Comparison of conserved structural and regulatory domains within divergent 16S rRNA–23S rRNA spacer sequences of cyanobacteria

Isabelle Iteman, Rosmarie Rippka, Nicole Tandeau de Marsac and Michael Herdman

Author for correspondence: Isabelle Iteman. Tel: +33 1 45 68 84 16. Fax: +33 1 40 61 30 42. e-mail: iiteman@pasteur.fr

Unité de Physiologie Microbienne (CNRS URA 1129), Département de Biochimie et Génétique Moléculaire, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris Cedex 15, France

PCR amplification of the internal transcribed spacer (ITS) between the 16S rRNA and 23S rRNA genes of the cyanobacterium *Nostoc* PCC 7120 gave three products. Two represented true ITS regions of different sizes, while the third was a heteroduplex. The longer spacer (ITS-L) contained 512 nucleotides and carried tRNA^{Ile} and tRNA^{Ala} genes, separated by a large stem–loop structure (V2) composed of short tandemly repeated repetitive sequences. Both tRNA genes, and the 5′ half of the intervening stem, were absent from the shorter spacer (ITS-S), of length 283 nucleotides, which was otherwise almost completely identical to ITS-L. The two spacer regions of *Nostoc* PCC 7120 were aligned to published ITS sequences of cyanobacteria, the cyanelle of *Cyanophora paradoxa* and *Escherichia coli*. Although the ITS regions of cyanobacteria vary in length from 283 to 545 nucleotides and contain either both tRNA^{Ile} and tRNA^{Ala} genes, only the tRNA^{Ile} gene, or neither, there is no correlation between ITS size and coding capacity for tRNAs. Putative secondary structures were determined for the deduced transcripts of the *rrn* operons of several cyanobacteria and were compared to that of *E. coli*. Highly conserved motifs important for folding and for maturation of the rRNA transcripts were identified, and regions homologous to bacterial antiterminators (box B–box A) were located. The conserved and variable regions of the cyanobacterial ITS are potential targets of PCR primers and oligonucleotide probes for detection and identification of cyanobacteria at different taxonomic levels.

Keywords: antitermination, cyanobacterium, ITS, rRNA, transfer RNA

INTRODUCTION

Bacterial rRNA genes are commonly organized in an operon in the order 16S rRNA–23S rRNA–5S rRNA, each rRNA gene being separated by an internal transcribed spacer (ITS) region (see Srivastava & Schlessinger, 1990, for a review). Secondary structure models of the complete operon are available for only a limited number of bacteria. In all cases examined, however, the primary transcript appears to undergo pairing of complementary regions to produce two double-stranded stems that carry the unprocessed 16S rRNA and 23S rRNA sequences, and contain sites necessary for processing and release of the mature rRNA molecules. Premature termination of transcription is prevented by antiterminator sites (box A); these may be present either as paired sequences within both double-stranded processing stems as in, for example, *Enterococcus* spp. (Naïmi *et al*., 1997) or as unpaired regions immediately adjacent to the processing stems as in *Escherichia coli* (Berg *et al*., 1989). The box A sequences are reasonably well conserved within bacteria and are immediately preceded by a box B stem of more variable sequence (see Condon *et al*., 1995) whose secondary structure appears to be more important than the primary sequence (Berg *et al*., 1989; Condon *et al*., 1995). In

Abbreviations: ITS, internal transcribed spacer; STRR, short tandemly repeated repetitive.

The GenBank accession numbers for the sequences reported in this paper are AF180968 and AF180969 for ITS-L and ITS-S, respectively.

addition to antitermination and processing sites, the 16S rRNA–23S rRNA spacer often contains one or two tRNA genes (either $tRNA^{Glu}$, $tRNA^{Ala}$, or both $tRNA^{Ala}$ and $tRNA^{Ile}$). The only known exceptions to this general rRNA operon organization occur in *Thermus thermophilus* (see Srivastava & Schlessinger, 1990, for a review), *Leptospira interrogans* (Fukunaga & Mifuchi, 1989), *Borrelia* spp. (Ojaimi *et al.*, 1994) and *Wolbachia pipientis* (Bensaadi-Merchermek *et al.*, 1995), where the rRNA genes are separated into distinct 16S rRNA and 23S rRNA operons.

