

MINIREVIEW

Evolutionary Perspective on the Structure and Function of Ribonuclease P, a Ribozyme

NORMAN R. PACE^{1*} AND JAMES W. BROWN²

Department of Biology and Institute for Molecular and Cellular Biology, Indiana University, Bloomington, Indiana 47405,¹ and Department of Microbiology, North Carolina State University, Raleigh, North Carolina 27695²

Ribonuclease P (RNase P) is a key enzyme in the biosynthesis of tRNA (for reviews, see references 3, 11, and 51). It is an RNA-processing endonuclease that specifically cleaves precursors of tRNA, releasing 5' precursor sequences and thereby forming the mature 5' ends of the tRNAs. RNase P is involved in processing all species of tRNA and is present in all cells and organelles that carry out tRNA synthesis. It is a particularly interesting enzyme because of its composition; it is a ribonucleoprotein (37). The nature of RNase P has been investigated for a broad diversity of organisms, including representatives of all the phylogenetic domains: *Archaea*, (eu)*Bacteria*, and *Eucarya* (see reference 72 for phylogenetic nomenclature). In the case of the enzyme from *Bacteria*, RNase P contains a catalytic RNA, or ribozyme (22, 62). The secondary structure of the bacterial RNase P RNA has now been determined, models for its three-dimensional structure have been proposed, and numerous correlations between catalytic function and structure are being discovered. The purpose of this minireview is to provide an overview of RNase P, with emphasis on the structure of the bacterial RNase P RNA and how it varies between different organisms. Any understanding of the function of this unusual type of enzyme will rest on knowledge of its structure and patterns of variation.

The RNase P Reaction

In *Bacteria*, as exemplified by *Escherichia coli* or *Bacillus subtilis*, the RNase P holoenzyme is composed of a large RNA (usually 350 to 400 nucleotides) and a single molecule of a small protein (ca. 120 amino acids in known instances). The bacterial RNase P RNA is clearly the catalyst in the reaction. At high ionic strengths *in vitro*, all bacterial RNase P RNAs tested, which collectively span the breadth of bacterial diversity, can cleave pre-tRNAs in the absence of the protein moiety. The affinity of the *E. coli* or *B. subtilis* RNase P for native pre-tRNAs is the same in the presence or absence of the corresponding RNase P protein (57), so all of the specific elements required for binding and catalysis must reside in the RNA moiety. In contrast, archaeal and eucaryal RNase P RNA subunits have not yet been found to exhibit catalytic activity after purification to remove proteins (see below). Because of the opportunity to correlate catalytic activity with structure, most experimental attention on RNase P RNA has focused on the bacterial version. Recent technological developments have been critical to progress in this field. With the availability of oligonucleotide-directed mutagenesis and *in vitro* transcription

systems (45), the design and synthesis of experimental RNase P and substrate RNAs have become relatively simple tasks.

RNase P cleaves precursors of tRNA and a few other small RNAs which probably resemble tRNA, for instance 4.5S RNA in *E. coli* (52). Much is known about how the bacterial RNase P interacts with its substrate. Modification-interference (31, 64, 20) and cross-linking experiments (8, 48, 50) have shown directly that RNase P mainly recognizes the mature domain of pre-tRNAs. The enzyme binds to the helix formed by coaxial stacking of the common arm and acceptor stem of the tRNA, which is adjacent to the site of cleavage. Short helices that mimic only the acceptor stem also are effective substrates for the enzyme (44). Such substrate mimics can sometimes be formed by annealing a short oligonucleotide to a longer RNA (16); thus, RNase P potentially is useful as a general nuclease, perhaps even as a tool for gene therapy (2, 18).

Cations, both monovalent (preferably K⁺ or NH₄⁺) and divalent (preferably Mg²⁺), are of critical importance in the RNase P reaction. Approximately 1 M monovalent cation and 10 mM Mg²⁺ elicit the maximum catalytic efficiency (k_{cat}/K_m) in the RNA-only reaction. The role of monovalent cations, and possibly the main role of the rather basic RNase P protein, in promoting the activity of the RNase P RNA seems to be in counteracting electrostatic repulsion between RNA phosphates (19, 22). In the absence of sufficient ionic screening, electronegative repulsion would interfere with formation of the enzyme-substrate complex and possibly distort the superstructure of the ribozyme. The protein probably makes specific stabilizing associations with the RNase P RNA, and possibly with the tRNA substrate. This is indicated by the fact that the protein can globally stabilize the structure of the RNA and thereby suppress the effects of many mutations manifest in the RNA-only reaction (42). Higher ionic strengths than those which are optimal for the native ribozyme can also suppress such mutants.

Divalent cations can serve as counterions for RNA phosphates, but more importantly, they are required for the reaction mechanism (23, 32, 54, 59, 61). Multiple Mg²⁺ ions participate in the reaction (59). Mg²⁺ is proposed to promote catalysis in two ways: by activating a water molecule to hydrolyze the susceptible phosphodiester bond in the substrate RNA and by coordinating on phosphate oxygens. The association of Mg²⁺ with phosphate oxygens would polarize the phosphorus center which is the target of the attacking water and stabilize the developing negative charge on the leaving group. It is imagined that the 3' O of the precursor chain is displaced in an S_N2-type nucleophilic reaction, forming the 5' phosphate of the mature tRNA domain (23, 59, 60). This mechanism is supported by kinetic analyses and experiments in which key tRNA residues are substituted with relevant analogs (53, 59). Of

* Corresponding author. Phone: (812) 855-6152. Fax: (812) 855-6705. Electronic mail address: PaceN@ucs.Indiana.edu.

course, all the catalytic players in the reaction must be organized by interactions with structural elements of the RNase P RNA. Therefore, understanding the overall process will require knowledge of the structure of the ribozyme.

Determination of the Bacterial RNase P RNA Secondary Structure

Functional RNAs tend to be compact in structure. Elements of "secondary" structure are arranged spatially by interactions of "tertiary" structure to achieve a functional display of binding and catalytic features. We define RNA secondary structure as resulting from two or more contiguous, canonical (Watson-Crick) or G·U base pairs, which are expected to form an A-type helix. Tertiary structure results from other intramolecular interactions, such as lone base pairs, base triples, base-backbone, and backbone-backbone interactions, and from more nebulous organization such as "stacking" interactions, in which bases or helices align on one another due to hydrophobic and van der Waals interactions between the purine and pyrimidine bases (73).

Although crystallography and spectroscopy are now routinely applied to the elucidation of protein structure, these methods have not yet proved very applicable to RNAs larger than tRNA (ca. 80 nt). Instead, the most successful method for determining higher-order structure in large RNAs has been phylogenetic comparative analysis (25). In this approach, "homologous" (i.e., of common ancestry) RNA sequences from different organisms are compared and structure is inferred on the basis of "covariations" in the sequences, changes that occur in concert. The most straightforward interpretation of concerted evolutionary change in the identity of particular bases is that they interact with one another in a specific way. Evolutionary change in one base therefore dictates change in the other in order to maintain higher-order structure required for in vivo function. In helical regions, paired bases covary to maintain complementarity. Thus, elements of secondary structure are identified by changes in a sequence collection that preserve the complementarity of putative pairings in any particular sequence. Specific base interactions in tertiary structure also can be detected by comparative analysis. However, changes in tertiary structure tend to be rarer than most compensatory changes in regular helices, so larger sets of sequences are required to detect them. Making comparative analysis of tertiary structure more difficult still, the specificities of bases that engage in tertiary pairs or triplets tend to be less rigid than the canonical complementarities that establish secondary structure.

The determination of the secondary structure of the bacterial RNase P RNA was a challenge because of the substantial sequence and length variation in the molecule from diverse organisms. The first RNase P RNA sequences determined, from *E. coli* (M1 RNA) (55) and *B. subtilis* (56), differ so much from one another that it was initially impossible to align the sequences over most of their lengths. The alignment process, ordering homologous nucleotides, usually in columns in a sequence set, is an essential step in comparative analysis; covariation of bases is meaningful only if the covarying bases are homologs. The alignment of these sequences was resolved by the determination and comparison of additional sequences from organisms of intermediate evolutionary relatedness to *E. coli* and *B. subtilis*.

Phylogenetic trees, such as those based on rRNA sequence comparisons, are important resources for the selection of source organisms for comparative sequence studies. In terms of rRNA sequence relatedness, *E. coli* and *B. subtilis* are about

as different from one another as are any bacteria. These two organisms are, respectively, representatives of the Proteobacteria ("Purple Bacteria and Relatives") and the "Gram-Positive Bacteria and Relatives," 2 of the approximately 12 main phylogenetic "kingdoms" (as proposed by Woese [71]) of *Bacteria* defined by rRNA sequence similarities. RNase P RNA sequences from other representatives of proteobacteria and gram-positive bacteria proved to be sufficiently similar within their respective groups that they could be aligned on the basis of conserved sequences. Intragroup variations provided evidence for some elements of secondary structure, but much of each sequence within the groups was invariant and so provided no tests of pairing. However, the invariant sequences within each of the two phylogenetic groups, the slowly evolving sequences, were sufficiently similar between the groups that they could be aligned and additional secondary structure could be detected by covariations (30). Comparative sequence analysis is an iterative process: as alignment is improved, additional structure is identifiable, which in turn can be useful in revision of the alignment, and so on.

