Phylogenetic Analysis and Evolution of RNase P RNA in Proteobacteria

JAMES W. BROWN, ELIZABETH S. HAAS, BRYAN D. JAMES,† DIRK A HUNT,
AND NORMAN R. PACE*

*Department of Biology and Institute for Molecular and Cellular Biology, Indiana University,
Bloomington, Indiana 47405*

The secondary structures of the eubacterial RNase P RNAs are being elucidated by a phylogenetic comparative approach. Sequences of genes encoding RNase P RNA from each of the recognized subgroups (α, β, γ, and δ) of the proteobacteria have now been determined. These sequences allow the refinement, to nearly the base pair level, of the phylogenetic model for RNase P RNA secondary structure. Evolutionary change among the RNase P RNAs was found to occur primarily in four discrete structural domains that are peripheral to a highly conserved core structure. The new sequences were used to examine critically the proposed similarity (C. Guerrier-Takada, N. Lumelsky, and S. Altman, Science 246:1578–1584, 1989) between a portion of RNase P RNA and the "exit site" of the 23S rRNA of *Escherichia coli*. Phylogenetic comparisons indicate that these sequences are not homologous and that any similarity in the structures is, at best, tenuous.

RNase P cleaves leader sequences from precursor tRNAs to generate the mature 5' end of tRNA (see references 1 and 19 for reviews). In the eubacteria *Bacillus subtilis* and *Escherichia coli*, RNase P is composed of a small protein (119 amino acids) and a large RNA (~400 nucleotides). Although RNase P functions as a ribonucleoprotein in vivo, the RNA alone is an efficient and accurate catalyst at elevated concentrations of salt in vitro. RNase P contains the only RNA known to act catalytically in vivo, in the sense that each molecule of RNase P RNA acts on many substrate molecules.

An accurate model for the structure of eubacterial RNase P RNA is a prerequisite for understanding the mechanism of substrate recognition, catalytic activity, and evolution of this ribozyme. A model for the secondary structure of RNase P RNA in eubacteria was originally derived by phylogenetic comparative analysis of the sequences of genes encoding RNase P RNA from several members of the γ subdivision of the proteobacteria and four species of gram-positive eubacteria of the genus *Bacillus* (12). On the basis of the criteria used for analysis of ribosomal RNA structure (8), covariation of two base pairings within a potential continuous helix was considered proof of the occurrence of that helix. In order to increase the resolution of this model, we have cloned and sequenced the genes for RNase P RNA from a phylogenetically diverse sampling of the following additional proteobacteria: the α proteobacteria *Agrobacterium tumefaciens* and *Rhodospirillum rubrum*, the β proteobacterium *Alcaligenes eutrophus*, the γ proteobacterium *Chromatium vinosum*, and the δ proteobacterium *Desulfovibrio desulfuricans* (Fig. 1). Together with the previously obtained gene sequences for RNase P RNA, these analyses complete a sampling of each of the major phylogenetic branches of proteobacteria.

MATERIALS AND METHODS

Cultures and growth conditions. *A. tumefaciens* DSM 30150, *A. eutrophus* DSM 531, and *C. vinosum* ATCC e17899 were grown at 37°C in 2× YT liquid media (24). *D. desulfuricans* ATCC 27771 was grown anaerobically at 37°C in medium 766 (5). *C. vinosum* 1.1.1, grown anaerobically in the light, was a gift from J. Favinger and C. Bauer (Indiana University).

Nucleic acid extraction. Extracts for the isolation of genomic DNAs were prepared by resuspending the cells in STE (10 mM Tris-Cl [pH 9], 100 mM NaCl, 1 mM EDTA) (15 ml/g of cells [wet weight]), treating them with lysozyme (1 mg/ml) for 5 min at 37°C, adding 2-aminosalicylate and sodium dodecyl sulfate (SDS) to 4.5 and 0.75%, respectively, and incubating the mixture at 70°C for 10 min. Following repeated extraction with STE-saturated phenol, phenol-chloroform, and chloroform, DNA was precipitated by the addition of sodium acetate to 0.3 M and an equal volume of 2-propanol. DNA was spooled out and allowed to dissolve overnight in STE (pH 8) containing 1% SDS and 50 μg of proteinase K per ml. The organic extractions and precipitations were repeated twice more, and the DNA was dissolved in 10 mM Tris (pH 8)–1 mM EDTA at a concentration of ~1 mg/ml.

Total RNA was isolated from cells lysed in a French press by organic extraction and precipitation as previously described (15).

Purification of RNase P RNA from *A. eutrophus*. Total RNA (10 mg) from *A. eutrophus* was fractionated on a Sepharose 4B-200 (Pharmacia, Piscataway, N.J.) column (1.4 by 65 cm) in 100 mM NaCl–10 mM Tris (pH 8)–1 mM EDTA–0.1% SDS. Fractions containing RNase P activity (see below) were pooled, ethanol precipitated, and further fractionated by electrophoresis in a 4% polyacrylamide–8 M urea gel. RNA bands visualized by UV shadow (18) were excised from the gel, eluted overnight at 4°C in 0.5 M ammonium acetate–1 mM EDTA–0.1% SDS and then ethanol precipitated. RNase P RNA was identified in RNase P activity assays and stored at −20°C. Fractions from the Sepharose 4B-200 column, which contained 16S and 23S rRNA but not RNase P activity, were pooled for use in

* Corresponding author.
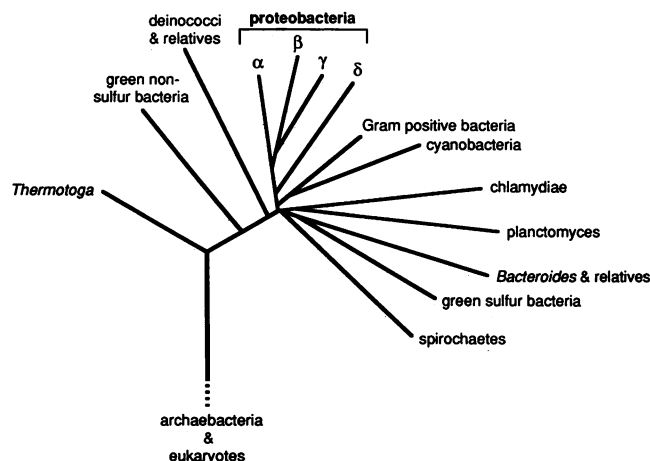† Present address: Abbott Laboratories, Abbott Park, IL 60064.

