Analysis of the gene encoding the RNA subunit of ribonuclease P from cyanobacteria

Agustín Vioque
Instituto de Bioquímica Vegetal y Fotosíntesis, Universidad de Sevilla-CSIC, Apartado 1113, 41080 Seville, Spain

ABSTRACT

The genes encoding the RNA subunit of ribonuclease P from the unicellular cyanobacterium *Synechocystis* sp. PCC 6803, and from the heterocyst-forming strains *Anabaena* sp. PCC 7120 and *Calothrix* sp. PCC 7601 were cloned using the homologous gene from *Anacystis nidulans* (*Synechococcus* sp. PCC 6301) as a probe. The genes and the flanking regions were sequenced. The genes from *Anabaena* and *Calothrix* are flanked at their 3'-ends by short tandemly repeated repetitive (STRR) sequences. In addition, two other sets of STRR sequences were detected within the transcribed regions of the *Anabaena* and *Calothrix* genes, increasing the length of a variable secondary structure element present in many RNA subunits of ribonuclease P from eubacteria. The ends of the mature RNAs were determined by primer extension and RNase protection. The predicted secondary structure of the three RNAs studied is similar to that of *Anacystis* and although some idiosyncrasies are observed, fits well with the eubacterial consensus.

INTRODUCTION

Ribonuclease P (RNase P) is an endonuclease responsible for the generation of the mature 5'-end of tRNAs from precursors of tRNAs (1, 2, 3). Every cell and every cellular compartment that synthesizes tRNAs has been shown to contain an RNase P activity. The enzyme is composed of RNA and protein in the majority of RNase P activities investigated so far (4). One important exception is the chloroplast, where there is no evidence at present for an RNA component in its RNase P (5). In many eubacteria the RNA component of the enzyme has been shown to be catalytically active in the absence of the protein subunit (6, 7, 8). The secondary structure of the catalytic RNA has been analyzed by different techniques, however the most powerful tool has been the phylogenetic comparative analysis of covariation of base pairs in possible helices (9, 10). From these studies a model has been proposed for the structure of bacterial RNase P RNA (11, 12). The refinement of the secondary structure models requires the analysis of as many sequences as possible. The sequences of representatives of the main eubacterial lineages

has been established. From the cyanobacterial lineage the only sequence available so far is the one from *Anacystis nidulans* (13). In addition the 5' and 3'-ends of the RNA component of RNase P has been determined experimentally in only a few instances (7, 14, 15, 16); in other cases they have been inferred from the determined ones by homology.

I have analyzed the RNase P RNA subunit from an unicellular (*Synechocystis sp.* PCC 6803) and two filamentous heterocyst-forming (*Anabaena sp.* PCC 7120 and *Calothrix sp.* PCC 7601) cyanobacteria. These represent, together with the previously characterized *Anacystis* RNA, a more relevant sampling of the evolutive radiation of the cyanobacteria. This new cyanobacterial sequences allow a refinement of the secondary structure model for these RNAs. Some corrections have been introduced in the *Anacystis nidulans* RNase P RNA secondary structure (13) based on the new information provided by the genes analyzed here. In addition the RNAs have been characterized and their authentic 5'-and 3'-end determined.

MATERIALS AND METHODS

Enzymes were purchased from Boehringer Manheim or Pharmacia. Radioactive nucleotides were obtained from Amersham or New England Nuclear.

Cyanobacterial strains

Synechocystis sp. PCC 6803, *Anabaena* sp. PCC 7120 and *Calothrix* sp. PCC 7601 were grown under constant illumination in BG11 medium supplemented with 1.7 mM Na_2CO_3 as previously described (17).

Cloning of the genes encoding the RNase P RNAs

All routine cloning techniques were performed as described (18). DNA was isolated from cultures of the three cyanobacterial strains, which had been grown in erlenmeyer flasks until stationary phase, by grinding with glass beads and phenol extraction (19). For cloning of the RNase P RNA genes, subgenomic libraries were constructed based on the information obtained from Southern hybridizations. Southern hybridization was carried out at 65°C in 0.75 M NaCl, 75 mM sodium citrate

(pH 7) (5×SSC), 5×Denhardt solution, 25 mM sodium phosphate (pH 7.5), 2.5 mM Na$_2$-EDTA, 0.1% SDS, 100 μg/ml sonicated herring sperm DNA and 10^6–10^7 cpm of a nick translated probe. The probe used was a PstI-XhoI fragment from plasmid Ani4-1Q. This plasmid contains the RNase P RNA gene from *Anacystis nidulans* (*Synechococcus* sp. PCC 6301) cloned in the polylinker site of pBluescript KS+ and was a gift from Amy Banta and Dr. Norman Pace (Bloomington, Indiana). Filters were washed twice in 0.3 M NaCl, 30 mM sodium citrate (pH 7) (2×SSC), 0.1% SDS at room temperature and once in the same solution at 65°C. The Southern blot indicated that a single band hybridized with the *Anacystis* probe in each case after digestion of the DNA with HindIII or DraI (Fig. 1). A 1.7 kb fragment from *Synechocystis* DNA, a 4.4 kb fragment from *Anabaena* DNA and a 6.6 kb fragment from *Calothrix* DNA, obtained after digestion with HindIII, hybridized with the probe and were used for the construction of the subgenomic libraries. The bands were cut from an agarose gel, eluted and cloned in pGEM-7zf(+) (Promega) that had been digested with HindIII and dephosphorylated. One hundred recombinant clones from the DNA of each cyanobacterial strain were plated on nitrocellulose filters. The filters, after lysis of the *Escherichia coli* colonies (18), were hybridized and washed in the same conditions used for Southern hybridization. Positive clones were confirmed by Southern hybridization.

