Nucleic Acids Research, Vol. 20, No. 7 1451–1456

Ribonuclease P RNA and protein subunits from bacteria

James W.Brown and Norman R.Pace*
Department of Biology and Institute for Molecular and Cellular Biology, Indiana University, Bloomington, IN 47405, USA

Received January 13, 1992; Revised and Accepted February 13, 1992

INTRODUCTION

Ribonuclease P (RNase P) is the site-specific endoribonuclease that cleaves leader sequences from pre-tRNAs (see 1, 2 for reviews). Cleavage by RNase P absolutely requires divalent cation, preferably Mg^{++} (3), and produces 3′-hydroxyl and 5′-phosphate ends (4). Although RNase P is thought to be present in all living cells, the enzyme is best understood in the bacteria *Escherichia coli* and *Bacillus subtilis*. In these organisms, RNase P is composed of two subunits, a large (ca. 400 nucleotides) RNA encoded by the *rnp*B gene (5,6) and a small protein (119 amino acids) encoded by the *rnp*A gene (7). Although bacterial RNase P functions as a ribonucleoprotein *in vivo*, the RNA alone is an accurate and efficient catalyst under the appropriate conditions *in vitro* (8). Bacterial RNase P RNA is the only naturally occurring RNA which is known to act catalytically in the sense that each RNA molecule cleaves many substrate molecules.

The purpose of this article is to present the aligned sequences of the currently available bacterial RNA and protein components of RNase P, and to provide an overview of RNase P structure in all organisms. For the purposes of this discussion, the nomenclature for the deepest phylogenetic branches proposed by Woese *et al.* (9) has been adopted: 'Bacteria' refers to organisms previously known as eubacteria, 'Archaea' to the archaebacteria, and 'Eucarya' to the eucaryotic nuclear lineage. We use the term 'homology' in it's strictest sense: homologous sequences are of common ancestry and function.

RNase P in Bacteria

The RNA component of bacterial RNase P varies in size among known examples from ca. 338 to ca. 444 nucleotides in length (the lengths are approximate because the mature 5′ and 3′ ends of most RNase P RNAs have not been precisely determined). The sequences of RNase P RNA-encoding genes currently are available from five of the eleven major phylogenetic branches (10) of Bacteria: purple bacteria and relatives (including *E. coli*) (11–16), Gram positive bacteria (including *B. subtilis*) (13,17), cyanobacteria (18), deinococci and relatives (including *Thermus*) (19–21), and *Thermotogales* (19,20). All of these RNAs have been shown to be catalytically active *in vitro* (12). A secondary structure for bacterial RNase P RNA has been proposed on the basis of comparative analyses of these sequences (Figure 1) (12,13,19). The RNase P RNA sequences from representatives of different bacterial groups are highly diverse in sequence (only 35–55% identical), yet the basic secondary structures of the RNAs are strikingly similar. Most of the structural variation, except in the case of the *Bacillus* RNA, is length-variation in helical regions of the molecule (22). The *Bacillus* structures are significantly altered with respect to the remaining RNase P RNAs; several large insertions and deletions have occurred in otherwise conservative regions of the molecule. It has been hypothesized that at least one of the novel insertions in the *Bacillus* sequences may compensate structurally for the loss of a long-range pairing elsewhere in the secondary structure (19). Nevertheless, the phylogenetically preserved regions of the molecule constitute a 'core' structure (Figure 1), a synthetic version of which is catalytically active *in vitro* (23).

The protein component of bacterial RNase P is less understood. RNase P protein gene sequences are available from only two species of purple bacteria and three species of Gram positive bacteria (7,24–27). Although important *in vivo*, the protein component can be supplanted *in vitro* by high salt concentrations (8,28). In the holoenzyme, the role of the protein may be as an electrostatic 'screen', allowing the highly negatively charged RNA enzyme-substrate complex to fold into the catalytic conformation under physiological (low salt) conditions (29). The abundance of charged residues in RNase P proteins is similar to that of many ribosomal proteins, to which RNase P protein, in a loose sense, may be analogous.

RNase P in Eucarya

The structure of RNase P from eucaryotic nuclei has been examined only in animals (*Xenopus* and human) (30,31) and fungi (budding and fission yeasts) (32–34). This is a relatively limited sampling of eucaryal diversity (35). These enzymes, like those of the bacteria, function as ribonucleoprotein complexes. Sensitivity of the enzymes to micrococcal nuclease indicates that the RNAs are essential components, yet none of the isolated RNAs has been shown to be catalytically active. No convincing similarity between the bacterial and eucaryal RNase P RNAs has been identified (but see reference 36). The variability of the RNA sequences within the few eucarya so far examined is sufficiently great that the secondary structures of these RNAs have yet to be determined; it is not yet clear that these RNAs in fact are homologous with each other, or to the RNase P RNAs of bacteria.

* To whom correspondence should be addressed

RNase P in organelles

It might be expected that the RNase P of mitochondria and chloroplasts would resemble that of their bacterial ancestors, as do the components of the translational machinery. The expected similarity is not yet apparent, however. Several mitochondrial RNase Ps have been shown to contain essential RNA and protein components, but the yeast RNA is best-characterized (37,38). That RNA is a product of the mitochondrial genome, whereas the protein component is derived from the nucleus (38). The RNA is not catalytically active in the absence of protein (37,38). The RNA components vary widely in size (from 140nt to at least 490nt) and they are very low in G+C content (ca. 23% G+C), as is the mitochondrial genome (39,40). The extremely high A+U content of these RNAs complicates the analysis of their structure because homologous sequences cannot be clearly identified. Consequently, the use of comparative analysis for identifying secondary structure has been limited. Because of the low sequence complexity, it seems unlikely that the RNA alone contains all the information required for its correct folding. Nonetheless, the few blocks of G+C-containing sequences in the mitochondrial RNase P RNA resemble the most highly conserved sequences in the bacterial RNAs (40). Although the mitochondrial RNase Ps are inactivated by treatment with micrococcal nuclease (37,38), the enzyme from *S. cerevisiae* mitochondria does not require intact RNase P RNA for activity. The RNA is fragmented during purification and some fragments appear to be entirely absent from purified, active enzyme (41).

