Characterization of ribonuclease P RNAs from thermophilic bacteria

James W.Brown, Elizabeth S.Haas and Norman R.Pace*

Department of Biology and Institute for Molecular and Cellular Biology, Indiana University, Bloomington, IN 47401, USA

ABSTRACT

The catalytic RNA component of bacterial RNase P is responsible for the removal of 5' leader sequences from precursor tRNAs. As part of an on-going phylogenetic comparative characterization of bacterial RNase P, the genes encoding RNase P RNA from the thermophiles *Thermotoga maritima*, *Thermotoga neapolitana*, *Thermus aquaticus*, and a mesophilic relative of the latter, *Deinococcus radiodurans*, have been cloned and sequenced. RNAs transcribed from these genes *in vitro* are catalytically active in the absence of other components. Active holoenzymes have been reconstituted from the *T.aquaticus* and *T.maritima* RNAs and the protein component of RNase P from *Escherichia coli*. The RNase P RNAs of *T.aquaticus* and *T.maritima*, synthesized *in vitro*, were characterized biochemically and shown to be inherently resistent to thermal disruption. Several features of these RNAs suggest mechanisms contributing to thermostabilty. The new sequences provide correlations that refine the secondary structure model of bacterial RNase P RNA.

INTRODUCTION

Ribonuclease P (RNase P) is a ribonucleoprotein which functions in tRNA biosynthesis by endonucleolytically removing 5' leader sequences from precursor tRNAs (see references 1–3 for review). Bacterial (formerly eubacterial [4]) RNase Ps so far examined are composed of two subunits, a small (ca. 120 amino acids) protein and a large (ca. 400 nucleotides) RNA (5). The RNA component of the enzyme is the catalytic moiety; at high ionic strength *in vitro*, the RNA is capable of cleaving precursor tRNA in the absence of the protein component (6). The protein component is thought to stabilize the enzyme or enzyme/substrate complex by screening electrostatic repulsion due to RNA phosphates (7).

Insight into the mechanisms of substrate recognition and catalysis by RNase P will require an understanding of the higher-order structure of the RNA component. Phylogenetic comparisons have been the most useful approach for determining the secondary structure of RNAs. In this method, base-paired nucleotides are identified by evolutionary variations in sequence that nevertheless

maintain complementarity. Sequences from organisms which are phylogenetically distant are particularly disparate, and so are especially important for inferring the structure of conservative portions of the RNA, in which sequence variation occurs rarely. However, most of the available information on RNase P RNA so far is based on genes cloned from close relatives of *Escherichia coli* (i.e. the purple Bacteria and relatives [8], also known as proteobacteria [9]) and *Bacillus subtilis* (low G+C Gram positive Bacteria [8]). In order to provide a collection of sequences from more diverse Bacteria, we cloned and determined the nucleotide sequences encoding RNase P RNAs from two additional branches of Bacteria, the deinococci and relatives (*Thermus aquaticus* and *Deinococcus radiodurans*) and Thermotogales (*Thermotoga maritima* and *Thermotoga neapolitana*) (Figure 1). In addition to their phylogenetic distinctness from the other Bacteria, representatives of the deinococci (and relatives) and Thermotogales include the most thermophilic Bacteria known, species of the genera *Thermus* and *Thermotoga*, which grow at temperatures in excess of 70°C. We therefore also examined the biochemical properties of the RNase P RNAs from *Thermus aquaticus* and *Thermotoga maritima*, with emphasis on those properties which might be related to their function at high temperatures. An understanding of RNA structural features that confer thermal stability on RNase P RNA should be of considerable use in the manipulation of the structure of this ribozyme.

MATERIALS AND METHODS

Cultures and growth conditions

Deinococcus radiodurans ATCC 13939 was grown in nutrient broth (Difco) with 1% glucose at 30°C. *Thermus aquaticus* YT-1 (ATCC 25104) was grown aerobically in Castenholz TYE medium (10) at 70°C. *Thermotoga maritima* strain MSB-8 (11) and *Thermotoga neapolitana* (12) were grown anaerobically in *Thermotoga* medium (11) at 75°C.

Cloning of RNase P RNA genes

Genomic DNAs were purified from cultured cells and DNA fragments containing RNase P RNA sequences were identified by Southern analysis of restriction digests of genomic DNAs as

previously described (13). An oligonucleotide (5'-GIIGAGG-AAAGTCCIIGCT-3': I is inosine) that contains the most highly conserved sequence from known RNase P RNAs was used as probe for cloning the RNase P RNA-encoding genes from *T.aquaticus* and *D.radiodurans*. A partially hydrolyzed, antisense transcript of the *E.coli* RNase P RNA-encoding gene (13) was used as probe for the *T.maritima* RNase P RNA-encoding gene, and a partially hydrolyzed transcript of the *T.maritima* RNase P RNA-encoding gene was used as probe to identify the *T.neapolitana* gene. Restriction endonuclease digests of genomic DNAs were separated by agarose gel electrophoresis, and fractions of the gels were excised that, based on Southern blot information, contained RNase P RNA-encoding genes. DNA was recovered from gel slices by disruption with NaI and was purified using glass powder (14). Size-selected DNAs were then cloned into pBluescript KS⁻ (Stratagene) and screened by hybridization to colonies (*D.radiodurans* and *T.neapolitana*) or plasmid DNAs (*T.aquaticus* and *T.maritima*) transferred to nitrocellulose membranes using the same probes used to identify the gene-containing fragments. Potential clones were then tested in RNase P RNA-alone assays of T7 and T3 RNA polymerase-generated rolling-circle transcripts (13). The RNase P RNA-encoding gene from *D.radiodurans* (GenBank accession number M64708 [15]) was cloned on a 1.8Kb *Hind* III-*Pst* I DNA fragment (pDra9-3). The RNase P RNA-encoding gene from *T.aquaticus* (GenBank accession number Z15006) was cloned on a 1.2Kb *Pst* I DNA fragment (pTA63-G). The RNase P RNA-encoding gene from *T.maritima* (GenBank accession number M64709 [15]) was cloned on a 2.3Kb *Hind* III DNA fragment (pTma1.62). The RNase P RNA-encoding gene from *T.neapolitana* (GenBank accession number Z15043) was cloned on a 3.4Kb *Pst* I/*Sal* I DNA fragment (pTne5 − 11). Series of nested deletions in each orientation of the cloned DNAs were generated using exonuclease III and S1 nuclease (16) and sequenced by the chain-termination method (17) using Sequenase version 2.0 (United States Biochemical, Inc.). Band compression in sequencing gels was alleviated by replacement of dGTP in sequencing reactions with 7-deaza-dGTP.

