

Comparative analysis of ribonuclease P RNA structure in Archaea

Elizabeth S. Haas, David W. Armbruster¹, Beverly M. Vucson, Charles J. Daniels¹ and James W. Brown*

Department of Microbiology, North Carolina State University, Raleigh, NC 27695, USA and ¹Department of Microbiology, The Ohio State University, Columbus, OH 43210, USA

Received December 26, 1995; Accepted February 12, 1996

GenBank accession nos[†]

ABSTRACT

Although the structure of the catalytic RNA component of ribonuclease P has been well characterized in Bacteria, it has been little studied in other organisms, such as the Archaea. We have determined the sequences encoding RNase P RNA in eight euryarchaeal species: *Halococcus morrhuae*, *Natronobacterium gregoryi*, *Halobacterium cutirubrum*, *Halobacterium trapanicum*, *Methanobacterium thermoautotrophicum* strains Δ H and Marburg, *Methanothermus fervidus* and *Thermococcus celer* strain AL-1. On the basis of these and previously available sequences from *Sulfolobus acidocaldarius*, *Haloferax volcanii* and *Methanosarcina barkeri* the secondary structure of RNase P RNA in Archaea has been analyzed by phylogenetic comparative analysis. The archaeal RNAs are similar in both primary and secondary structure to bacterial RNase P RNAs, but unlike their bacterial counterparts these archaeal RNase P RNAs are not by themselves catalytically proficient *in vitro*.

INTRODUCTION

Ribonuclease P (RNase P) is an endoribonuclease present in all cells that cleaves the leader sequences from precursor tRNAs to generate their mature 5'-ends (for reviews see 1–3). RNase P has been most well studied in Bacteria, in which an RNA of ~130 kDa (400 nt) and a protein of ~14 kDa (120 amino acids) make up the holoenzyme (4). The RNA is the catalytic moiety and enzymatic activity by the RNA in the absence of protein has been demonstrated *in vitro* from a wide phylogenetic range of Bacteria (5–7). RNase P from Archaea and Eucarya also contain an RNA component, but unlike their bacterial counterparts, enzymatic activity by these RNase P RNAs when separated from the other components of the holoenzyme or when synthesized *in vitro* has not been demonstrated (8,9).

Previous studies have shown significant biochemical diversity among the archaeal RNase P enzymes. RNase P from the halophilic Archaea *Haloferax volcanii* is sensitive to micrococcal nuclease and has a high buoyant density (1.61 g/cm³) in cesium sulfate gradients (10). The halophilic RNase P appears, therefore,

to be composed largely of RNA and thus resembles the bacterial enzyme. RNase P from *Sulfolobus acidocaldarius* (misclassified as *S.solfataricus* in Darr *et al.*; 11) is not sensitive to micrococcal nuclease, has a low buoyant density (1.27 g/cm³) in cesium sulfate and is much larger than the size of the RNA alone (9). RNase P from *S.acidocaldarius* is, therefore, predominantly protein and thus more closely resembles the eucaryal RNase P. The genes encoding the RNAs associated with RNase P RNA enzymes in *H.volcanii* (8), *S.acidocaldarius* (9) and *Methanosarcina barkeri* (D.W.Armbruster and C.J.Daniels, unpublished data, Genbank accession no. U42984) have been cloned and characterized. All three contain recognizable sequence similarity to one another and to bacterial RNase P RNAs, but not to the eucaryal RNAs. Nevertheless, these archaeal RNase P RNAs are not capable of catalysis in the absence of other components (presumed to be protein) of the holoenzyme.

Secondary structure models of the archaeal RNase P RNAs presented previously represent little more than the archaeal sequences 'forced' into the model developed for bacterial RNase P RNA secondary structure. In order to develop a secondary structure model based on archaeal sequences we have cloned the genes encoding RNase P RNA from eight additional species (see Fig. 1): *Halococcus morrhuae*, *Natronobacterium gregoryi*, *Halobacterium cutirubrum*, *Halobacterium trapanicum*, *Methanobacterium thermoautotrophicum* strain Δ H, *M. thermoautotrophicum* strain Marburg, *Methanothermus fervidus* and *Thermococcus celer* strain AL-1 and used these sequences in a phylogenetic comparative analysis of secondary structure.

MATERIALS AND METHODS

Nucleic acid purification

Organisms were grown in the following media: *N.gregoryi*, natronobacteria medium (ATCC medium 1590); *H.morrhuae*, halobacterium medium (ATCC medium 1270); *H.cutirubrum* and *H.trapanicum*, yeast medium (ATCC medium 112) containing 25% (w/v) NaCl; *M.thermoautotrophicum* strains Marburg and Δ H and *M.fervidus*, ER medium (12). All cultures of extreme halophiles were grown aerobically at 37°C with continuous shaking; methanogens were grown without shaking under 40 p.s.i. 60% H₂/40% CO₂ at 65 (*M.thermoautotrophicum*) or 80°C

* To whom correspondence should be addressed

[†]U42980, U42981, U42982, U42983, U42985, U42986, U42987 and U42988

control reactions. Internally ^{32}P -labeled precursor tRNA^{Asp} from *Bacillus subtilis* was synthesized from plasmid pDW128 (18).