RNase P enzyme assays

To further confirm the identity of the cloned fragments, the RNase P activity of RNAs generated by *in vitro* transcription from plasmids containing the appropriate fragments was measured at high ionic strength as described (12). A precursor tRNA^{Tyr} was used as substrate (2).

DNA sequencing

Fragments from the recombinant plasmids that contained portions of the RNase P RNA genes were subcloned in pBluescript SK+ (US Biochemicals) and both strands were sequenced using the M13/pUC sequencing primer and the M13 reverse sequencing primer (Boehringer). DNA sequencing was performed with double stranded DNA using the dideoxy chain termination method (20) with Sequenase version 2.0 (US Biochemicals) and [^{35}S]dATP[αS]. When necessary a set of nested deletions was generated with exonuclease III and nuclease S1 (21). A kit from Pharmacia was used for this purpose. dITP was used as required to alleviate band compression in sequencing gels.

Preparation of total RNA

RNA was extracted from cultures grown to mid-logarithmic phase by a modification of the procedure of Mohamed and Jansson (22). The cells (from 100 ml cultures) were harvested by centrifugation in the cold and immediately frozen with liquid nitrogen. The cell pellet was thawed on ice and the freeze-thaw cycle was repeated once. The cells were suspended in resuspension buffer [0.3 M sucrose, 10 mM sodium acetate (pH 4.5)], transferred to an Eppendorf tube and pelleted at 12000×g for 5 min. The pellet was suspended in 250 μl of resuspension buffer supplemented with 75 μl 250 mM Na$_2$-EDTA, and the suspension was incubated on ice for 5 min. 375 μl of lysis buffer [2% (w/v) SDS, 10 mM sodium acetate (pH 4.5)] was added followed by incubation at 65°C for 3 min. 700 μl of hot (65°C) phenol (molecular biology grade, Sigma) was added to the lysed cells,

followed by incubation at 65°C for 3 min. The suspension was centrifuged, the upper phase was collected and the hot phenol treatment was repeated twice, followed by an extraction with hot phenol:chloroform (1:1). 1/10 volume of 3 M sodium acetate (pH 4.5) and 2 volumes of ethanol were added and the nucleic acids were precipitated at −20°C for 30 min. The pellet was washed twice with 75% ethanol, dried and suspended in 10 mM Tris-HCl (pH 7.5), 10 mM MgSO$_4$ and treated with 100 units of DNase I (RNase free, Boehringer). The solution was extracted twice with phenol:chloroform (1:1) and once with chloroform. 1/10 volume of 3 M sodium acetate (pH 4.5) and 2.5 volumes of ethanol were added and the nucleic acids were precipitated at −20°C for 30 min. The pellet was washed twice with 75% ethanol, dried and suspended in 50 μl of 10 mM Tris-HCl (pH 7.5), 1 mM Na$_2$-EDTA. Glassware and solutions used for RNA extraction were treated with 0.1% (v/v) diethyl pyrocarbonate (Sigma) overnight at 37°C followed by autoclaving.

Northern hybridization

About 5 μg of total RNA were separated on a 5% polyacrylamide, 7M urea gel and transferred to nylon membranes (Zetaprobe, BioRad) by electroblotting. The membranes were hybridized with the *Anacystis* probe and washed in the same conditions used for Southern hybridization. As size markers a set of ^{32}P-labelled DNA fragments were generated by digestion of plasmid pBR322 with HpaII and filling the recessed 3'-ends with Klenow polymerase in the presence of [α-^{32}P]dCTP (18).

Primer extension

Primer extension was used to map the 5'-end of the RNase P RNAs. As primer the synthetic oligonucleotide 5'-GG(G,A)A-G(C,T)CCGGACTTTCCTCA-3' was used. This oligonucleotide is complementary to a highly conserved region of bacterial RNase P RNAs, close to the 5'-end. The oligonucleotide was labeled at the 5'-end with [γ-^{32}P]ATP and polynucleotide kinase (18). 10^6 cpm of labelled oligonucleotides were mixed with 5 μg of total RNA in 30 μl of 40 mM PIPES-Na (pH 6.4), 1 mM Na$_2$-EDTA, 0.4 M NaCl, heated at 85°C and annealed overnight at 50°C. The hybridization mix was precipitated with ethanol and extension reactions were performed as described (18) with 40 units of AMV reverse transcriptase (Boehringer). For the preparation of size markers, the same oligonucleotide (but not phosphorylated at the 5'-end) was used to prime DNA sequencing reactions with the appropriate templates. DNA sequencing was performed with double stranded DNA using the dideoxy chain termination method with Sequenase version 2.0 (US Biochemicals) and [^{35}S]dATP[αS].

RNase protection

RNase protection assays were used to map the 5' and 3' ends of the RNase P RNAs. Labelled RNA probes complementary to the RNase P RNAs were synthesized by *in vitro* run-off transcription with T7 or T3 RNA polymerase of appropriate templates in the presence of [α-^{32}P]CTP (18) and purified on acrylamide gels. The probes were used within 24 hours of their purification. Figure 2 shows the different probes used for mapping the ends of the three RNAs studied here.

— probe S5 was used to map the 5'-end of the *Synechocystis* RNase P RNA and extends from the HincII site into the gene (nucleotide position 271 of the sequenced region) to a HincII site at nucleotide position 1 of the sequenced region.

– probe S3 was used to map the 3'-end of the *Synechocystis* RNase P RNA and extends from an RsaI site downstream the sequenced region to an RsaI site into the gene (nucleotide position 367 of the sequenced region).

– probe A5 was used to map the 5'-end of the *Anabaena* RNase P RNA and extends from nucleotide position 273 of the sequenced region to an upstream RsaI site at nucleotide position 2.