The current secondary structure models for the RNase P RNAs of *E. coli* and *B. subtilis*, based on comparisons of about 50 sequences from representatives of the main bacterial kingdoms (5), are shown in Fig. 1. These and other RNase P RNAs are composed of a collection of base-paired and "single-stranded" sequence elements. The paired sequences are expected to form helices, which are sometimes irregular due to the occurrence of bulges or noncanonical pairs. The single-stranded regions probably are tightly engaged in tertiary structural interactions, which as of yet are poorly defined (see below). The general architectural features of the RNase P RNAs, for instance the occurrence of "tetraloops," "pseudoknots," noncanonical covariations, etc., are similar to those seen in the rRNAs, which have been reviewed recently (25, 73).

The *E. coli* and *B. subtilis* RNase P RNAs superficially appear quite different because each RNA contains blocks of sequence, usually helical elements, that have no homologs in the other RNA. These two structures are among the most disparate types of bacterial RNase P RNAs yet encountered. However, both types of RNA contain a core of similar (and homologous) sequence and secondary structure. This core structure is the foundation of all bacterial RNase P RNAs so far examined, which are summarized diagrammatically in Fig. 2. As with the *E. coli* and *B. subtilis* RNAs, the core structure is embellished in different organisms by the insertion of (usually) helical elements, which can differ dramatically in length and sequence. Most types of the RNA are similar in form to that of *E. coli*. The distribution of helices in the *E. coli* structure is essentially the same as that seen in the genus *Thermotoga*, one of the earliest branchings in the phylogeny of *Bacteria*. This commonality may indicate that the *E. coli* type of structure is the ancestral form of RNase P RNA. However, some of the variably present regions in these two types of RNA are not necessarily homologs.

The general secondary structure of the bacterial RNase P RNA is essentially determined (26). There are no remaining unpaired, phylogenetically consistent, dinucleotide complements in the conserved core of the structure model. Of course, considerable paired structure remains to be determined that does not fall into the category of contiguous, Watson-Crick base pairs. The secondary structure model engages 64% of the *E. coli* nucleotide bases in pairs, a fraction comparable to the extent of pairing in other well-defined secondary structure models of RNAs, such as the small subunit (16S) rRNA (60% of bases paired) or tRNA (55% of bases paired). At the current stage of resolution, nearly every base pair in comparable

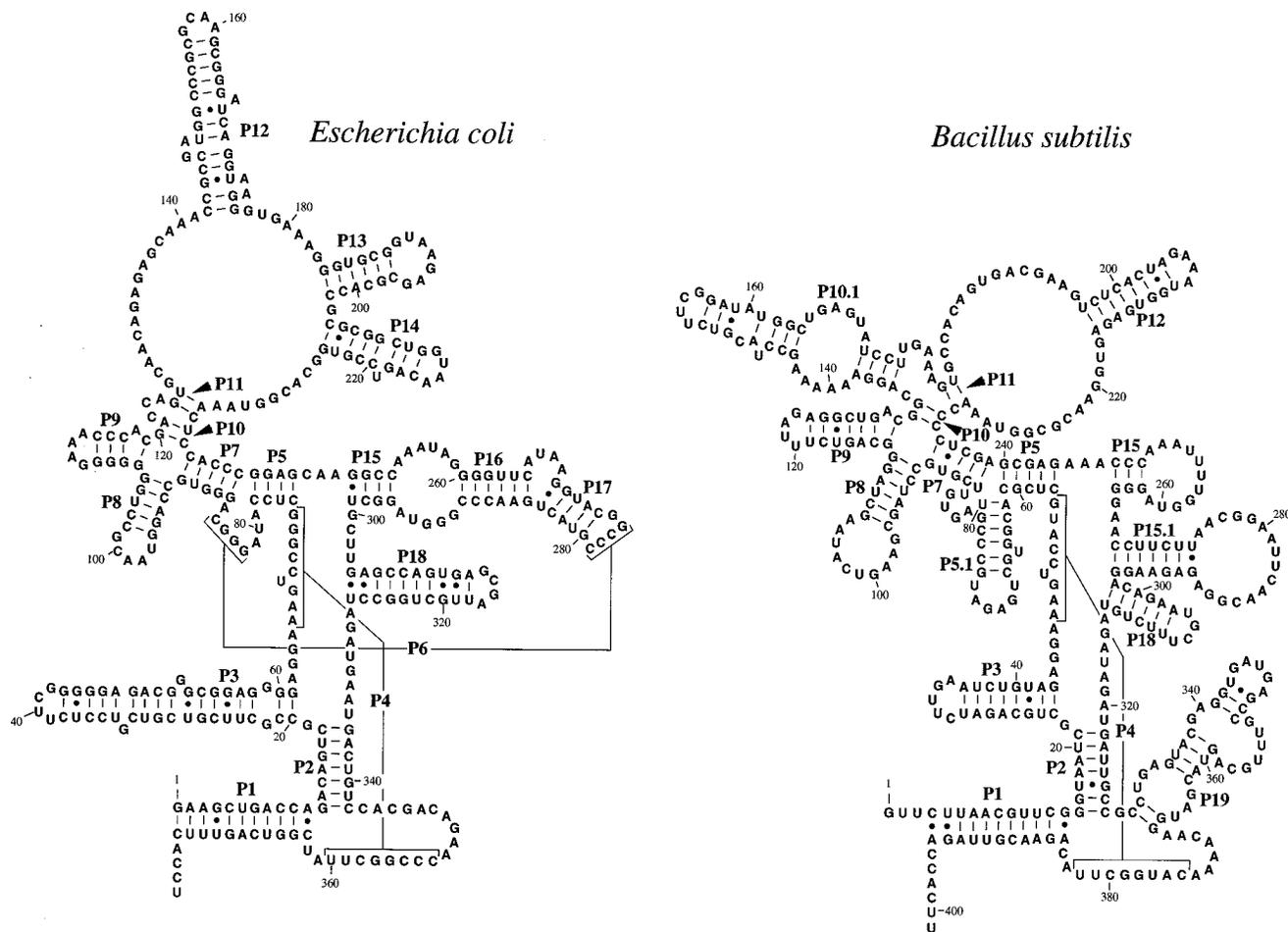


FIG. 1. Bacterial RNase P RNA secondary structures. Secondary structures of *E. coli* and *B. subtilis* RNase P RNAs were determined by phylogenetic comparative analysis as described in the text. The helix nomenclature used is that developed for group I introns (9) as described for use with RNase P RNA (26). Helices are numbered as they occur 5' to 3', preceded by the designation P (paired; e.g., P3 is the third helix encountered from the 5' end). The capping loops of particular helices are numbered as the helix with which they are associated, with the prefix L (loop; e.g., L3 is the capping loop of helix P3). Sequences that occur between helices are numbered with the helices that they connect, with the prefix J (joining; e.g., J5/15 is the sequence that links helices P5 and P15 [nucleotides 247 to 249 in the *E. coli* RNase P RNA]). Base pairings in P4 and P6 are shown as brackets and lines.

helices in different RNase P RNAs is "proven" by the encounter of two or more instances of the RNA with sequence covariations consistent with the complementarity.

The Phylogenetic Minimum Bacterial RNase P RNA

Much of the length of each native RNase P RNA consists of patches of sequence, usually helical domains, that do not occur in the RNAs of some other organism. Since all bacterial RNase P RNAs are catalytically proficient in the absence of the protein, the variably present sequences may not be required for substrate binding and catalysis. On the other hand, evolutionarily conserved sequences and structures are potentially crucial to catalytic function. It is thus useful to articulate a "phylogenetic minimum" structure of the RNA, the minimum collection of nucleotide positions and structural elements present in all known instances of the bacterial RNase P RNA. A phylogenetic minimum structure is expected to embody the most important parts of the ribozyme and thereby provide a framework for experiments directed at its essential functions.

The current phylogenetic minimum structure of the bacterial RNase P RNA, a theoretical construct based on the minimum

segment lengths in known sequences (26), is shown in Fig. 3. This consensus structure, at 225 nt, is only about half the size of the typical native RNase P RNA. A synthetic version of the minimum structure has not yet been tested for catalytic activity. However, a similar structure, a chimera of *E. coli* and *Bacillus megaterium* sequences, cleaves pre-tRNAs with nearly the catalytic efficiency (k_{cat}/K_m) of the native RNAs (68). This synthetic RNase P RNA contains helix P18, not present in the minimum structure shown in Fig. 3, but helix P18 is now known to be absent from some native RNase P RNAs and not required for catalytic activity (26).