The RNase P of chloroplasts may be unique in the lack of an RNA component. An RNA has not been found in the chloroplast RNase P, nor is the enzyme activity sensitive to treatment with micrococcal nuclease (42). The lack of an RNA component also is suggested by the low density of the enzyme, about that of protein-alone, in Cs_2SO_4 bouyant-density gradients. Moreover, computer (see below) and Southern hybridization (18) searches have failed to identify sequences resembling the bacterial RNase P RNA in the chloroplast genome. If confirmed, the apparent lack of an RNA component in chloroplast RNase P would be surprising; the RNase P of *Anacystis nidulans*, a cyanobacterium (to which chloroplasts are specifically related), contains a conventional bacterial-type, catalytic RNA (18). Moreover, the RNase P of *Sulfolobus* (see below) also has the qualities of resistance to nuclease-digestion and low density, suggesting the absence of RNA, yet it nonetheless contains an RNA element (43,44).

RNase P in Archaea

The RNase P holoenzymes of the thermophile *Sulfolobus* and of the extremely halophilic Archaea differ significantly in biochemical properties. The enzyme from *Sulfolobus* is resistant to micrococcal nuclease treatment and of low density (43,44), whereas the RNase P from *Haloferax volcanii* is sensitive to nuclease and dense in Cs_2SO_4 gradients (45,46). Nonetheless, both enzymes contain RNA components with striking sequence and secondary structure similarities to the bacterial RNase P