Plasmid constructions

A 483bp *Rma* I-*Acc* I DNA fragment from pTA63-G containing the *T.aquaticus* RNase P RNA gene was end-filled with T4 DNA polymerase and dNTPs (18) and ligated into the *Sma* I site of pBluescript KS⁻. Sequences between the T7 RNA polymerase transcription initiation site and the desired 5' end of the RNase P RNA-encoding region were removed in a 69 bp oligonucleotide-directed deletion (18,19). The first three nucleotides of the transcribed DNA were converted in the process from CCC to GGG in order to accommodate the T7 promoter. A *Sty* I site downstream of the RNase P RNA-encoding gene was then moved to the desired 3' end of gene by a 34bp oligonucleotide-directed deletion, generating plasmid pTaq5'3'. This deletion also changes the GGGACC sequence at the downstream end of the transcribed DNA to CCCAAG, restoring complementarity of the 5' and 3' ends of the RNA transcript.

A 414bp *Hae* II-*Eco* RI fragment from pTma1.62 containing the *T.maritima* RNase P RNA-encoding gene was subcloned into *Hinc* II-*Eco* RI linearized pBluescript SK⁻. An oligonucleotide-directed 79 bp deletion was used to place the T7 RNA polymerase transcription initiation site at the desired 5' end of the RNase
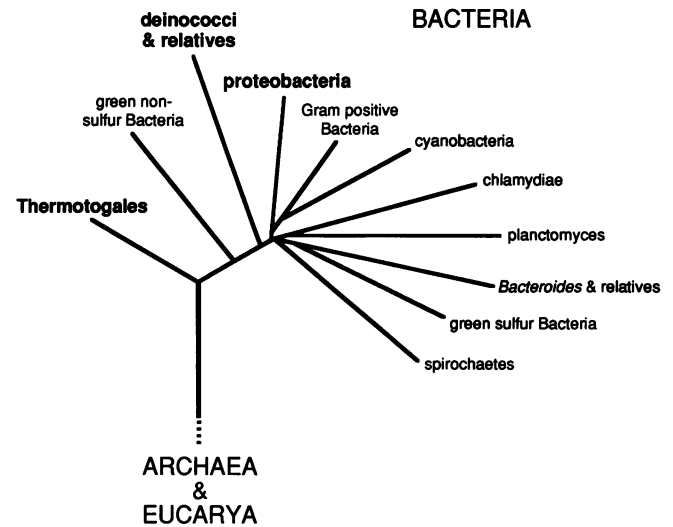

Figure 1. Phylogenetic tree of Bacteria based on 16S rRNA sequences (8). Each of the 11 recognized 'kingdoms' of Bacteria are shown (4). Evolutionary divergence is represented by total line length. *E.coli* is a member of the proteobacteria (also known as the 'purple Bacteria and relatives' [8]); *Thermus* spp. and *Deinococcus* spp. are members of the 'deinococci and relatives'; and *Thermotoga* spp. are members of the Thermotogales.

P RNA-encoding region and three oligonucleotide-directed substitutions (GTTATA to GGTACC) created a *Kpn* I site at the 3' end of the RNase P RNA-encoding region. No changes were made in the region of the gene transcribed by T7 RNA polymerase.

RNase P RNAs were synthesized *in vitro* using T7 RNA polymerase and purified by denaturing polyacrylamide gel electrophoresis as described (20). *E.coli* RNase P RNA and *B.subtilis* pre-tRNA^Asp were synthesized using *Sna* BI-linearized pDW98 and *Bst* NI-linearized pDW128, respectively, as templates (21).

Measurement of enzyme activity

RNase P RNA-alone enzyme activity was measured in the presence of 50mM Tris−HCl (pH 8 at 22°C), 25mM MgCl₂, 0.1% sodium dodecylsulfate (SDS), 0.05% nonidet P-40, ammonium chloride (concentration as indicated for each experiment), and 2nM uniformly ³²P-labeled *B.subtilis* tRNA^Asp precursor. Enzyme concentrations were adjusted in each experiment so that steady-state rates were measured, and so that less than ca. 30% of the substrate was cleaved in the assays. Enzyme reactions were stopped by the addition of an equal volume of 8M urea in gel electrode buffer, and the products were separated by electrophoresis through 8% polyacrylamide gels in the presence of 8M urea. Gels were fixed and dried prior to autoradiography. Radioactive substrate and product bands were excised from dried gels and Cerenkov radiation was quantitated by scintillation spectrophotometry.