About 20% of the nucleotides in the current phylogenetic minimum structure are conserved in identity throughout the *Bacteria*. Only a few of these conserved nucleotides occur in double-helical elements; the evolutionary constraints on such structures are mainly to preserve complementarity. The only helix that contains a substantial number of conserved nucleotides is P4, a central element in the structure model (Fig. 1). The high degree of sequence conservation in P4 indicates that it has structural constraints beyond helix formation, perhaps involvement in triple-strand or other irregular structure. Although most of the conserved nucleotides in the minimum

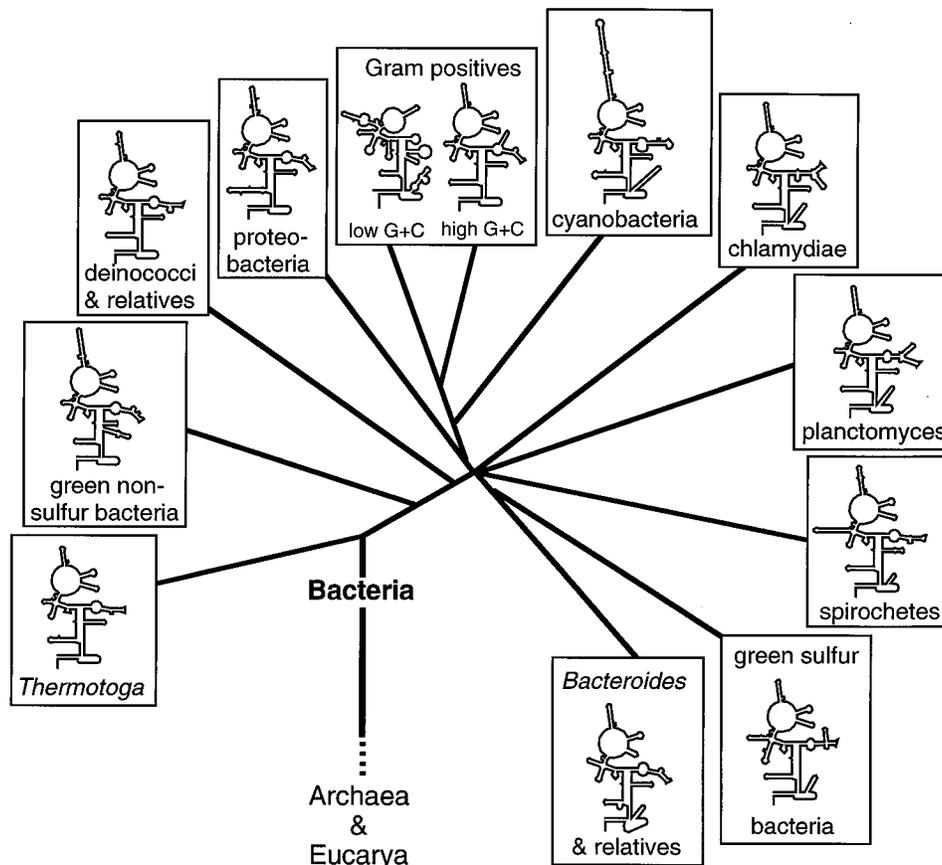


FIG. 2. Varieties of bacterial RNase P RNA. Outlines of the secondary structures of the RNase P RNAs typically found in the indicated phylogenetic groups of *Bacteria* (5) are shown associated with the corresponding phylogenetic lines of descent based on rRNA sequence comparisons (71). The specific RNase P RNAs represented are from the following organisms: *Thermotoga*, *T. maritima*; green nonsulfur bacteria, *Thermoleophilum album*; deinococci and relatives, *Thermus aquaticus*; proteobacteria, *E. coli*; low G+C gram positives, *B. subtilis*; high G+C gram positives, *Streptomyces bikiniensis*; cyanobacteria, *Anabena* PCC 7120; chlamydiae, *Chlamydia psittaci*; planctomyces, *Planctomyces maris*; spirochetes, *Leptospira weilii*; green sulfur bacteria, *Chlorobium limicola*; *Bacteroides* species and relatives, *Bacteroides thetaiotaomicron*.

structure are not associated with identified helices, it is likely that they are nevertheless solidly engaged in tertiary structures. This is indicated by their general insensitivity to chemical modifying agents such as kethoxal or dimethyl sulfate, which attack nonpaired bases (36, 39). A few conserved bases are susceptible to chemical agents in the absence of substrate but are protected in its presence, indicating that they probably interact directly with pre-tRNA.

Native RNase P RNAs differ from the phylogenetic minimum structure by their content of extra, hairpin-type helices. Although not always present and, therefore, unlikely to be essential for binding or catalysis, organism-specific helices can have structural roles. Their deletion from a native RNA typically results in an increased ionic strength requirement for optimal activity and a lower thermal stability than those exhibited by the native RNA (11, 68). These properties indicate that the deleted helices contribute to the global stability of the native RNA structure. Where there is information available, the stabilizing influence of the phylogenetic group-specific helices can be understood by their service as long-range struts, which are anchored on both ends and presumably buttress the RNA superstructure. One example of such a strut is the compound helix P16/P17. When this structural element is deleted from the *E. coli* RNase P RNA, the remaining molecule is thermally less stable than the native RNA and its affinity for

the substrate is reduced (13). This destabilization is probably due to the loss of the structural span created by pairing between the terminal loop (L17) of helix P16/P17 and nucleotides 82 to 85, which forms P6 (27). Another example of a variably present, stabilizing strut is helix P18. The deletion of this helix from the *E. coli* RNA also results in destabilization of the remaining structure and reduction in its affinity for the substrate, correctable in this case with very high ionic strength (3 M monovalent cation). A recent finding (6) that the loop of helix P18 binds to another helix, P8, accounts for the instability: removal of P18 results in the loss of a structural span between different parts of the molecule.

RNase P RNAs that lack variably present, stabilizing helices presumably compensate in some way for their absence. There is currently no information on the structural nature of such compensations. It might be expected, however, that evolutionary redesign will result in similar local three-dimensional structure, even though markedly different sequences and interactions are involved. One possible example of this is a compensation for the evolutionary loss of helix P6, one of the long-range pairings just discussed. Helix P6 occurs in most RNase P RNAs, including that of *E. coli*, but not in the RNA from the (phylogenetically defined) low G+C gram-positive bacteria, for instance *B. subtilis* (Fig. 1). The *B. subtilis* type of RNase P RNA cannot form P6 because it lacks the P16/P17

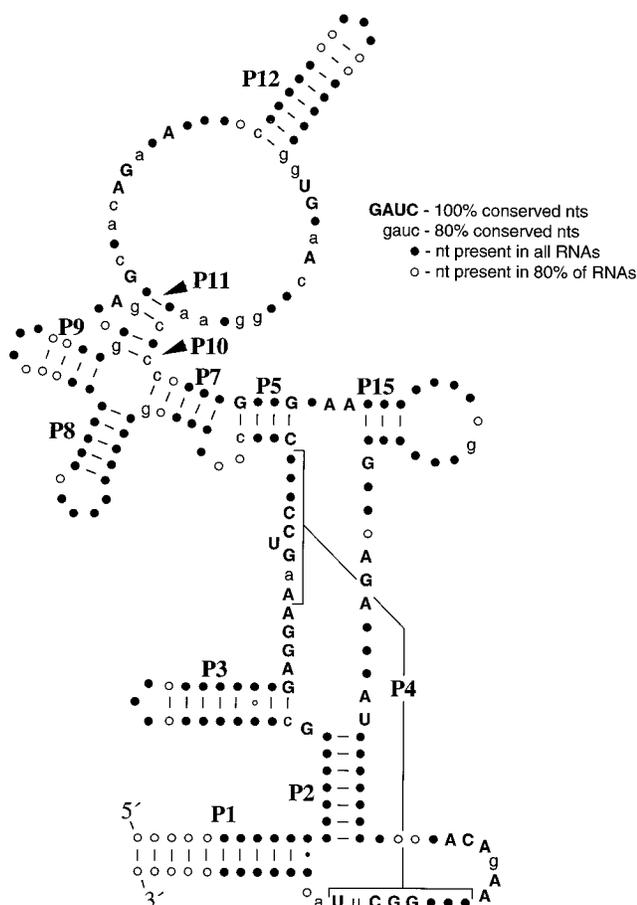


FIG. 3. Phylogenetic minimum bacterial consensus RNase P RNA. Helices are labeled as described in the legend to Fig. 1. Base pairs in P4 are shown as brackets and lines. As described in the text, only the structural elements present in all known bacterial RNase P RNAs are included in the minimum structure. Universally conserved nucleotides are in uppercase letters; those which are at least 80% conserved, but are not invariant, are in lowercase letters. Nucleotides which are not conserved in identity but are present in all sequences are indicated by filled circles; those which are present in at least 80% of sequences, but are absent in at least one, are indicated by open circles. The base pairs indicated by closed and open dots are, respectively, a conserved noncanonical (G·G or A·C) interaction and a pairing that is frequently G·G (e.g., in *E. coli*) rather than canonical.

helix, the loop (L17) of which interacts to form P6. Instead, the *B. subtilis* type of RNA contains a novel hairpin helix, P5.1, which in three dimension resembles P6 of the *E. coli* RNA, the ancestral type of the molecule. This is illustrated in Fig. 4. Analogy with the *E. coli* type of RNA would suggest that the loop of helix P5.1 docks elsewhere in the *B. subtilis* RNA to anchor the span. The resistance of bases in that loop to chemical modifying agents is consistent with this speculation (39).

