Conserved and variable domains within divergent RNase P RNA gene sequences of Prochlorococcus strains

Astrid Schön,† Christiane Fingerhut and Wolfgang R. Hess

Author for correspondence: Wolfgang R. Hess. Tel: +49 030 2093 8144. Fax: +49 030 2093 8139. e-mail: wolfgang.hess@rz.hu-berlin.de

RNase P RNA gene (rnpB) sequences were PCR-amplified from different members of the Prochlorococcus group. Aligned nucleotide sequences revealed a variance of up to 27% for rnpB. Comparative secondary structure analysis showed that domains P12, P18 and P19 of these novel ribozyme sequences in particular are highly divergent. Thus, these regions in RNase P RNA might serve as potential targets for deoxyoligonucleotide primers for the identification of specific genotypes of Prochlorococcus and for probing field populations. Phylogenetic trees constructed from RNase P RNA sequences were similar to, but not fully congruent with, those derived previously using sequences of the 16S rRNA gene. However, the application of rnpB sequences allowed a better resolution within clades of very closely related genotypes. As is known from 16S rRNA-based phylogenetic trees, sequences from individual strains clustered according to their physiology and the conditions at the original site of isolation, rather than their geographical origin. All sequences obtained from high-light-adapted strains formed a single coherent clade, as did the four sequences from low-light-adapted strains that were previously isolated from the North Atlantic and the subtropical North Pacific. This suggests a remarkable genetic stability of Prochlorococcus genotypes that thrive under identical ecological conditions.

Keywords: ribonuclease P RNA gene sequence, cyanobacteria, Prochlorococcus, Synechococcus, phylogeny

INTRODUCTION

RNase P is a ubiquitous enzyme essential for tRNA maturation. In bacteria, the catalytic site resides in the RNA subunit, which is able to perform precise endonucleolytic cleavage of tRNA precursors to yield the mature 5' end (Frank & Pace, 1998; Pace & Brown, 1995; Schön, 1999), whereas the small, basic protein subunit improves catalytic efficiency by binding the pre-tRNA 5' flank (Crary et al., 1998; Niranjankumari et al., 1998). The RNA component of RNase P is encoded by rnpB, the protein by rnpA.

Sequences of rnpB from different bacterial groups vary considerably in sequence and length. In addition to several short patches containing invariant nucleotides, only two stretches of sequence conservation are consistently present. Despite this variability, co-variation analysis of sequences covering the whole range of phylogenetic groups has made possible the definition of a universally conserved minimal consensus structure for these catalytically active RNAs (Haas et al., 1994). According to this model, the two conserved regions constitute a long-range base pairing (P4) involved in maintaining the three-dimensional structure of the ribozyme (Chen et al., 1998; Harris et al., 1994, 1997; Westhof & Altman, 1994; Westhof et al., 1996).

Published online ahead of print on 3 December 2001 as DOI 10.1099/ijs.0.01983-0.

A structure-optimized alignment of Prochlorococcus RNase P RNA sequences is available as supplementary material in IJSEM Online (http://ijs.sgmjournals.org/).

† Present address: Institut fur Biochemie, Universitat Leipzig, Talstrasse 33, D-04103 Leipzig, Germany.

Abbreviations: HL, high-light-adapted; LL, low-light-adapted.

The EMBL accession numbers for the RNase P RNA gene sequences reported in this paper are AJ272218–AJ272226 (see Fig. 3).
RNase P RNA sequences of related species are more variable than their 16S rRNAs. Thus, this ubiquitous RNA could be a very useful tool for a more detailed classification of bacteria, including species still unclassified. Previous comparative analyses of *rnpB* sequences from natural samples provided important information on the diversity of natural microbial populations (Brown *et al.* 1996). Sequences and structures of RNase P RNAs were suggested as distinctive markers within *Saccharomonospora*, the *Chlamydiales* and actinomycetes and allowed the inference of evolutionary relationships (Cho *et al.*, 1998; Herrmann *et al.*, 2000; Yoon & Park, 2000). In addition, information about the existing variants of RNase P RNAs provided clues towards the functional relevance of the different domains for the function of the catalytic core (Brown *et al.*, 1996).