FIG. 1. Phylogenetic tree of the eubacteria. Evolutionary relationships of the eubacteria, which are based on the 16S rRNA sequence (assembled from reference 31). Each of the 11 recognized "phyla" of eubacteria is shown. Separate branches are shown for the α, β, γ, and δ groups of proteobacteria. Evolutionary divergence is represented by total line length.

genomic Southern hybridization experiments and for sequence confirmation of the *A. eutrophus* strain.

Genomic Southern blots. Genomic DNAs were digested with restriction endonucleases in buffer containing 33 mM Tris-acetate (pH 7.9), 6.6 mM potassium acetate, 1 mM magnesium acetate, 100 μg of bovine serum albumin per ml, and 5 mM dithiothreitol. DNA fragments were separated by electrophoresis in 1% agarose gels, blotted to Hybond N sheets (Amersham, Arlington Heights, Ill.), and fixed to the sheets by UV irradiation (13). Prehybridization (65°C for >2 h) and hybridization (65°C overnight) were performed in 0.01% polyvinylpyrrolidone–0.01% bovine serum albumin–0.01% Ficoll–20 mM Na-PO$_4$ (pH 7.6)–1 mM dithiothreitol–5× SSC (1× SSC is 150 mM NaCl–15 mM sodium citrate [pH 7])–100 μg of poly(A) per ml–0.1% SDS and 1 to 5 μCi of probe. The sheets were washed twice at room temperature for 5 min in 2× SSC–0.1% SDS and once for 20 min at 65°C in 1× SSC–0.1% SDS before autoradiography.

Probes for *R. rubrum*, *A. tumefaciens*, *C. vinosum*, and *D. desulfuricans* genomic Southern blots were partially hydrolyzed, uniformly ^{32}P-labeled, antisense transcripts of the *E. coli* RNase P RNA gene (12). The labeled RNAs were generated by T7 RNA polymerase runoff transcription and partially hydrolyzed by incubation at 95°C in 50 mM sodium carbonate (pH 9). Aliquots removed at 5, 10, 15, and 20 min were pooled, and unincorporated nucleotides were removed by Sephadex G-50 spin-column gel filtration (24). The probe used for *A. eutrophus* genomic Southern blots was native *A. eutrophus* RNase P RNA; this RNA was partially hydrolyzed in the same way, and then the fragments were 5′ end labeled by using T4 polynucleotide kinase and [γ-^{32}P]ATP (16). Prehybridizations and hybridizations contained 25 μg of *A. eutrophus* 16S and 23S rRNA per ml whenever the *A. eutrophus* native probe was used.

Cloning of RNase P RNA genes. Restriction endonuclease digests of genomic DNAs were separated by agarose gel electrophoresis, and fractions of gels containing RNase P RNA genes were excised on the basis of previously obtained Southern blot information. The gel slices were minced, and DNA was recovered either by passive elution in 10 mM

Tris–1 mM EDTA (pH 8) overnight at 4°C or by disruption of the gel with NaI and purification of the DNA with glass powder (28). The size-selected DNAs were then cloned into transcription vectors and screened by colony hybridization (24) with the same probe as for the Southern analyses. Confirmation of potential clones was by Northern (RNA) blot analysis, RNase P assays of T7 and T3 RNA polymerase-generated transcripts from the clones, and sequence analysis. The RNase P RNA genes from *R. rubrum* (the gene was contained on a 2.4-kb *Xho*I DNA fragment), *A. eutrophus* (the gene was contained on a 2.3-kb *Pst*I DNA fragment), *D. desulfuricans* (the gene was contained on a 3.0-kb *Pst*I DNA fragment), and *C. vinosum* (the gene was contained on a 2.8-kb *Sal*I-*Pst*I DNA fragment) were cloned into pBluescript KS$^+$ (26). The RNase P RNA gene from *A. tumefaciens*, contained on a 2.3-kb *Hind*III-*Pst*I DNA fragment, was cloned into pTZ18U (17).

Northern blot analysis. Total cellular RNAs were separated by electrophoresis through 5% polyacrylamide–8 M urea gels, stained with ethidium bromide, and transferred to Hybond N sheets by electroelution in 25 mM Na-PO$_4$ (pH 6.8). RNA was fixed to the sheets by UV irradiation. Hybridization was carried out as described for Southern blots. Probes were generated by nick translation (24) or in vitro synthesis of antisense RNAs from clones containing the RNase P RNAs of interest.

RNase P enzyme assays. T7 and T3 RNA polymerase runoff transcripts from potential RNase P RNA gene-containing clones were synthesized in 20-μl reaction mixtures containing 20 mM Na-PO$_4$ (pH 7.7); 10 mM dithiothreitol; 8 mM MgCl$_2$; 5 mM spermidine; 0.5 mM each ATP, GTP, CTP, and UTP; ~1 μg of template DNA; and 50 U of either T7 or T3 RNA polymerase (23). The clones, which contain large inserts (up to 3.0 kb), were transcribed by using *Pvu*II-fragmented or covalently closed circular DNAs. After incubation at 37°C for 3 h, samples (up to 5 μl) from transcription reactions were assayed for RNase P RNA enzyme activity in 10-μl reactions containing 600 mM NH$_4$Cl, 30 mM MgCl$_2$, 50 mM Tris-Cl (pH 8), 0.1% (wt/vol) SDS, 0.05% (vol/vol) Nonidet P-40, and 25 to 100 nM substrate precursor tRNA^{Asp}. After incubation for 1 h at 37°C, the samples were dried and separated by electrophoresis through 6% acrylamide–8 M urea gels (24), and the products were detected by autoradiography.

