapelle, and R. De Wachter. 1993. Compilation of small ribosomal subunit RNA structures. *Nucleic Acids Res.* **21**:3025-3049. (Data base issue.)
 47. Noller, H. F. 1991. Ribosomal RNA and translation. *Annu. Rev. Biochem.* **60**:191-227.
 48. 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.
 49. Noller, H. F., D. Moazed, S. Stern, T. Powers, P. N. Allen, J. M. Robertson, B. Weiser, and K. Triman. 1990. Structure of rRNA and its functional interaction in translation, p. 73-92. In W. E. Hill, A. Dahlberg, R. A. Garrett, P. B. Moore, D. Schlessinger, and J. R. Warner (ed.), *The ribosome: structure, function, and evolution*. American Society for Microbiology, Washington, D.C.
 50. Olsen, G. J. 1983. Comparative analysis of nucleotide sequence data. Ph.D. thesis. University of Colorado, Denver.
 51. Powers, T., and H. F. Noller. 1991. A functional pseudoknot in 16S ribosomal RNA. *EMBO J.* **10**:2203-2214.
 52. Puglisi, J. D., J. R. Wyatt, and I. Tinoco. 1990. Solution conformation of an RNA hairpin loop. *Biochemistry* **29**:4215-4226.
 53. Puglisi, J. D., J. R. Wyatt, and I. Tinoco, Jr. 1990. Conformation of an RNA pseudoknot. *J. Mol. Biol.* **214**:437-453.
 54. Rajbhandary, U. L., A. Stuart, R. D. Faulkner, S. H. Chang, and H. G. Khorana. 1966. Nucleotide sequence studies on yeast phenylalanine sRNA. *Cold Spring Harbor Symp. Quant. Biol.* **31**:425-434.

55. SantaLucia, J., R. Kierzek, and D. H. Turner. 1991. Stabilities of consecutive A-C, C-C, G-G, U-C, and U-U mismatches in RNA internal loops: evidence for stable hydrogen-bonded U-U and C-C+ pairs. *Biochemistry* **30**:8242-8251.
56. Stern, S., B. Weiser, and H. F. Noller. 1988. Model for the three-dimensional folding of 16S ribosomal RNA. *J. Mol. Biol.* **204**:447-481.
57. Stiege, W., C. Glotz, and R. Brimacombe. 1983. Localisation of a series of intra-RNA cross-links in the secondary and tertiary structure of 23S RNA, induced by ultraviolet irradiation of *Escherichia coli* 50S ribosomal subunits. *Nucleic Acids Res.* **11**:1687-1706.
58. 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.
59. ten Dam, E., K. Pleij, and D. Draper. 1992. Structural and functional aspects of RNA pseudoknots. *Biochemistry* **31**:11665-11676.
60. Tuerk, C., P. Gauss, C. Thermes, D. R. Groebe, M. Gayle, N. Guild, G. Stormo, Y. D'Aubenton-Carafa, O. C. Uhlenbeck, I. Tinoco, E. N. Brody, and L. Gold. 1988. CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc. Natl. Acad. Sci. USA* **85**:1364-1368.
61. Varani, G., C. Cheong, and I. Tinoco, Jr. 1991. Structure of an unusually stable RNA hairpin. *Biochemistry* **30**:3280-3289.
62. Waugh, D. S., C. J. Green, and N. R. Pace. 1989. The design and catalytic properties of a simplified ribonuclease P RNA. *Science* **244**:1569-1571.
63. Winker, S., R. Overbeek, C. R. Woese, G. J. Olsen, and N. Pfluger. 1990. Structure detection through automated covariance search. *CABIOS* **6**:365-371.
64. Winker, S., and C. R. Woese. 1991. A definition of the domains Archaea, Bacteria and Eucarya in terms of small subunit ribosomal RNA characteristics. *Syst. Appl. Microbiol.* **14**:305-310.
65. Woese, C. R., and R. R. Gutell. 1989. Evidence for several higher order structural elements in ribosomal RNA. *Proc. Natl. Acad. Sci. USA* **86**:3119-3122.
66. Woese, C. R., R. Gutell, R. Gupta, and H. F. Noller. 1983. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol. Rev.* **47**:621-669.
67. Woese, C. R., O. Kandler, and M. L. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* **87**:4576-4579.
68. Woese, C. R., L. J. Magrum, R. Gupta, R. B. Siegel, D. A. Stahl, J. Kop, N. Crawford, J. Brosius, R. Gutell, 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.
69. Woese, C. R., S. Winker, and R. R. Gutell. 1990. Architecture of ribosomal RNA: constraints on the sequence of tetra-loops. *Proc. Natl. Acad. Sci. USA* **87**:8467-8471.
70. Yonath, A., W. Bennett, S. Weinstein, and H. G. Wittmann. 1990. Crystallography and image reconstructions of ribosomes, p. 134-147. *In* W. E. Hill, A. Dahlberg, R. A. Garrett, P. B. Moore, D. Schlessinger, and J. R. Warner (ed.), *The ribosome: structure, function, and evolution*. American Society for Microbiology, Washington, D.C.
71. Zachau, H. G., D. Dutting, H. Feldmann, F. Melchers, and W. Karau. 1966. Serine specific transfer ribonucleic acids. XIV. Comparison of nucleotide sequences and secondary structure models. *Cold Spring Harbor Symp. Quant. Biol.* **31**:417-424.
72. 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.