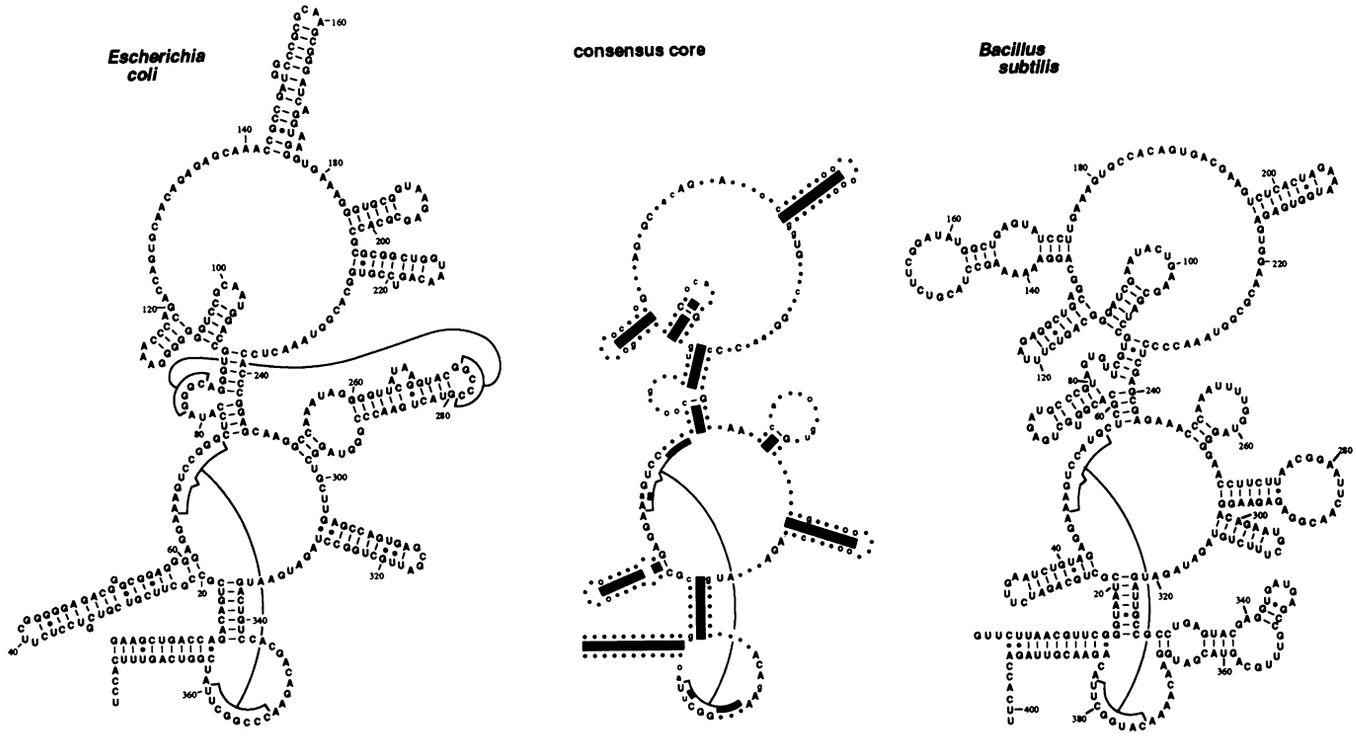
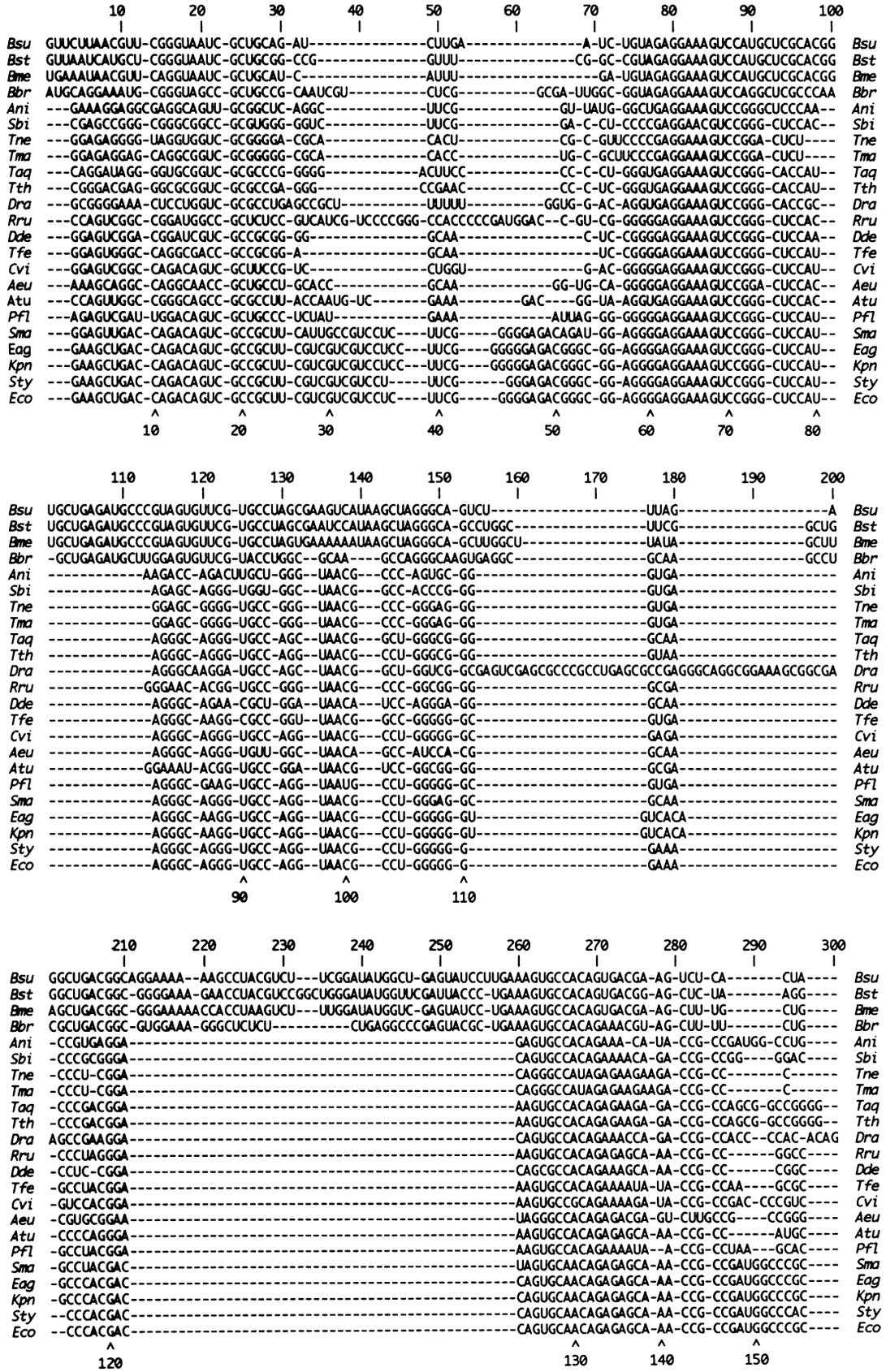
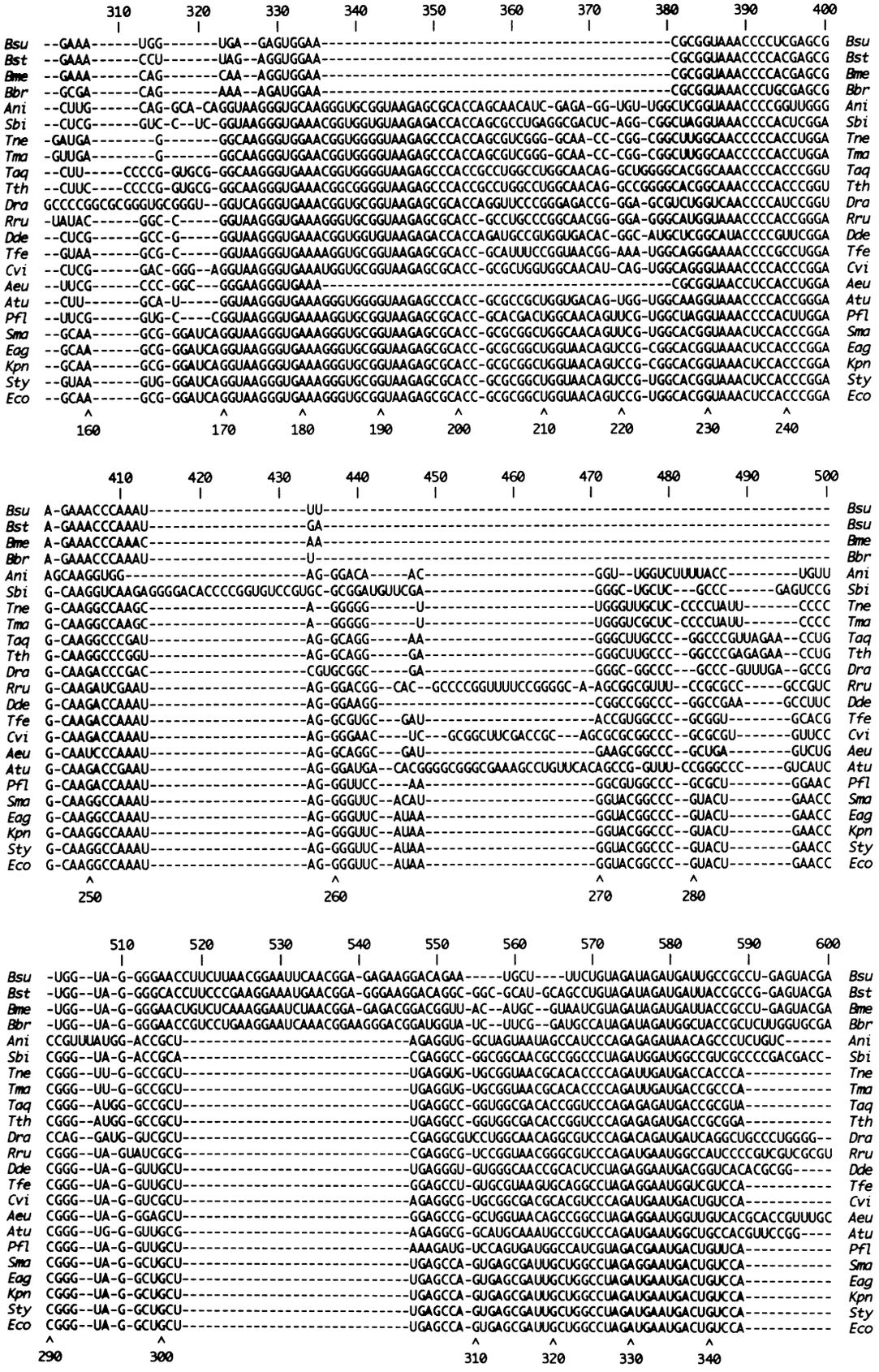