Nucleic acid sequencing. Single- and double-stranded plasmid DNAs were sequenced by the dideoxy chain termination method (25) with Sequenase version 2.0 (US Biochemicals, Cleveland, Ohio). RNase P RNA genes from *R. rubrum* and *D. desulfuricans* were sequenced after a series of nested deletions was generated by using exonuclease III and nuclease S1 (10); all other RNase P RNA genes were sequenced from restriction fragment subclones. Deaza-dGTP and dITP were used as required to alleviate band compression in sequencing gels. Sequence from the 16S rRNA of *A. eutrophus* was obtained from high-molecular-weight RNA by using reverse transcriptase and oligonucleotides complementary to universally conserved 16S rRNA sequences (15).

Nucleotide sequence accession numbers. The newly determined RNase P RNA sequences have been assigned the following EMBL accession numbers: *A. tumefaciens*, M59354; *A. eutrophus*, M59353; *C. vinosum*, M59356; *R. rubrum*, M59355; and *D. desulfuricans*, M59357. The accession number of the revised *Thiobacillus ferrooxidans* RNase P RNA sequence is X16580 (14).

RESULTS

Genes encoding RNase P RNAs. The RNase P RNA genes from the α proteobacteria *R. rubrum* (429 nucleotides [nt]) and *A. tumefaciens* (402 nt), the β proteobacterium *A. eutrophus* (341 nt), the γ proteobacterium *C. vinosum* (369 nt), and the δ proteobacterium *D. desulfuricans* (360 nt) have been cloned and sequenced (Fig. 2). The RNase P RNA probe hybridized to single genomic DNA fragments in Southern analyses of each case, indicating that RNase P RNA is encoded by a single-copy gene in these organisms (data not shown). The identities of these genes as those that express RNase P RNA have been confirmed by Northern analysis (data not shown). In addition, T7 and T3 RNA polymerase-generated transcripts from linearized or covalently closed circular DNA were tested for RNase P enzyme activity in vitro under conditions in which RNase P RNA is active in the absence of the RNase P protein component. In each case, transcripts from the sense DNA strand contained RNase P enzymatic activity (see below).

RNase P RNA from *A. eutrophus*. Because the *E. coli* antisense RNase P RNA probe failed to hybridize to *A. eutrophus* genomic DNA, native RNase P RNA from *A. eutrophus* was purified for use as probe. Following purification from total cellular RNA by gel filtration and electrophoresis, a single band containing RNase P RNA activity was used to probe genomic digests. The probe hybridized to single bands in genomic digests; however, clones of this DNA fragment were found to contain a gene which apparently encodes 10Sa RNA (an abundant, stable RNA of unknown function that in *E. coli* nearly comigrates with RNase P RNA [11]), transcripts of which lacked RNase P activity (3). Inclusion of the 10Sa RNA-encoding plasmid DNA (25 μg/ml, *Hin*dIII linearized and denatured by boiling) in the genomic hybridizations resulted in diminution of the hybridization signal from the 10Sa-encoding fragments and the appearance of single novel bands, which were cloned and found to encode functional RNase P RNA.

The sequence of the RNase P RNA gene from *A. eutrophus* seemed unusually different from that of the other β proteobacterium, *T. ferrooxidans* (27), considering the structural diversity seen among the α and γ groups. To confirm that the identity of the strain from which DNA and RNA were isolated was correct, rRNA-specific oligonucleotides were used with reverse transcriptase to obtain a sequence from the 16S rRNA from the batch of RNA which was used in the Northern analyses and from which *A. eutrophus* RNase P RNA was purified. The ~600 nucleotides of sequence analyzed clearly indicate that this strain is a member of the β-2 subgroup of proteobacteria, as defined by Woese (31) (data not shown).

Activity of transcripts containing extra 5' and 3' nucleotides. RNase P RNA sequences are highly variable, so hybridization signals obtained with heterologous RNase P RNA probes sometimes were weak and difficult to distinguish from background hybridization to the gene in the host *E. coli* chromosome. It was therefore advantageous during the isolation of new genes to verify by enzymatic activity that putative clones indeed contained an RNase P RNA gene. It proved unnecessary for detection of activity to trim the RNase P RNA gene-containing DNA to a size that would produce near-native-sized RNA when transcribed. In vitro transcripts containing an RNase P RNA sequence embedded within several kilobases of extraneous sequence, even within transcripts of covalently closed circular DNA, exhibit activity. This is documented in Fig. 3, which compares assay results obtained with native-length *E. coli* RNase P RNA transcripts and results obtained with transcripts of covalently closed circular DNA containing an RNase P RNA gene. The latter transcripts would consist of tandem arrays of RNase P RNA (or its complement) separated by 3 to 5 kb of plasmid-specified RNA. As seen in Fig. 3, the tandem arrays are approximately 0.1 to 1% as active as the native-length RNase P RNA. The fact that precise generation of the 5' and 3' ends of RNase P RNAs is not essential for enzymatic activity allows the rapid screening of putative RNase P RNA gene clones.

DISCUSSION

Refinement of the secondary structure model of RNase P RNA. The previous model for the secondary structure of RNase P RNA was based on the sequences from six γ proteobacteria and four gram-positive species of the genus *Bacillus*. On the basis of the criteria used for analysis of ribosomal RNA structure (32), covariation of two base pairings within a potential continuous helix was considered proof of the presence of that helix. The resolution of the RNase P RNA model can now be refined from the helix to nearly the base pair level (Fig. 4). Evidence, in the form of sequence covariation to maintain complementarity, now exists for nearly every base pair in the core structure model. Further verification of the model comes from the fact that covariations exist for most of the base pairings within just the proteobacterial sequences. Questionable assignments of homologies between the disparate sequences and structures of the proteobacterial and gram-positive RNase P RNAs are thus largely eliminated.