Simplified, synthetic versions of the native RNase P RNAs have held promise as experimental tools for some studies. Because they are smaller in size than the native RNAs, it seemed possible that they might be useful models in studies with analytical methods for which the native RNAs are too large, for instance spectroscopy or crystallography. Our current understanding of RNase P RNA structure suggests, however, that simplified versions of the molecule inevitably will prove to be intrinsically less stable than the native RNAs, and consequently would not be ideal subjects for most physical studies. Nevertheless, the conceptual simplification of RNase P RNA

through comparative analysis provides us with a fairly clear view of the essential structural domains that make up the ribozyme. It may be possible to tease apart some of the workings of RNase P by study of particular structural domains independently from the rest of the molecule.

Association of Function with Structure

Considerable progress has been made in localizing some functions of RNase P within its structure. Cross-linking (1, 8, 24, 48, 50), footprinting (36, 39, 67), and binding experiments with mutant RNAs (28, 63) have identified RNase P nucleotides and sequence blocks that possibly interact with the substrate and the RNase P protein. These potentially interacting elements tend to be distributed throughout the conserved core of the secondary structure model but presumably are drawn together by tertiary structure. For instance, an arylazide photoaffinity agent attached to the substrate phosphate in tRNA cross-links to *E. coli* nucleotides 248 and 330 to 333, indicating that these nucleotides are in the vicinity of, and possibly constitute part of, the active site of the ribozyme (8). Although separate in the sequence and secondary structure model, these sites of cross-linking must be proximal in three dimensions. Other cross-linking (50) and chemical footprint studies (36, 39) have localized the site in RNase P RNA that binds the CCA sequence which occurs at the 3' ends of all tRNAs. It has long been recognized that interaction between RNase P RNA and the CCA of tRNA contributes significantly to their binding (44). Cross-linking to a photoaffinity agent on the 3' end of tRNA localized the binding site of the CCA to the internal loop (J15/16, probably an irregular helix) that joins P15 and P16 in the secondary structure model of the *E. coli* RNA (50). In the *B. subtilis* RNA, the 3' end of tRNA cross-links to the loop that caps helix P15, proposed in the structure model to be (partially) homologous to the cross-linked region in the *E. coli* RNA. The patterns of cross-linking of both types of RNase P RNA with tRNAs lacking sequential nucleotides of the CCA sequence orient the CCA parallel to nucleotides 254 to 259 and antiparallel to nucleotides 291 to 295 in the *E. coli* RNA (50). Perhaps the CCA sequence forms a triple helix with that substructure of the ribozyme.

RNase P nucleotides in the spatial vicinity of sites that selectively bind divalent cations also are scattered in the secondary structure model. These nucleotides are revealed by the particular sensitivity of their associated phosphodiester bond to divalent cation-catalyzed hydrolysis promoted by Mg^{2+} , Pb^{2+} , or other metal ions (10, 32, 74). The hydrolysis is due to undefined motifs in tertiary structure that bind divalent cations and activate hydrolysis of a phosphodiester bond nearby in space (7). No relationship between these sites of cleavage and the mechanism of action of RNase P has been established. The sites of cleavage do not necessarily correspond to the sequence elements responsible for binding the hydrolytic divalent cations, and in turn the cation-binding sites do not necessarily correspond to the ones used by RNase P RNA in its mechanism of action. On the practical side, however, the divalent cation-catalyzed cleavage pattern is a sensitive test for the integrity of tertiary structure and so is useful for monitoring the structural consequences of mutations (74).

Many RNase P RNA mutants have been created and analyzed in vivo and in vitro (e.g., see references 27, 34, 42, 46, and 58). The results of such studies have provided a number of insights into the structure of the RNA, bolstering the comparative studies. However, experimental alterations of bases have not yet been very informative in locating the elements of RNase P that are important in catalysis. Site-directed mutation

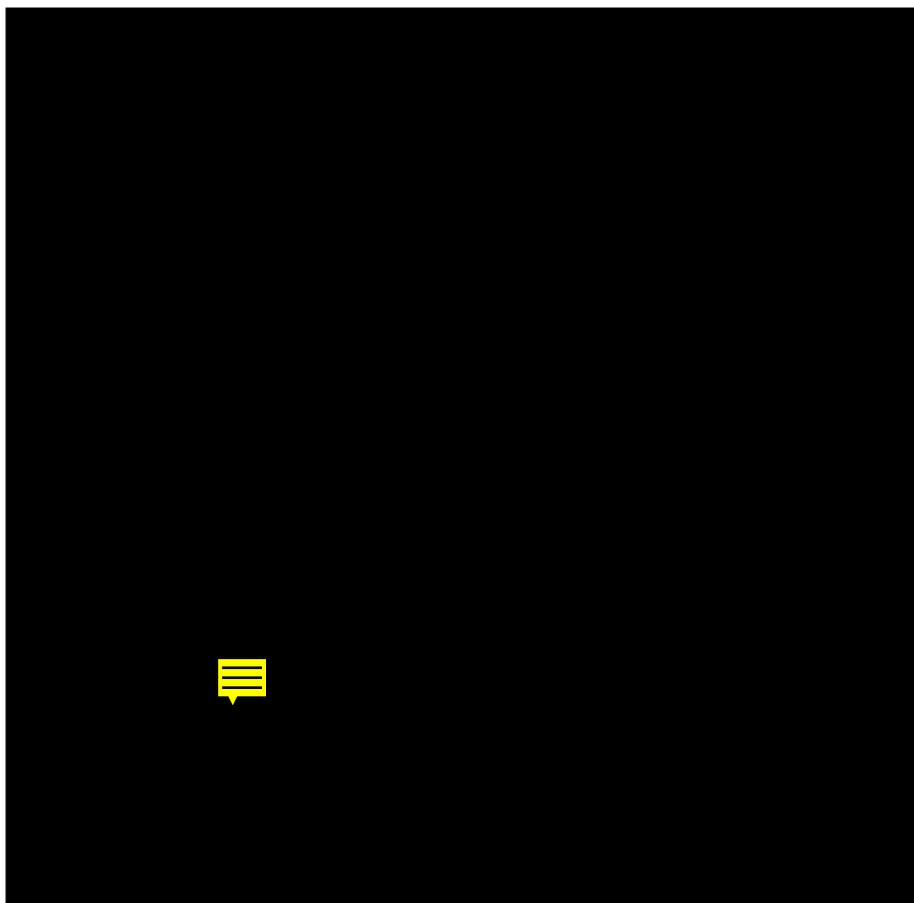


FIG. 4. Possible evolutionary substitution of structural elements in two bacterial RNase P RNAs. As discussed in the text, helix P6 and its vicinity in the *E. coli* RNase P RNA (left) resembles helix P5.1 and its vicinity in the *B. subtilis* RNA (right). (A) Helix P6 of the *E. coli* RNA is drawn as in Fig. 1 and is disjointed in order to portray P6. Yellow sequences correspond to the homologous helices P5 and P7 in both the *E. coli* and *B. subtilis* RNAs. Sequences indicated by red or blue in both RNAs are proposed to occupy analogous positions in the folded structures, as shown in the three-dimensional portrayals in panel B (top, "side" view; bottom, "top" view).

of even universally conserved bases usually results in either no discernible effect on the in vitro phenotype or only a modest effect that can be accounted for in terms of the structural stability of the ribozyme. The involvement of these nucleotides in structure is suggested by the fact that such mutants can generally be suppressed by higher ionic strength than that required for the optimal activity of the native RNA. Mutations affecting the chemistry of the reaction are expected to be independent of ionic strength. Moreover, it seems that no single feature of RNase P RNA is absolutely required for catalysis. Among a collection of mutant *B. subtilis* RNase P RNAs that had been manipulated to lack part of their sequences it was possible to delete every nucleotide in the RNA without completely abolishing catalytic activity (69). Although some of the deletion mutants were severely damaged (to yield 10^{-4} to 10^{-6} of their native activity), the general retention of some activity indicates that the elements of RNase P RNA responsible for binding and catalysis are distributed throughout the sequence and act cooperatively. It is possible that the chemistry of RNase P action does not depend on functional groups associated with specific bases. Rather, the identity of particular bases may be important only for the structure of the RNA. The important RNase P elements in the chemical steps of the reaction may be general features, 2'-OH and phosphate

groups, that bind Mg^{2+} and manipulate the substrate phosphodiester.

Toward the Three-Dimensional Structure of RNase P RNA

Further study of the function of RNase P will be facilitated by knowledge of its three-dimensional structure. In fact, much of the three-dimensional structure of RNase P RNA is implicit in the secondary structure; with reasonable confidence we can predict that paired sequences form helices with known structure at the atomic level. However, the helical barrels and their interconnecting sequences need to be ordered in space. Several laboratories have attempted the crystallization of this large RNA for diffraction analysis, but no success has been reported. In lieu of other methods for determining the detailed structure of large RNAs, molecular modeling based on available information is being used to develop a global view of the three-dimensional structure of RNase P RNA.