In reconstitution experiments, RNase P RNA (0.5nM) and *E.coli* RNase P protein (C5 protein, 20nM; a gift from B. Pace) were preincubated in RNase P assay buffer containing 100mM ammonium chloride, and lacking SDS, for 15 minutes at 40°C. Enzyme activity at 40°C was measured as described above after the addition of ³²P-labeled precursor tRNA^Asp.
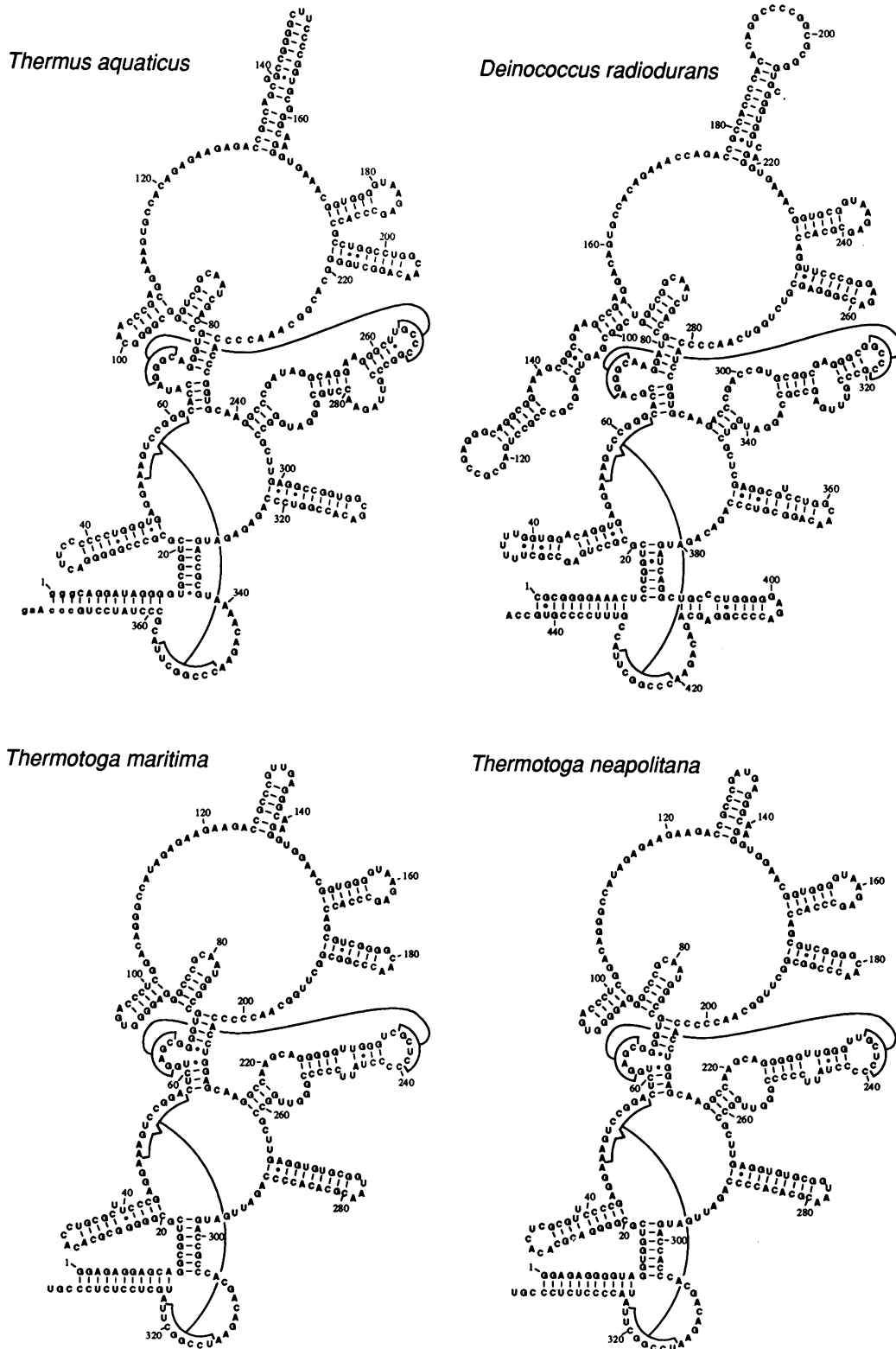
Figure 2. Secondary structures of RNase P RNAs from the Thermotogales and deinococci. The structures are based on phylogenetic comparative analysis (see text). Watson-Crick base-pairs are shown with lines (−), non-Watson-Crick base-pairs with filled circles (●). Lines and brackets indicate long-range pairings in the structures. Nucleotides in lower case have been changed from the native sequence to accomodate synthesis of the RNAs *in vitro*; the sequences and termini shown represent the synthetic transcripts rather than the native RNAs (see text). The structure of the region between nucleotides 104 and 146 of the *D.radiodurans* sequence is based on thermodynamic predictions (30), because no homolog is available for comparative analysis.
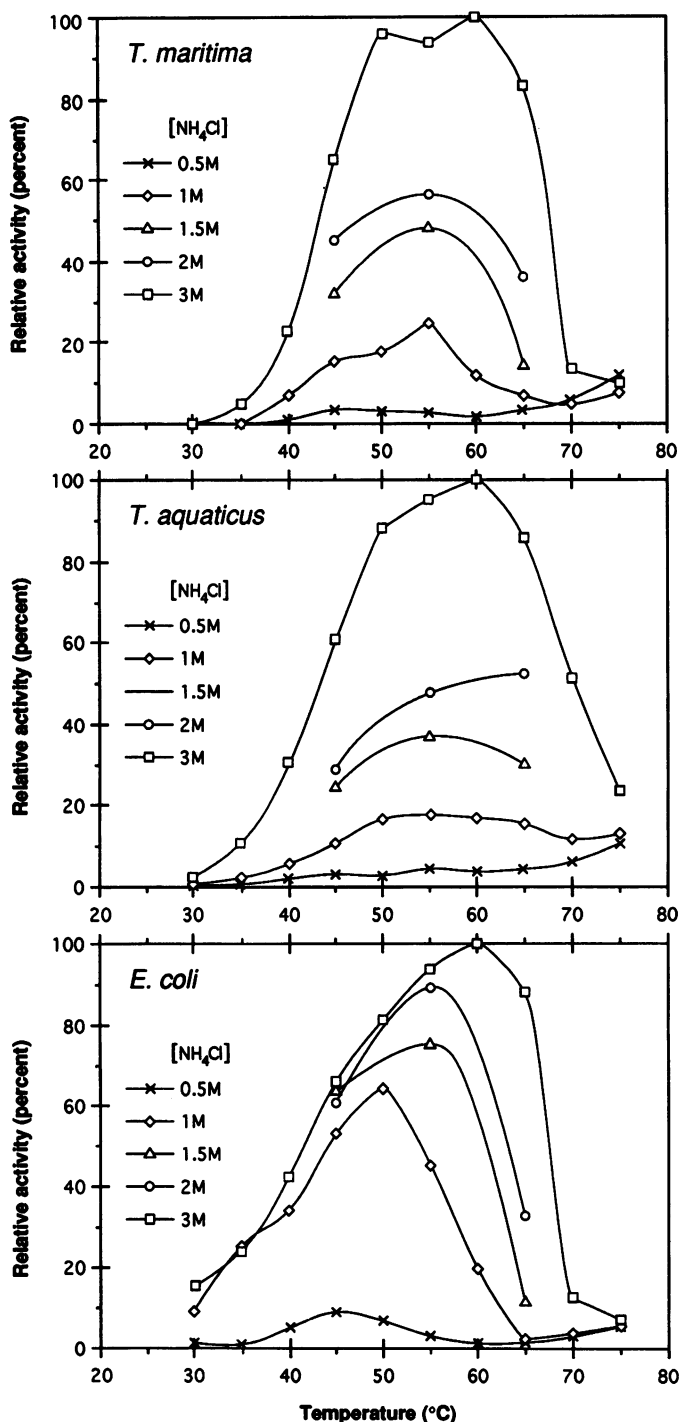
Figure 3. Effect of salt concentration and temperature on RNase P RNA activity. (A) *T.maritima* RNase P RNA-alone activity at various temperatures assayed in the presence of 0.5M-3M ammonium chloride; (B) *T.aquaticus* RNase P RNA-alone activity, and (C) *E.coli* RNase P RNA-alone activity assayed under the same conditions. Activity was normalized to the maximum value in each case. Actual maximal rates of cleavage were:*T.maritima*, 2.2×10^8 M^{-1} min^{-1}; *T.aquaticus*, 3.1×10^8 M^{-1} min^{-1}; *E.coli*, 0.75×10^8 M^{-1} min^{-1}.