Characterization of *rnpB* and the enzymic activity of RNase P RNA from the cyanobacterium *Prochlorococcus marinus* subsp. *marinus* CCMP 1375\(^T\) (strain SS120\(^T\); Hess *et al.*, 1998) has been reported previously. Picophytoplankton species belonging to the genera *Synechococcus* and *Prochlorococcus* are thought to have evolved rapidly and nearly simultaneously (Urbach *et al.*, 1998). Genetic analyses of different *Prochlorococcus* strains indicated the existence of a strong correlation between phylogeny and physiology in this genus (Moore *et al.*, 1998; Rocap *et al.*, 1999; Scanlan *et al.*, 1996; Urbach & Chisholm, 1998; Urbach *et al.*, 1998; West & Scanlan, 1999). The strains of *Prochlorococcus* that occupy the most distal part of 16S rRNA trees are all adapted to high-light conditions (high-light-adapted, HL) and are characterized by a relatively low chlorophyll *a*/*b* ratio. According to such a 16S rRNA-based phylogeny, low-light-adapted (LL) strains belong to at least four different phylogenetic groups, whereas the HL strains can be subdivided into two groups, sometimes called HLI and HLII (West & Scanlan, 1999).

The knowledge of additional RNase P RNA sequences from members of the *Synechococcus/Prochlorococcus* group are described, suggesting variable and conserved regions possibly useful for an analysis of natural samples, and some of the secondary structural elements potentially relevant for catalytic activity are discussed.

### METHODS

**Bacterial strains.** Cultures of various *Prochlorococcus* strains (Table 1) were maintained as described previously (PCC 9511\(^T\), TAK9803-2, *P. marinus* subsp. *marinus* SS120\(^T\), NATL1A, NATL2A; Partensky *et al.*, 1999) or DNA was extracted from cells obtained from F. Partensky at the Biological Station in Roscoff, France (TATL1B, TATL2 and *Synechococcus MINOS1*), or from L. Campbell at the Texas A&M University, College Station, TX, USA (PAC1). DNA extraction and *rnpB* amplification from PAC1 was performed with two different DNA samples, one isolated in 1994 and the other in 1998, after clonal diversity of this culture had been recognized on the basis of phycoerythrin sequences (Penno *et al.*, 2000).

**DNA amplification and sequencing.** Genomic DNA from the different *Prochlorococcus* strains (Table 1) was amplified either with the primer pair cprp5'/cprp3' (Baum *et al.*, 1996) or the alternative primer set cprp5'-OL (CTGAGGAAA-GTCCGGGCT) and cprp3'-OL (GTAAGCCGGGTTC-CTCCGCT). The 5' flanking region of PAC1 *rnpB* was obtained with the primer pair Pm-F5' (CCCCAGACCTACGGTTTAGG) and cprp3'-OL. PCRs were performed with the proof-reading *Pfu* DNA polymerase (Stratagene) in a Hybaid OmniGene PCR incubator (0.5 µM each primer, 200 µM each dNTP, 0.05 U *Pfu* DNA polymerase µl\(^{-1}\) and 1–10 ng template DNA in a total volume of 20 µl). After denaturation for 2 min at 94°C, 35 cycles of PCR were performed (30 s at 94°C, 30 s of annealing at 61°C, 1 min at 72°C). PCR products were purified by agarose gel electrophoresis and eluted with a JETquick kit (Genomed), followed by cloning into pUC19 and *Escherichia coli* DH5α or JM109 (Baum *et al.*, 1996). DNA was sequenced using an ABI 373 DNA sequencer (Applied Biosystems). Several independent clones were sequenced from all PCR-derived constructs to ensure that the native sequences were obtained. The GCG program package version 9.0 (University of Wisconsin, Madison, USA) was used for data analysis. Sequence alignments were obtained with **LINEUP** and **CLUSTAL W**. Homology searches were performed with the programs **FASTA** or **BESTFIT**. Preliminary sequence data