Molecular modeling of RNase P RNA structure has so far relied mainly on the secondary structure model of the RNA and a growing set of distance constraints imposed by photoaffinity cross-linking results (8, 29, 48, 50). A typical photoagent spans 9 to 12 Å, so cross-linked nucleotides must reside within that distance from one another. Intermolecular cross-links be-

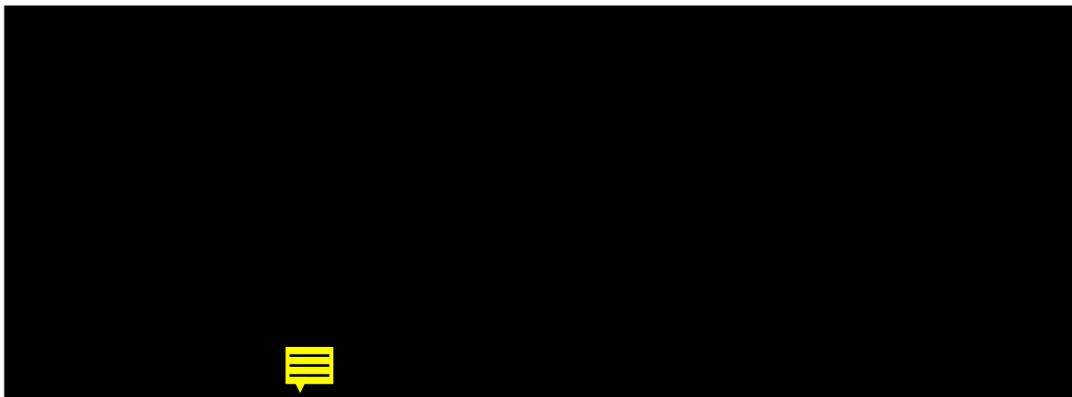


FIG. 5. Three-dimensional model of *E. coli* RNase P RNA with bound pre-tRNA substrate. As described by Harris et al. (29), the available secondary structure and cross-linking information for RNase P RNA and the known three-dimensional structure of tRNA (33) were merged into a three-dimensional model using a molecular mechanics RNA structure refinement protocol (43). Sequences including P12 to 14 and P18 (Fig. 1) were omitted from the model. Panel A shows the positions of the indicated helices. Helices which are (nearly) invariant in length are in red. Helices that vary significantly in length are green. The tRNA chain is in blue, with the substrate bond indicated in yellow. Panel B shows the path of the phosphodiester chain in the model. Nucleotide positions are indicated with numbers, corresponding to those in Fig. 1. Nucleotide bases that are invariant in identity are indicated in red. Nucleotides in regions that vary significantly in length are indicated in green.

tween tRNA and RNase P RNA provide an absolute positioning because the three-dimensional structure of tRNA is known (33). Intramolecular cross-links from one site to another in the RNase P RNA provide relative positions for the two sites. In the modeling studies, assumptions are made about the stacking orientations of adjacent helices, and the helical barrels and sequences connecting them are positioned to satisfy optimally the combination of the secondary structure features, the confines of the phosphodiester chain, available distance constraints, footprint results, and the avoidance of steric clashes. Balancing all the data is a computer-assisted process, but developing a model with the current, limited data set requires a good deal of speculation about the path of the RNA chain.

Two models for the three-dimensional structure of the *E. coli* RNase P RNA have been proposed (29, 70). One of these (29), based on a molecular mechanics RNA structure refinement protocol, is portrayed in Fig. 5. A number of phylogenetic-comparative aspects of RNase P are consistent with this model. For instance, the helices (Fig. 5A) and most of the nucleotides (Fig. 5B) that are conserved among all bacterial RNase P RNAs are clustered in the core of the structure model, in the vicinity of the substrate bond in the pre-tRNA. Moreover, regions of the molecule that vary in length among different RNase P RNAs occur on the surface of the model (Fig. 5B), as is to be expected. Such models inevitably contain inaccuracies at the present stage of development, but they provide an important context for further experimentation.

The RNase P of *Archaea* and *Eucarya*

The RNase P holoenzymes from *Archaea* and *Eucarya* differ significantly from their bacterial counterparts, and in their gross properties seem to resemble one another more closely than either resembles the bacterial version. This is consistent with the phylogenetic relationships indicated by rRNA and other sequence comparisons, i.e., that *Archaea* and *Eucarya* are specific relatives (14). The archaeal and eucaryal RNase P holoenzymes contain RNA subunits of about the same sizes as the bacterial RNAs, 300 to 400 nucleotides. However, they contain a much larger complement of protein (ca. 50 to 70%) than the bacterial holoenzyme (ca. 10%). None of the protein(s) from the archaeal or eucaryal RNase P holoenzymes

has yet been characterized. The RNA moieties of the archaeal and eucaryal holoenzymes are required for catalytic activity, as evidenced by the fact that the enzymes are inactivated by treatment with an RNase. However, the RNAs have not been found to be catalytically active following deproteinization (12, 15, 21, 38, 41, 47).

RNase P RNAs have been examined from a broad variety of *Archaea*, including representatives of all the main phylogenetic groups. The proposed secondary structures of the archaeal RNAs, as illustrated by that of *Sulfolobus acidocaldarius* in Fig. 6, show less variability throughout the group than is exhibited among the bacterial RNAs. This is consistent with the generally more slowly evolving nature of *Archaea* (71). The structural resemblance of the archaeal and bacterial RNase P RNAs is striking, clear testimony to their homology. The archaeal RNAs contain most of the consensus nucleotide identities, structures, and spacing between consensus features that occur in the bacterial phylogenetic minimum structure (above). In the light of such similarity, we find it surprising that the archaeal RNase P RNAs are not catalytically active in the absence of the protein element(s) of the holoenzyme.

Eucaryal types of RNase P holoenzyme or RNA that have been inspected include only limited phylogenetic diversity, from a few animals (e.g., see references 4 and 15) and fungi (yeasts) (e.g., see references 35, 41, and 66). However, the structural characters of the RNAs from animals and fungi differ markedly from one another and from the corresponding archaeal and bacterial RNAs. Indeed, the fungal (*Saccharomyces cerevisiae* [Fig. 6]) and animal types of sequences are so different that they cannot be convincingly aligned over most of their lengths. Despite this extreme variability, a few important, bacterial RNA-like sequence and structural elements are recognizable in the eucaryal RNAs (17, 66). This homology and the clear resemblance of the archaeal RNA to the bacterial version indicate that the RNA components of the eucaryal and archaeal RNase Ps, although independently inactive, remain intrinsically the catalytic seat of the enzymes. The extreme degeneracy of the eucaryal RNA suggests that the elements required for catalysis by RNase P, even though they are perhaps assisted by proteins, are far simpler than those embodied in the bacterial minimum structure.

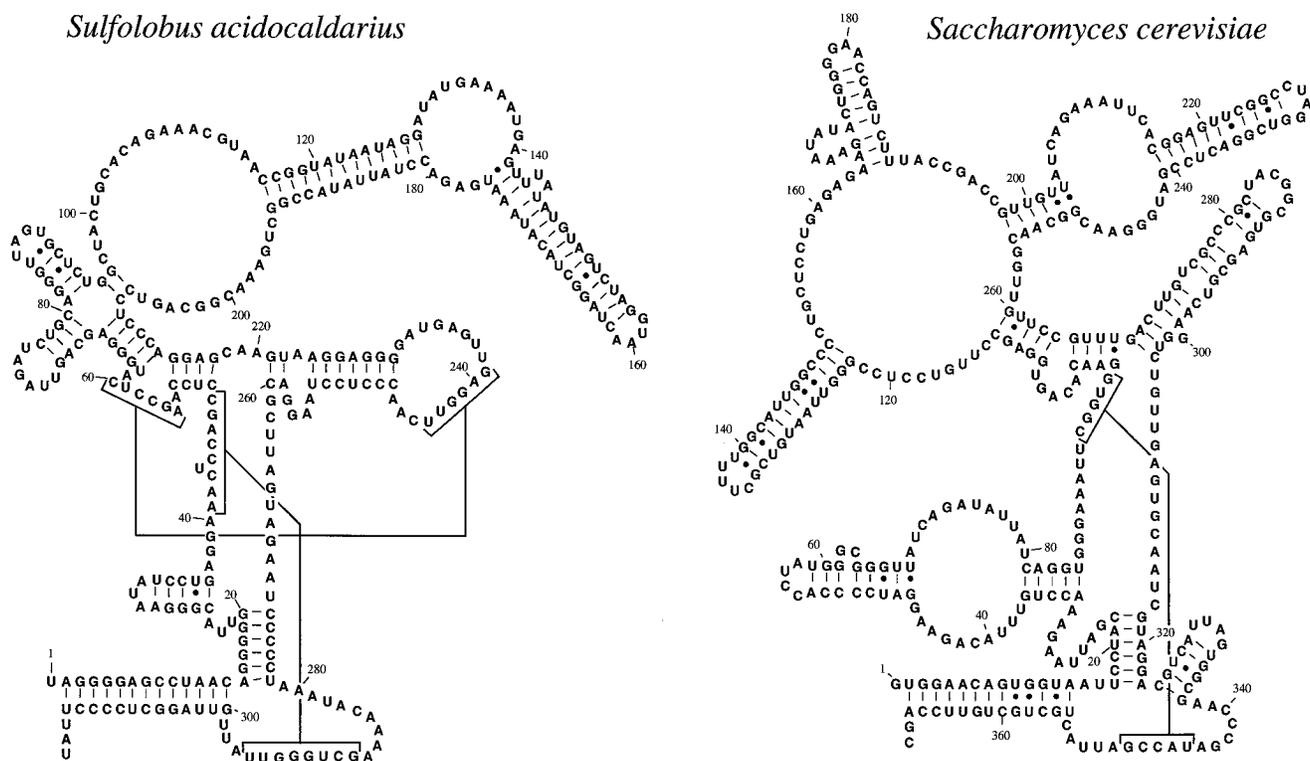


FIG. 6. Models for the secondary structures of archaeal and eucaryal RNase P RNAs. The nucleotide sequences of RNase P RNAs from the archaeon *Sulfolobus acidocaldarius* (40) and the eucaryote *Saccharomyces cerevisiae* (41), as indicated, are drawn to resemble the secondary structure of the bacterial RNase P RNA, as suggested by available comparative (40, 65) and other (66) information. Nucleotides 90 to 93 and 351 to 353 in the model of the *S. cerevisiae* RNA are not paired, which would complete an equivalent to helix P4 in the bacterial and archaeal RNAs, because of some noncanonical variations in other fungal sequences (65).