Thermal denaturation assays

Thermal denaturation of RNAs was measured by hyperchromicity at a wavelength of 260nm in 5.3mM sodium phosphate (pH7) and in 5.3mM sodium phosphate (pH7), 990mM NaCl as previously described (22).
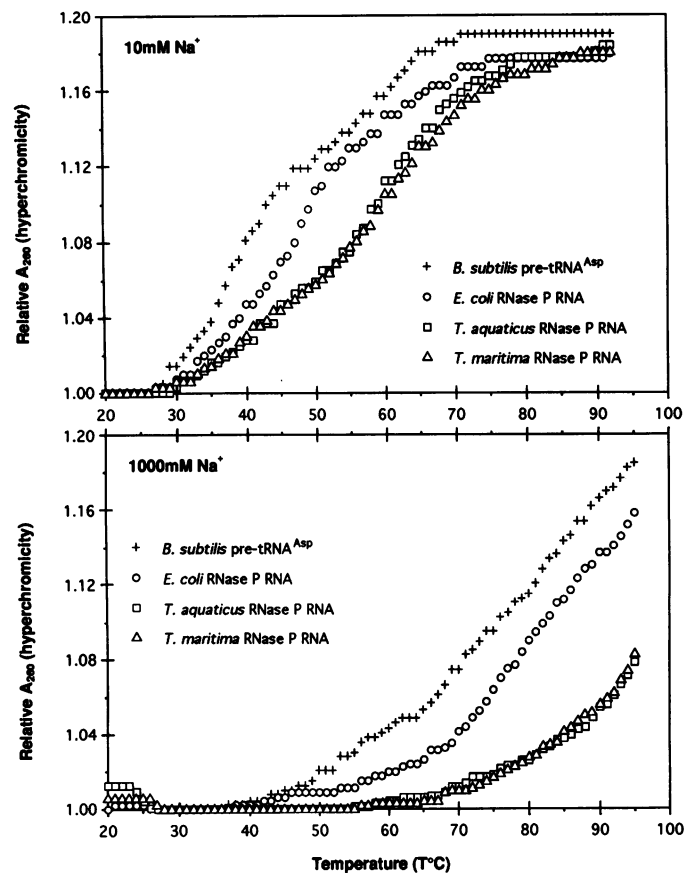

Figure 4. Thermal denaturation curves of RNase P RNAs. Change in absorbance at 260nm was measured with increasing temperature. (A) 5.3mM sodium phoshate, pH 7 (10mM with respect to Na^+); (B) 5.3mM sodium phosphate, pH 7/990mM NaCl (1000mM with respect to Na^+).

RESULTS

RNase P RNA-encoding genes

The RNase P RNA-encoding genes from *D.radiodurans*, a mesophile, and a specific relative *T.aquaticus*, a thermophile, and two thermophilic species of *Thermotoga*, *T.maritima* and *T.neapolitana*, were cloned and sequenced. (Figure 2). The probes used to identify RNase P RNA sequences in Southern blots hybridized in each case to single DNA fragments (data not shown). RNase P RNA is therefore encoded by single-copy genes in these organisms, as in all Bacteria so far tested (5). Transcripts of the 'sense' strand of the genes synthesized *in vitro* using T7 (*T.aquaticus*, *T.maritima*, and *T.neapolitana*) or T3 (*D.radiodurans*) RNA polymerase from covalently closed circular plasmid DNA (rolling-circle transcripts) contain RNase P activity, confirming that the cloned DNA fragments encode functional RNase P RNAs (data not shown). The locations of the 5′ and 3′ termini of the RNAs have not been determined experimentally, but are based (in the secondary structure models) on the known mature ends of the RNAs from other bacteria (5,20,23) and homology to conserved elements of RNase P RNA structure. Previous experiments have not shown detectable influence of short extensions of, or sequence changes in, the termini of RNase P RNAs on enzymatic activity *in vitro* (21,24). It is possible that rho-independent terminator-like sequences at the 3′ ends of the RNAs may be retained *in vivo*, as seems to be the case in *Thermus*

thermophilus (25); however these sequences, if present, are not conserved and are clearly dispensable.

Response of ribozyme activity to salt and temperature

Like other bacterial RNase P RNAs, the *T.aquaticus* and *T.maritima* RNase P RNAs require divalent cation for enzymatic activity. At a saturating concentration of monovalent salt, magnesium chloride at concentrations higher than 25mM had little or no additional stimulatory effect (data not shown). The addition of spermine or spermidine (1mM) had no stimulatory or thermal stabilizing effect, nor did the polyamines alter the requirement of the activity for monovalent salt (data not shown).

Although the *T.aquaticus* and *T.maritima* RNAs require high temperatures (and high monovalent salt concentrations) for activity, their thermophilic character is not readily evident by comparison with mesophilic RNase P RNAs (e.g. that of *E.coli*) using standard assays for cleavage of *B.subtilis* pre-tRNA^{Asp}. Measurements of optimal and maximal temperatures for activity are complicated by differing ionic strength optima for the compared ribozymes, and are compromised by thermal denaturation of substrate (a mesophilic precursor tRNA) and the properties of RNase P kinetics. Under RNA-alone reaction conditions standard for *E.coli* or *B.subtilis* RNAs (reaction buffer containing 1M NH_4Cl), the optimal reaction temperatures of the RNase P RNAs from *T.aquaticus* and *T.maritima* are only 5–10°C higher than that of *E.coli* (Figure 3). Maximal activity is obtained from the thermophilic and *E.coli* RNase P RNAs under identical conditions, i.e. 3M NH_4Cl at 60°C. The catalytic efficiency of the thermophilic RNase P RNAs, expressed in terms of k_{cat}/K_m (*T.maritima*, 2.2×10^8 $M^{-1}min^{-1}$; *T.aquaticus*, 3.1×10^8 $M^{-1}min^{-1}$) are 3–4 times higher than that of the *E.coli* RNA (7.5×10^7 $M^{-1}min^{-1}$) under the same conditions. Nevertheless, it is clear that *E.coli* RNase P RNA functions at nearly maximal activity in the presence of 3M NH_4Cl at temperatures up to 65°C, even though the optimal growth temperature of *E.coli* is 37°C. The rapid turn-over of product at high temperatures by the thermophilic RNase P RNAs, and to a lesser extent the the *E.coli* RNA, is most likely due to facilitated release of the product, which is the rate-limiting step of the reaction at 37°C.