Two base pairings in the previous structure model have been removed in the refined model. The base pairing of nucleotides 19 and 335 (numberings used throughout this manuscript are based on the *E. coli* sequence, which is shown in Fig. 2) is eliminated from the model because these nucleotides are universally conserved; no covariation evidence for their pairing exists. Because this pairing would represent the simple continuation of a proven helix by 1 bp, it would be retained in the structure model if the potential pairing were not a noncanonical (G · U) interaction. Pairing between nucleotides 92 (C in *E. coli*) and 106 (G in *E. coli*) is also removed in the refined model because variation in these nucleotides is seen only in *A. eutrophus*, in which both residues are U. Because U · U pairs in rRNAs often do not disrupt the structure of a helix (21), it remains possible that nucleotides 92 and 106 interact by base pairing, except in the case of *A. eutrophus*. This would be especially likely if helix 93–96/102–105 forms a coaxial stack with either helix 87–91/238–242 or helix 108–110/115–117, thereby internalizing the U · U interaction in *A. eutrophus* in a continuous helix. The presence in *A. eutrophus* of a U at position 92 is noteworthy because the identity of this nucleotide, a C in all other RNase P RNAs examined so far, has been suggested to be critical for both enzymatic activity and the interaction of the RNA and protein components of RNase P RNA (6).

In the previous structure model, there was ambiguity in the interaction of nucleotide 284 with either nucleotide 266 or 270 in the proteobacterial structures. Parsimonious reconstruction of the evolution of this structural element reveals the conserved base pairing of nucleotide 284 with nucleotide 270, creating a 6-bp stem proximal to the bulge in γ, β, and δ proteobacteria and a 7-bp stem in the α proteobacteria.

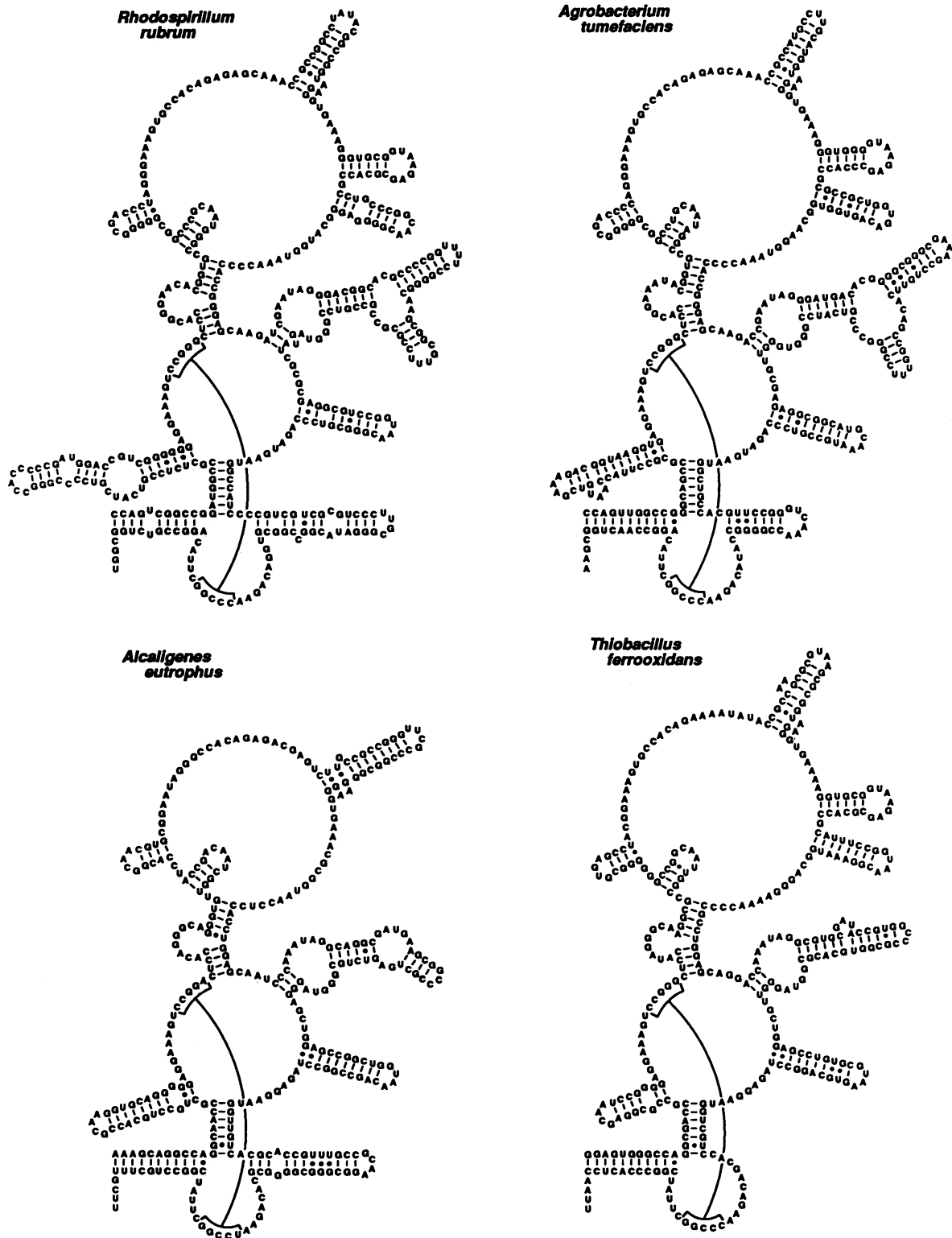Two noncanonical base pairings are added to the refined