– probe A3 was used to map the 3'-end of the *Anabaena* RNase P RNA and extends from nucleotide position 809 to an RsaI site at nucleotide position 433 of the sequenced region.

– probe C5 was used to map the 5'-end of the *Calothrix* RNase P RNA and extends from nucleotide position 311 of the sequenced region to an upstream RsaI site.

– probe C3 was used to map the 3'-end of the *Calothrix* RNase P RNA and extends from a downstream RsaI site to nucleotide position 251 of the sequenced region.

All the probes contained in addition nucleotides transcribed from the polylinker region of the plasmids used as templates.

About 500.000 cpm of the purified probes were mixed with 5 μg of total RNA from the appropriate organism in 30 μl of 40 mM PIPES-Na (pH 6.4), 1 mM Na$_2$-EDTA, 0.4 M NaCl, 80% formamide, heated at 85°C for 10 min and annealed overnight at 45°C. The hybridization mix was precipitated with ethanol and treated with 10 μg/ml RNase A and 2 μg/ml RNase T$_1$ as described (18). A control with 15 μg of *E.coli* tRNA was included in all the experiments to rule out artifactual protection of the probes. The size of the protected fragments was analyzed on 0.4 mm sequencing gels using labelled pBR322 digested with HpaII as size markers. The DNA fragments used as markers migrated in the gel conditions used alike RNA molecules of the same length.
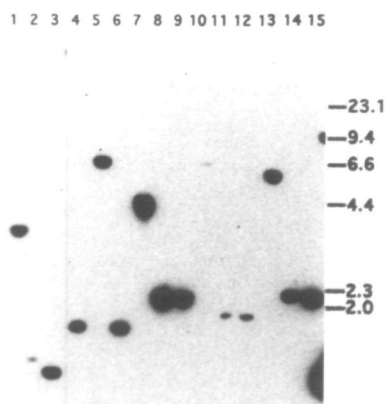
Figure 1. Southern analysis of cyanobacterial genomic DNA using the RNase P RNA gene from *Anacystis nidulans* as a probe. DNA from *Fischerella muscicolla.* UTEX 1829 (lanes 1 to 3), *Synechocystis* sp. PCC 6803 (lanes 4 to 6), *Anabaena* sp. PCC 7120 (lanes 7 to 9), *Calothrix* sp. PCC 7601 (lanes 10 to 12) and *Anabaena variabilis* ATCC 29413 (lanes 13 to 15) was digested with HindIII (lanes 1, 4, 7, 10 and 13), DraI (lanes 2, 5, 8, 11 and 14) or both enzymes (lanes 3, 6, 9, 12 and 15) and run on a 0.6% agarose gel. After transfer to a nylon membrane and hybridization with a ^{32}P-labelled DNA fragment containing the *Anacystis* RNase P RNA gene (see Materials and Methods), the bands which hybridized with the probe were visualized by autoradiography. Differences in intensity are due to differences in loading of the gel rather than in hybridization efficiency. Size of DNA markers (in kilobase pairs) are indicated to the right of the panel.

RESULTS

I have cloned DNA fragments encoding the RNase P RNA from three diverse cyanobacteria using as a probe the homologous gene from the cyanobacterium *Anacystis nidulans*. Southern hybridization analysis indicated that there is a single copy gene present in all the cyanobacteria analyzed (Fig. 1). An unicellular (*Synechocystis*) and two herocyst-forming strains (*Anabaena* and *Calothrix*) of cyanobacteria were chosen for study. These three strains, together with the previously analyzed *Anacystis nidulans*, are a significative representation of the evolutive radiation of this eubacterial group (23).

Enzymatic activity of the RNase P RNAs

RNAs were generated by *in vitro* transcription from plasmids containing the RNase P RNA genes under control of T7 or T3 promoters. These RNAs had RNase P enzymatic activity when assayed, at high ionic strength, with an *E.coli* precursor tRNA^{Tyr} substrate. The products of the reaction were identical to those obtained with the *E.coli* RNase P RNA (results not shown). This demonstrates that the cloned genes codify for the RNA subunit of RNase P.

Nucleotide sequence of the RNase P RNA genes

The genes and flanking regions of the three cyanobacterial RNase P RNAs have been sequenced (Fig. 3). The RNAs seem to be transcribed as a single transcript in the mature form without any further processing, as has been determined by primer extension and RNase protection (see below).

Promoters. In cyanobacteria, many promoters depart significantly from the *E.coli* consensus. In *Synechocystis* there are instances where a −35 region can not be identified (24). However, an *in vitro* study in *Anabaena* indicates that the RNA polymerase prefers promoter sequences that are close to the *E.coli* consensus sequence (25). In the *Anabaena* and *Calothrix* genes studied here, regions with similarity to −10 and −35 regions can be identified at appropriate positions upstream the transcription start point. In *Synechocystis* a region with similarity to −10 and −35
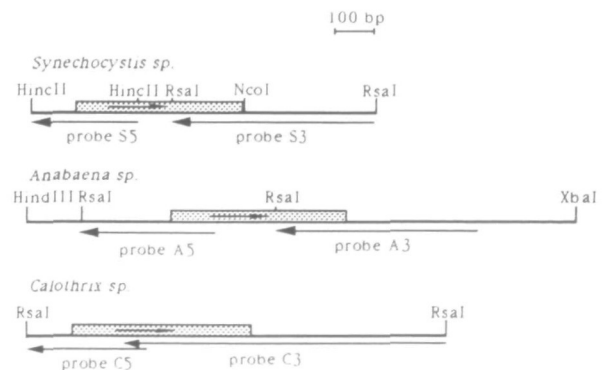
Figure 2. Schematic representation of DNA fragments containing the cloned genes and of the RNA probes used in the RNase protection assays. The dotted boxes represent the RNase P RNA coding sequences, which are transcribed from left to right. RNA probes used for RNase protection assays are represented by the arrows below the genes and are described in detail in Materials and Methods. Only relevant restriction sites have been included.