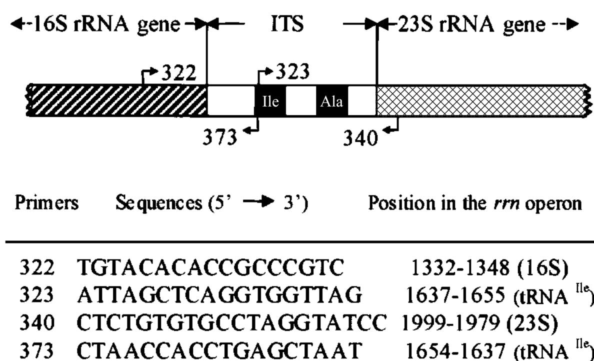
Cyanobacteria have colonized many ecological niches and have developed diverse strategies to survive in greatly different environments. Early endosymbiotic forms gave rise to the photosynthetic organelles (chloroplasts and cyanelles) of eukaryotes (see Giovannoni *et al.*, 1988), thus sharing their ability to perform oxygenic photosynthesis. Extant free-living forms exhibit varied physiological properties and range in morphology from simple unicellular to complex filamentous organisms (Rippka *et al.*, 1979; Castenholz & Waterbury, 1989). Under conditions of combined nitrogen limitation, many filamentous strains such as *Nostoc* PCC 7120 are able to differentiate heterocysts, specialized cells responsible for the aerobic fixation of molecular N_2 (Wolk, 1996). This morphological and physiological diversity is mirrored by extensive genetic variability, cyanobacterial genomes ranging in mean DNA base composition from 32 to 71 mol% G+C (Herdman *et al.*, 1979b; M. Herdman & R. Rippka, unpublished) and in complexity from 2·0 to 13·2 Mbp (Herdman *et al.*, 1979a; M. Herdman & R. Rippka, unpublished). Consequently, it cannot be excluded that transcriptional control and processing mechanisms may vary between different members of the cyanobacterial phylum. However, little is known about the organization of their rRNA operons. Limited data have been obtained from the physical or genetic maps of the unicellular strains *Synechococcus* PCC 6301 (Tomioka *et al.*, 1981), *Synechococcus* PCC 7002 (Chen & Widger, 1993) and *Synechocystis* PCC 6803 (Kaneko *et al.*, 1996). Further information is available from the sequences of the 16S rRNA–23S rRNA ITS domain in members of several cyanobacterial genera that include the unicellular *Synechococcus* PCC 6301 (Tomioka & Sugiura, 1984), *Synechocystis* PCC 6803 (Kaneko *et al.*, 1996) and 47 strains of the genus *Microcystis* (Otsuka *et al.*, 1999), the filamentous (non-heterocystous) *Arthrospira* PCC 7345 (Nelissen *et al.*, 1994), *Spirulina* PCC 6313 (Nelissen *et al.*, 1994) and *Trichodesmium* NIBB 1067 (Wilmotte *et al.*, 1994), the heterocystous *Nodularia* BCNOD 9427 (Hayes & Barker, 1997) and a strain of uncertain generic identity, '*Mastigocladus* HTF' strain PCC 7518 (G. Van der Auwera & A. Wilmotte, personal communication). In addition, the sequence of the ITS of the photosynthetic cyanelle of *Cyanophora paradoxa* has been determined (Janssen *et al.*, 1987).

The cyanobacterial ITS regions investigated to date vary in size from 354 to 545 nucleotides (287 in the cyanelle) and, for those that have been sequenced, only a single ITS species was found in each strain. This is surprising, since sequence and length information on the 16S rRNA–23S rRNA spacer in other bacterial groups suggests that considerable variation can occur not only between species but also between the alleles of the rRNA operon within a single strain (Gürtler & Stanisich, 1996). Consistent with other investigations (Lu *et al.*, 1997; Neilan *et al.*, 1997), we have indeed observed multiple ITS products in PCR amplifications of many cyanobacteria, in particular among the filamentous heterocystous strains (Iteman *et al.*, 1999, and unpublished data). However, no comparative analyses are available for the ITS regions that have been sequenced so far, and the differences between spacer regions of different lengths within a single cyanobacterial genome have not been examined. Furthermore, the regions involved in coordinated transcription of cyanobacterial *rrn* operons and maturation of their products have not yet been identified, and anti-terminator box B–box A motifs have been reported to be undetectable in *Synechococcus* PCC 6301 (Berg *et al.*, 1989). We have therefore sequenced and analysed the PCR products corresponding to two 16S rRNA–23S rRNA spacer domains of different sizes in *Nostoc* PCC 7120, which contains four *rrn* operons (Ligon *et al.*, 1991). The two distinct sequences have been aligned with a selection of those available from other cyanobacteria, *E. coli* and the photosynthetic cyanelle of *C. paradoxa*. The putative secondary structures of some deduced rRNA transcripts have been compared, and anti-terminator and structural motifs potentially implicated in the control of *rrn* transcription and maturation in cyanobacteria have been identified.

METHODS

Organisms and culture conditions. The two axenic strains of cyanobacteria from the Pasteur Culture Collection of Cyanobacteria (PCC), maintained in our laboratory, were grown to an OD_{750} (measured in a Kontron Uvikon 933 spectrophotometer) of 0·8–1·0 in 500 ml volumes at 28 °C under white light (Osram Universal White) with a photosynthetic photon flux density of 30 µmol quanta m^{-2} s^{-1} (measured with a LICOR LI-185B quantum/radiometer/photometer equipped with a LI-190SB quantum sensor) in liquid media supplemented with 10 mM $NaHCO_3$ and 0·19 mM $NaCO_3$ and gassed with 1% v/v CO_2 in air. Medium BG-11 was employed for the unicellular *Synechocystis* PCC 6803, and the filamentous heterocystous strain *Nostoc* PCC 7120 was cultivated in both BG-11 and $BG\text{-}11_0$ (Rippka *et al.*, 1979). The latter organism, previously proposed as *Anabaena* sp. (Rippka *et al.*, 1979) has been renamed as *Nostoc* sp. (Rippka & Herdman, 1992) on the basis of the DNA–DNA hybridization results of Lachance (1981). *E. coli* JM109 used in the cloning experiments was grown at 37 °C in Luria–Bertani medium supplemented with ampicillin (50 µg ml^{-1}).