Templates for the synthesis by T7 RNA polymerase of halophile RNase P RNAs were PCR amplification products from plasmid DNAs. Amplification was from pIBI31::MB (*H.volcanii*), pB-171 (*H.morrhuae*), p-28 (*N.gregoryi*), pRNA (*H.cutirubrum*) and pT7Trap (*H.trapanicum*) using primers T7 P-PCR 5' and P-PCR 3' (see above). PCR-generated templates were extracted with phenol/chloroform and treated with the Klenow fragment of *E.coli* DNA polymerase I in the presence of excess dNTPs prior to *in vitro* transcription using T7 RNA polymerase.

Partial RNase P transcripts (~1 µg) from the methanogens and *T.celer* were tested after reconstitution, by heating to 80°C and slow cooling, with 1 µg of an oligonucleotide designed to anneal to up- and downstream polylinker-derived sequences and the downstream strand of P2 to artificially generate secondary structure absent in the partial clones (see Fig. 2). The homologous *E.coli* partial RNA for use in control reactions was generated by transcription from a clone containing the partial sequence amplified from *E.coli* DNA using 59FBam and 347REco (the bacterial versions of A59FXba and A347RBam) (19).

Sequence analysis

Sequences were aligned manually using SeqApp (Don Gilbert, Indiana University). Comparative analysis of secondary structure was performed as previously described (21). Comparative support for secondary structure was identified using Covariation (22). Sequences derived from primers in clones obtained by PCR amplification were excluded from the analysis. The structures of regions unique to each sequence (P12, P15/P16 and P19) and therefore not amenable to comparative analysis were predicted thermodynamically using Mulfold (22). All sequences, alignments and secondary structures are available from the Ribonuclease P Database (<http://jwbrown.mbio.ncsu.edu/RNaseP/>). The Genbank accession nos for these sequences are U42980, U42981, U42982, U42983, U42985, U42986, U42987 and U42988.

RESULTS

Cloning RNase P RNA-encoding genes

A subclone of the *H.volcanii* RNase P RNA gene (8) was used as probe in Southern analyses to identify DNA fragments encoding RNase P RNA in genomic DNAs of *H.cutirubrum*, *H.morrhuae* and *N.gregoryi*. The *H.volcanii* probe hybridized to single fragments in each of several restriction digests of the genomic DNAs, indicating that RNase P RNA is encoded by a single copy gene in each of these halophiles, as is also the case for previously cloned archaeal genes and those of Bacteria (4). The DNAs containing the *H.morrhuae* (2.2 kb *SaI*I fragment), *H.cutirubrum* (7.6 kb *Sph*I fragment) and *N.gregoryi* (2 kb *Xma*I fragment) RNase P RNA-encoding genes were cloned and the nucleotide sequences of the RNase P RNA-encoding regions contained within these fragments were determined. The 5'- and 3'-terminal sequences of the *H.volcanii*, *H.morrhuae*, *H.cutirubrum* and *N.gregoryi* RNase P RNA-encoding genes are highly conserved and so these sequences were used to clone the gene from *H.trapanicum* genomic DNA by PCR amplification. Each of these cloned genes hybridized to RNAs of the expected sizes in Northern analyses and to the corresponding single DNA fragments of the source genomic DNA (data not shown), confirming

that the cloned fragments encode functional RNase P RNAs. Weak hybridization in Northern analyses to additional larger RNAs in the case of *H.trapanicum* (data not shown) suggests that this RNA may be synthesized as a larger precursor molecule, as occurs in *E.coli* (23).

The RNase P RNA-encoding genes from *T.celer*, *M.thermoautotrophicum* strains Marburg and ΔH and *M.fervidus* were cloned by PCR amplification using primers complementary to highly conserved sequences near the 5'- and 3'-termini of archaeal and bacterial RNAs. Analogous primers have previously been used to obtain bacterial RNase P RNA-encoding sequences (19,24). In each case PCR resulted in the amplification of a single band in the predicted size range (~300 bp). Each of the cloned genes hybridized to single DNA fragments of the source genomic DNA (data not shown) in Southern analyses and phylogenetic trees of the sequences are consistent with those based on 16S rRNAs (data not shown).

Comparative analysis of secondary structure

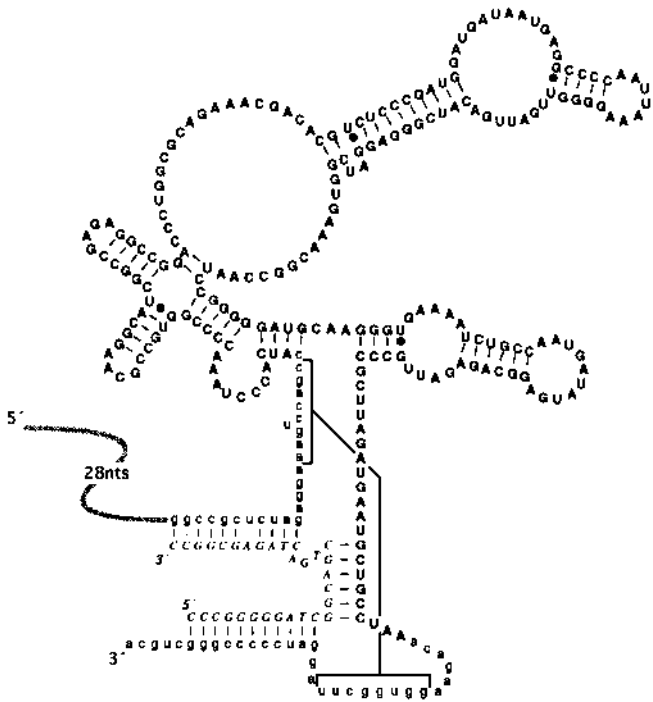
The archaeal sequences were analyzed by the comparative method (20) to construct a model for the secondary structure of archaeal RNase P RNA. Compensatory changes were first identified amongst the sequences from the extreme halophiles, which are sufficiently similar over the majority of their lengths to be readily aligned on the basis of sequence similarity. This initial secondary structure information was used to aid in addition of the methanogen *T.celer* and *S.acidocaldarius* sequences to the alignment. Several rounds of alignment/structure refinement were performed, resulting in the construction of individual and consensus secondary structures (Figs 2 and 3). The consensus secondary structure is supported by compensatory changes at at least two positions (the generally accepted indication that a helix exists) in all helices except P5 and P10. P5 is supported by co-variation at a single base pair (U-A in *S.acidocaldarius*, A-U in all other cases). Evidence for or against the presence of P10 is lacking; all four of the nucleotides in this helix are invariant in the archaeal sequences available. Of 75 base pairings in the consensus secondary structure 51 are individually supported by compensatory substitutions of the paired nucleotides.