High concentrations of NH_4Cl were required by the *E.coli* RNA only for activity at high temperatures. Lower NH_4Cl concentrations were optimal at lower temperatures, with no decrease in reaction rate relative to assays at the same temperature carried out at higher concentrations of NH_4Cl. The thermophilic RNAs, however, required high concentrations of NH_4Cl for activity at any temperature. The optimal temperature for activity remained essentially unchanged at different concentrations of NH_4Cl, but activity at lower salt concentrations was drastically reduced. The difference between the temperature optima of the RNase P RNAs from the thermophiles and that of *E.coli* is small at moderate NH_4Cl concentrations and nonexistent at high NH_4Cl concentrations. However, optical melting curves (below) indicate that the upper temperature limits of the activities of all these RNase P RNAs reflect thermal disruption of the precursor tRNA^{Asp} (from *B.subtilis*), rather than the RNase P RNAs.

Thermal denaturation profiles

The thermal stabilities of the RNase P RNAs from these thermophilic Bacteria are evident in thermal 'melt' curves of the RNAs (Figure 4). The T_ms of the RNase P RNAs from *T.aquaticus* and *T.maritima* at low ionic strength (10mM Na$^+$)

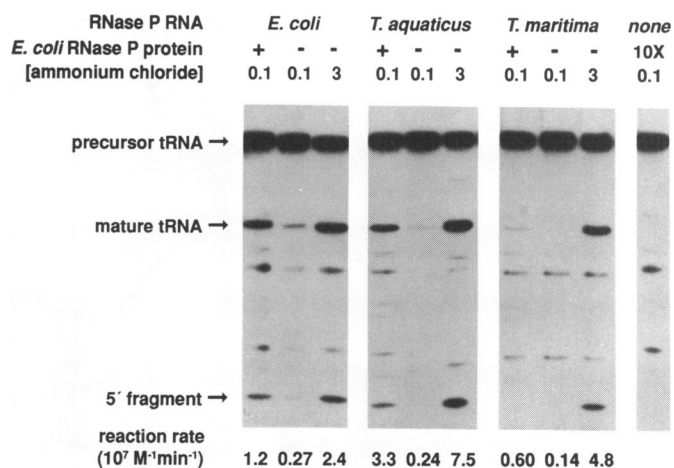| RNase P RNA	E. coli			T. aquaticus			T. maritima			none
E. coli RNase P protein	+	-	-	+	-	-	+	-	-	10X
[ammonium chloride]	0.1	0.1	3	0.1	0.1	3	0.1	0.1	3	0.1
precursor tRNA →										
mature tRNA →										
5′ fragment →										
reaction rate (10^7 M^{-1}min^{-1})	1.2	0.27	2.4	3.3	0.24	7.5	0.60	0.14	4.8	

Figure 5. Reconstitution of RNase P RNAs and *E.coli* C5 protein. RNase P RNAs were preincubated in either 0.1M or 3M ammonium chloride in the presence or absence of an 40-fold molar excess of purified RNase P protein from *E.coli* for 15 min. at 40°C, and then were assayed in 15 min. reactions at the same temperature after the addition of pre-tRNA^{Asp} substrate. Reconstitution of holoenzyme is indicated by activation of the RNA in the presence of only 0.1M ammonium chloride; in the absense of protein, the RNA alone requires high salt concentrations (3M was used in these experiments) for maximal activity. A small amount of RNA-alone activity is evident in the case of the *E.coli* RNA at the low salt concentration. Reaction rates were quantitated and are indicated below each lane.

are 8°C higher than the *E.coli* RNA. At high ionic strength (1000mM Na$^+$), the difference in T_ms of the thermophilic RNAs relative to *E.coli* RNA is ca. 16°C (the differences in T_ms are less certain at high ionic strength because the upper limits of the curves are not defined). At both high and low ionic strength, melting of the substrate precursor tRNA^{Asp} preceeded that of the RNase P RNAs from either the thermophiles or *E.coli*. Despite the differences in the optimal and maximal growth temperatures of *T.aquaticus* and *T.maritima*, and the phylogenetic distance separating them, the melting profiles of the thermophilic RNAs in both high and low salt concentrations are indistinguishable.