FIG. 2. Secondary structures of RNase P RNAs in proteobacteria. The sequences of the newly determined RNase P RNAs are drawn according to the refined structure model (see the text). The *E. coli* and *B. subtilis* RNase P RNAs are shown for comparison (22, 23). The pairings indicated by the brackets and line (*E. coli* 70–74/354–358) complete a pseudoknot in the secondary structure. The sequence of the RNase P RNA from *T. ferrooxidans* (27) contains unpublished sequence corrections (14). The 5' and 3' termini of the RNAs from *A. tumefaciens*, *R. rubrum*, *A. eutrophus*, *T. ferrooxidans*, *C. vinosum*, and *D. desulfuricans* were not determined directly but are based on the established mature ends of the *E. coli* RNase P RNA (22).
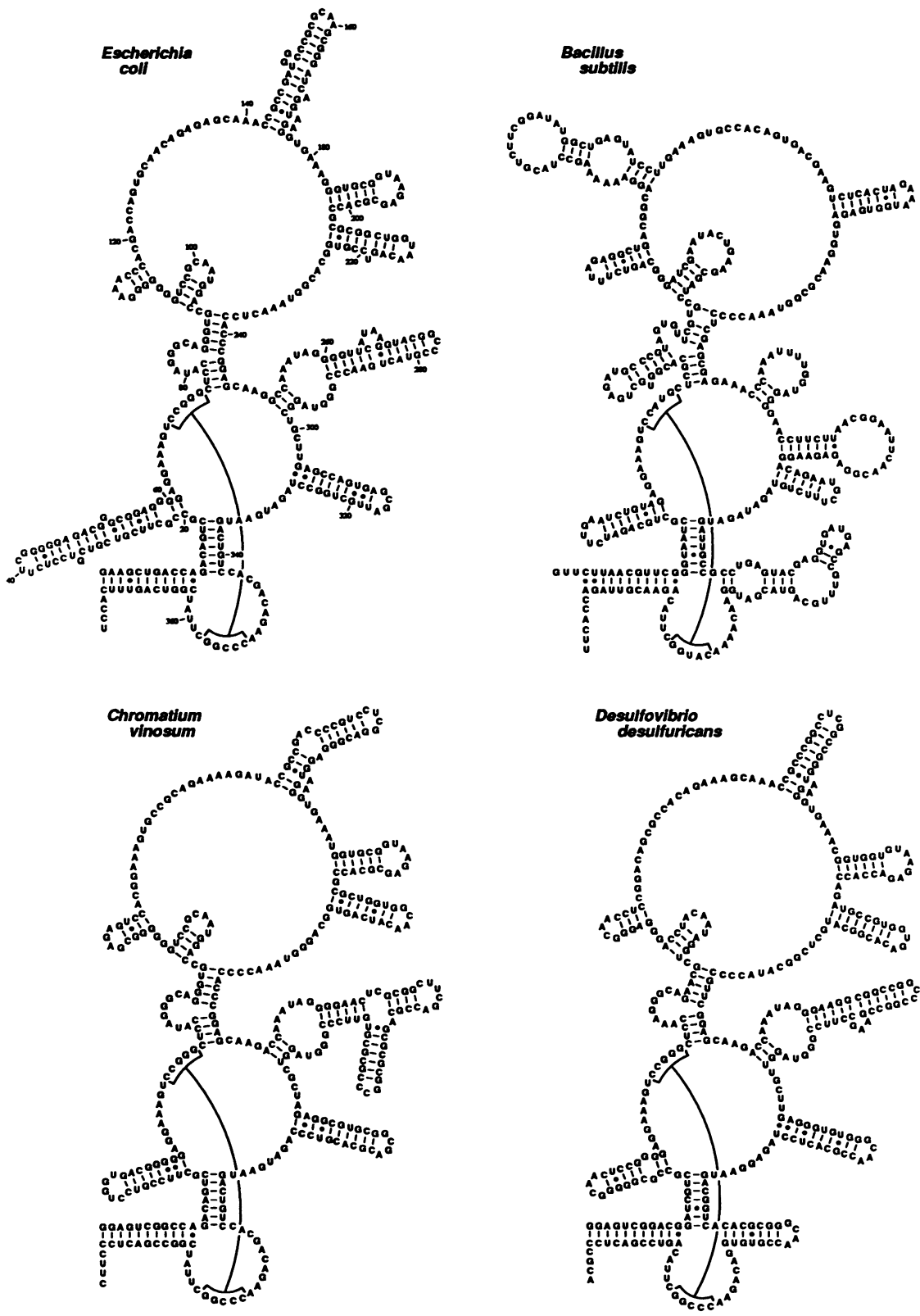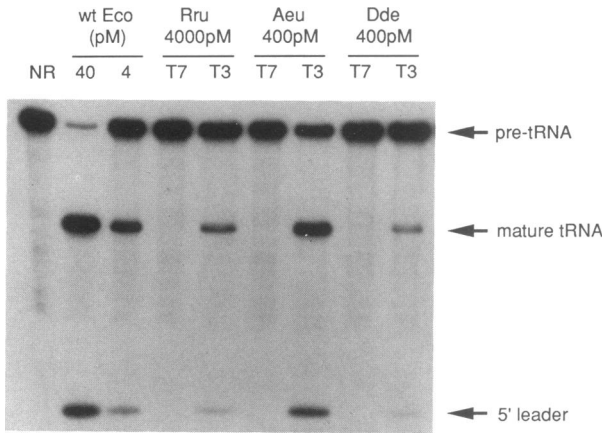
FIG. 2—*Continued.*

FIG. 3. RNase P enzyme assays of transcripts from covalently closed circular template DNAs. Transcripts of RNase P RNA-encoding clones, synthesized in vitro with T7 or T3 RNA polymerase, were assayed as described in Materials and Methods with 50 nM precursor tRNA^{Asp} as the substrate. Lane NR, reaction lacking added RNase P RNA. A T7 runoff transcript from a cloned $E.$ $coli$ RNase P RNA containing the mature 5' and 3' ends (wt Eco) is included for comparison. Assays of transcripts from the 5.4-kb $R.$ $rubrum$ clone (Rru), 5.3-kb $A.$ $eutrophus$ clone (Aeu), and 6-kb $D.$ $desulfuricans$ clone (Dde) are shown. In each of these cases, the sense DNA strand corresponds to the T3 transcript. Bands corresponding to the substrate (precursor tRNA [pre-tRNA]) and products (mature tRNA and 5' leader) of the reaction are indicated on the right. The transcripts from covalently closed circular DNAs are approximately 0.1 to 1% as active as the native-length $E.$ $coli$ RNase P RNA.