Synechocystis sp. PCC 6803
```
                                                              -35
ACTCCTAGCACCAATTTCCCATGACTACGGAGGGGGCAATGAAGTTTCAATTAATTGCCGTCACAACCACAGCGGCCTA   80
   -10
TGCCTCTAATCAATGCACACTAGAAAAAGAGAGTTAGGGAGGGAGTTGCCGGATTCCTGTCACAGGGAATCTGAGGAAAGT  160
CCCGGCTTCCCAAAGGCCAAACTTGCTCGGTAACTCCCAGTCGCGCGCGAGGGTCAGGACAGTGCCACAGAAAAATACCGC  240
CCTTTTTAGAAAACAGCAACCAGTAAACAGTTAACAGGGTTTTTCTGCTGACTAACTGGTAACTGACCACTGAAAGGGT    320
AAGGGTGCAAAGGTGCGGTAAGACGGCACCAGCAGTATCGAGAGGTACTGGCTCGGTAAACCCCGGTTGGAAGCAAGGTC  400
GGAGGGGCAAAGGTTGGTCTTTTTCCTGCCCCATGATTGGTGGAACCGCTTGAGGAATTTGGTAACAAATTTCCCAGATA  480
GATAACTCCCCAAGGGTGCCTCGCATCCTGGAACAGAACCCGGCTTAGGACTAACTCTCTTTTTTGTCCATGGGTCCAG    560
TGGCATGCTGTCAAAAACTATTCATGAAGGGCCTGGTT                                           598
```

Anabaena sp. PCC 7120
```
TACGAGCAGGTTTTGTTCCTGGGTCAGAATTTCCTCCCATATCCATTATTAATTGTCCCTTCACTATCAAAAAACTACCT   80
                  -35               -10
TGGACTTATGCCCTACCCTGGAATAAAAAGAAATAAGCAAAACAGACACAAGACACCAACGAAGATTACACTGGAGCTAG   160
GGAGAGAGTAGGCGTTGGCGGTTGCAGACCAGTTAGCTTAACTGATTTGAGGAAAGTCGGGACTCCCGAAAGACCAGACT  240
TGCTGGATAACGTCCAGTGCGAGCGGATCGTGAGGATAGTGCCACAGAAAAATACCGCCAAGATTGGGGACTGGCGACTAG  320
GGGTTGGGGACTGGCGAAGAAACTTCCCAATCCCTAATCCCCCATACCCAATACCCAACTCTTGGTAAGGGTGCAAAGGT  400
GGGGTAAGACGCCACCAGCAGTATCGAGAGGTACTGGCTCGGTAAACCCCGGTTGGGAGCAAGCCGAAGAACTATGGTT  480
GGTCTTTTACCAGTTCCCCTATCAGAGAGCCGCTAGACGGGTTTGGTAACAAACGTCCCAGATAGATAATCCCCCTCGTG  560
TAAAGCAATTTACACAGAGAACAGAACCCGGCTTACTACCAACTCTCTGCTCTTTTAGTCAATAGTCATTAGTCGATAGT  640
CATTAGTTATTAGTCAATAGTCATTAGTCATTAGTCCCTAGTCAAAAGAAACGACCAATGACAACTGACAACTGACAACT  720
GACAACTGACAACTGACAACTGACAACTGGCAACTGACAACTGACAACTAACAAATGACTCTTGTGTATCCCCGCCGTCA  800
GAGGCAACTTGAAAGTTTAGTAAGAAAATTAGGTCTGCCAATAACAGCACCCATTAAGTGGGAATTGCTGGATTTAGCAT  880
TGACTCATCCTACTGTGTCTGAATCAGCTAATTATGAACAGTTGGAGTTTGTTGGG                          936
```

Calothrix sp. PCC 7601
```
CATTATTAATTGCCCTTTCACAAGGAGAAAACTACCCTGGACATACGCCCTACCTTGGAATAAAAGGAAGTAACTTAAAT   80
        -35              -10
CAGACACACAGCTCCATCCAAGATTACACTGGATCTAAGGAGAGAGTAGGCCGTAGGCGGTTGCAGATCAGTCACAAAAA  160
CTGAATTTGAGGAAAGTCGGGGCTCCCGAAAGACCCAAACTTGCTGGATAACGTCCAGTGCGAGCGGATCGTGAGGATAGTG  240
CCACAGAAAGATACGGCCAATTAGTCATTAGTCATTTGTCATTAGTCATTTGGAAAAACACAAGGACAAAGGACAAAGA   320
ACAAATGACCAAATTGGTAAGGGTGCAAAGGTGGGGTAAGACGCCACCAGCAACGTCGGAGGCGGTTGGCTCGGTAAACC  400
CCGGTTGGGAGCAAGGCGAAAGGAACTATGGTTGGTCTTTTACCAGTTCCCCTAAAACAGAGCGCTAGAGCGGTTTGGT  480
AACAAACGTCCCAGATAGATAACCGCCCTCATGTAGAGAAGTTACTTGTAACATTTCTACATTCGAGAACAGAACCCGGC  560
TTACTACCAACTCTTTCCCTTTTTCGTCATTTGGTGTTTGTCATTTGTAATTTGTCATTTGGTAATGGGTAATGGGTAAT  640
GGGTAATGGGTAATGGGTAATGGG                                                         664
```

Figure 3. Nucleotide sequence of the RNase P RNA genes. The transcribed sequences are underlined. The limits of the transcribed sequences were deduced from the primer extension and RNase protection assays. The region to which the oligonucleotide used for primer extension hybridizes is double underlined. Arrows above the sequence of the *Anabaena* and *Calothrix* genes indicate the STRR sequences. The sequences that best approximate the *E.coli* promoter consensus in the −35 and −10 regions are boxed.

consensus sequences is present, but it is 18 nucleotides upstream of the transcription start point and thus its significance is doubtful.