Reconstitution of heterologous holoenzymes

The RNase P RNAs from *T.aquaticus* and *T.maritima* respond to the protein subunit of RNase P from *E.coli* (C5 protein) (Figure 5). The protein component is thought to be essential in *E.coli* for RNase P enzymatic activity *in vivo*, and is required for substantial activity at low ionic strength *in vitro* (7). The activities of *E.coli* and *T.aquaticus* RNase P RNAs, assayed in 100mM NH_4Cl at 40°C, are stimulated by reconstitution with *E.coli* C5 protein to nearly the level of activity of the RNA-alone assayed in the presence of high salt concentrations (3M). The *T.maritima* RNase P RNA also responds to reconstitution with *E.coli* RNase P protein, albeit much less efficiently than the other RNAs. This is not surprising given the great evolutionary distance between *T.maritima* and *E.coli*; the Thermotogales are the least-related of known Bacteria to *E.coli*. Plasmid DNAs containing the *T.aquaticus* RNase P RNA-encoding gene (but not that of *T.maritima*) are also capable of maintaining the growth of *E.coli* strain DW2 (26) at 42°C, at which temperature the native *E.coli* gene is eliminated and survival of the strain depends on the expression of the plasmid-born RNase P RNA-encoding gene (data not shown).
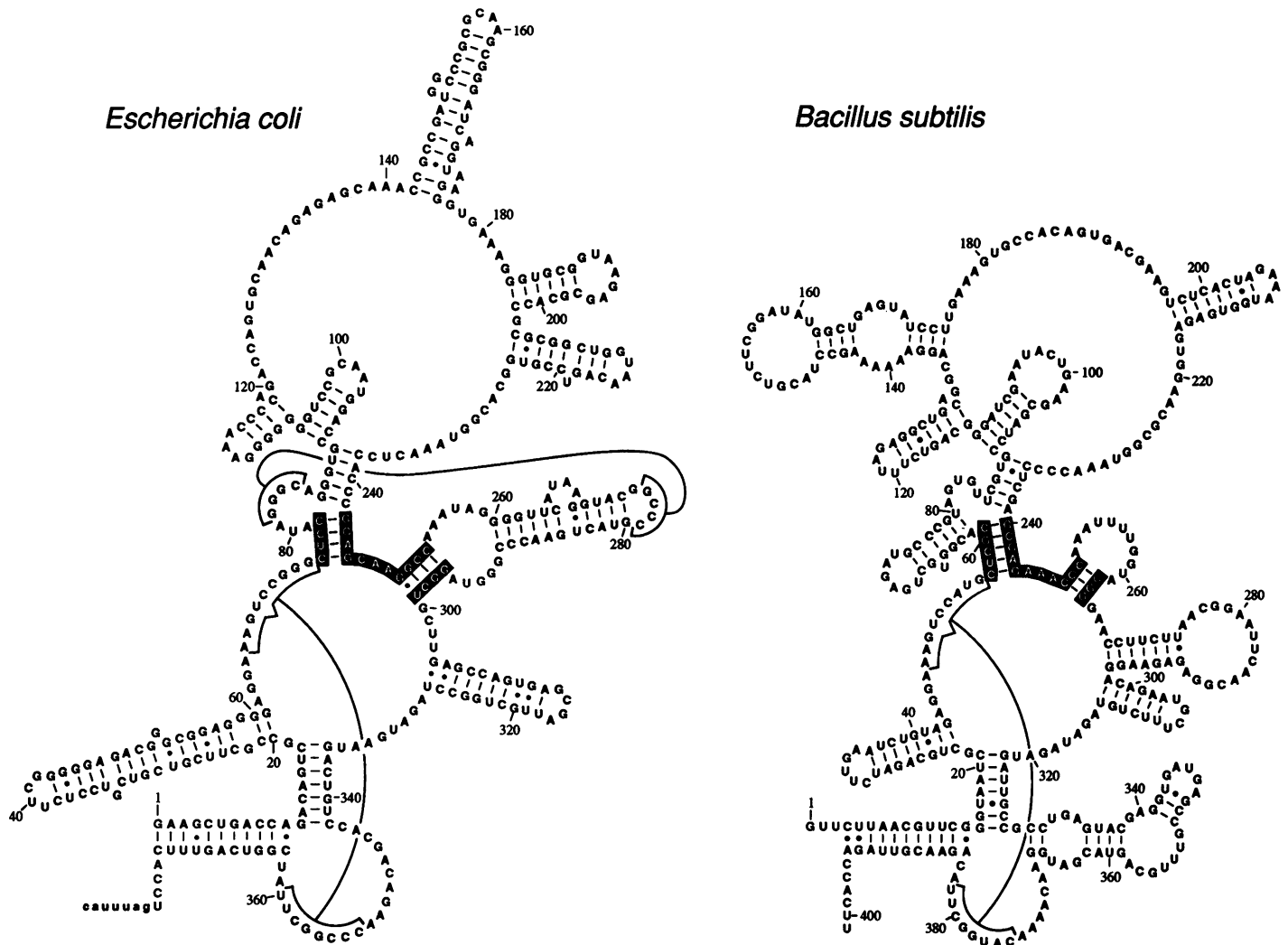
Figure 6. Secondary structures of *E.coli* and *B.subtilis* RNase P RNAs. The structures are based on phylogenetic comparative analysis (see text). Watson-Crick base-pairs are shown with lines (—), non-Watson-Crick base-pairs with filled circles (●). Lines and brackets indicate long-range pairings in the structures. Nucleotides shown in lower case at the 3' terminus of the *E.coli* RNA are additional nucleotides in the *in vitro* transcript used in this study which are not present in the native RNA. The nucleotides which have been realigned and restructured in the model (see text) are highlighted in black.

DISCUSSION

Structures of the RNase P RNAs

Although the new RNase P RNA sequences are only 42–55% identical to that of *E.coli*, their proposed secondary strutures are highly similar (Figs 1 and 2). Nevertheless, the RNAs have several unusual features. The *D.radiodurans* RNA contains a unique structure (nt 104–146) at the end of a stem that previously has been seen to vary only slightly in length. The stem-loop formed by residues 177–219 also is exceptionally long compared to other known RNase P RNAs. These elements render the *D.radiodurans* RNase P RNA the largest bacterial RNase P RNA so far identified. The RNase P RNAs from *Thermotoga* (96% similar to one another) are, on the other hand, the smallest bacterial RNase P RNAs so far identifed. The small sizes of these RNAs are primarily the result of the small sizes of hairpin elements. In addition, nucleotides corresponding to *E.coli* nucleotides 80, 81, 118, and 362 are absent from the *Thermotoga* RNAs.

The new sequences provide information which dictates the adjustment of the alignment of the RNase P RNA sequences from the proteobacteria and those of *Bacillus*, and the refinement of the secondary structure models of the RNAs (Fig. 6). The previous alignments were unsatisfactory in the region of *E.coli* helix 75–78/243–246, because the most similar sequences in this region were not directly coincident in the secondary structure model (5, 27). Although the misalignment between sequence and secondary structures was small (1 nucleotide position), the identification of homologous nucleotides in this region was questionable, suggesting that the model for the secondary structure in this region needed further refinement. This apparent incongruity between primary and secondary structure has now been resolved. The new RNase P RNA sequences, as well as that of the cyanobacterium *Anacystis nidulans* (28), provide convincing evidence for the pairing of the nucleotides corresponding to *E.coli* nucleotides 250 with 299. The length of the helix formed by *E.coli* nucleotides 250–253/296–299

Table 1. Base-pair occurences in bacterial RNase P RNAs
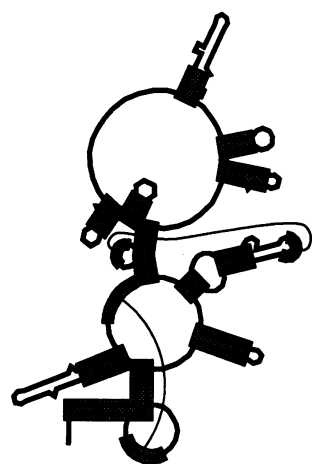
RNase P RNA(s)	Growth Topt (°C)	total %G+C	% within shaded helices *				
			G-C	A-U	G•U	A•C G•A	mismatched
Escherichia coli	37	61.8	67.4	20.7	8.7	2.2	1.1
proteobacteria (average)	26-37	62.6	67.9	20.9	7.8	2.4	1.1
Thermus aquaticus	70	68.9	76.1	15.2	6.5	1.1	1.1
Thermus thermophilus	75	71.4	80.4	13.0	3.3	2.2	1.1
Deinococcus radiodurans	30	68.9	72.8	18.5	6.5	1.1	1.1
Thermotoga maritima	80	67.7	77.2	16.3	4.3	1.1	0
Thermotga neapolitana	80	66.5	76.1	18.5	3.3	1.1	0