Figure 1. Secondary structure of RNase P RNA. Secondary structures of the RNase P RNAs of *E. coli* and *B. subtilis*, and a consensus core structure, are based on phylogenetic comparisons (12,13,19). The helices formed from nucleotides 66–74/353–360 and 82–85/276–279 (*E. coli* numbering) complete two pseudoknots in the secondary structures. Absolutely conserved nucleotides in bacterial RNase P RNAs are shown in the consensus structure in upper case; nucleotides which are not invariant, but are conserved in at least 90% of the available sequences, are shown in lower case. Nucleotides which are less than 90% conserved are shown with filled circles (●). Nucleotides which are not present in all sequences, but are absent in less than 10% of the available sequences, are indicated with hollow circles (○); those which are absent in greater than 10% of the sequences are not shown. Base-pairings for which phylogenetic evidence exists are indicated by heavy black bars.




310 320 330 340 350 360 370 380 390 400
Bsu --GAAA------UGG-------UGA--GAGUGGAA----------CGCGGUAAACCCCUCGAGCG Bsu
Bst --GAAA------CCU-------UAG--AGGUGGAA----------CGCGGUAAACCCCACGAGCG Bst
Bme --GAAA------CAG-------CAA--AGGUGGAA----------CGCGGUAAACCCCACGAGCG Bme
Bbr --GCGA------CAG-------AAA--AGAUGGAA----------CGCGGUAAACCCUGCGAGCG Bbr
Ani --CUUG------CAG-GCA-CAGGUAAGGGUGCAAGGGUGCGGUAAGAGCGCACCAGCAACAUC-GAGA-GG-UGU-UGGCUCGGUAAACCCCGGUUGGG Ani
Sbi --CUCG------GUC-C--UC-GGUAAGGGUGAAACGGUGGUGUAAGAGACCACCAGCGCCUGAGGCGACUC-AGG-CGGCUAGGUAAACCCCACUCGGA Sbi
Tne -GAUGA--------G-------GGCAAGGGUGGAACGGUGGGGUAAGAGCCCACCAGCGUCGGG-GCAA-CC-CGG-CGGCUUGGCAACCCCCACCUGGA Tne
Tma -GUUGA--------G-------GGCAAGGGUGGAACGGUGGGGUAAGAGCCCACCAGCGUCGGG-GCAA-CC-CGG-CGGCUUGGCAACCCCCACCUGGA Tma
Taq --CUU-----CCCCG-GUGCG-GGCAAGGGUGAAACGGUGGGGUAAGAGCCCACCGCCUGGCCUGGCAACAG-GCUGGGGCACGGCAAACCCCACCCGGU Taq
Tth --CUUC----CCCCG-GUGCG-GGCAAGGGUGAAACGGCGGGGUAAGAGCCCACCGCCUGGCCUGGCAACAG-GCCGGGGCACGGCAAACCCCACCCGGU Tth
Dra GCCCCGGCGCGGGUGCGGGU--GGUCAGGGUGAAACGGUGCGGUAAGAGCGCACCAGGUUCCCGGGAGACCG-GGA-GCGUCUGGUCAACCCCAUCCGGU Dra
Rru -UAUAC------GGC-C-----GGUAAGGGUGAAAGGGUGCGGUAAGAGCGCACC-GCCUGCCCGGCAACGG-GGA-GGGCAUGGUAAACCCCACCGGGA Rru
Dde --CUCG------GCC-G-----GGUAAGGGUGAAACGGUGGUGUAAGAGACCACCAGAUGCCGUGGUGACAC-GGC-AUGCUCGGCAUACCCCGUUCGGA Dde
Tfe --GUAA------GCG-C-----GGUAAGGGUGAAAAGGUGCGGUAAGAGCGCACC-GCAUUUCCGGUAACGG-AAA-UGGCAGGGAAAACCCCGCCUGGA Tfe
Cvi --CUCG------GAC-GGG--AGGUAAGGGUGAAAUGGUGCGGUAAGAGCGCACC-GCGCUGGUGGCAACAU-CAG-UGGCAGGGUAAACCCCACCCGGA Cvi
Aeu --UUCG------CCC-GGC---GGGAAGGGUGAAA----------CGCGGUAACCUCCACCUGGA Aeu
Atu --CUU-------GCA-U-----GGUAAGGGUGAAAGGGUGGGGUAAGAGCCCACC-GCGCCGCUGGUGACAG-UGG-UGGCAAGGUAAACCCCACCGGGA Atu
Pfl --UUCG------GUG-C----CGGUAAGGGUGAAAAGGUGCGGUAAGAGCGCACC-GCACGACUGGCAACAGUUCG-UGGCUAGGUAAACCCCACUUGGA Pfl
Sma --GCAA------GCG-GGAUCAGGUAAGGGUGAAAGGGUGCGGUAAGAGCGCACC-GCGCGGCUGGCAACAGUUCG-UGGCACGGUAAACUCCACCCGGA Sma
Eag --GCAA------GCG-GGAUCAGGUAAGGGUGAAAGGGUGCGGUAAGAGCGCACC-GCGCGGCUGGUAACAGUCCG-CGGCACGGUAAACUCCACCCGGA Eag
Kpn --GCAA------GCG-GGAUCAGGUAAGGGUGAAAGGGUGCGGUAAGAGCGCACC-GCGCGGCUGGUAACAGUCCG-CGGCACGGUAAACUCCACCCGGA Kpn
Sty --GUAA------GUG-GGAUCAGGUAAGGGUGAAAGGGUGCGGUAAGAGCGCACC-GCGCGGCUGGUAACAGUCCG-UGGCACGGUAAACUCCACCCGGA Sty
Eco --GCAA------GCG-GGAUCAGGUAAGGGUGAAAGGGUGCGGUAAGAGCGCACC-GCGCGGCUGGUAACAGUCCG-UGGCACGGUAAACUCCACCCGGA Eco
160 170 180 190 200 210 220 230 240
410 420 430 440 450 460 470 480 490 500
Bsu A-GAAACCCAAAU----------UU---------- Bsu
Bst A-GAAACCCAAAU----------GA---------- Bsu
Bme A-GAAACCCAAAC----------AA---------- Bme
Bbr A-GAAACCCAAAU----------U---------- Bbr
Ani AGCAAGGUGG----------AG-GGACA-----AC----------GGU--UGGUCUUUUACC---------UGUU Ani
Sbi G-CAAGGUCAAGAGGGGACACCCCGGUGUCCGUGC-GCGGAUGUUCGA----------GGGC-UGCUC---GCCC------GAGUCCG Sbi