structure model. Nucleotides 11 and 363 covary as noncanonical A · C or G · A base pairs in the proteobacteria and $Bacillus$ spp. and occur as canonical pairs in some extremely thermophilic RNase P RNAs (9). Nucleotides 305 and 326, which have the potential to form either A · C or A · G noncanonical pairs in the proteobacteria, are internal in a proven helix and covary normally as canonical base pairs in $Bacillus$ spp. The significance of these noncanonical pairings
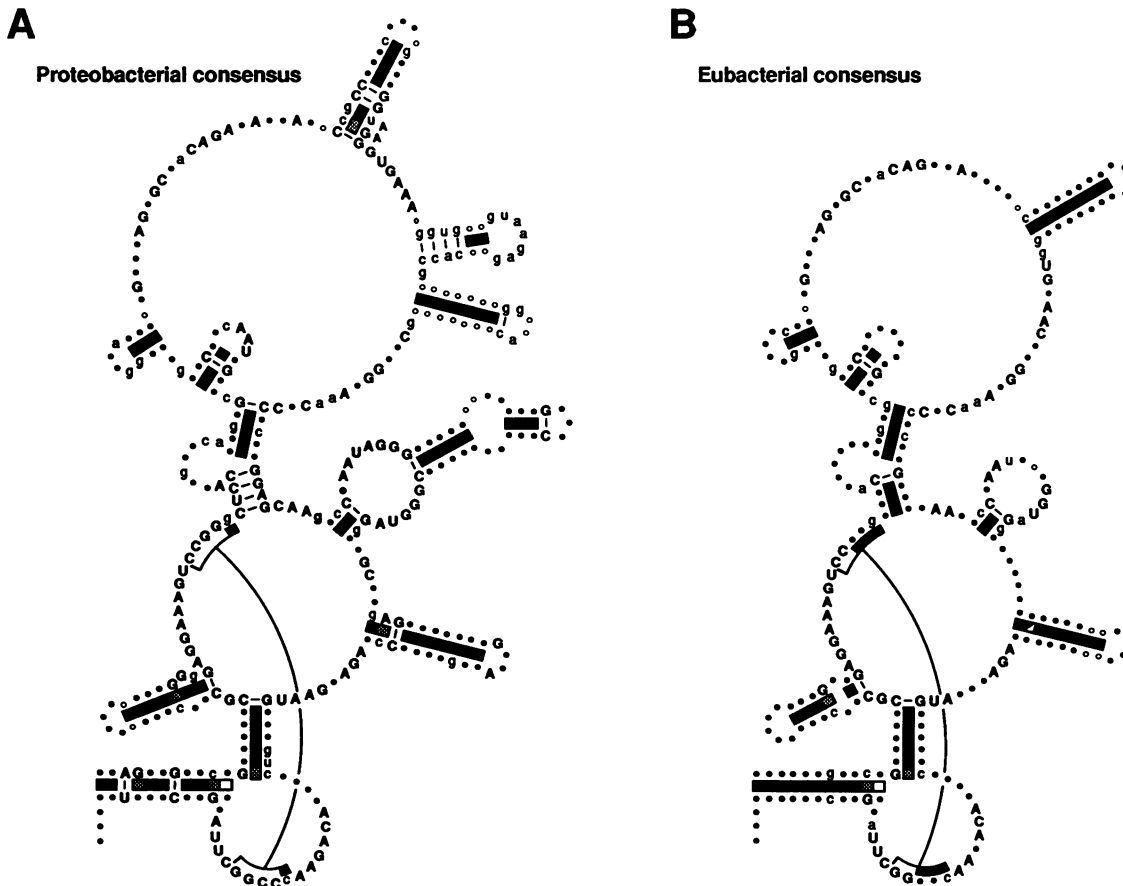

FIG. 4. Conserved RNase P RNA structure and covariation evidence within the proteobacteria (A) and within the eubacteria (B). Only nucleotides which are present in all or all but one of the available sequences are included. Nucleotides which vary are indicated by closed circles, invariant nucleotides are indicated in uppercase letters, and nucleotides which vary in only one known sequence are shown in lowercase letters. Variable nucleotides which are absent in one sequence are indicated by open circles. Bars indicate pairing for which there is covariation evidence, as follows: ■, canonical or G · U pairing; □, conserved noncanonical G · A or A · C pairing; ▨, pairing at position 305/326 in panel B, which is canonical in the $Bacillus$ structures but noncanonical in the proteobacterial structures; ▦, base pairing which contains only G-C and G · U variants or A · C and A · G variants. Base pairings which lack phylogenetic evidence but are included in the model because they represent the continuation of a proven helix are indicated by lines.

is unknown; perhaps they are required for the formation of distinct local structures.

The region of conserved nucleotides 248 and 249 in the RNA is of particular interest, because it seems to be associated with the active site. Deletions in this region have more severe effects on catalytic activity than deletions in any other portion of the molecule (29), and these nucleotides react with a photoactivated cross-linking agent attached to the substrate phosphate in tRNA (4). In the previous structure model, nucleotide 74 was paired with nucleotide 247; there was apparent covariation of these nucleotides between the proteobacterial and *Bacillus* sequences. The alignment of the sequences from the two groups, however, is particularly difficult in this region. The RNase P RNA sequence from *A. eutrophus* clearly demonstrates covariation of nucleotide 74 with nucleotide 353 rather than with nucleotide 247. In the refined model, therefore, helix 70–74/353–357 is extended by a single base pair, at the expense of helix 75–78/243–246. It is possible, by assuming that nucleotide U69 is bulged, to extend helix 70–74/353–357 by an additional 3 bp (66–68/358–360). These potential base pairs are, however, entirely composed of invariant nucleotides. Because no covariations have been seen to either confirm or refute these pairings and since they would not constitute the uninterrupted continuation of a phylogenetically confirmed helix, they are not included in the structure model.

A helical element previously encountered in the *Bacillus* but not in the *E. coli* RNA (at position 344–345) proved to be common in the RNase P RNAs of proteobacteria. Only *T. ferrooxidans* and the γ proteobacteria lack this structure among the known sequences. The sequences in this stem are so divergent that homology is frequently impossible to define at the nucleotide level, but complementarity is maintained in the different RNAs.