* Only base-paired regions of the RNA which are present in all of the organisms compared and for which homology can be clearly defined (90 pairings, shown diagramatically as shaded blocks in the line representation of *E. coli* RNase P RNA) were counted. The sequence from the proteobacterium *Alcaligenes eutrophus* was excluded from the analysis because of structural differences between it and the RNase P RNAs from other proteobacteria (13).

is therefore 4 base-pairs in all known RNase P RNAs except in *Bacillus* spp. and *A. nidulans*, in which the helix contains only 3 base-pairs. The length of the helix corresponding to *E.coli* nucleotides 75−78/243−246 is 4 base-pairs, except in the case of *Bacillus* spp. and *A. nidulans*, which contain an additional base-pair. The lengths of these helices must have changed in concert at least twice during evolution, because both helices are four base-pairs in length in the RNase P RNA of the high G+C Gram positive bacterium *Streptomyces bikiniensis* (15), which is more closely related to *Bacillus* than are the cyanobacteria. Covariation of the length of these helices is evidence that they form a single coaxial stack, because the sum of the lengths of the two helices is constant.

Evolutionary implications

The Thermotogales are the deepest known divergence in the evolutionary tree of Bacteria, and additionally seem to be the most primitive of the Bacteria (8). Because the last common ancestor of the Thermotogales and the other Bacteria gave rise to all of the known phylogenetic diversity in Bacteria, features common to the Thermotogales and other lineages of Bacteria must also have existed in this ancestral bacterium. The ability of the RNase P RNAs from *Thermotoga* to cleave precursor tRNAs in the absence of protein implies that this ability is ancestral in the Bacteria, and all modern bacterial RNase P RNAs examined retain this catalytic activity. The secondary structure of RNase P RNA is likewise highly conserved in the Bacteria; all organisms so far examined except *Bacillus* spp. retain the ancestral secondary structure of the RNA. The *Bacillus* RNAs retain the functional 'core' of the secondary structure (29), but have undergone deletion and addition of some helical elements. The fact that chimeric RNase P holoenzymes can be reconstituted from the thermophilic RNAs (including *T.maritima*) and *E.coli* protein implies that these RNAs are sufficiently similar in global structure

to correctly bind to *E.coli* protein and respond normally to its function, and further suggests that structurally and functionally similar protein components also exist in these species.

Inherent thermostability of RNase P RNAs from thermophiles

The RNase P RNAs from these thermophiles are remarkably resistant to thermal denaturation. Although factors such as protein binding and base modification (not known to occur in RNase P RNA) may assist in stabilization of structure *in vivo*, the thermostability of the RNase P RNAs from the thermophiles is to a large extent inherent in their nucleotide sequences. There are several features of the thermophilic RNAs that may be involved in thermal stability: *i*, an increase in the number of hydrogen bonds in helices; *ii*, a minimization of irregularities in helices; *iii*, additional Watson-Crick base pairs at the bases of stem-loops; *iv*, shortened connections between helices; and *v*, minimization of alternative foldings.

Increased G+C-richness and increased utilization of G-C base-pairs in the secondary structure are often invoked to explain thermophilicity of RNAs, and indeed both are apparent to a modest extent in the RNase P RNA sequences from the thermophilic bacteria (Table 1). More striking is the apparent bias *against* non-Watson-Crick base-pairs (G·U, G·A, and A·C) and mismatches. These differences are especially evident in the sequences from *T.maritima* and *T.neapolitana*. The scarcity of mismatches and non-Watson-Crick base-pairs in the thermophilic sequences is consistent with the hypothesis that increased hydrogen bonding in the secondary structure can confer thermostability.

Helices in the thermophilic RNAs contain fewer bulged nucleotides and other irregularities which might be destabilizing at high temperatures. Additionally, many of the bases which flank helices, at either the proximal or distal ends, clearly have the potential to form base-pairs in the thermophilic RNase P RNAs,

but these are not evident in the mesophilic RNAs. Examples of this are nucleotides 73/263, 173/193, 245−246/289−290 of the *T.aquaticus* sequence and nucleotides 62/240, 67/235, and 219−220/256−257 of the *T.maritima* sequence. If the bases flanking these various helices are indeed paired, it would suggest that the homologous bases in the mesophilic RNAs form non-canonical pairs or are stacked directly onto the end of the helices.

In addition to stability conferred by base-pairs, the tertiary structure of an RNA might be stabilized at high temperatures by decreasing to a minimum the lengths of connections between helices, thereby reducing the vibrational flexibility of the structure. Three regions of the *Thermotoga* RNAs contain shorter connections between helices than in other RNAs. The *Thermotoga* RNAs lack nucleotides in the regions corresponding to *E.coli* nucleotides 79−81 (1 nucleotide in *Thermotoga* rather than 3 or 4 as in other RNase P RNAs), 361−362 (1 nucleotide rather than 2 or 3), and the loop at the base of helix 260−290 (10 nucleotides in all, rather than 11 to >13). Although these features are not obvious in the *Thermus* RNAs, it is noteworthy that these same connections, with the exception nucleotides 361−362, are shorter in *Thermus* than in its closest mesophilic relative, *D.radiodurans*, with which comparison is most meaningful. The unexpectedly high requirement for ionic strength by the thermophilic RNAs may be a result of such minimal connections between regions of the molecule. The short connections may force some phosphates into particularly close proximity, resulting in electronegative repulsion that distorts the ribozyme in the absence of very high ionic strength to screen the repulsion. The fact that the RNAs from *T.aquaticus* and *T.maritima* respond to the *E.coli* RNase P protein implies that the protein also can serve as such an ionic screen in the thermophilic RNAs.