Tne G-CAAGGCCAAGC----------A--GGGGG------U----------UGGGUUGCUC-CCCCUAUU-------CCCC Tne
Tma G-CAAGGCCAAGC----------A--GGGGG------U----------UGGGUCGCUC-CCCCUAUU-------CCCC Tma
Taq G-CAAGGCCCGAU----------AG-GCAGG-----AA----------GGGCUUGCCC--GGCCCGUUAGAA--CCUG Taq
Tth G-CAAGGCCCGGU----------AG-GCAGG-----GA----------GGGCUUGCCC--GGCCCGAGAGAA--CCUG Tth
Dra G-CAAGACCCGAC----------CGUGCGGC-----GA----------GGGC-GGCCC---GCCC-GUUUGA--GCCG Dra
Rru G-CAAGAUCGAAU----------AG-GGACGG--CAC--GCCCCGGUUUUCCGGGGC-A-AGCGGCGUUU--CCGCGCC-----GCCGUC Rru
Dde G-CAAGACCAAAU----------AG-GGAAGG----------CGGCCGGCCC--GGCCGAA-----GCCUUC Dde
Tfe G-CAAGACCAAAU----------AG-GCGUGC---GAU----------ACCGUGGCCC--GCGGU--------GCACG Tfe
Cvi G-CAAGACCAAAU----------AG-GGGAAC----UC---GCGGCUUCGACCGC---AGCGCGCGGCCC--GCGCGU-------GUUCC Cvi
Aeu G-CAAUCCCAAAU----------AG-GCAGGC---GAU----------GAAGCGGCCC--GCUGA--------GUCUG Aeu
Atu G-CAAGACCGAAU----------AG-GGAUGA--CACGGGGCGGGCGAAAGCCUGUUCACAGCCG-GUUU-CCGGGCCC-----GUCAUC Atu
Pfl G-CAAGACCAAAU----------AG-GGUUCC----AA----------GGCGUGGCCC--GCGCU--------GGAAC Pfl
Sma G-CAAGGCCAAAU----------AG-GGGUUC--ACAU----------GGUACGGCCC--GUACU--------GAACC Sma
Eag G-CAAGGCCAAAU----------AG-GGGUUC--AUAA----------GGUACGGCCC--GUACU--------GAACC Eag
Kpn G-CAAGGCCAAAU----------AG-GGGUUC--AUAA----------GGUACGGCCC--GUACU--------GAACC Kpn
Sty G-CAAGGCCAAAU----------AG-GGGUUC--AUAA----------GGUACGGCCC--GUACU--------GAACC Sty
Eco G-CAAGGCCAAAU----------AG-GGGUUC--AUAA----------GGUACGGCCC--GUACU--------GAACC Eco
250 260 270 280
510 520 530 540 550 560 570 580 590 600
Bsu -UGG--UA-G-GGGAACCUUCUUAACGGAAUUCAACGGA-GAGAAGGACAGAA-----UGCU----UUCUGUAGAUAGAUGAUGAUUGCCGCCU-GAGUACGA Bsu
Bst -UGG--UA-G-GGGCACCUUCCCGAAGGAAAUGAACGGA-GGGAAGGACAGGC-GGC-GCAU-GCAGCCUGUAGAUAGAUGAUGAUUACCGCCG-GAGUACGA Bst
Bme -UGG--UA-G-GGGAACUGUCUCAAAGGAAUCUAACGGA-GAGACGGACGGUU-AC--AUGC--GUAAUCGUAGAUAGAUGAUGAUUACCGCCU-GAGUACGA Bme
Bbr -UGG--UA-G-GGGAACCGUCCUGAAGGAAUCAAACGGAAGGGACGGAUGGUA-UC--UUCG--GAUGCCAUAGAUAGAUGGCUACCGCUCUUGGUGCGA Bbr
Ani CCGUUUAUGG-ACCGCU----------AGAGGUG-GCUAGUAAUAGCCAUCCCAGAGAGAUAACAGCCCUCUGUC------ Ani
Sbi CGGG--UA-G-ACCGCA----------CGAGGCC-GGCGGCAACGCCGGCCCUAGAUGGAUGGCCGUCGCCCCGACGACC- Sbi
Tne CGGG--UU-G-GCCGCU----------UGAGGUG-UGCGGUAACGCACACCCCAGAUUGAUGACCACCCA---------- Tne
Tma CGGG--UU-G-GCCGCU----------UGAGGUG-UGCGGUAACGCACACCCCAGAUUGAUGACCGCCCA---------- Tma
Taq CGGG--AUGG-GCCGCU----------UGAGGCC-GGUGGCGACACCGGUCCCAGAGAGAUGACCGCGUA---------- Taq
Tth CGGG--AUGG-GCCGCU----------UGAGGCC-GGUGGCGACACCGGUCCCAGAGAGAUGACCGCGGA---------- Tth
Dra CCAG--GAUG-GUCGCU----------CGAGGCGUCCUGGCAACAGGCGUCCCAGACAGAUGAUCAGGCUGCCCUGGGG-- Dra
Rru CGGG--UA-GUAUCGCG----------CGAGGCG-UCCGGUAACGGGCGUCCCAGAUGAAUGGCCAUCCCCGUCGUCGCGU Rru
Dde CGGG--UA-G-GUUGCU----------UGAGGGU-GUGGGCAACCGCACUCCUAGAGGAAUGACGGUCACACGCGG----- Dde
Tfe CGGG--UA-G-GUUGCU----------GGAGCCU-GUGCGUAAGUGCAGGCCUAGAGGAAUGGCUCGUCCA---------- Tfe
Cvi CGGG--UA-G-GUCGCU----------AGAGGCG-UGCGGCGACGCACGUCCCAGAUGAAUGACUGUCCA---------- Cvi
Aeu CGGG--UA-G-GGAGCU----------GGAGCCG-GCUGGUAACAGCCGGCCUAGAGGAAUGGUUGUCACGCACCGUUUGC Aeu
Atu CGGG--UG-G-GUUGCG----------AGAGGCG-GCAUGCAAAUGCCGUCCCAGAUGAAUGGCUGCCACGUUCCGG---- Atu
Pfl CGGG--UA-G-GUUGCU----------AAAGAUG-UCCAGUGAUGGCCAUCGUAGACGAAUGACUGUUCA---------- Pfl
Sma CGGG--UA-G-GCUGCU----------UGAGCCA-GUGAGCGAUUGCUGGCCUAGAGGAAUGACUGUCCA---------- Sma
Eag CGGG--UA-G-GCUGCU----------UGAGCCA-GUGAGCGAUUGCUGGCCUAGAUGAAUGACUGUCCA---------- Eag
Kpn CGGG--UA-G-GCUGCU----------UGAGCCA-GUGAGCGAUUGCUGGCCUAGAUGAAUGACUGUCCA---------- Kpn
Sty CGGG--UA-G-GCUGCU----------UGAGCCA-GUGAGCGAUUGCUGGCCUAGAUGAAUGACUGUCCA---------- Sty
Eco CGGG--UA-G-GCUGCU----------UGAGCCA-GUGAGCGAUUGCUGGCCUAGAUGAAUGACUGUCCA---------- Eco
290 300 310 320 330 340