Repeated sequences. Different sets of imperfect heptanucleotide repeats, called short tandemly repeated repetitive (STRR) sequences are identified just downstream of the transcription termination site in the *Anabaena* and *Calothrix* genes (Fig. 3). Similar observations have been made for other genes in these heterocyst-forming cyanobacteria (26−35). More surprising, two other sets of STRR sequences are present within the coding sequences of the *Anabaena* and *Calothrix* genes. These repeated sequences interrupt a secondary structure element of the RNAs (see below). The STRR sequences described here are related to families of previously characterized STRR sequences (26).

Table 1. Analysis of the RNase protection experiment.

|	Size of the probe	Size of protected fragment	Deduced RNA ends	
probe S5	271	163	*Synechocystis* 5' end	G109
probe S3	550 aprox.	180	*Synechocystis* 3' end	T546
probe A5	272	115	*Anabaena* 5' end	A159
probe A3	377	183	*Anabaena* 3' end	T616
probe C5	340 aprox.	193	*Calothrix* 5' end	A119
probe C3	790 aprox.	334	*Calothrix* 3' end	T584

The size of the protected fragments was deduced from the experiment shown in Fig. 7. The RNA ends are indicated according to the numbering of the nucleotide sequence in Fig. 3. The accuracy of the estimate is of 1−2 nucleotides, except for the *Calothrix* 3'-end, where the accuracy is only of 3−4 nucleotides. However the 3'-end was assigned to T584 by homology with the other determined 3'-ends.

Structure of the RNase P RNAs

The secondary structure of the three characterized RNAs was determined as described (9, 10, 11), by alignment with previously analyzed bacterial RNase P RNAs. The secondary structures (Fig. 4) fit into the model proposed for proteobacteria (11, 12). The structure includes all the helices present in proteobacteria, including the variable helix 391/415 (numbering refers to the *Synechocystis* sequence) which is not present in many proteobacteria. The most surprising feature of the structure is the length and sequence of helix 130/225 (*Synechocystis* numbering). While this helix is about 35 nucleotides long in *E.coli* and 33 nucleotides long in *Anacystis*, it is 88 nucleotides long in *Synechocystis*, 97 nucleotides long in *Anabaena*, and 89 nucleotides long in *Calothrix*. The sequences of this extra long region are apparently unrelated, except for the base of the helices which contain 5 base pairs and two buldged A residues, which are almost universally conserved among the bacterial RNase P RNAs. Beyond this conserved region the rest of the helix is unrelated in the three RNAs analyzed here, and in *Anabaena* and *Calothrix* it is composed of two sets of STRR sequences. In addition, the STRR sequences are different in *Anabaena* and *Calothrix* . In *Anabaena* they might be adscribed to the STRR1 family and in *Calothrix* to the STRR2 family (26). Thus it seems like this extended helix has been generated by independent insertions in the three organisms analyzed here.

Another surprising idiosyncrasy in the *Anabaena* RNase P RNA is the absence of a canonical base pair in the conserved tertiary interaction between sequence 56−64 and sequence 429−436, which forms a helix of 9 base pairs. Nucleotide 64 is almost always a G that pairs with a C at position 429. Two exceptions are *Alcaligenes eutrophus* (12) and *Thermotoga maritima* (36) where the G-C base pair is replaced by an A-U base pair. However, in *Anabaena*, nucleotide 64 is an A and nucleotide 429 is a C. Thus *Anabaena* is the only example found so far where there is A-C at these positions instead of a canonical base pair.

Helix 241/284 of the *Anacystis nidulans* RNA has been modified from the previously proposed structure (13). The previous model contains a stem formed by nucleotides 247−256 that pair with nucleotides 266−278. Nucleotides 269, 273 and 274 are unpaired. The structure proposed here (Fig. 4) has a stem formed by nucleotides 247−256 that pair with nucleotides 266−273. Nucleotides 252 and 253 are unpaired. The basic modification introduced is that nucleotides 247−250 are paired with nucleotides 270−273 rather than with nucleotides 275−278.