Amplification of the ITS regions. PCR amplifications were performed with DNA of *Nostoc* PCC 7120 and *Synechocystis* PCC 6803 isolated by the mini extraction method (Cai & Wolk, 1990) or directly with cell lysates obtained by five alternating cycles of freezing in liquid nitrogen and thawing at 50 °C. A set of primers (322 and 340) designed initially for sequencing was used to specifically amplify the part of the rRNA operon containing the ITS region. The sequence and



Primers	Sequences (5' → 3')	Position in the *rrn* operon
322	TGTACACACCGCCCGTC	1332-1348 (16S)
323	ATTAGCTCAGGTGGTTAG	1637-1655 (tRNA Ile)
340	CTCTGTGTGCCTAGGTATCC	1999-1979 (23S)
373	CTAACCACCTGAGCTAAT	1654-1637 (tRNA Ile)

Fig. 1. Schematic representation of an rRNA operon containing an ITS with two tRNA genes. The table shows the primers used for amplification and sequencing (322 and 340), or uniquely for sequencing (323 and 373). Their positions are relative to the nucleotide sequence of the *Synechocystis* PCC 6803 rRNA operon. The two black boxes represent the tRNA genes.

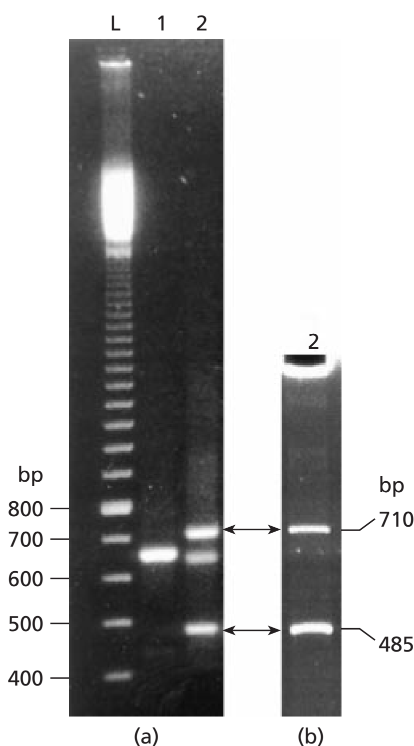


Fig. 2. Analysis of PCR products corresponding to the ITS amplification (primers 322–340) in agarose gel (a) and acrylamide denaturating gel (b). L, molecular mass ladder (100 bp, Pharmacia); lane 1, *Synechocystis* PCC 6803; lane 2, *Nostoc* PCC 7120.

the position of each primer are indicated in Fig. 1. The PCR mixture contained 10 µl *Taq* (10×) commercial buffer, 10 µl lysate or purified DNA (100–500 ng), 150 µM of each dNTP, 500 ng of each primer and 2·5 U *Taq* polymerase (Appligene). The total reaction volume was 100 µl. After an initial cycle consisting of 3 min at 95 °C, 2 min at 55 °C and 30 s at 72 °C, 30 cycles of amplification were started (1·5 min at 95 °C, 2·5 min at 55 °C and 3 min at 72 °C). The termination cycle was 7 min at 72 °C. The PCR products were migrated either on 1·5% (w/v) agarose gels or on denaturing gels containing 8% (w/v) polyacrylamide and urea (6 M) in 0·25×Tris/borate/EDTA buffer (Sambrook *et al.*, 1989), and visualized by staining with ethidium bromide.

Cloning and sequencing of the ITS of *Nostoc* PCC 7120. The PCR products of three independent reactions were mixed and cloned in the pGEM-T vector (Promega) after purification with the Wizard PCR kit (Promega). Competent *E. coli* JM109 cells were transformed and recombinant plasmids were purified from white colonies by the alkaline method (Birnboim & Doly, 1979). Recombinant plasmids, containing either of the two ITS regions observed in *Nostoc* PCC 7120 by gel electrophoresis of the PCR products, were identified by the size of their inserts following amplification using primers complementary to the flanking M13 primer sites located on the vector, and by partial sequencing. Three clones of each were mixed and sequencing of both strands of the inserts was performed (T7 sequencing kit, Pharmacia) using the M13 primer sites and the four primers indicated (Fig. 1). Primer 322 is identical to primer 14 of Wilmotte *et al.* (1993); primer 323 is primer 17 of Wilmotte *et al.* (1993) and 373 is complementary to it; primer 340 corresponds to primer 18 of the latter authors with the *Not*I site and TTT extension omitted.