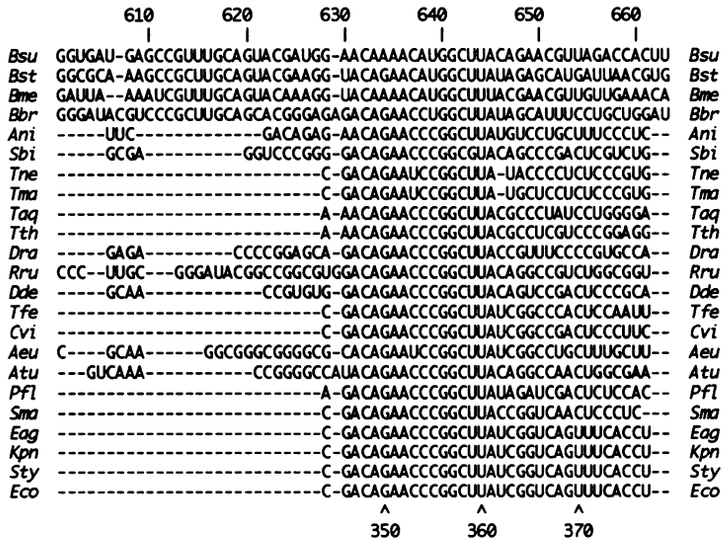


Figure 2. Bacterial RNase P RNA nucleotide sequence alignment. RNase P RNA sequences are from *Bacillus subtilis* (*Bsu*) (17), *Bacillus stearothermophilus* (*Bst*) (13), *Bacillus megaterium* (*Bme*) (13), *Bacillus brevis* (*Bbr*) (13), *Anacystis nidulans* (*Ani*) (18), *Streptomyces bikiniensis* (*Sbi*) (19), *Thermotoga neapolitana* (*Tne*) (20), *Thermotoga maritima* (*Tma*) (19), *Thermus aquaticus* (*Taq*) (20), *Thermus thermophilus* (*Tth*) (21), *Deinococcus radiodurans* (*Dra*) (19), *Rhodospirillum rubrum* (*Rru*) (12), *Desulfovibrio desulfuricans* (*Dde*) (12), *Thiobacillus ferrooxidans* (*Tfe*) (16), *Chromatium vinosum* (*Cvi*) (12), *Alcaligenes eutrophus* (*Aeu*) (12), *Agrobacterium tumefaciens* (*Atu*) (12), *Pseudomonas fluorescens* (*Pfl*) (13), *Serratia marcescens* (*Sma*) (14), *Erwinia agglomerulans* (*Eag*) (14), *Klebsiella pneumoniae* (*Kpn*) (14), *Salmonella typhimurium* (*Sty*) (11), and *Escherichia coli* (*Eco*) (15). The alignment is numbered at the top; numbers based on the *E.coli* RNA are indicated with carets (^) below the *Eco* sequence.