Although the secondary structure is for the most part well defined, the region distal to P12 is extremely variable in length and sequence (even between close relatives) and so no particular structure in this region is well supported. These structures were therefore predicted thermodynamically (22), but it should be realized that no comparative evidence for these particular structures is available. The region distal to P16 is also variable in sequence length, although less dramatically than in the case of P12, and structure in this region was likewise predicted thermodynamically, within the constraints of the comparative data. The *M.fervidus* RNA lacks Watson-Crick complementarity in the nucleotides expected to form P6; this helix is well supported amongst the remaining sequences, so this seems to be an idiosyncrasy of this RNA.

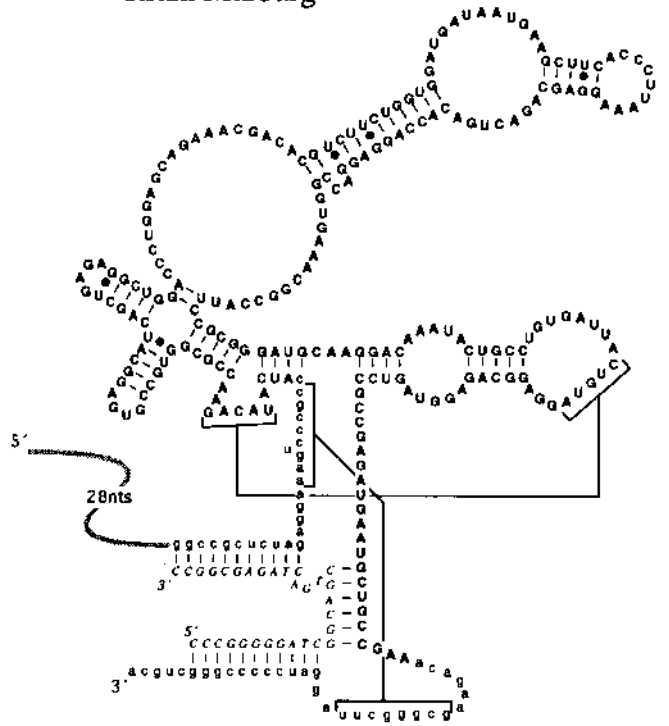
RNase P RNA alone enzyme assays

The previously characterized RNase P RNAs from *H.volcanii*, *Methanosarcina barkeri* and *S.acidocaldarius* do not have the ability to cleave precursor tRNAs in the absence of other enzyme components (8,9; D.W.Armbruster and C.J.Daniels, unpublished data, Genbank accession no. U42984). The products of the RNase P RNA-encoding genes from *H.cutirubrum*, *H.morrhuae*,