Anacystis nidulans Synechocystis sp. PCC 6803 Anabaena sp. PCC 7120 Calothrix sp. PCC 7601
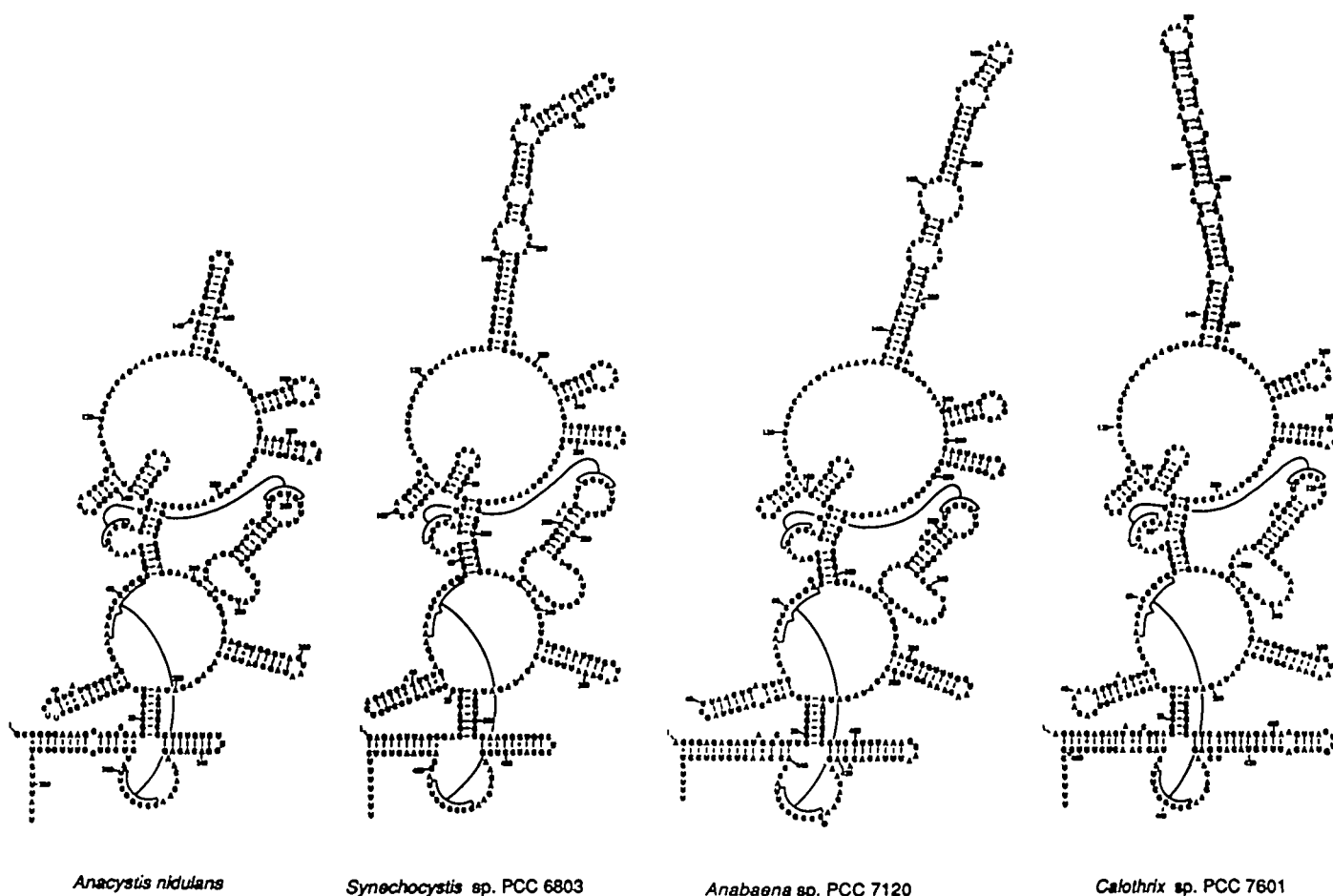
Figure 4. Proposed secondary structures of cyanobacterial RNase P RNAs. The *Anacystis nidulans* model has been modified in relation to reference 13 as is described in the text. The * in the *Anabaena* structure mark nucleotides A64 and C429, which form a canonical base pair in all the bacterial RNase P RNAs characterized so far.

This modification has been introduced because nucleotides 270−273 instead of 275−278 are homologous to the nucleotides that occupy the equivalent position in the structure of the *Synechocystis*, *Anabaena* or *Calothrix* RNAs. The stem, as proposed here, is fully supported by compensatory substitution at most of the positions.

Analysis of the RNase P RNAs

The RNase P RNAs were analyzed by Northern hybridization, primer extension and RNase protection assays, in order to determine their size, ends and possible processing steps.

Northern analysis. A single hybridizing band of the expected size, or minor breakdown products, are detected in the three species analyzed (Fig. 5).

Primer extension. A single extended product was observed in *Synechocystis* and *Anabaena* (Fig. 6). In *Calothrix*, extension terminated at two contiguous nucleotide positions and also shorter extended products were detected, due probably to premature termination. From these data the position of the transcription start point can be assigned to G109 in *Synechocystis*, to G159 in *Anabaena* and to A119 or G120 in *Calothrix*. These results are in agreement with the RNase protection assays (see below) and

there is also good agreement with the 5'-end of other characterized RNase P RNA genes.

RNase protection. This analysis detected a main protected band (Fig. 7) with all the probes used, in addition to smaller breakdown products. From the size of the main bands the 5'- and 3'-termini of the three RNAs could be determined with a precision of 1−2 nucleotides (Table 1). The exception is the 3'-end of the *Calothrix* RNA where the precision is lower due to the larger size of the protected fragment and the gel system used. The 5'-ends determined by this procedure agree quite well with the results obtained by primer extension. The RNAs terminate with a run of U nucleotides, however this sequence does not seem to be part of a rho-independent terminator because no stable stem and loop can be formed by the sequence preceding the run of U nucleotides. In *Anabaena* and *Calothrix* the 3'-end of the RNA is immediately followed by the STRR sequences.