<table>
<thead>
<tr>
<th>Strain</th>
<th>Site of isolation</th>
<th>Ecological type</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCMP 1375(^T) (= SS120(^T))</td>
<td>Sargasso Sea</td>
<td>LL</td>
<td>Chisholm <em>et al.</em> (1992)</td>
</tr>
<tr>
<td>PCC 9511(^T)</td>
<td>Mediterranean Sea</td>
<td>HL</td>
<td>Rippka <em>et al.</em> (2000)</td>
</tr>
<tr>
<td>NATL1A</td>
<td>North Atlantic</td>
<td>LL</td>
<td>Scanlan <em>et al.</em> (1996)</td>
</tr>
<tr>
<td>NATL2A</td>
<td>North Atlantic</td>
<td>LL</td>
<td>Scanlan <em>et al.</em> (1996)</td>
</tr>
<tr>
<td>TAK9803-2</td>
<td>South Pacific Ocean, Takapoto Lagoon</td>
<td>HL</td>
<td>Garczarek <em>et al.</em> (2000)</td>
</tr>
<tr>
<td>TATL1B</td>
<td>South Atlantic</td>
<td>HL</td>
<td>Scanlan <em>et al.</em> (1996)</td>
</tr>
<tr>
<td>TATL2</td>
<td>South Atlantic</td>
<td>HL</td>
<td>Scanlan <em>et al.</em> (1996)</td>
</tr>
<tr>
<td>PAC1</td>
<td>Subtropical North Pacific Ocean</td>
<td>LL</td>
<td>Penno <em>et al.</em> (2000)</td>
</tr>
<tr>
<td><em>Synechococcus</em> sp. MINOS1</td>
<td>Mediterranean Sea</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

### Table 1. List of strains used in this study
for *Prochlorococcus* sp. MIT 9313 was obtained from the DOE Joint Genome Institute (http://spider.jgi-psf.org/JGI_microbial/html/).

**RESULTS AND DISCUSSION**

Eight different sequences were obtained from the seven cultures of *Prochlorococcus* investigated and one from a hitherto uncharacterized marine *Synechococcus* strain (MINOS1). The sequences were 307–335 nt and encompassed about 85% of the full-length *rnpB* gene. The *Prochlorococcus* sequences shared 74–99% similarity. Interestingly, all *Prochlorococcus* sequences have about the same degree of identity (70–75% in the raw alignment) if compared to the two *Synechococcus* sequences, MINOS1 and *Synechococcus elongatus* PCC 6301 (formerly *Anacystis nidulans* PCC 6301; for details of reclassification, see Herdman et al., 2001).

The total genome sequence of *Prochlorococcus* strain MED4 is now available. Its RNase P RNA sequence is identical to that from *P. marinus* subsp. *pastoris* PCC 9511T except for an A-to-T transversion at position 289 of the PCC 9511T *rnpB* sequence. Thus, this situation is very similar to the one nucleotide difference found between the two strains in the 16S rRNA gene (Rippka et al., 2000; GenBank acc. no. AF180967) and shows their very high degree of relatedness.

An improved, structure-based alignment was obtained taking into account conserved and variable elements of secondary and tertiary structure. The alignment is available as supplementary material in IJSEM Online (http://ijs.sgmjournals.org/). It should be noted here that our alignment of the P12 region differs slightly from that suggested in the RNase P database (Brown, 1999); however, these alternative alignments do not lead to significant differences in phylogenetic trees (not shown).

Two different sequences, PAC1A and PAC1B (identity 95%), were amplified from a single culture in the case of PAC1. The multiclonal state of this culture is in agreement with a previous study investigating the variability of phycoerythrin genes within the genus *Prochlorococcus* (Penno et al., 2000) and has also been observed during analysis of 16S rRNA sequences from other *Prochlorococcus* strains, such as TATL1 or

![Fig. 1. RNase P RNA structures. Typical structural elements of *Prochlorococcus* RNAs are represented. The complete structure of PAC1B RNase P RNA (the 3' terminus has been deduced from the known P1 sequence and is drawn in grey) and the core structure of the TAK9803-2 RNA (5' and 3' ends were not determined and are drawn schematically in grey) are shown.](http://ijs.sgmjournals.org)
NATL2. All differences between PAC1A and PAC1B can be localized to the helices P15/16, P18 and P19. In addition, the 5' end of PAC1B rnpB was amplified using primers Pm-F5' and cprp3'-OL. This experiment confirmed a genome arrangement similar to that in SS120T (Hess et al., 1998), with a gene for tRNAArg located immediately upstream on the complementary strand (Hess & Schön, 1999). Interestingly, the PAC1A sequence is more similar to NATL2A than NATL2A is to NATL1A and PAC1A is to PAC1B, obtained from the same culture.