Sequence alignment. The ITS sequences of *Nostoc* PCC 7120 were first aligned to a selection of the available ITS sequences of free-living cyanobacteria, the cyanelle of *C. paradoxa* and the *rrnA* operon of *E. coli* by large block identity using the software Macaw v. 2.05 (Multiple Alignment Construction and Analysis Workbench, NCBI). The alignment was then manually refined using Genedoc v. 2.4 (Nicholas & Nicholas, 1997) by reference to our secondary structure models (see below). The aligned sequences (Fig. 3), with GenBank accession numbers in parentheses, are: PCC 7120, *Nostoc* PCC 7120 (this study); BCNOD 9427, *Nodularia* BCNOD-9427 (AJ224448); PCC 7518, '*Mastigocladus* HTF' strain PCC 7518 (G. Van der Auwera & A. Wilmotte, personal communication); PCC 6313; *Spirulina* PCC 6313 (X75045); NIBB1067, *Trichodesmium* NIBB1067 (X72871); PCC 7345, *Arthrospira* PCC 7345 (X75044); PCC 6301, *Synechococcus* PCC 6301 (K01983); PCC 6803, *Synechocystis* PCC 6803 (D90916); TC8, *Microcystis* TC8 (AB015386); Cyanelle, *Cyanophora paradoxa* cyanelle (M19493); *E. coli*, *E. coli* K-12 MG1655 *rrnA* operon (AE000460, AE000461). The sequence of *Microcystis* strain TC8 was chosen as representative of the 47 virtually identical (Otsuka *et al.*, 1999) sequences currently available.

Secondary structure determination. To simplify the construction of secondary structure models, the 16S rRNA, 23S rRNA, 5S rRNA and the 3′ end of the operon were deleted from the predicted transcript sequences. All of the resulting sequences were folded using RNAdraw (Matzura & Wennborg, 1996) and the secondary structures were employed to improve the sequence alignment. For *Synechococcus* PCC 6301, *Synechocystis* PCC 6803 and the cyanelle of *C. paradoxa*, the only cyanobacterial strains for which the sequence of an entire operon was available, and for the *E. coli* K-12 MG1655 *rrnA* transcript, the secondary structures were exported from RNAdraw in mfold ct format for further refinement with the program RnaViz (De Rijk & De Wachter, 1997). These models permitted us to define the structure of the transcript in the 16S rRNA–23S rRNA spacer region of *Nostoc* PCC 7120.