Evolutionary Perspective on RNase P RNA Function

With the possible exception of the ribosome (49), RNase P is the only known example of an RNA that *in vivo* truly acts as an enzyme in the sense that it reacts with multiple substrates. Other known catalytic RNAs, for instance self-cleaving introns or satellite RNAs, naturally perform only a single intramolecular reaction (62). Why is the catalytic element of RNase P composed of RNA? Is it a chance remnant of some precellular "RNA world," or is it a consequence of some property of RNA that makes it particularly suitable for the reaction that it carries out?

RNase P RNA is certainly an ancient molecule. This is indicated by the homology of the RNA in all organisms, with no evidence for lateral transfer in evolution. The numerous common sequences and structures among the most similar bacterial and archaeal RNAs must also have been present in the ancestor of all life. The ancestral RNase P RNA, like the modern bacterial version, probably would have been catalytically active without any protein moiety, since the "root" of the tree of all life seems to have been in the bacterial line of descent (for a review, see reference 14). The venerable nature of RNase P RNA is consistent with, but not evidence for, an RNA world. Regardless of its origin, the uniqueness of RNase P as a ribozyme suggests that some property of RNA makes it advantageous in the RNase P reaction.

Conjecture as to why RNA might be more suitable than protein in the RNase P reaction is probably best directed at its electrostatic properties, which strongly govern all interactions involving RNA and clearly distinguish RNA from protein. One notion (51) is that the highly charged character of RNA sur-

faces could facilitate the formation of a catalytic pocket with high ionic or water activity. Another idea is that the polyanionic character of RNA could promote the condensation of the multiple Mg^{2+} ions required for the RNase P mechanism into a small, catalytic pocket. These roles would not seem unique to RNA, however, so there remains an interesting question: why is RNase P composed of RNA? The answer may lend insight into the special qualities of RNA that have made that type of molecule central to the origin and evolution of life.

ACKNOWLEDGMENTS

We thank our colleagues for many useful suggestions and delightful discussions. Particular thanks go to Michael Harris, for providing Fig. 5. We beg the indulgence of all our colleagues who have contributed to our knowledge of RNase P but were not cited here due to lack of space.

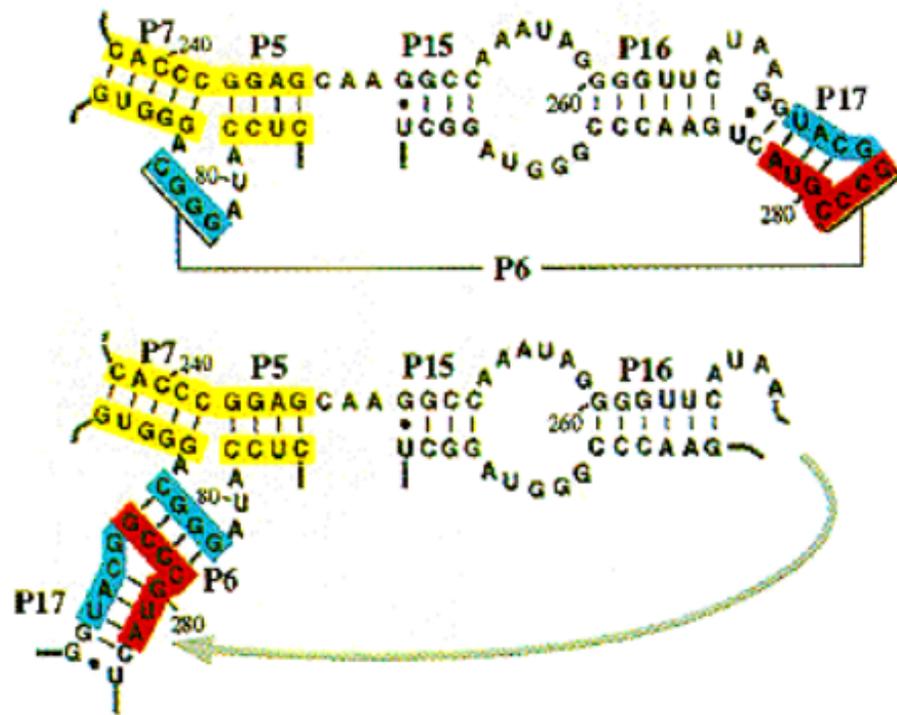
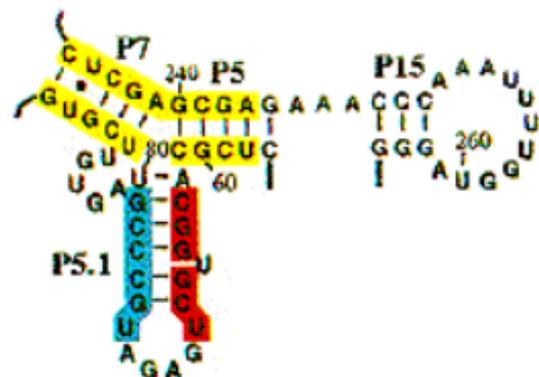
Research in the laboratory of N.R.P. is supported by grants from the NIH (GM34527) and DOE (92ER20088).

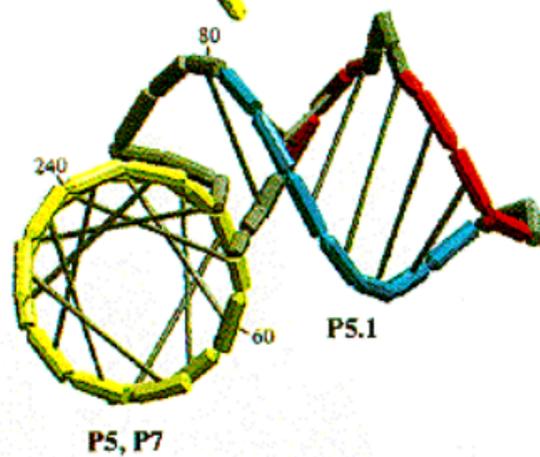
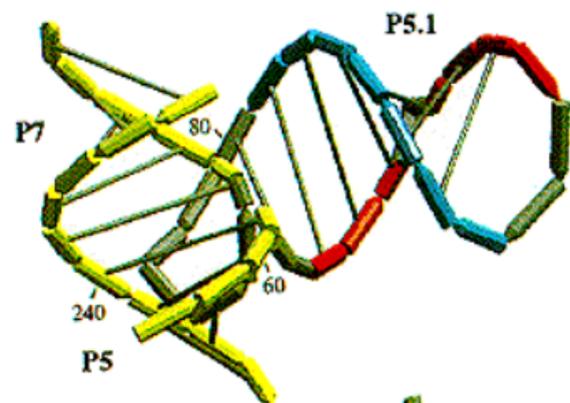
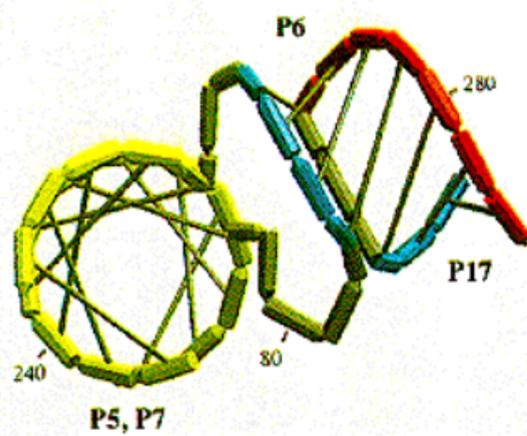
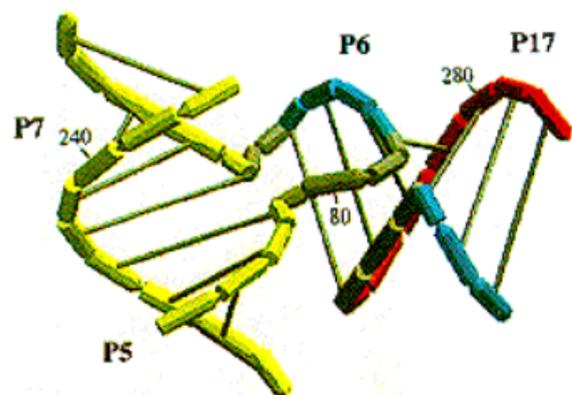
REFERENCES