**Structural comparison**

Despite the high variability at the sequence level, the main structural determinants are fully conserved. The postulated secondary structures (Fig. 1) indicate clearly that all investigated RNAs belong to the cyanobacterial type (Banta et al., 1992; Fingerhut & Schön, 1998; Pascual & Vioque, 1994; Vioque, 1992). The long-distance base pairing P6 (Haas et al., 1994), which might be potentially extended to 6 bp in P. marinus subsp. marinus SS120T (Hess et al., 1998), consists of only 5 bp in the PAC1B RNA (Fig. 1). In TAK9803-2 (Fig. 1), and also in TATL1B and TATL2, P6 may be extended to 7 bp if the distal end of P16 is opened (Fig. 2). The RNA of Synechococcus MINOS1, a closely related marine cyanobacterium, may even employ 8 bp for P6 (Fig. 2).

An important structural feature, the extended asymmetrical loop joining stems P15 and P16, is present in all the novel RNase P RNAs described in this paper. It has also been found in other cyanobacterial RNase P RNAs, except those of *Pseudanabaena*, *Oscillatoria* and *Prochlorothrix*. The corresponding domain in *E. coli* contains the canonical 5'-GGU-3' motif, binding the substrate via its 3' CCA end. However, despite the obvious replacement of this binding domain by the P15/P16 loop, the pre-tRNA CCA end remains crucial for efficient turnover in *Prochlorococcus*, as shown previously (Hess et al., 1998). The presence of three consecutive U–U ‘mismatches’ in helix P16 of *P. marinus* subsp. *marinus* SS120T RNase P RNA (Hess et al., 1998) is unique within the cyanobacterial group, as none of the other RNAs contains a similar stretch of unpaired nucleotides within P16 (Fig. 2). Most cyanobacterial RNase P RNAs possess a 2- to 4-base bulge within an otherwise fully paired P16, and this is exactly what was found for *Prochlorococcus* strains PAC1A, PAC1B, TATL1B, TATL2, NATL1A and NATL2A and *Synechococcus* sp. MINOS1, whereas *P. marinus* subsp. *pastoris* PCC 9511T has a fully paired P16 domain. Another peculiarity of the *P. marinus* subsp. *marinus* SS120T RNA, an unpaired U–U in helix P5 (Hess et al., 1998), is restricted to this species and PCC 9511T, whereas P5 is fully paired in all other ribozyme variants.

A structural feature that correlates well with the position of the respective strains in the phylogenetic
The PAC1 culture has previously been characterized as belonging to the LL clade of *Prochlorococcus* (Penno *et al*., 2000; Urbach *et al*., 1998), as is the case for the type strain *P. marinus* subsp. *marinus* CCMP 1375^T^ (equivalent to strain SS120^T^). To obtain the sequence information required to include the complete 5′ domain in our analysis, the 5′ flank of PAC1B, including part of the adjacent *trnR* gene, was amplified. From a total of 37 positions in P3, 13 differ between these two strains (Fig. 1; Hess *et al*., 1998). The marine *Synechococcus* strain MINOS1 was included in this analysis because marine *Synechococcus* are phylogenetically closer to *Prochlorococcus* than to freshwater cyanobacteria. Interestingly, its RNase P RNA structure is much more similar to that of freshwater *Synechococcus* PCC 6301 than to that of any *Prochlorococcus*; in contrast with all *Prochlorococcus* ribozymes, domain P19 is missing. P16 has a ‘double bulge’ and P6 can be potentially extended to 8 bp, both very similar to *Synechococcus* PCC 6301. Furthermore, P12 from MINOS1 closely resembles its counterpart from *Synechococcus* PCC 6301.

Fluorescent in situ hybridization of primers targeted to the 16S rRNA has recently become a powerful tool for the analysis of natural populations of phytoplankton (West & Scanlan, 1999). RNase P RNA might constitute a potential alternative target, given its comparatively high intracellular abundance. Hence, it might be envisaged that RNase P RNA sequences may be used to determine directly the spatial and temporal distribution of particular geno- and/or ecotypes in natural populations. Based on the findings of this study and existing three-dimensional models (Chen *et al*., 1998; Harris *et al*., 1997, 1994; Westhof & Altman, 1994; Westhof *et al*., 1996), P12, P18, P19 and possibly P3 are suggested as the most promising regions, by virtue of their extremely high variability and their easy accessibility, for the design of genotype-specific probes.