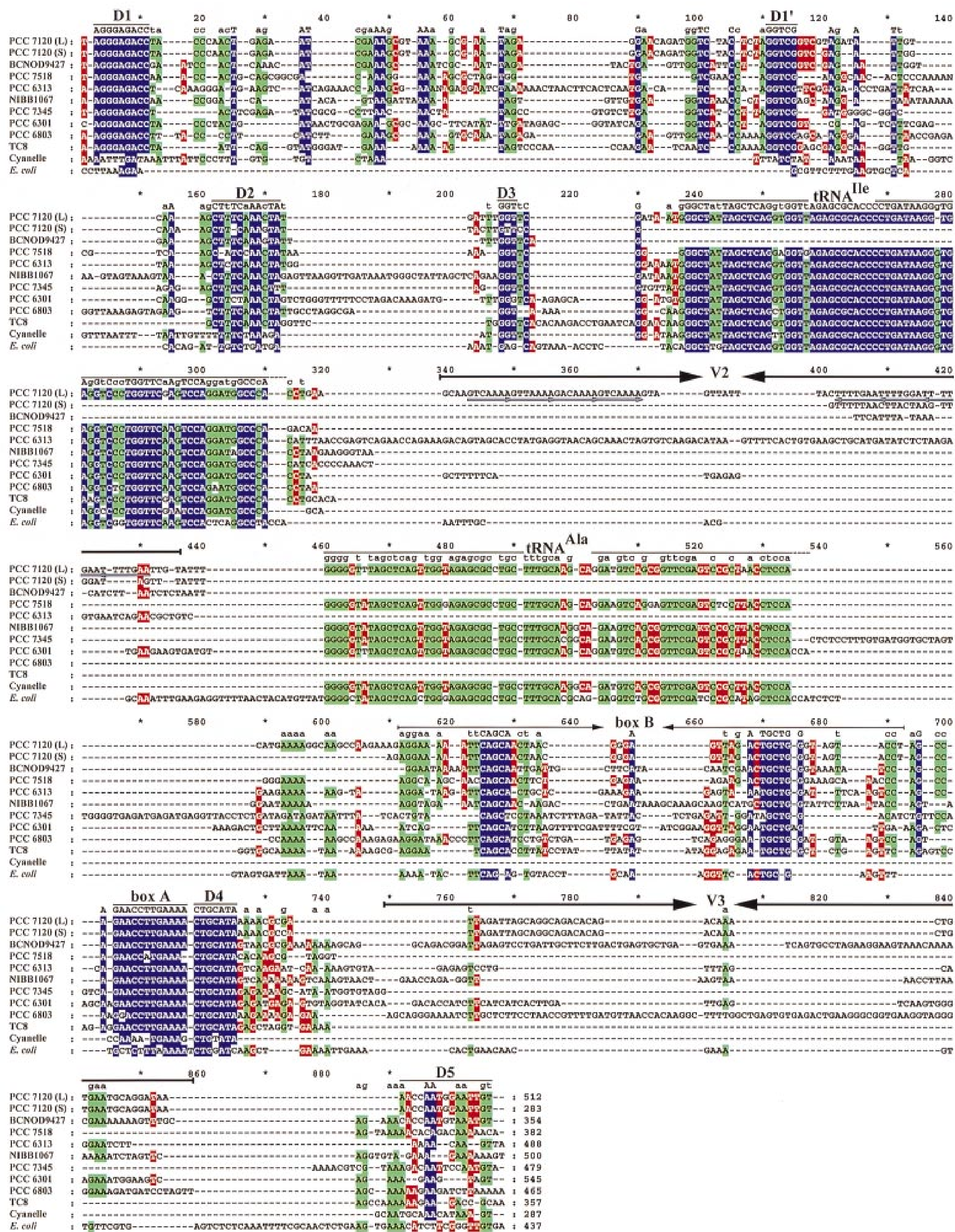


Fig. 3. For legend see facing page.

RESULTS

Amplification of the ITS

The amplified fragments corresponded in size to the length of the ITS plus a total of 200 bp from the 3′ end of the 16S rRNA gene and from the 5′ end of the 23S rRNA gene (Fig. 1). The PCR products, separated on agarose gels, showed two different profiles (Fig. 2a). The unicellular *Synechocystis* PCC 6803 exhibited a single band (665 bp), in agreement with sequence data available from GenBank (accession numbers D64000 and D90916 for the two identical *rrn* operons of this strain). In contrast, three bands (485, 650 and 710 bp) were consistently observed in each of 10 individual PCR amplifications for the filamentous *Nostoc* PCC 7120, the intermediate fragment (650 bp) being in all cases less intense. Since PCR products of 16S rRNA–23S rRNA spacer amplifications may contain artificial bands corresponding to heteroduplexes that result from association of single strands of the different ITS species (Jensen *et al*., 1993), we also separated the PCR amplicons in a denaturing gel. Under these conditions only two bands (485 and 710 bp) were visible (Fig. 2b), indicating that the intermediate fragment of 650 bp was indeed a heteroduplex. Therefore, the four *rrn* operons of *Nostoc* PCC 7120 (Ligon *et al*., 1991) comprise two sets of clearly distinguishable ITS species. To verify that differentiated cells exhibit the same pattern as vegetative cells, cultures were grown in the presence (medium BG-11) or absence (medium BG-11_0) of a source of combined nitrogen, the former condition repressing heterocyst differentiation completely; the two ITS species were detected in the same stoichiometry in both cultures (data not shown).

Sequence alignment of the ITS regions

The sequenced 16S rRNA–23S rRNA spacers of *Nostoc* PCC 7120 have an exact length of 283 bp (ITS-S) and 512 bp (ITS-L), confirming the sizes obtained by DNA migration. Although it has been suggested that the extensive variations in sequence render the alignment of cyanobacterial 16S rRNA–23S rRNA spacer regions impossible (Wilmotte, 1994), the two new sequences, together with those of eight cyanobacterial ITSs of unicellular and filamentous strains, the ITS of the *rrnA* operon of *E. coli* and the intergenic domain of the cyanelle of *C. paradoxa*, were successfully aligned (Fig. 3) by using a combination of primary and secondary structural features. The tRNA genes and other conserved blocks (indicated in colour in Fig. 3) were easily identified by analysis with Macaw and this permitted the construction of a first sequence alignment, in which the non-conserved regions were not treated. Secondary structure models were then made with RNAdraw for each sequence. These models enabled us to validate the alignment of conserved domains such as D1, D1′, D2, D3, D4 and box A, and to deduce their roles in the formation of the structure of the transcript or in the control of transcription (see below). The models also revealed that extensive unaligned variable regions of the ITS are involved in the formation of stem–loop structures (regions V2, V3 and box B; see below). It was thus possible to manually improve the alignment by positioning such structures with respect to their 5′ basal extremities, their complementary paired regions and 3′ extremities, and their respective terminal loops. The three sets of opposing arrows in Fig. 3 indicate the two sides of each stem–loop structure; the gap between the arrows contains the terminal loop of each structure. As a consequence of their high degree of sequence divergence and variation in length, the alignment of the remaining regions is more subjective, although it is again based on the secondary structure models.