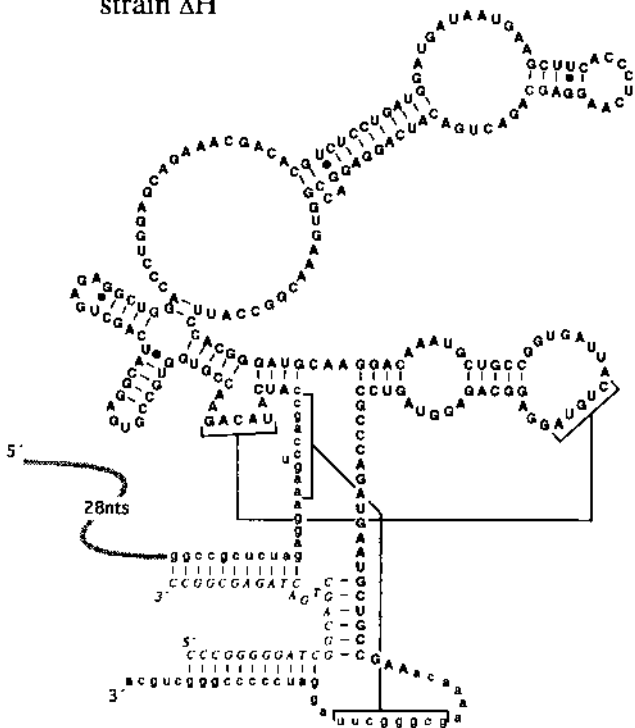
Methanothermus fervidus



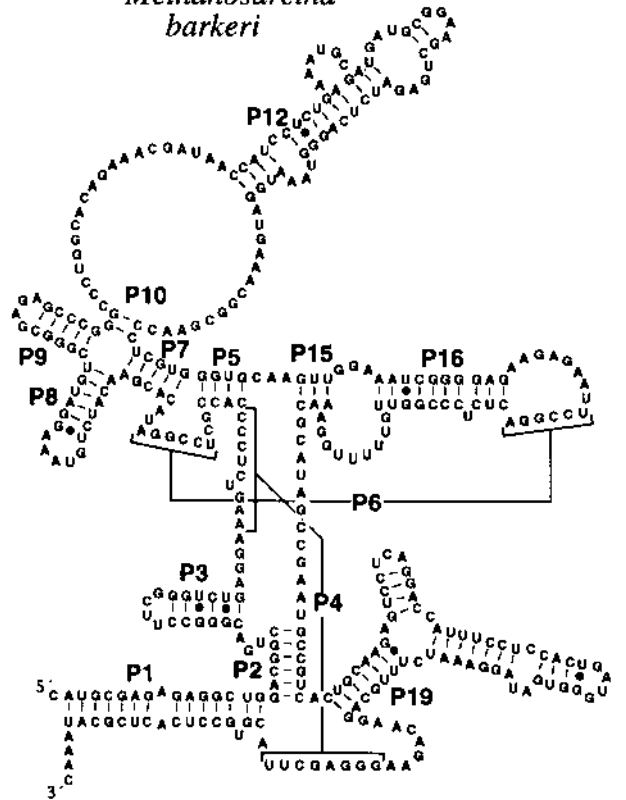
Methanobacterium thermoautotrophicum strain Marburg



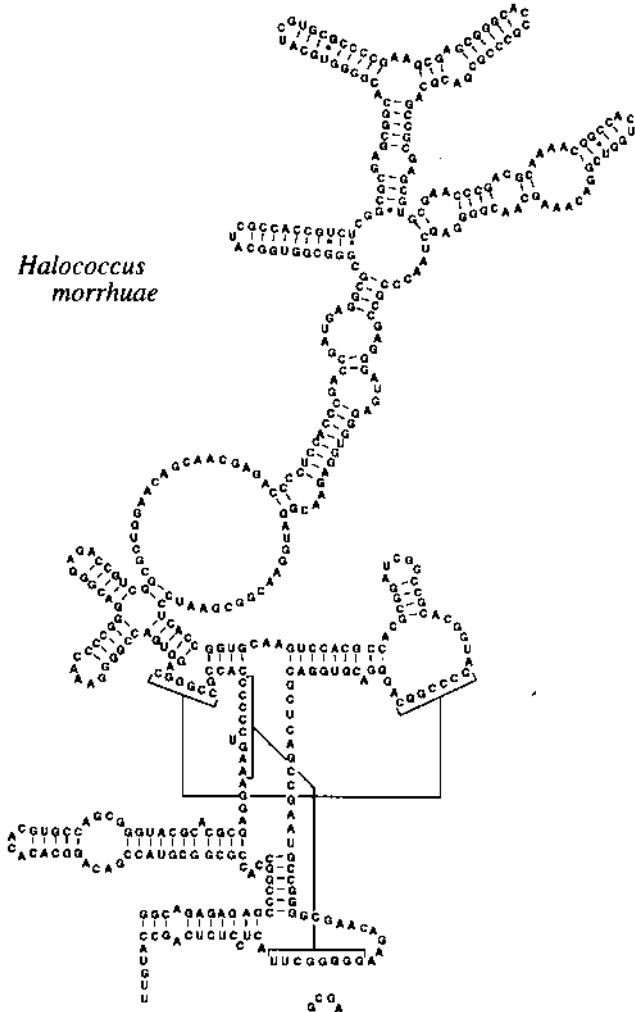
Methanobacterium thermoautotrophicum strain ΔH