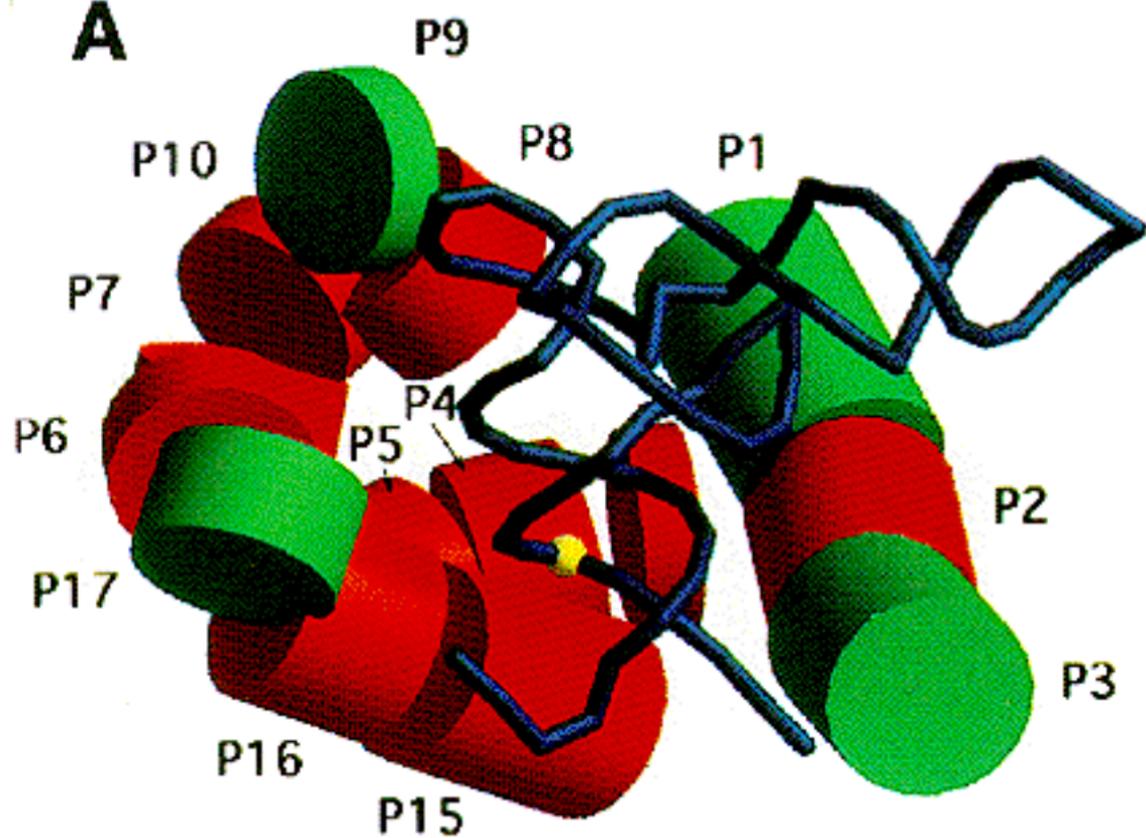
- Altman, S. 1984. Aspects of biochemical catalysis. *Cell* **36**:237-239.
- Altman, S. 1993. RNA enzyme-directed gene therapy. *Proc. Natl. Acad. Sci. USA* **90**:10898-10900.
- Altman, S., L. Kirsebom, and S. Talbot. 1993. Recent studies of ribonuclease P. *FASEB J.* **7**:7-14.
- Altman, S., D. Wesolowski, and R. S. Puranam. 1993. Nucleotide sequences of the RNA subunit of RNase P from several mammals. *Genomics* **18**:418-422.
- Brown, J. W., E. S. Haas, D. G. Gilbert, and N. R. Pace. 1994. The ribonuclease P database. *Nucleic Acids Res.*, in press.
- Brown, J. W., J. M. Nolan, E. S. Haas, M. A. Rubio, F. Major and N. R. Pace. 1994. Novel tertiary structure in ribonuclease P RNA detected by covariation in natural microbial populations. *Proc. Natl. Acad. Sci. USA*, in press.
- Brown, R. S., J. C. Dewan, and A. Klug. 1985. Crystallographic and biochem-