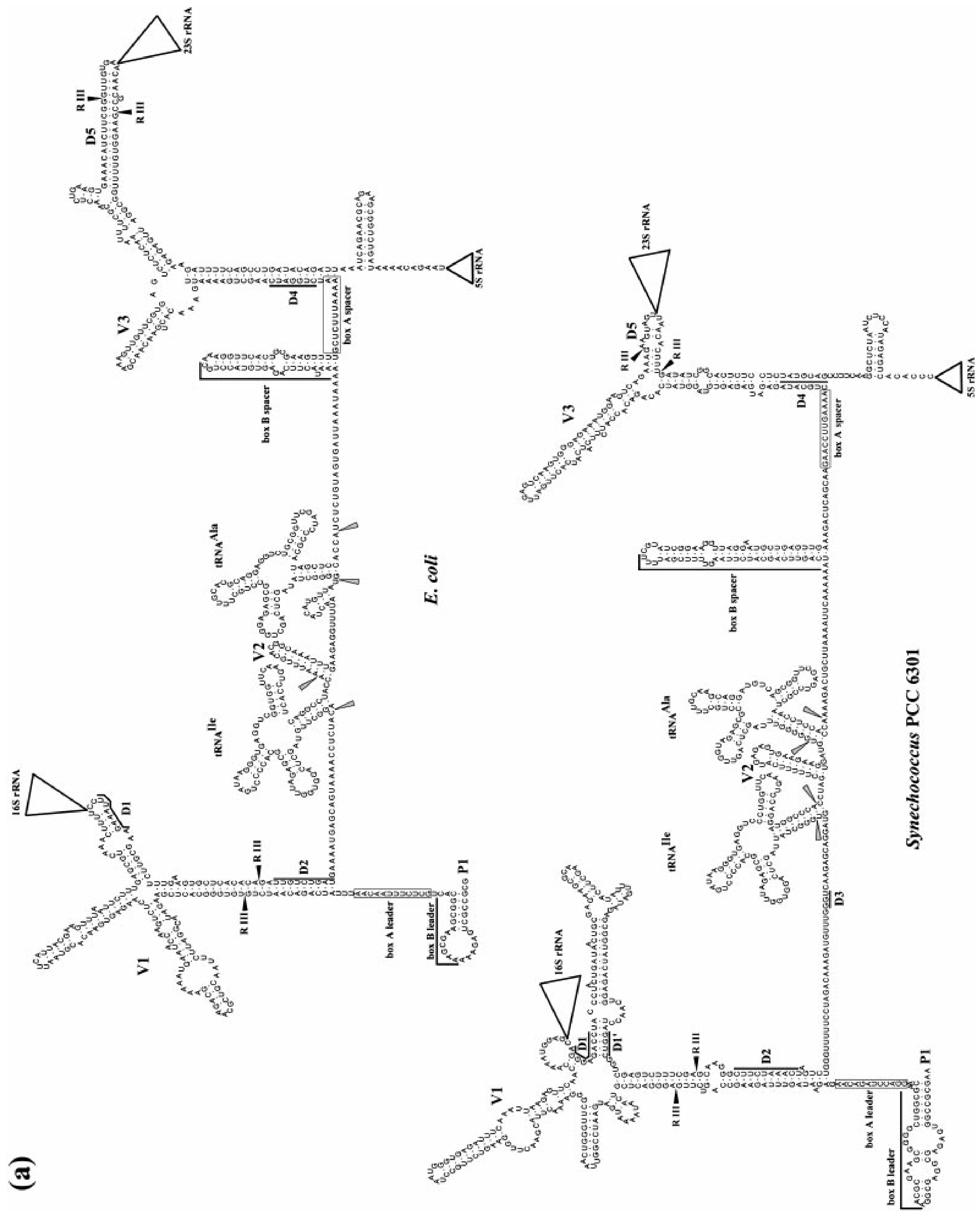
tRNA genes of the ITS regions

The two largest conserved blocks in the sequence alignment, located between positions 237–310 and 460–534, represent respectively the $tRNA^{Ile}$ and $tRNA^{Ala}$ genes (Fig. 3). Both are present in the ITS-L of *Nostoc* PCC 7120 and in the ITS of '*Mastigocladus* HTF' strain PCC 7518, *Trichodesmium* NIBB1067, *Arthrospira* PCC 7345, *Synechococcus* PCC 6301 and the cyanelle of *C. paradoxa*. Other strains lack either $tRNA^{Ala}$ (*Spirulina* PCC 6313, *Synechocystis* PCC 6803 and *Microcystis* TC8) or both tRNA genes (*Nodularia* BCNOD 9427 and the ITS-S of *Nostoc* PCC 7120). In a survey of the sizes of the ITS of six strains of *Nodularia* from the PCC, however, we have found large and small ITS amplicons in all (data not shown). For four of these strains the size of the small ITS was similar (330–350 nucleotides) to that of *Nodularia* BCNOD 9427 (354 nucleotides), whereas the large ITS ranged from 630 to 680 nucleotides. This suggests that *Nodularia* BCNOD 9427 may contain an equivalent to ITS-L and that only the ITS-S has been sequenced.