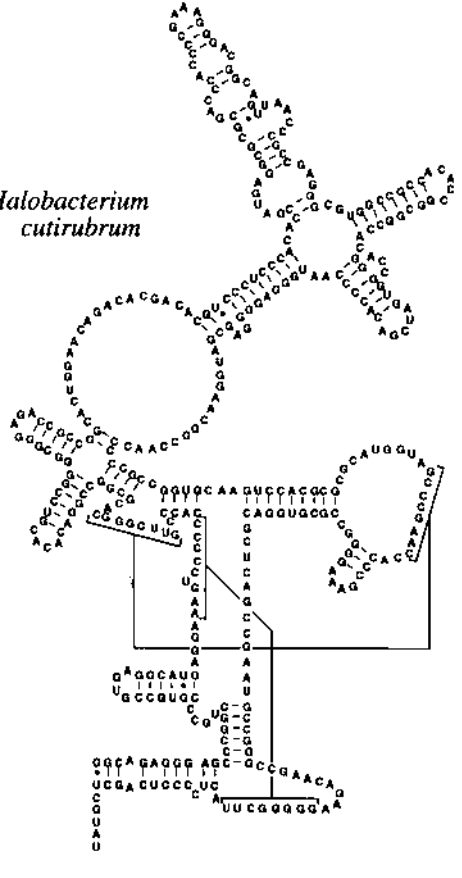
Methanosarcina barkeri



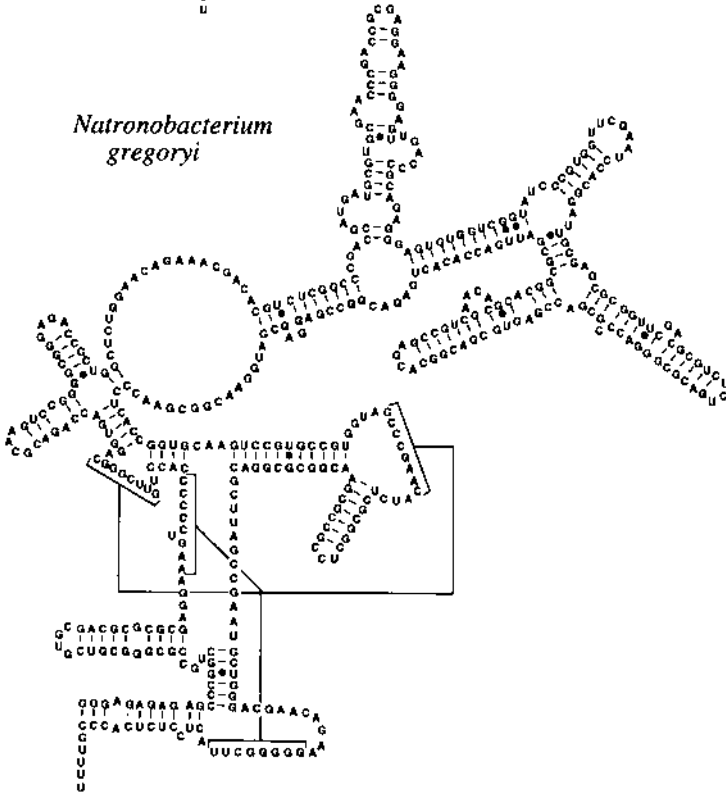
Halococcus morrhuae



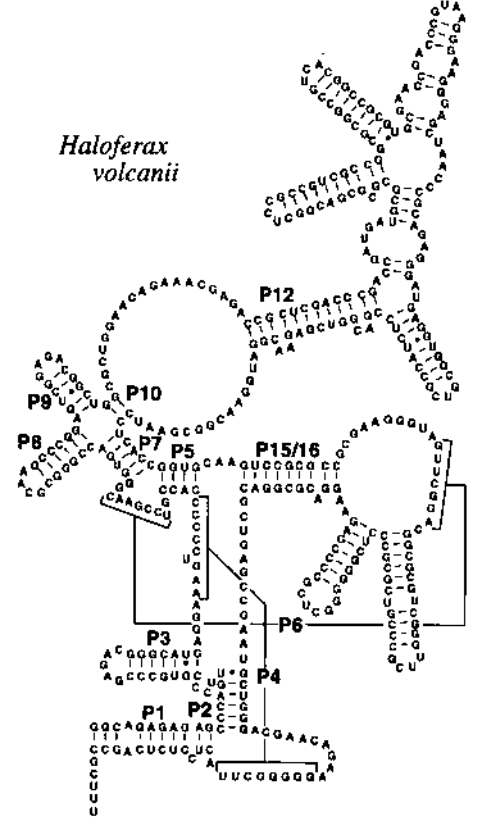
Halobacterium cutirubrum



Natronobacterium gregoryi



Haloferax volcanii



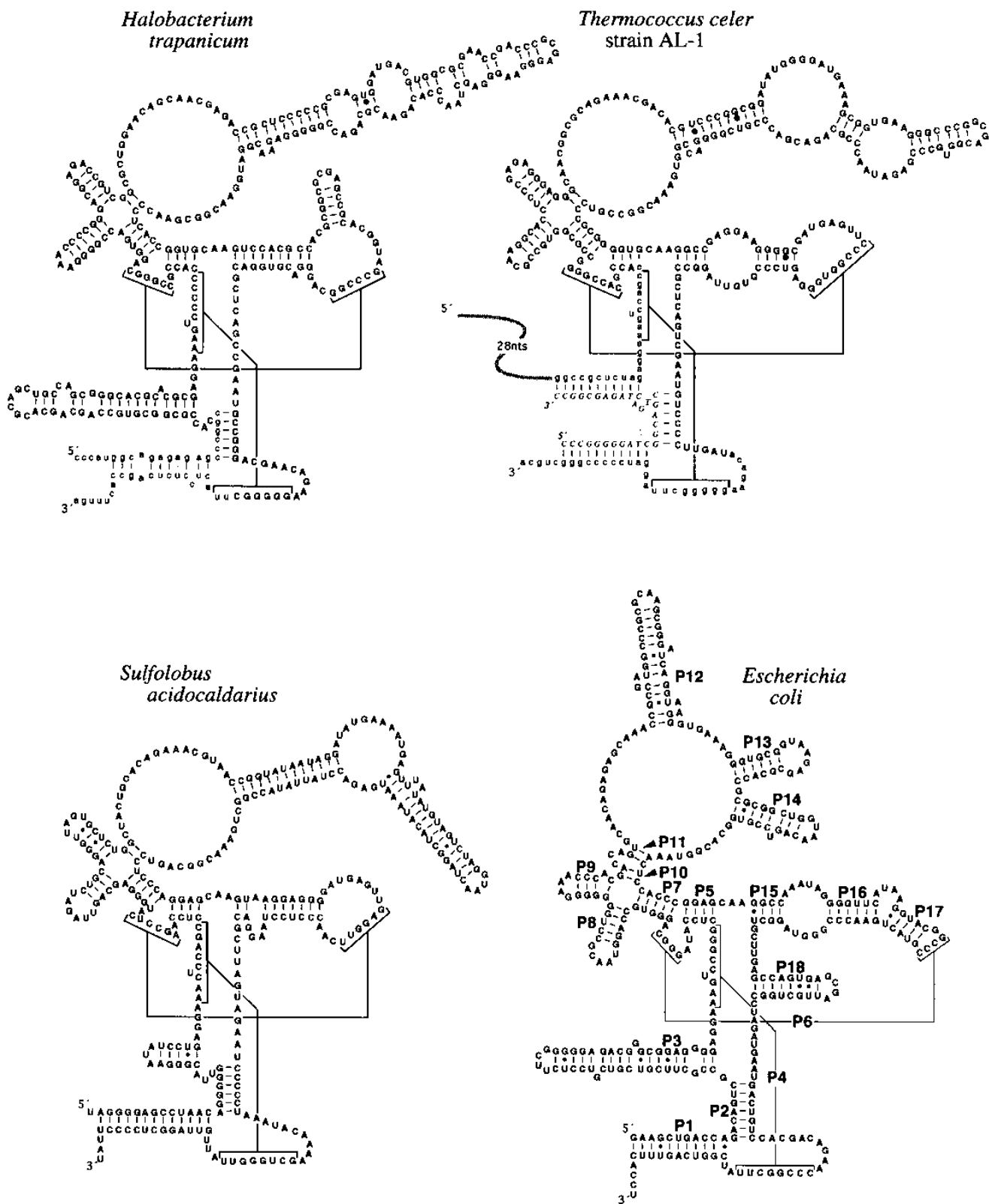


Figure 2. Proposed secondary structures of archaeal RNase P RNAs. The secondary structures were determined by comparative analysis (see text). Helices nomenclature is presented on the *M.barkeri*, *H.volcanii* and *E.coli* structures (shown as an example bacterial RNA for comparison). Sequences in lower case (P1-3) of some of the RNAs are derived from the oligonucleotide primers used to obtain the gene using PCR amplification, as well as flanking vector-derived sequences present in the *in vitro* synthesized transcripts used in the analysis of catalytic function. The oligodeoxynucleotides used in these experiments to regenerate 'missing' structure