DISCUSSION

The RNase P RNA subunit of the cyanobacteria studied in this work are transcribed from single copy monocistronic transcription units. No further processing of the primary transcript seems to occur as shown by Northern hybridization, RNase mapping and
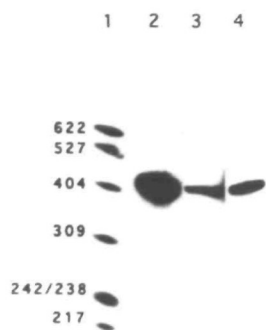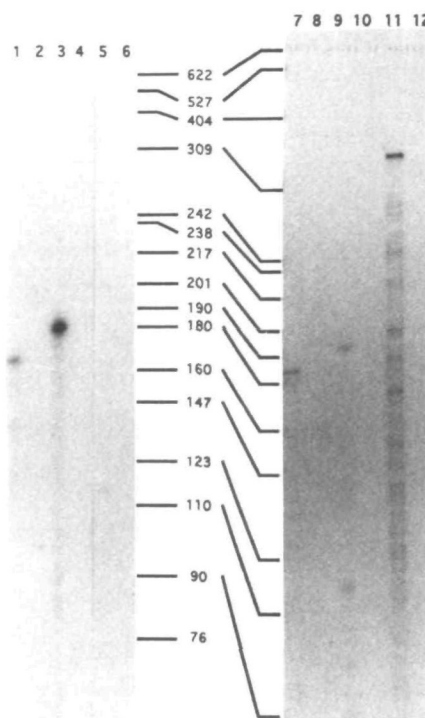
Figure 5. Northern analysis. RNA extracted from *Synechocystis* (lane 2), *Anabaena* (lane 3) and *Calothrix* (lane 4) was analyzed by Northern hybridization on a 5% polyacrylamide, 7 M urea gel with a probe from the *Anacytis* RNase P RNA gene as described in Materials and Methods. Numbers on the left of the panel indicate the size in base pairs of labelled fragments of pBR322 digested with HpaII (lane 1).
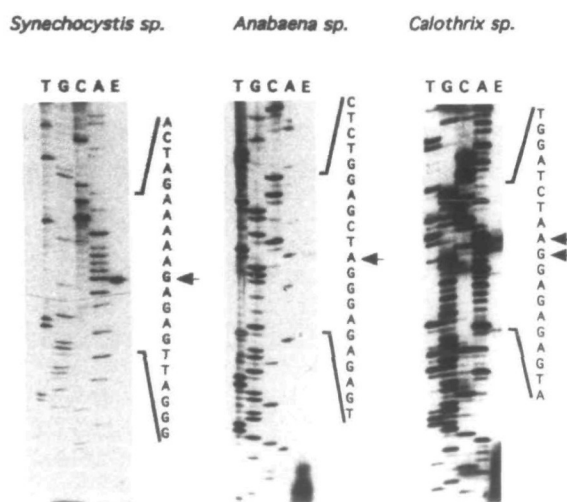
Figure 6. Primer extension analysis. The primer extended products (E) are presented along dideoxy sequencing ladders which lack, in contrast with the extended product, the 5' terminal phosphate. Sequence lanes are presented as the complement of the DNA coding sequences. Arrows indicate potential transcription start sites.

Figure 7. RNase protection analysis. 5 μg of RNA extracted from *Synechocystis* (lanes 1 and 3), *Anabaena* (lanes 5 and 7) and *Calothrix* (lanes 9 and 11) were hybridized to probe S5 (lane 1), probe S3 (lane 3), probe A5 (lane 5), probe A3 (lane 7), probe C5 (lane 9) and probe C3 (lane 11). 15 μg of *E.coli* tRNA was hybridized to probe S5 (lane 2), probe S3 (lane 4), probe A5 (lane 6), probe A3 (lane 8), probe C5 (lane 10) and probe C3 (lane 12). The size in base pairs of labelled fragments of pBR322 digested with HpaII is shown between lane 6 and 7.

primer extension (Fig. 5, 6 and 7). This is the general situation for the reported RNase P RNA genes. However in *E.coli* a maturation step at the 3' end is necessary to generate the mature RNA (14).

The RNase P RNA genes from *Anabaena* and *Calothrix* are followed by short tandemly repeated repetitive (STRR) sequences. This kind of sequences have been described as specific of heterocyst-forming cyanobacteria (26) and generally appear in intergenic regions, immediately upstream or downstream of a gene. In the RNase P RNA genes from *Anabaena* and *Calothrix* there are STRR sequences also within the transcribed region, and thus are present in the functional mature RNA. This is a unique characteristic of these genes. Although STRR sequences are transcribed in many cases, they have never been reported to interrupt coding sequences. As far as I know, the only exception

is the STRR sequence found within an open reading frame present downstream of the *het*A gene from *Anabaena* (27). Although this open reading frame is transcribed, there are no data about its function if any. As a consequence of the insertion of these STRR sequences within the RNase P RNA genes, an helix is increased in length. In *Synechocystis*, the same helix is increased in length, but in this case the insertion is not a STRR sequence. Thus it can be thought that this modified helix is not functionally important for the activity of the enzyme. This idea is supported by the fact that the helix is rather variable in bacteria and is not present in bacteria of the genus *Bacillus*. However, *in vitro* studies indicate that deletion of this helix perturbs the activity of the enzyme (37), probably through the creation of an unstable structure. It is possible that only nucleotides present at the base of the helix, as the highly conserved unpaired A residues, are involved in stabilizing interactions within the core of the enzyme, and the rest of the helix extends to the outside of the structure and can tolerate insertions of unrelated sequences without affecting the activity. It can not be ruled out that these STRR sequences within the RNase P RNAs play some specific functional or regulatory role in the heterocyst-forming cyanobacteria as has been suggested for the intergenic STRR sequences (26, 27). Further *in vitro* studies will be required to solve these questions. It is noteworthy that in the three organisms analyzed here there is a large insertion of a sequence in the same position (although in *Synechocystis* it is not an STRR sequence)

and that this sequence is different in the three organisms, suggesting that it has happened independently. *Anacystis*, which has no equivalent insertion, is in this respect more similar to proteobacteria. In fact, *Anacystis* branches closer to the root of the phylogenetic tree of cyanobacteria than *Synechocystis* or the heterocyst-forming species do (23,38).

A surprising idiosyncrasy of the *Anabaena* sequence is the replacement by A-C of of a canonical base pair at one of the ends of an otherwise highly conserved helix between sequence 56−64 and sequence 429−436. The presence of this helix, and the functional importance of the base pair between nucleotides 64 and 429 has been demonstrated by covariation data and mutagenesis (36), but in every case it is a canonical G-C or A-U base pair.