- ical investigation of the lead (II)-catalyzed hydrolysis of yeast phenylalanine tRNA. *Biochemistry* **24**:4785–4801.
8. **Burgin, A. B., and N. R. Pace.** 1990. Mapping the active site of ribonuclease P RNA using a substrate containing a photoaffinity agent. *EMBO J.* **9**:4111–4118.
 9. **Burke, J. M., M. Belfort, T. R. Cech, R. W. Davies, R. J. Schwyen, D. A. Shub, J. W. Szostak, and H. F. Tabak.** 1987. Structural conventions for group I introns. *Nucleic Acids Res.* **15**:7217–7221.
 10. **Ciesiolka, J., W. D. Hardt, J. Schlegl, V. A. Erdmann, and R. K. Hartmann.** 1994. Lead-ion-induced cleavage of RNase P RNA. *Eur. J. Biochem.* **219**:49–56.
 11. **Darr, S. C., J. W. Brown, and N. R. Pace.** 1992. The varieties of ribonuclease P. *Trends Biochem. Sci.* **17**:178–82.
 12. **Darr, S. C., B. Pace, and N. R. Pace.** 1990. Characterization of ribonuclease P from the archaeobacterium *Sulfolobus solfataricus*. *J. Biol. Chem.* **265**:12927–12932.
 13. **Darr, S. C., K. Zito, D. Smith, and N. R. Pace.** 1992. Contributions of phylogenetically variable structural elements to the function of the ribozyme ribonuclease P. *Biochemistry* **31**:328–333.
 14. **Doolittle, W. F., and J. R. Brown.** 1994. Tempo, mode, the progenote, and the universal root. *Proc. Natl. Acad. Sci. USA* **91**:6721–6728.
 15. **Doria, M., G. Carrara, P. Calandra, and G. P. Tocchini-Valentini.** 1991. An RNA molecule copurifies with RNase P activity from *Xenopus laevis* oocytes. *Nucleic Acids Res.* **19**:2315–2320.
 16. **Forster, A. C., and S. Altman.** 1990. External guide sequences for an RNA enzyme. *Science* **249**:783–786.
 17. **Forster, A. C., and S. Altman.** 1990. Similar cage-shaped structures for the RNA components of all ribonuclease P and ribonuclease MRP enzymes. *Cell* **62**:407–409.
 18. **Frank, D. N., M. E. Harris, and N. R. Pace.** 1994. Remodeling active-site structure in a ribozyme: rational design of self-cleaving pre-tRNA-ribonuclease P conjugates. *Biochemistry* **33**:10800–10808.
 19. **Gardiner, K. J., T. L. Marsh, and N. R. Pace.** 1985. Ion dependence of the *Bacillus subtilis* RNase P reaction. *J. Biol. Chem.* **260**:5415–5419.
 20. **Gaur, R. K., and G. Krupp.** 1993. Modification interference approach to detect ribose moieties important for the optimal activity of a ribozyme. *Nucleic Acids Res.* **21**:21–26.
 21. **Gold, H. A., and S. Altman.** 1986. Reconstitution of RNAase P activity using inactive subunits from *E. coli* and HeLa cells. *Cell* **44**:243–249.
 22. **Guerrier-Takada, C., K. Gardiner, T. Marsh, N. Pace, and S. Altman.** 1983. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**:849–857.
 23. **Guerrier-Takada, C., K. Haydock, L. Allen, and S. Altman.** 1986. Metal ion requirements and other aspects of the reaction catalyzed by M1 RNA, the RNA subunit of ribonuclease P from *Escherichia coli*. *Biochemistry* **25**:1509–1515.
 24. **Guerrier-Takada, C., N. Lumelsky, and S. Altman.** 1989. Specific interactions in RNA enzyme-substrate complexes. *Science* **246**:1578–1584.
 25. **Gutell, R. R., N. Larsen, and C. R. Woese.** 1994. Lessons from an evolving rRNA - 16S and 23S rRNA structures from a comparative perspective. *Microbiol. Rev.* **58**:10–26.
 26. **Haas, E. S., J. W. Brown, C. Pitulle, and N. R. Pace.** 1994. Further perspective on the catalytic core and secondary structure of ribonuclease P RNA. *Proc. Natl. Acad. Sci. USA* **91**:2527–2531.
 27. **Haas, E. S., D. P. Morse, J. W. Brown, F. J. Schmidt, and N. R. Pace.** 1991. Long-range structure in ribonuclease P RNA. *Science* **254**:853–856.
 28. **Hardt, W. D., J. Schlegl, V. A. Erdmann, and R. K. Hartmann.** 1993. Role of the D arm and the anticodon arm in tRNA recognition by eubacterial and eukaryotic RNase P enzymes. *Biochemistry* **32**:13046–13053.
 29. **Harris, M. E., J. M. Nolan, A. Malhotra, J. W. Brown, S. C. Harvey, and N. R. Pace.** 1994. Use of photoaffinity crosslinking and molecular modeling to analyze the global architecture of ribonuclease P RNA. *EMBO J.* **13**:3953–3963.
 30. **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.
 31. **Kahle, D., U. Wehmeyer, S. Char, and G. Krupp.** 1990. The methylation of one specific guanosine in a pre-tRNA prevents cleavage by RNase P and by the catalytic M1 RNA. *Nucleic Acids Res.* **18**:837–844.
 32. **Kazakov, S., and S. Altman.** 1991. Site-specific cleavage by metal ion cofactors and inhibitors of M1 RNA, the catalytic subunit of RNase P from *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **88**:9193–9197.
 33. **Kim, S. H., G. J. Quigley, F. L. Suddath, A. McPherson, D. Sneden, J. J. Kim, J. Weinzierl, and A. Rich.** 1973. Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain. *Science* **179**:285–288.
 34. **Kirsebom, L. A., and S. G. Svard.** 1993. Identification of a region within M1 RNA of *Escherichia coli* RNase P important for the location of the cleavage site on a wild-type tRNA precursor. *J. Mol. Biol.* **231**:594–604.
 35. **Kline, L., S. Nishikawa, and D. Söll.** 1981. Partial purification of RNase P from *Schizosaccharomyces pombe*. *J. Biol. Chem.* **256**:5058–5063.
 36. **Knap, A. K., D. Wesolowski, and S. Altman.** 1990. Protection from chemical modification of nucleotides in complexes of M1 RNA, the catalytic subunit of RNase P from *E. coli*, and tRNA precursors. *Biochimie* **72**:779–790.
 37. **Kole, R., M. F. Baer, B. C. Stark, and S. Altman.** 1980. *E. coli* RNase P has a required RNA component in vivo. *Cell* **19**:881–887.
 38. **Krupp, G., B. Cherayil, D. Frendewey, S. Nishikawa, and D. Söll.** 1986. Two RNA species co-purify with RNase P from the fission yeast *Schizosaccharomyces pombe*. *EMBO J.* **5**:1697–1703.
 39. **LaGrandeur, T. E., A. Hüttenhofer, H. F. Noller, and N. R. Pace.** 1994. Phylogenetic comparative footprint analysis of the interaction between ribonuclease P RNA and tRNA. *EMBO J.* **13**:3945–3952.
 40. **LaGrandeur, T. E., S. C. Darr, E. S. Haas, and N. R. Pace.** 1993. Characterization of the RNase P RNA of *Sulfolobus acidocaldarius*. *J. Bacteriol.* **175**:5043–5048.
 41. **Lee, J. Y., and D. R. Engelke.** 1989. Partial characterization of an RNA component that copurifies with *Saccharomyces cerevisiae* RNase P. *Mol. Cell. Biol.* **9**:2536–2543.
 42. **Lumelsky, N., and S. Altman.** 1988. Selection and characterization of randomly produced mutants in the gene coding for M1 RNA. *J. Mol. Biol.* **202**:443–454.
 43. **Malhotra, A., R. Tan, and S. C. Harvey.** 1994. Modeling large RNAs and ribonucleoprotein particles using molecular mechanics techniques. *Biophys. J.* **66**:1777–1795.
 44. **McClain, W. H., C. Guerrier-Takada, and S. Altman.** 1987. Model substrates for an RNA enzyme. *Science* **238**:527–530.
 45. **Milligan, J. F., and O. C. Uhlenbeck.** 1989. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.* **164**:51–62.
 46. **Morse, D. P., and F. J. Schmidt.** 1993. Suppression of loss-of-function mutations in *Escherichia coli* ribonuclease P RNA (M1 RNA) by a specific base-pair disruption. *J. Mol. Biol.* **230**:11–14.
 47. **Nieuwlandt, D. T., E. S. Haas, and C. J. Daniels.** 1991. The RNA component of RNase P from the archaeobacterium *Haloferax volcanii*. *J. Biol. Chem.* **266**:5689–5695.
 48. **Nolan, J. M., D. H. Burke, and N. R. Pace.** 1993. Circularly permuted tRNAs as specific photoaffinity probes of ribonuclease P RNA structure. *Science* **261**:762–765.
 49. **Noller, H. F., W. Hoffarth, and L. Zimiak.** 1992. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* **256**:1416–1419.
 50. **Oh, B.-K., and N. R. Pace.** 1994. Interaction of the 3'-end of tRNA with ribonuclease P RNA. *Nucleic Acids Res.*, in press.
 51. **Pace, N. R., and D. Smith.** 1990. Ribonuclease P: function and variation. *J. Biol. Chem.* **265**:3587–3590.
 52. **Peck-Miller, K. A., and S. Altman.** 1991. Kinetics of the processing of the precursor to 4.5 S RNA, a naturally occurring substrate for RNase P from *Escherichia coli*. *J. Mol. Biol.* **221**:1–5.
 53. **Perreault, J. P., and S. Altman.** 1992. Important 2'-hydroxyl groups in model substrates for M1 RNA, the catalytic RNA subunit of RNase-P from *Escherichia coli*. *J. Mol. Biol.* **226**:399–409.
 54. **Perreault, J. P., and S. Altman.** 1993. Pathway of activation by magnesium ions of substrates for the catalytic subunit of RNase P from *Escherichia coli*. *J. Mol. Biol.* **230**:750–756.
 55. **Reed, R. E., M. F. Baer, C. Guerrier-Takada, H. Donis-Keller, and S. Altman.** 1982. Nucleotide sequence of the gene encoding the RNA subunit (M1 RNA) of ribonuclease P from *Escherichia coli*. *Cell* **30**:627–636.
 56. **Reich, C., K. J. Gardiner, G. J. Olsen, B. Pace, T. L. Marsh, and N. R. Pace.** 1986. The RNA component of the *Bacillus subtilis* RNase P: sequence, activity, and partial secondary structure. *J. Biol. Chem.* **261**:7888–7893.
 57. **Reich, C., G. J. Olsen, B. Pace, and N. R. Pace.** 1988. Role of the protein moiety of ribonuclease P, a ribonucleoprotein enzyme. *Science* **239**:178–181.
 58. **Shiraishi, H., and Y. Shimura.** 1988. A rapid and efficient method for targeted random mutagenesis. *Gene* **64**:313–319.
 59. **Smith, D., and N. R. Pace.** 1993. Multiple magnesium ions in the ribonuclease P reaction mechanism. *Biochemistry* **32**:5273–5281.
 60. **Steitz, T. A., and J. A. Steitz.** 1993. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. USA* **90**:6498–6502.
 61. **Surratt, C. K., B. J. Carter, R. C. Payne, and S. M. Hecht.** 1990. Metal ion and substrate structure dependence of the processing of tRNA precursors by RNase P and M1 RNA. *J. Biol. Chem.* **265**:22513–22519.
 62. **Symons, R. H.** 1994. Ribozymes. *Curr. Opin. Struct. Biol.* **4**:322–330.
 63. **Talbot, S. J., and S. Altman.** 1994. Gel retardation analysis of the interaction between C5 protein and M1 RNA in the formation of the ribonuclease P holoenzyme from *Escherichia coli*. *Biochemistry* **33**:1399–1405.
 64. **Thurlow, D. L., D. Shilowski, and T. L. Marsh.** 1991. Nucleotides in precursor tRNAs that are required intact for catalysis by RNase P RNAs. *Nucleic Acids Res.* **19**:885–891.
 65. **Tranguch, A. J., and D. R. Engelke.** 1993. Comparative structural analysis of nuclear RNase P RNAs from yeast. *J. Biol. Chem.* **268**:14045–14055.
 66. **Tranguch, A. J., D. W. Kindelberger, C. E. Rohlfman, J. Y. Lee, and D. R. Engelke.** 1994. Structure-sensitive RNA footprinting of yeast nuclear ribonuclease P. *Biochemistry* **33**:1778–1787.
 67. **Vioque, A., J. Arnez, and S. Altman.** 1988. Protein-RNA interactions in the RNase P holoenzyme from *Escherichia coli*. *J. Mol. Biol.* **202**:835–848.

68. **Waugh, D. S., C. J. Green, and N. R. Pace.** 1989. The design and catalytic properties of a simplified ribonuclease P RNA. *Science* **244**:1569–1570.
69. **Waugh, D. S., and N. R. Pace.** 1993. Gap-scan deletion analysis of *Bacillus subtilis* RNase P RNA. *FASEB J.* **7**:188–195.
70. **Westhof, E., and S. Altman.** 1994. Three-dimensional working model of M1 RNA, the catalytic RNA subunit of ribonuclease P from *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **91**:5133–5137.
71. **Woese, C. R.** 1987. Bacterial evolution. *Microbiol. Rev.* **51**:221–271.
72. **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.
73. **Wyatt, J. R., and I. Tinoco.** 1993. RNA structural elements and RNA function, p. 465–496. *In* R. F. Gesteland and J. F. Atkins (ed.), *The RNA world*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
74. **Zito, K., A. Hüttenhofer, and N. R. Pace.** 1993. Lead-catalyzed cleavage of ribonuclease P RNA as a probe for integrity of tertiary structure. *Nucleic Acids Res.* **21**:5916–5920.

A*Escherichia coli**Bacillus subtilis*

B

A

B