**Phylogenetic comparisons**

Phylogenetic trees based on 16S rRNA sequences have become a widely accepted standard for inferring evolutionary relationships. Their comparison to other possible markers, i.e. mainly *rpoC1* (Palenik, 1994; Palenik & Haselkorn, 1992) in the case of *Prochlorococcus*, has firmly confirmed the value of 16S rRNA phylogenetic trees. Thus, alternative markers might be useful only if they allowed a better resolution of closely related species or the application of novel techniques. In cyanobacteria, the 16S rRNA–23S rRNA internal transcribed spacer region (Iteman *et al*., 2000) and, within *Prochlorococcus*, the petB–petD intergenic region (Urbach *et al*., 1998), have previously been investigated towards this goal. Thus, *rnpB*-based phylogenies might be of potential interest because of the high level of variability within the coding region. Phylogenetic trees have previously been inferred successfully from the comparison of RNase P RNA sequences of various species (Cho *et al*., 1998; Herrmann *et al*., 1996, 2000; Yoon & Park, 2000). The sequence identity between the different *rnpB* sequences is 74–99%. Thus, these sequences are considerably less similar to each other than are 16S rRNA sequences from the same organisms (> 97%; Moore *et al*., 1998), and a much higher level of resolution might be expected for *rnpB*-based phylogenetic trees.

A phylogenetic tree was constructed by neighbour-joining analysis with the two *Synechococcus* sequences as outgroups (Fig. 3). All *Prochlorococcus* HL strains belong to a single large cluster; the LL strains PAC1A, 1B, NATL1A and 2A fall into a second single clade, whereas the two other LL strains are located in separate places. The same basic branching order was obtained when the optimality criterion was set to maximum-likelihood or maximum-parsimony. However, in some analyses, the positions of NATL1A,
NATL2A and PAC1A could not be resolved. In principle, the trees obtained on the basis of rnpB sequences are of similar topology to those constructed from rRNA sequences. The HL strains form a coherent cluster and all LL strains branch off earlier and fall into different groups. Thus, the strains cluster according to their ecophysiological and ecological characteristics, rather than their geographical sites of isolation. This supports results obtained earlier with 16S rRNA (Rocap et al., 1999; Urbach & Chisholm, 1998) or rpoCI sequences (Palenik, 1994). The clear resolution within the two main clusters in the tree is very interesting. Among deep (LL) strains, the North Atlantic and Pacific isolates are surprisingly similar. In the tree shown in Fig. 3, the PAC1A sequence is even closer to NATL1A from the North Atlantic than to PAC1B obtained from the same culture. This suggests a remarkable genetic stability of Prochlorococcus genotypes that thrive under identical ecological conditions. Within the HL cluster, TATL2 and TAK9803-2 represent the two most closely related forms. In 16S rRNA phylogenies, both strains have occasionally been included with additional sequences into HL cluster II, which is separate from HL cluster I containing strains PCC 9511³ and TATL1B (West & Scanlan, 1999). It is intriguing that RNase P phylogenies suggest such a possible branching of these HLI sequences from within the other HL-adapted genotypes, but further analyses including additional strains might help to substantiate this observation further.

Thus, our data show the high versatility of RNase P as a potential marker for the phylogeny and molecular ecology of Prochlorococcus and specifically for the investigation of closely related genotypes.

ACKNOWLEDGEMENTS

This work was supported by grants from the Deutsche Forschungsgemeinschaft to A. S. (Scho 515/4-1) and the European Union (PROMOLEC; MAS3-CT-97-0128) to W. R. H. We thank R. Ripka, Paris, for providing a culture of PCC 9511³, F. Partensky, Roscoff, for providing samples of MINOS1, TATL1B, TATL2B, TAK9803-2, NATL1A, NATL2A and SS120², L. Campbell, College Station, for an aliquot of PAC1, and F. Larimer and S. W. Chisholm for permission to include the preliminary sequence data obtained from the DOE Joint Genome Institute at http://www.jgi.doe.gov/JGI_microbial/html/prochlorococcus/prochlo_pickastrain.html for Prochlorococcus strain MIT 9313.

REFERENCES