The new RNase P RNA sequences described in this work have allowed the refinement of the secondary structure model of cyanobacterial RNase P RNA, that was based solely on the *Anacystis nidulans* sequence. Minor corrections on the proposed model have been introduced in order to agree better with the new data (Fig. 4).

The new data presented in this work provide additional sequences to the growing data base of bacterial RNase P RNAs. New additional sequences from diverse cyanobacteria will aide in the study of the evolution of this group of bacteria and in the analysis of the structure and function of RNase P. Interesting information could be obtained from the study of the enzymatic activity of these RNAs and specifically of the role if any of the peculiar extended helix found in the RNAs analyzed here. Work in those directions is in progress.

ACKNOWLEDGEMENTS

REFERENCES

1. Altman,S. (1989) *Adv. Enzymol. Relat. Areas Mol. Biol.* 62, 1−36.
2. Altman,S., Baer,M., Guerrier-Takada,C. and Vioque,A. (1986) *Trends Biochem. Sci.* 11, 515−518.
3. Pace,N. and Smith,D. (1990) *J. Biol. Chem.* 265, 3587−3590.
4. Darr,S.C., Brown,J.W. and Pace,N.R. (1992) *Trends Biochem. Sci.* 17, 178−182.
5. Wang,M.J., Davis,N.W. and Gegenheimer,P. (1988) *EMBO J.* 7, 1567−1574.
6. Guerrier-Takada,C., Gardiner,K., Marsh,T., Pace,N. and Altman,S. (1983) *Cell.*, 35, 849−857.
7. Baer,M. and Altman,S. (1985) *Science* 228, 999−1002.
8. Reich,C., Gardiner,K.J., Olsen,G.J., Pace,B., Marsh,T.L. and Pace,N.R. (1986) *J. Biol. Chem.* 261, 7888−7893.
9. Brown,J.W. and Pace,N.R. (1991) *Biochimie* 73, 689−697.
10. Pace,N.R., Smith,D.K., Olsen,G.J. and James,B.D. (1989) *Gene* 82, 65−75.
11. James,B.D., Olsen,G.J., Liu,J. and Pace,N.R. (1988) *Cell* 52, 19−26.
12. Brown,J.W., Haas,E.S., James,B.D., Hunt,D.A. Liu,J. and Pace,N.R. (1991) *J. Bacteriol.* 173, 3855−3863.
13. Banta,A.B., Haas,E.S., Brown,J.W. and Pace, N.R. (1992) *Nucleic Acids Res.*, 20, 911.
14. Reed,R., Baer,M.F., Guerrier-Takada,C., Donis-Keller,H. and Altman,S. (1982) *Cell* 30, 627−636.
15. Nieuwland,D.T., Haas,E.S. and Daniels,C.J. (1991) *J. Biol. Chem.*, 266, 5689−5695.
16. Hartmann,R.K. and Erdmann,V.A. (1991) *Nucleic Acids Res.* 19, 5957−5964.
17. Rippka,R., Deruelles,J., Waterbury,J.B., Herdman,M. and Stanier,R.Y. (1979) *J. Gen. Microbiol.* 111, 1−61.
18. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor University Press, Cold Spring Harbor.
19. Cai,Y. and Wolk,C.P. (1990) *J. Bacteriol.* 172, 3138−3145.
20. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463−5467.
21. Henikoff,S. (1984) *Gene* 28, 351−359.
22. Mohamed,A. and Jansson,C. (1989) *Plant Mol. Biol.* 13, 693−700.
23. Giovannoni,S.J., Turner,S., Olsen,G.J., Barns,S., Lane,D.J. and Pace,N. (1988) *J. Bacteriol.* 170, 3584−3592.
24. Lill,H. and Nelson,N. (1991) *Plant Mol. Biol.* 17, 641−652.
25. Schneider,G.J., Lang,J.D. and Haselkorn,R. (1991) *Gene* 105, 51−60.
26. Mazel,D., Houmard,J., Castets,A.M. and Tandeau de Marsac,N. (1990) *J. Bacteriol.* 172, 2755−2761.
27. Holland,D. and Wolk,C.P. (1990) *J. Bacteriol.* 172, 3131−3137.
28. Curtis,S.E. (1987) *J. Bacteriol.* 169, 80−86.
29. Bergsland,K.J. and Haselkorn,R. (1991) *J. Bacteriol.* 173, 3446−3455.
30. McCarn,D.F., Whitaker,R.A., Alam,J., Vrba,J.M. and Curtis,S.E. (1988) *J. Bacteriol.* 170, 3448−3458.
31. Houmard,J., Capuano,V., Coursin,T. and Tandeau de Marsac,N. (1988) *J. Bacteriol.* 170, 5512−5521.
32. Mulligan,M.E. and Haselkorn,R. (1989) *J. Biol. Chem.* 264, 19200−19207.
33. Houmard,J., Capuano,V., Colombano,M.V., Coursin,T. and Tandeau de Marsac,N. (1990) *Proc. Natl. Acad. Sci. USA* 87, 2152−2156.
34. Ewart,G.D., Reed,K.C. and Smith,G.D. (1990). *Eur. J. Biochem.* 187, 215−223.
35. Floriano,B., Herrero,A. and Flores,E. (1992) *Mol. Microbiol.* 6, 2085−2094.
36. Haas,E.S., Morse,D.P., Brown,J.W., Schmidt,F.J. and Pace,N.R. (1991) *Science* 254, 853−856.
37. Darr,S.C., Zito,K., Smith,D. and Pace,N.R. (1992) *Biochemistry* 31, 328−333.
38. Bonen,L., Doolittle,W.F. and Fox,G.E. (1979) *Can. J. Biochem.* 57, 879−888.