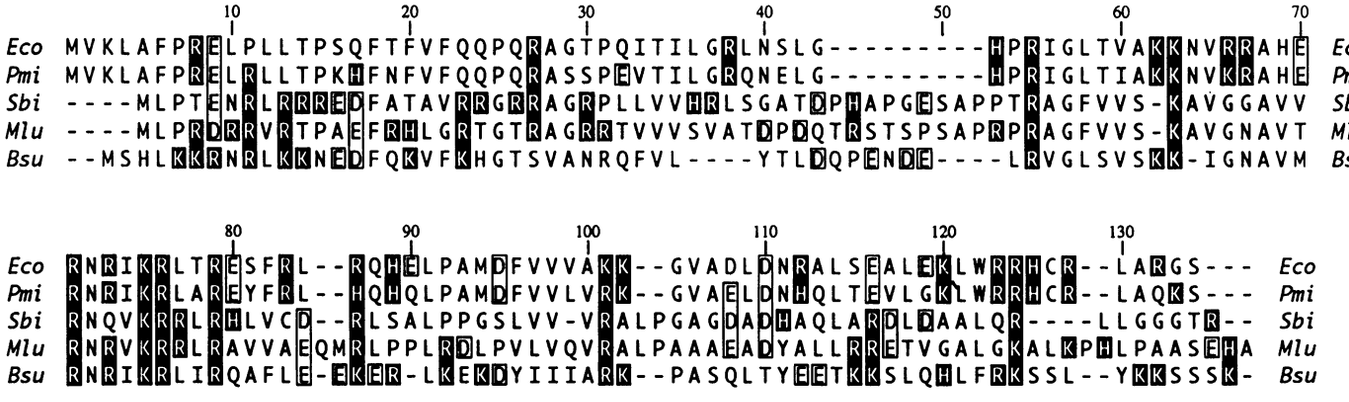


Figure 3. Bacterial RNase P protein amino acid sequence alignment. RNase P protein sequences, deduced from gene sequences, are from *Escherichia coli* (*Eco*) (7), *Proteus mirabilis* (*Pmi*) (27), *Streptomyces bikiniensis* (*Sbi*) (25), *Micrococcus luteus* (*Mlu*) (24), and *Bacillus subtilis* (*Bsu*) (26). Alignment gaps added to maximize sequence similarity are shown as dashes (−). Basic amino acid residues (histidine, lysine, and arginine) are enclosed by black squares; acidic residues (aspartic acid and glutamic acid) are enclosed by white squares. The alignment is numbered at the top.

RNAs. In contrast to the bacterial RNA, however, the archaeal RNase P RNAs apparently are not catalytically active when isolated from the holoenzyme. The resistance of the *Sulfolobus* RNase P to nuclease digestion apparently results from inaccessibility of the RNA in the holoenzyme (44); such masking of the RNA also could be in part responsible for the low density of the enzyme in Cs_2SO_4 gradients, if cesium cations were somehow excluded from binding to the RNA.

SEQUENCE ALIGNMENTS

The RNA components of the available bacterial RNase P RNA sequences are aligned on the basis of their secondary structures, which are in turn based on phylogenetic comparisons (Figure 2). Parsimonious reconstruction of the evolutionary path of the RNA sequence and structure has been used to align portions of the molecule that cannot be aligned solely on the basis of structure. The alignment of nucleotides within regions 21−60, 260−290, and 343−344 (except in the case of the purple bacteria and relatives) is ambiguous because of the extreme variability in length (even presence) and sequence. These regions have therefore been aligned, somewhat arbitrarily, from the proximal to distal end, along the lengths of their continuous helices. Base-paired nucleotides are assigned homologous positions in the alignment in every case.

Because homologies of nucleotides in the bacterial, mitochondrial and eucaryal RNAs cannot yet be assigned, the latter sequences are not included in the alignment. Portions of the archaeal RNase P RNAs are strikingly similar to those of the bacteria; however, large regions of the *Haloferax* and

Sulfolobus RNAs cannot be aligned with each other, or to the bacterial RNAs. These RNAs, therefore, also are not included in the alignment.

The structure of the bacterial RNase P protein has not been investigated beyond determination of gene sequences. The alignment of the presumptive amino acid sequences is, therefore, based only on maximized sequence similarity. The UWGCG BESTFIT and GAP programs (47) were used to generate the multiple alignments, by iterative pair-wise alignments, which then were compressed manually into a more contiguous alignment (Figure 3).

A fundamental difference between the alignments of the RNA and protein subunit sequences that we present must be emphasized; the RNA alignment is based on homology (common ancestry) of sequences, whereas the protein alignment is based on similarity.

SEARCH FOR RNASE P SEQUENCES IN CHLOROPLAST GENOMES

We have searched the complete chloroplast genome sequences of rice (48), tobacco (49), liverwort (50), and *Epiphagus* (51), in an attempt to identify bacterial-like RNase P RNA- or protein-encoding sequences. Short, redundant, consensus sequence 'probes' were identified from alignments of the bacterial sequences. In the case of the RNA alignment, the sequences from the enterobacteria were represented only by *E. coli* and those of the thermotogales only by *T. maritima*, so that closely related sequences would not monopolize the consensus. Because the dramatic structural changes of the *Bacillus* RNase P RNAs appear unique to this group (22), these sequences were likewise omitted from the alignment used in the construction of consensus RNA probes. The UWGCG WORDSEARCH, FASTA, and TFASTA programs (47) were used to search the chloroplast genomes for similarity to the probe sequences. Regions of similarity to the probes were compiled with respect to their relative orientation and sequential order (relative to the bacterial sequence consensus) and by dot-matrix ('dot plot') analysis. No sequences within the chloroplast genome with the potential to encode bacterial-like RNase P RNA or protein subunits were identified. This does not, of course, discount the possibility that bacterial-like RNase P exists in the chloroplast. It is possible that bacterial-like RNase P RNA and protein subunits could be encoded by the nuclear genome and imported into the organelle, or could be so divergent from their bacterial counterparts that they could not be identified on the basis of sequence similarity. Nevertheless, these results strengthen the suggestion (42) that the chloroplast RNase P does not seem to contain a bacterial-like RNA component.

ACKNOWLEDGMENTS

The authors thank John Donello for assistance in the computer search of the chloroplast genomes for RNase P-like sequences, and Drs. Elizabeth Haas and Sylvia Darr for helpful discussion and comments during the preparation of this manuscript. This work was supported by grant GM34527 from the National Institutes of Health to NRP.