Another variably present element is helix 184–189/197–202. The sequences involved, previously invariant, are now shown to form a helix by the observation of covariation in two pairs (188/198 and 189/197). The remainder of this helix is inferred from the potential of the invariant sequences to form a helix continuous with these covarying nucleotides. The helices 184–189/197–202 and 204–211/216–224 and the adjacent nucleotides 183, 203, and 225 covary in their presence or absence, so it seems likely that this region constitutes a distinct, dispensable (at least in terms of enzyme activity) domain of the RNA.

Although most of the pairings in the conserved core of the structure model have now been confirmed by covariation, a few remain invariant. These pairs are retained in the structure model because they are uninterrupted continuations of helices proven by canonical covariations.

Evolution of RNase P RNA structure in proteobacteria. Most length variation in RNase P RNA structure occurs in four helical regions: the distal portion of stem 20–38/43–61, the medial portion of helix 142–156/161–176, the medial portion of helix 260–275/279–290, and the whole of helix 344–345. Helices 260–275/279–290 and 344–345 are group-specific elements (i.e., occur only in some species), and the length (and therefore the presence of the distal region) of helix 20–38/43–61 varies widely from organism to organism. It might be expected, therefore, that these regions are not directly related to enzyme activity, and in fact the variably present regions have been shown to be dispensable in vitro by the construction of a minimal RNase P RNA which lacks these elements (30). Because these helices are so variable in length and structure, they are probably oriented so that the

E. coli RNase P RNA　　　　　**E. coli 23S rRNA E-site**

B. subtilis RNase P RNA　　　　**B. subtilis 23S rRNA E-site**

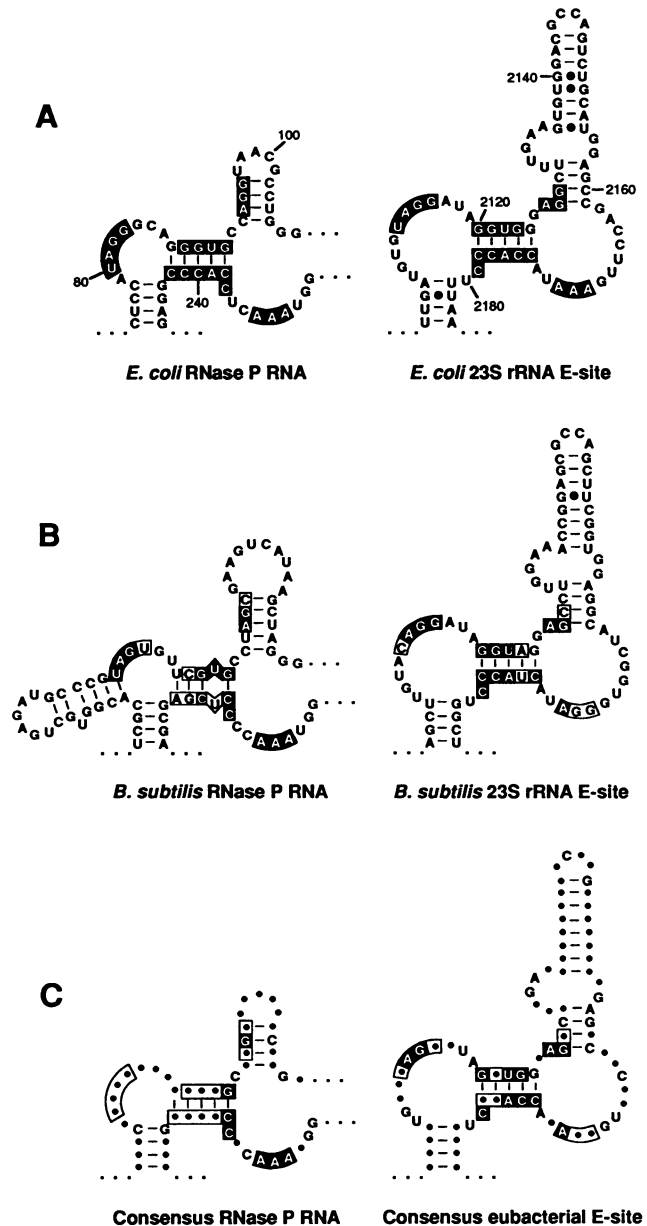**Consensus RNase P RNA**　　　**Consensus eubacterial E-site**

FIG. 5. Comparison of RNase P RNA and ribosomal exit site structure. The secondary structures of the regions of the 23S rRNA and RNase P RNA which have been proposed to be homologous (6) are shown. Regions of sequence similarity between the two RNAs in *E. coli* are boxed; nucleotides which are identical to those of *E. coli* are shown by white letters on black. The homologous structures from *E. coli* (A) and *B. subtilis* (B) are shown. In the consensus structures (C), letters indicate nucleotides which are invariant or which differ from the consensus in only one sequence. Variable nucleotides are designated by closed circles. Sequences used to derive the RNase P RNA consensus were the 16 currently available sequences from the proteobacteria and gram-positive eubacteria and those of *Deinococcus radiodurans* and *Thermus aquaticus* (9). Sequences used to derive the 23S rRNA exit site (E-site) consensus were the 19 eubacterial and chloroplast (excluding that of *Euglena gracilis*) sequences from the current 23S rRNA sequence compilation (7).

variable portion of each helix is peripheral to the more conservative core of the tertiary structure.

Sequence length variation outside of these four helical regions is rare and scattered throughout the molecule. Most of these changes are minor, such as single bulged nucleotides in helices, the shortening of a helix (108–110/115–117) by 1 bp, or the enlargement or reduction of loops by 1 or 2 nt. An exception is the absence from the *A. eutrophus* RNA of helices 184/202 and 204/224. These structures also are absent from the RNase P RNAs of *Bacillus* species and are dispensable in vitro (30).

RNase P RNA and the ribosomal exit site. The 16S and 23S rRNAs interact with tRNA, as does RNase P RNA. It is conceivable that rRNAs and RNase P RNA might have similar or even homologous tRNA binding sites. An early model of RNase P RNA secondary structure was proposed on the basis of sequence and potential structural similarity between RNase P RNA and a tRNA binding center of the 16S rRNA of *E. coli* (2). However, the similarity is not supported phylogenetically or by the current structure model for RNase P RNA.