Secondary structure of putative *rrn* transcripts

With the exception of *Nostoc* PCC 7120, for which only the sequence of the 16S rRNA–23S rRNA ITS region is available, the secondary structures of the transcripts (Fig. 4a, b) commence at the putative transcription start site of the promoter P1 and include the 5′ leader sequence upstream from the 16S rRNA, plus the 16S rRNA–23S

Fig. 3. Alignment of the nucleotide sequences of the ITS regions of free-living cyanobacteria, the photosynthetic cyanelle of *C. paradoxa* and the *rrnA* operon of *E. coli*. Percentage identity is represented by three colours: blue for at least 80%, green for 60% and red for 40%, gaps being treated as differences. The different conserved domains (D1–D5), the tRNA genes and the antiterminator (box B and box A) are indicated and named as in the secondary structures (Fig. 4). The opposing arrows indicate the positions and orientation of the complementary sides of box B and of the major variable stems V2 and V3, the loops of these structures being aligned between the arrow heads. The short grey arrows below the sequence of the ITS-L of *Nostoc* PCC 7120 indicate the complementary STRR motifs (4×7 bases) in region V2. PCC 7120 (L) and PCC 7120 (S) are the long and the short ITS, respectively.


(a)
E. coli
Synechococcus PCC 6301
16S rRNA
23S rRNA
5S rRNA
tRNA^Ile
tRNA^Ala
V1
V2
V3
D1
D1'
D2
D3
D4
D5
R III
box A leader
box B leader
box A spacer
box B spacer
P1