are shown in italic. The location of the 5'- and 3'-terminii of the *H.cutirubrum*, *H.trapanicum*, *H.morrhuae* and *N.gregoryi* RNAs are predicted based on those determined experimentally for *H.volcanii* and represent the terminii of the RNAs tested *in vitro* for catalytic activity.

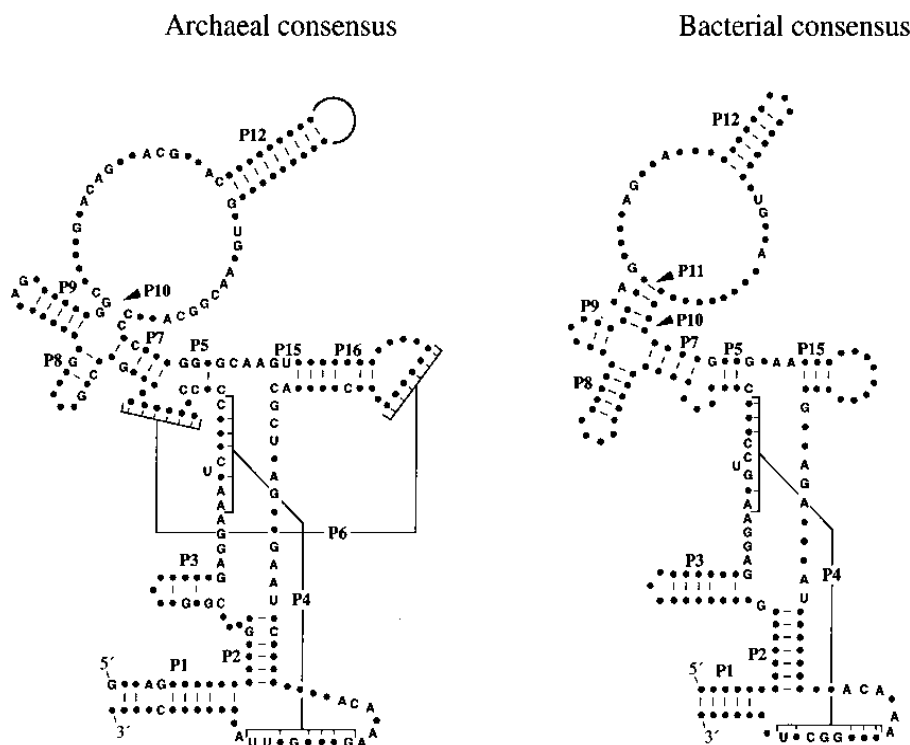


Figure 3. Archaeal and bacterial consensus RNase P RNAs. Consensus structures include only those nucleotides which are present in all of the available RNase P RNA sequences for each group. Invariant nucleotides are indicated by letter (G, A, U or C); nucleotides that are universally present but vary in identity are indicated by filled circles. Only the base pairings for which specific comparative evidence exists are marked by short lines. Helices are labeled according to Haaset *al.* (7). The archaeal and bacterial consensus structures are derived from 11 and 116 RNase P RNA sequences, respectively.