REFERENCES

1. Altman, S. (1990) *J. Biol. Chem.*, **265**, 20053–20056.
2. Pace, N. R. and Smith, D. (1990) *J. Biol. Chem.*, **265**, 3587–3590.
3. Guerrier-Takada, C., Haydock, K., Allen, L. and Altman, S. (1986) *Biochemistry*, **25**, 1509–1515.
4. Altman, S. and Smith, J. D. (1971) *Nature (London) New Biol.*, **233**, 35–39.
5. Stark, B. C., Kole, R., Bowman, E. J. and Altman, S. (1977) *Proc. Natl. Acad. Sci. U. S. A.*, **75**, 3719–3721.
6. Gardiner, K. and Pace, N. R. (1980) *J. Biol. Chem.*, **255**, 7507–7509.
7. Hansen, F. G. and Hansen, E. B. and Atlung, T. (1985) *Gene*, **38**, 85–93.
8. Guerrier-Takada, C., Gardiner, K., Marsh, T. L., Pace, N. R. and Altman, S. (1983) *Cell*, **35**, 849–857.
9. Woese, C. R., Kandler, O. and Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci. U. S. A.*, **87**, 4576–4579.
10. Woese, C. R. (1987) *Microbiol. Rev.*, **51**, 221–271.
11. Baer, R. E. and Altman, S. (1985) *Science*, **228**, 999–1002.
12. Brown, J. W., Haas, E. S., James, B. D., Hunt, D. A, Liu, J. and Pace, N. R. (1991) *J. Bacteriol.* **173**, 3855–3963.
13. James, B. D., Olsen, G. J., Liu, J. and Pace, N. R. (1988) *Cell*, **52**, 19–26.
14. Lawrence, N. P., Richman, A., Amini, R. and Altman, S. (1987) *Proc. Natl. Acad. Sci. U. S. A.*, **84**, 6825–6829.
15. Reed, R. E., Baer, M. F., Guerrier-Takada, C., Donis-Keller, H. and Altman, S. (1982) *Cell*, **30**, 627–636.
16. Takeshima, T., Inoue, C., Kitagawa, Y. and Kusano, T. (1989) *Nucl. Acids Res.*, **17**, 9482.
17. Reich, C., Gardiner, K. J., Olsen, G. J., Pace, B., Marsh, T. L. and Pace, N. R. (1986) *J. Biol. Chem.*, **261**, 7888–7893.
18. Banta, A. B., Haas, E. S., Brown, J. W. and Pace, N. R. (unpublished data). X63566
19. Haas, E. S., Morse, D. P., Brown, J. W., Schmidt, F. J. and Pace, N. R. (1991) *Science*, **254**, 853–856.
20. Brown, J. W., Haas, E. S. and Pace, N. R. (unpublished data). X63821 X63822
21. Hartmann, R. K. and Erdmann, V. A. (1991) *Nucl. Acids Res.*, **19**, 5957–5964.
22. Brown, J. W. and Pace, N. R. (1991) *Biochimie*, **73**, 689–697.
23. Waugh, D. S., Green, C. J. and Pace, N. R. (1989) *Science*, **244**, 1569–1571.
24. Fujita, M. Q., Yoshikawa, H. and Ogasawara, N. (1990) *Gene*, **93**, 73–78.
25. Morse, D. P. and Schmidt, F. J. (personal communication). M83112
26. Ogasawara, N., Moriya, S., von Meyenburg, K., Hansen, F. G. and Yoshikawa, H. (1985) *EMBO J.*, **4**, 3345–3350.
27. Skovgaard, O. (1990) *Gene*, **93**, 27–34.
28. Gardiner, K. J., Marsh, T. L. and Pace, N. R. (1985) *J. Biol. Chem.*, **260**, 5415–5419.
29. Reich, C., Olsen, G. J., Pace, B. and Pace, N. R. (1988) *Science*, **239**, 178–181.
30. Carrara, G., Calandra, P., Fruscoloni, P., Doria, M. and Tocchini-Valentini, G. P. (1989) *Cell*, **58**, 37–45.
31. Gold, H. A. and Altman, S. (1986) *Cell*, **44**, 243–249.
32. Cherayil, B., Krupp, G., Schuchert, P., Char, S. and Söll, D. (1987) *Gene*, **60**, 157–161.
33. Krupp, G., Cherayil, B., Frendewey, D., Nishikawa, S. and Söll, D. (1986) *EMBO J.*, **5**, 1697–1703.
34. Lee, J., Rohlman, C. E., Molony, L. A., and Engelke, D. R. (1991) *Mol. Cell. Biol.*, **11**, 721–730.
35. Sogin, M. L., Gunderson, J. H., Elwood, H. J., Alonso, R. A. and Peattie, D. A. (1989) *Science*, **243**, 75–77.
36. Forster, A. C. and Altman, S. 1990 *Cell*, **62**, 407–409.
37. Doerson, C., Guerrier-Takada, C., Altman, S. and Attardi, G. (1985) *J. Biol. Chem.*, **260**, 5942–5949.
38. Hollingsworth, M. J. and Martin, N. C. (1986) *Mol. Cell Biol.*, **6**, 1058–1064.
39. Miller, D. L. and Martin, N. C. (1983) *Cell*, **34**, 911–917.
40. Wise, C. A. and Martin, N. C. (1991) *J. Biol. Chem.*, **266**, 19154–19157.
41. Morales, M. J., Wise, C. A., Hollingsworth, M. J. and Martin, N. C. (1989) *Nucl. Acids. Res.*, **17**, 6865–6881.
42. Wang, M. J., Davis, N. W. and Gegenheimer, P. (1988) *EMBO J.*, **7**, 1567–1574.
43. Darr, S. C., Pace, B. and Pace, N. R. (1990) *J. Biol. Chem.*, **265**, 12927–12932.
44. LaGrandeur, T., Darr, S. and Pace, N. R. (unpublished data).
45. Lawrence, N., Wesolowski, H., Gold, H., Bartkiewicz, M., Guerrier-Takada, C., McClain, W. H. and Altman, S. (1987) *Cold Spring Harbor Symp. Quant. Biol.*, volume LII. Cold Sping Harbor Press. Cold Spring Harbor.
46. Nieulandt, D. T., Haas, E. S. and Daniels, C. J. (1991) *J. Biol. Chem.*, **266**, 5689–5695.
47. Devereux, J., Haeberli, P. and Smithies O. (1984) *Nucl. Acids Res.* **12**, 387–395.
48. Sugiura, M. (unpuplished). X15901.
49. Sugiura, M. (unpublished). Z00044.
50. Ohyama, K. (unpublished). X04465. X01647.
51. Wolfe, K. H., Morden, C. W. and Palmer, J. D. (personal communication).