Still another feature of thermophilic RNAs that may be involved in thermostability is the minimization of alternative foldings. The number of folding alternatives of an RNA can be reduced in two ways: *i*, minimization of pairing possibilities; and *ii*, minimization of sequence length. The secondary structures of thermophilic RNAs in general are more accurately predicted by thermodynamic calculations, and such sequences are predicted (by computer algorithm) to have fewer alternative structures near the minimal free energy (30,31). It has already been shown that the RNase P RNA secondary structure from the moderately thermophilic *Bacillus stearothermophilus* is more accurately predicted, with fewer alternatives near the minimal free energy, than its mesophilic relatives (27). It is not surprising, therefore, that the *Thermus* and *Thermotoga* secondary structures also are more accurately predicted from fewer alternatives than those of *D.radiodurans* or *E.coli* (data not shown). The potential for formation of alternative foldings may limit the utility of increased G+C content for thermal stability. As the fraction of A+U residues decreases, the sequence complexity decreases and the number of alternative foldings increases. The G+C content of an RNA therefore might reach some optimum value, below 100%, trading-off maximization of hydrogen bonding with minimization of alternative foldings.

The other factor affecting the number of alternative foldings is sequence length; shorter sequences will have fewer competing alternative secondary structures. Indeed, the RNase P RNAs from *Thermotoga* are the shortest so far known (ca. 334 nucleotides as compared to 377 for *E.coli*). Helices which are phylogenetically variable in length are generally quite short in the *Thermotoga* sequences. Although the *Thermus* sequences are

comparable in length to the *E.coli* RNA, they are nonetheless much smaller than that of their mesophilic relative, *D.radiodurans*.

There may be fundamental differences in the ways that thermal stability is achieved by the *Thermus* and *Thermotoga* RNAs. The *Thermus* RNase P RNAs appear to rely primarily on strengthening the secondary structure of the RNA by the addition of base-pairs to the ends of helices, whereas the *Thermotoga* RNAs seem instead to minimize alternative foldings and the length of connections between helices. It is believed that the ancestors of the bacterial lineage and the last common ancestor of the three major phylogenetic lineages (Bacteria, Archaea and Eucarya) were thermophilic (8). Species from primitive lineages of the Bacteria that are comprised solely of thermophiles (such as the Thermotogales and green non-sulfur Bacteria) seem never to have adapted to mesophily; these organisms are 'ancestral' thermophiles. Thermophilic Bacteria related to mesophilic species (such as thermophilic Gram-positive Bacteria and *Thermus* spp.) are likely to have derived from mesophilic ancestors, and to have adapted more recently from mesophily to thermophily. If thermal stability in *Thermus* is indeed recently derived (a point that has yet to be well-established), such 'jury-rigging' of weak points in the RNA structure might be expected as the organism adapted to life at high temperatures. The Thermotogales, however, seem to have evolved only from thermophilic ancestors (8) and so the mechanisms for thermostability of RNase P RNA in these organisms are not adaptive. The global mechanisms outlined above that may be involved in the stability of the *Thermotoga* RNase P RNAs are consistent with an ancestral thermophilic character.

ACKNOWLEDGEMENTS

REFERENCES

1. Pace, N. R. and Smith, D. (1990) *J. Biol. Chem.*, **265**, 3587−3590.
2. Altman, S. (1990) *J. Biol. Chem.*, **265**, 20053−20056.
3. Darr, S. C., Brown, J. W. and Pace, N. R. (1992) *TIBS*, **17**, 178−182.
4. Woese, C. R., Kandler, O. and Wheelis, M. L. (1990) *Proc. Natl. Acad, Sci. USA.*, **87**, 4576−4579.
5. Brown, J. W. and Pace, N. R. (1992) *Nucl. Acids Res.*, **20**, 1451−1456.
6. Guerrier-Takada, C., Gardiner, K., Marsh, T. L., Pace, N. R. and Altman, S. (1983) *Cell*, **35**, 849−857.
7. Reich, C., Olsen, G. J., Pace, B. and Pace, N. R. (1988) *Science*, **239**, 178−181.
8. Woese, C. R., (1987) *Microbiol. Rev.*, **51**, 221−271.
9. Stackebrandt, E., Murray, R. G. E. and Trüper, H. G. (1988) *Int. J. Sys. Bacteriol.*, **38**, 321−325.
10. ATCC Catalog of Bacteria and Bacteriophages, 17th edition. (1989) Gherna, R., Pienta, M. S. and Cote, M. A., eds., pp 307.
11. Huber, R., Langworthy, T. A., König, H., Thomm, M., Woese, C. R., Sleytr, U. B. and Stetter, K. O. (1986) *Arch. Microbiol.*, **144**, 324−333.
12. Jannasch, H. W., Huber, R., Belkin, S. and Stetter, K. O. (1988) *Arch. Microbiol.*, **150**, 103−104.
13. Brown, J. W., Haas, E. S., James, B. D., Hunt, D. A, Liu, J. and Pace, N. R. (1991) *J. Bacteriol.*, **173**, 3855−3963.
14. Vogelstein, B. and Gillespie, D. (1979) *Proc. Natl. Acad. Sci., USA.*, **76**, 615−619.
15. Haas, E. S., Morse, D. P., Brown, J. W., Schmidt, F. J. and Pace, N. R. (1991) *Science*, **254**, 853−856.
16. Henikoff, S. (1984) *Gene*, **28**, 351−358.
17. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463−5467.

18. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
19. Kunkle, T. A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 488−492.
20. Reich, C., Gardiner, K. J., Olsen, G. J., Pace, B., Marsh, T. L. and Pace, N. R. (1986) *J. Biol. Chem.*, **261**, 7888−7893.
21. Waugh, D. S. (1989) Ph.D. Thesis, Indiana University.
22. Darr, S. C., Zito, K., Smith, D. and Pace, N. R. (1992) *Biochemistry*, **31**, 328−333.
23. Reed, R. E., Baer, M. F., Guerrier-Takada, C., Donis-Keller, H. and Altman, S. (1982) *Cell*, **30**, 627−636.
24. Waugh, D. S. and Pace, N. R. (1992) *FASEB J.* (in press).
25. Hartmann, R. and Erdmann, V. A. (1991) *Nucl. Acids Res.* **19**, 5957−5964.
26. Waugh, D. S. and Pace, N. R. (1990) *J. Bacteriol.*, **172**, 6316−6322.
27. Pace, N. R., Smith, D. K., Olsen, G. J. and James, B. D. (1989) *Gene*, **82**, 65−75.
28. Banta, A. B., Haas, E. S., Brown, J. W. and Pace, N. R. (1992) *Nucl. Acids Res.*, **20**, 911.
29. Waugh, D. S., Green, C. J. and Pace, N. R. (1989) *Science*, **244**, 1569−1571.
30. Zuker, M. (1989) *Science*, **244**, 48−52.
31. Haas, E. S., Brown, J. W., Daniels, C. J. and Reeve, J. N. (1990) *Gene*, **90**, 51−59.