H. trapanicum and *N. gregoryi*, synthesized *in vitro* using T7 RNA polymerase, were tested for the ability to cleave precursor tRNA^{Asp} from *B. subtilis*. Assays were performed at 37°C in standard assay buffer (18) at 'low' (1 M ammonium acetate; optimal for the *E. coli* or *B. subtilis* RNAs) and 'high' (3 M ammonium acetate; required by many mutant RNAs) ionic strength. No specific cleavage of the substrate by the halophilic RNase P RNAs could be detected, even upon extended incubation (24 h) and at an equimolar RNase P RNA:substrate ratio (20 nM each) (data not shown).

In vitro transcripts of partial bacterial RNase P RNA-encoding genes cloned after amplification using 59F and 347R PCR primers (19,24) can be functionally reconstituted using an oligonucleotide that anneals to the 3'-strand of P2 and the upstream and downstream vector-derived transcribed sequences; these oligonucleotides are designed to regenerate the secondary structure absent in these partial RNAs (unpublished data). However, no catalytic activity could be detected from analogous transcripts of the methanogens or *T. celer* annealed with the appropriate oligonucleotides under a variety of reaction conditions (data not shown).

DISCUSSION

The previous proposal for the secondary structure of an archaeal RNase P RNA, that of *H. volcanii*, was constructed essentially by 'forcing' the archaeal sequence into the model for bacterial RNase P RNA structure; no other archaeal sequences were then available for comparative analysis (8). In order to provide the information required to develop a model for RNase P RNA

secondary structure in Archaea we have determined the sequences of the genes encoding RNase P RNA from eight members of the Euryarchaea: four species of extreme halophiles, three species of methanogens and one sulfur-metabolizing thermophile. These sequences and those of *H. volcanii*, *S. acidocaldarius* and *M. barkeri* were analyzed by the comparative method (20) and a model for the secondary structure of archaeal RNase P RNA was developed independently of the bacterial model.

In a comparative analysis of secondary structure potential helices in an RNA molecule (i.e. complementary sequences) are tested by examination of the equivalent nucleotides in homologous RNAs with different sequences. The occurrence of sequence differences that nevertheless maintain complementarity (i.e. compensatory changes) is evidence for the existence of a helix. The consensus archaeal RNase P RNA secondary structure (Fig. 3) is well supported by the comparative data. Of the 13 helices present in the consensus model only two (P5 and P10) are not supported by compensatory changes in two or more base pairings in the helix, the generally accepted indication that a helix exists. In fact, of the proposed base pairings in the consensus model 68% are specifically supported by compensatory sequence changes. There are several regions of significant sequence length variation in the archaeal RNase P RNAs; these correspond primarily to length variations in P3 and structural variation in P12, P19 and the joining regions between P15/16 and P6. Variation in these regions is not surprising, given that this is likewise seen in the homologous regions in bacterial RNase P RNAs, however, the degree of variation in P12 amongst the RNAs from the halophiles is striking given their relatively close evolutionary affiliation.

The primary sequences and secondary structures of the archaeal RNase P RNAs largely resemble their bacterial counterparts and those nucleotides which are most conservative in the bacterial RNAs are also generally conserved in the archaeal sequences. Despite these similarities, the archaeal RNAs are unlike their bacterial counterparts in that they are not capable of catalysis in the absence of other enzyme components. The structural basis of this deficiency must reside in differences in the bacterial and archaeal sequences or higher order structure. The archaeal consensus secondary structure contains all but one helix present in the bacterial consensus structure, the 2 bp helix P11. This structure is adjacent to a large region in which the structure remains unknown even in bacterial RNase P RNAs. As in Bacteria, the presence of P19 is sporadic, occurring only in the *M.barkeri* RNA amongst the archaeal RNAs.

The structure of the P15/16 region also varies amongst the archaeal RNAs (Fig. 2), despite the conservation of this region in Bacteria, in which it has been implicated in recognition of the substrate pre-tRNA 3'-terminus. In the halophile RNAs this region has the potential to form a continuous helix of Watson-Crick (and G-U) base pairs. However, in the methanogens and *T.celer* P15 and P16 are separated by an internal loop that conforms to the sequence and structure consensus of the homologous region in the bacterial RNAs. Because of its role in substrate recognition (at least in Bacteria; 25,26), it is reasonable to imagine that structural divergence in this region, like that seen in some of the Archaea, might be responsible for the absolute dependence of these RNAs on the other (presumably protein) components of the holoenzyme for catalytic activity. However, because even the methanogen RNase P RNAs apparently lack catalytic proficiency *in vitro*, it seems unlikely that this region is entirely responsible for this 'defect' in the archaeal RNAs.

P18, a conservative element in bacterial RNase P RNAs, is entirely absent in the archaeal structures. However, this element is also absent in the RNase P RNAs of *Chlorobium tepidum* and *Chlorobium limicola*, two species of 'green sulfur Bacteria'. The *Chlorobium* RNase P RNAs are nevertheless active in the absence of protein and deletion of P18 from the *E.coli* RNA affects the optimal ionic strength, but not the kinetic properties, of that RNA (7). Bacterial sequences lacking the usual 8 bp/GNRA loop form of P18 vary in P8 structure, which is otherwise highly conserved; this may, at least in part, be attributed to a tertiary interaction between the P18 GNRA loop sequence and P8 (24). Archaeal P8 structure is also unusual, varying in both length and loop structure, in accordance with the absence of P18.