More recently, similarity between the tRNA exit site of 23S rRNA and RNase P RNA in *E. coli* has been suggested (6). An identification of homology (common ancestry) between a tRNA binding site in the ribosomal RNA and RNase P RNA would be a significant observation, because it would relate the origins of the protein-synthesizing and RNA-processing machineries. We examined the proposed similarities in 23S rRNA and RNase P RNA from the perspective of the corresponding sequences from other organisms.

We use the term homology in its strictest sense: sequences that are homologous have common ancestry. If similarity in sequence is to be taken to indicate homology, the similar sequences should occur in the same structural context; the sequence similarities are expected to be superimposable on other structural similarities. Although the exit site of 23S rRNA and nucleotides 75–107/229–246 of RNase P RNA seem grossly similar in secondary structure in *E. coli*, the similar nucleotide sequences do not occur in exactly analogous parts of the two structures (Fig. 5A). This indicates that the similar sequences are not homologous. In addition, if similarity in these RNAs were significant, such similarity would not be idiosyncratic to *E. coli*. If the sequence similarities in 23S and RNase P RNAs were conserved in *E. coli* from some common ancestral RNA, they would also be highly conserved in the corresponding RNAs of other organisms. Sequences for both 23S rRNAs and RNase P RNAs are available only for close relatives of *E. coli* and *B. subtilis*; even so, it is clear that these nucleotides and their similarities are not highly conserved (Fig. 5B). Comparison of eubacterial consensus structures of these two RNAs further shows this lack of conservation (Fig. 5C). Only five of the nucleotide similarities seen in the *E. coli* 23S and RNase P RNAs are shared in the consensus sequences, and these, again, do not occur in corresponding positions in the structures. We conclude that there is no evidence for homology between the 23S rRNA exit site and RNase P RNA in eubacteria and that any similarity in structures is at best tenuous.

REFERENCES

1. **Altman, S.** 1989. Ribonuclease P: an enzyme with a catalytic RNA subunit. Adv. Enzymol. Relat. Areas Mol. Biol. **62:**1–36.
2. **Boehm, S.** 1987. Similarities between a predicted secondary structure for the M1 RNA ribozyme and the tRNA binding center of 16S rRNA from *E. coli*. FEBS Lett. **220:**283–287.
3. **Brown, J. W., D. A. Hunt, and N. R. Pace.** 1990. Nucleotide sequence of the 10Sa RNA gene of the β-purple eubacterium *Alcaligenes eutrophus*. Nucleic Acids Res. **18:**2820.
4. **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.
5. **Gherna, R., P. Pienta, and R. Cote (ed).** 1989. American Type Culture Collection catalogue of bacteria and bacteriophages, 17th ed. American Type Culture Collection, Rockville, Md.
6. **Guerrier-Takada, C., N. Lumelsky, and S. Altman.** 1989. Specific interactions in RNA enzyme-substrate complexes. Science **246:**1578–1584.
7. **Gutell, R. R., M. N. Schnare, and M. W. Gray.** 1990. A compilation of large subunit (23S-like) ribosomal RNA sequences presented in a secondary structure format. Nucleic Acids Res. **18(Suppl.):**2319–2330.
8. **Gutell, R. R., B. Weiser, C. R. Woese, and H. F. Noller.** 1985. Comparative anatomy of 16-S-like ribosomal RNA. Prog. Nucleic Acid Res. **32:**155–215.
9. **Haas, E. S., J. W. Brown, and N. R. Pace.** Unpublished data.
10. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28:**351–359.
11. **Jaine, S. K., M. Gurevitz, and D. Apirion.** 1982. A small RNA that complements mutants in the RNA processing enzyme ribonuclease P. J. Mol. Biol. **162:**515–533.
12. **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.
13. **Khandjian, E. W.** 1986. UV crosslinking of RNA to nylon membrane enhances hybridization signals. Mol. Biol. Rep. **11:**107–115.
14. **Kusano, T. (Akita Prefectural College of Agriculture).** 1990. Personal communication.
15. **Lane, D. J., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin, and N. R. Pace.** 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analysis. Proc. Natl. Acad. Sci. USA **82:**6955–6959.
16. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**499–560.
17. **Mead, D. A., E. Szczesna-Skorupa, and B. Kemper.** 1986. Single-stranded DNA "blue" T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. Protein Eng. **1:**67–74.
18. **Ogden, R. C., and D. A. Adams.** 1987. Electrophoresis in agarose and acrylamide gels. Methods Enzymol. **152:**61–87.
19. **Pace, N. R., and D. Smith.** 1990. Ribonuclease P: function and variation. J. Biol. Chem. **265:**3587–3590.
20. **Pace, N. R., D. K. Smith, G. J. Olsen, and B. D. James.** 1989. Phylogenetic comparative analysis and the secondary structure of ribonuclease P RNA—a review. Gene **82:**65–75.
21. **Papanicolaou, C., M. Gouy, and J. Ninio.** 1984. An energy model that predicts the correct folding of both the tRNA and 5S RNA molecules. Nucleic Acids Res. **12:**31–44.
22. **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.
23. **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.

24. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

25. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.

26. Short, J. M., J. M. Fernandez, J. A. Sorge, and W. D. Huse. 1988. λ ZAP: a bacteriophage λ expression vector with in vivo excision properties. Nucleic Acids Res. 16:7583–7600.

27. Takeshima, T., C. Inoue, Y. Kitagawa, and T. Kusano. 1989. Nucleotide sequence of a Thiobacillus ferrooxidans chromosomal gene which encodes putative RNA component of RNase P. Nucleic Acids Res. 17:9482.

28. Vogelstein, B., and D. Gillespie. 1979. Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. USA 76:615–619.

29. Waugh, D. S. 1989. Mutational analysis of catalytic function in ribonuclease P RNA. Ph.D. dissertation. Indiana University, Bloomington.

30. 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.

31. Woese, C. R. 1987. Bacterial evolution. Microbiol. Rev. 51:221–271.

32. 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.