There are also fine scale differences between conserved structures in the archaeal and bacterial RNase P RNAs. P2, which is invariantly 7 bp in length in Bacteria (not including an unproven G-U pair possible in all cases), is 6 bp in the Archaeal RNAs (in which the additional pair is never possible). Perhaps associated with this change is the difference in the length of the P2/P3 joining region (J2/3), which is invariantly 1 nt in Bacteria (always G) but 3 or 4 in Archaea. In addition, in bacterial RNAs lacking P13/14, P12 lacks an otherwise highly conserved AA bulge 2 bp from the base of the helix. The euryarchaeal RNAs, which also lack P13/14, are also interrupted near the base of the helix.

RNase P enzymes from all organisms share common ancestry; fundamental features of the structure and biochemistry should

therefore be preserved in all of its modern forms. Because the RNA components of archaeal RNase Ps are not capable of catalysis and yet are sufficiently similar to those of Bacteria for the identification of homologous sequences and higher order structures, they provide an important avenue for the investigation of essential functions and structures in this RNA enzyme.

ACKNOWLEDGEMENTS

We thank N.R.Pace for helpful discussion during these investigations. This work was supported by NIH grant 1-R29-GM52894-01 to JWB and NIH grant R01-GM48665-03 to CJD. BMV was supported by a North Carolina Agriculture Foundation graduate fellowship. CJD is an associate of the Canadian Institutes for Advanced Research.

REFERENCES

- 1 Pace,N.R. and Brown,J.W. (1995) *J. Bacteriol.*, **177**, 1919–1928.
- 2 Altman,S., Kirsebom,L. and Talbot,S. (1993) *FASEB J.*, **7**, 7–14.
- 3 Darr,S.C., Brown,J.W. and Pace,N.R. (1992) *Trends Biochem. Sci.*, **17**, 178–182.
- 4 Brown,J.W. and Pace,N.R. (1992) *Nucleic Acids Res.*, **20**, 1451–1456.
- 5 Guerrier-Takada,C., Gardiner,K., Marsh,T., Pace,N. and Altman,S. (1983) *Cell*, **35**, 849–857.
- 6 Brown,J.W., Haas,E.S. and Pace,N.R. (1993) *Nucleic Acids Res.*, **21**, 671–679.
- 7 Haas,E.S., Brown,J.W., Pitulle,C. and Pace,N.R. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 2527–2531.
- 8 Nieuwlandt,D.T., Haas,E.S. and Daniels,C.J. (1991) *J. Biol. Chem.*, **266**, 5689–5695.
- 9 LaGrandeur,T.E., Darr,S.C., Haas,E.S. and Pace,N.R. (1993) *J. Bacteriol.*, **175**, 5043–5048.
- 10 Lawrence,N., Wesolowski,D., Gold,H., Bartkiewicz,M., Guerrier-Takada,C., McClain,W.H. and Altman,S. (1987) *Cold Spring Harbor Symp. Quant. Biol.*, **52**, 233–238.
- 11 Darr,S.C., Pace,B. and Pace,N.R. (1990) *J. Biol. Chem.*, **265**, 12927–12932.
- 12 Hook,L., Corder,R.E., Hamilton,P.T., Frea,J.I. and Reeve,J.N. (1984) In Strohl,W.R. and Touvinen,O.H. (eds), *Microbial Chemoautotrophy*. Ohio State University Press: Columbus, OH. pp. 275–289.
- 13 Brown,J.W., Haas,E.S., James,B.D., Hunt,D.A. and Pace,N.R. (1991) *J. Bacteriol.*, **173**, 3855–3863.
- 14 Saiiki,R.K., Gelfand,D.H., Stoffel,S., Scharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B. and Erlich,H.A. (1988) *Science*, **239**, 487–491.
- 15 Vogelstein,B. and Gillespie,D. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 615–619.
- 16 Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
- 17 Henikoff,S. (1984) *Gene*, **28**, 351–359.
- 18 Waugh,D. (1989) PhD Thesis, Indiana University, Bloomington, IN.
- 19 Haas,E.S., Brown,J.W., Pitulle,C. and Pace,N.R. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 2527–2531.
- 20 James,B.D., Olsen,G.J., Liu,J. and Pace,N.R. (1988) *Cell*, **52**, 19–26.
- 21 Brown,J.W. (1991) *CABIOS*, **7**, 391–393.
- 22 Zuker,M., Jaeger,J.A. and Turner,D.H. (1991) *Nucleic Acids Res.*, **19**, 2707–2714.
- 23 Lawrence,N.P., Richman,A., Amini,R. and Altman,S. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 6825–6829.
- 24 Brown,J.W., Nolan,J.M., Haas,E.S., Rubio,M.A., Major,F. and Pace,N.R. (1996) *Proc. Natl. Acad. Sci. USA*, in press.
- 25 Oh,B.-K. and Pace,N.R. (1994) *Nucleic Acids Res.*, **22**, 4087–4094.
- 26 Hardt,W.D., Schlegel,J., Erdmann,V.A. and Hartmann,R.K. (1995) *J. Mol. Biol.*, **247**, 161–172.
- 27 Woese,C.R., Kandler,O. and Wheelis,M.L. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 4576–4579.
- 28 Woese,C.R. (1987) *Microbiol. Rev.*, **51**, 221–271.