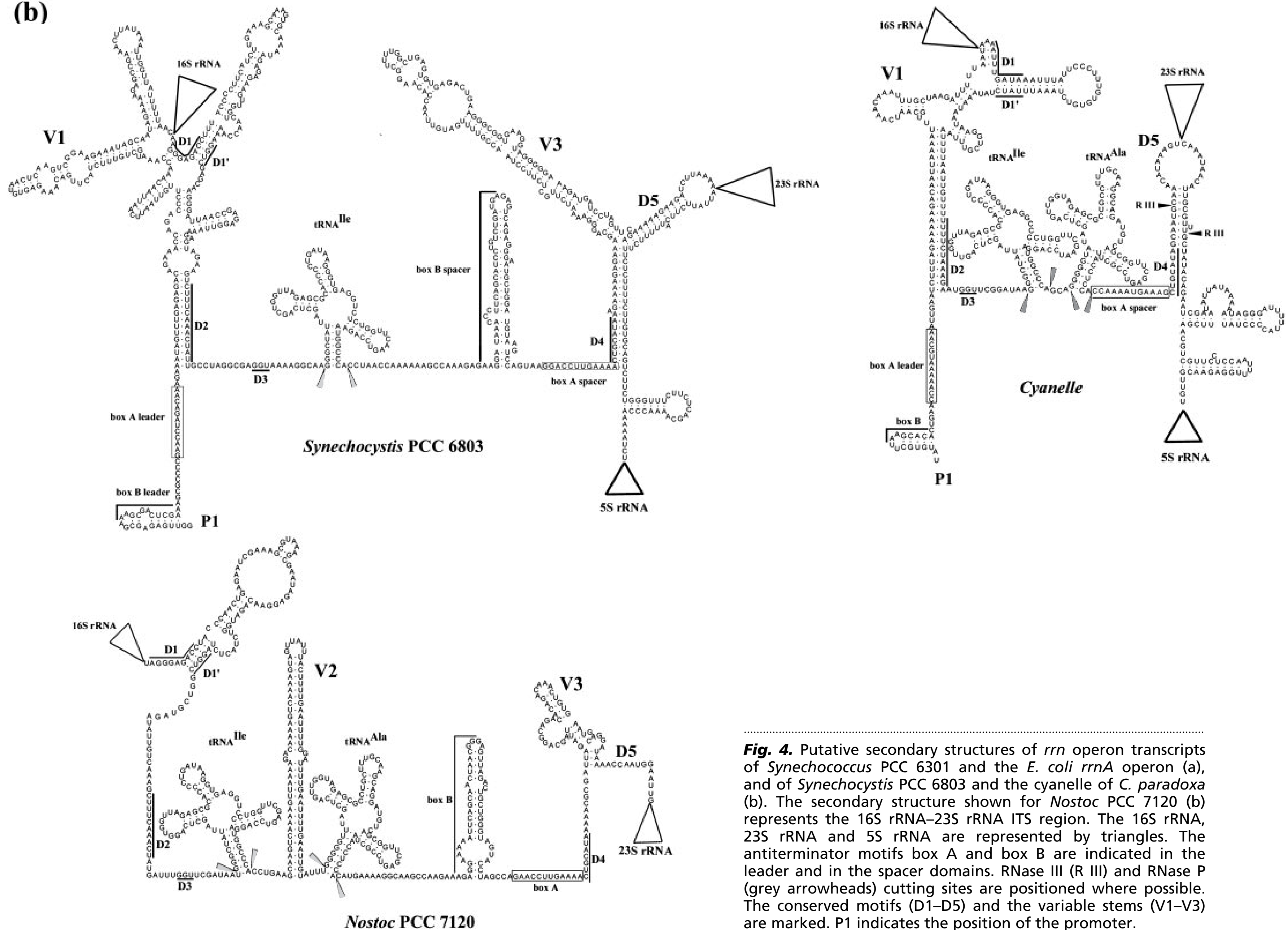


Fig. 4. Putative secondary structures of *rrn* operon transcripts of *Synechococcus* PCC 6301 and the *E. coli rrnA* operon (a), and of *Synechocystis* PCC 6803 and the cyanelle of *C. paradoxa* (b). The secondary structure shown for *Nostoc* PCC 7120 (b) represents the 16S rRNA–23S rRNA ITS region. The 16S rRNA, 23S rRNA and 5S rRNA are represented by triangles. The antiterminator motifs box A and box B are indicated in the leader and in the spacer domains. RNase III (R III) and RNase P (grey arrowheads) cutting sites are positioned where possible. The conserved motifs (D1–D5) and the variable stems (V1–V3) are marked. P1 indicates the position of the promoter.

rRNA ITS and the 23S rRNA–5S rRNA spacer. Despite large variations in the length of the transcripts, the secondary structures show many highly conserved regions (see also Fig. 3). Near the 5′ end of the transcripts, a stem that carries the 16S rRNA is formed by pairing of part of the ITS with the 5′ leader sequence of the operon (Fig. 4a, b). In all organisms examined, this stem has a conserved domain (D2) at the base. However, the remainder of this promoter region of the operon is highly variable (Fig. 4a, b), since it contains one to three stem–loop structures (region V1). At the beginning of the ITS, immediately following the 16S rRNA, a conserved domain (D1) containing 9 bases is observed in all free-living cyanobacterial strains examined (Fig. 3). This domain pairs both with part of the V1 region preceding the 16S rRNA, and with a second conserved domain (D1′) of the ITS, to form the basal part of a stem–loop structure (Fig. 4a, b). This stem, except in the D1/D1′ region, is highly variable and ranges in length from a minimum of 35 bases in the cyanelle to 85 bases in *Spirulina* PCC 6313 (Fig. 3). In *E. coli*, the sequence corresponding to D1 is entirely paired with a 5′ leader region (Fig. 4a).

The unpaired domain situated between D2 and the tRNA region contains a short conserved sequence (D3) flanked upstream and downstream by variable-sized regions of 2–35 bases and 4–26 bases, respectively (Fig. 3). The stem V2 (Figs 3, 4a and 4b) observed immediately after the $tRNA^{Ile}$ in *Nostoc* PCC 7120 (ITS-L), *Spirulina* PCC 6313 and *Synechococcus* PCC 6301 varies in length from 24 to 96 bases and, although shorter, is also present in the *E. coli* sequence. In *Nostoc* PCC 7120, this stem is composed of four copies of short tandemly repeated repetitive (STRR) sequences: 5′-G(A/T)(C/T)AAAA-3′ in the 5′ part of the helix and complementary 5′-TTTTG(A/G)A-3′ sequences in the 3′ part. The tRNA domain is followed by conserved antiterminator regions (boxes A and B; see below) and then by another stem that carries the 23S rRNA and associated structures. This processing stem is formed by pairing of the ITS with part of the 23S rRNA–5S rRNA spacer region and has a highly conserved domain (D4) at the base (Fig. 4a, b). Domain D5, situated at the 3′ extremity of the ITS and immediately preceding the 23S rRNA, is less conserved than the D1/D1′ region adjacent to the 16S rRNA and, as in *E. coli*, is paired with part of the 23S rRNA–5S rRNA spacer (Fig. 4a). The D5 region is preceded in 8 of the 12 sequences examined by a variable stem–loop structure (V3) formed entirely by pairing between parts of the ITS and containing up to 108 bases; this is absent from '*Mastigocladus* HTF' strain PCC 7518, *Arthrospira* PCC 7345, *Microcystis* TC8 and the cyanelle (Fig. 3).

Antitermination sequences and maturation sites

The bacterial *rrn* operon normally contains antiterminator sites that are essential for maintaining transcription of the rRNA genes in the correct stoichiometry (Condon *et al*., 1995); box A sequences occur either within each of the two processing stems or immediately adjacent to them, and in both cases they are preceded by a box B stem structure. By comparison of the secondary structures (Fig. 4a, b), we identified a motif of 12 nucleotides in the cyanobacterial spacer regions that corresponds to antiterminator box A. This is immediately adjacent to domain D4 at the base of the 23S rRNA processing stem in all sequences analysed and is extremely well conserved in the free-living cyanobacteria (consensus 5′-GAACCTTGAAAA-3′) but differs in the cyanelle of *C. paradoxa* (Fig. 3). Box A is immediately preceded in all cases, except the cyanelle, by a stem–loop structure equivalent to box B of *E. coli*, situated 7–64 bases 3′ to the tRNA (Figs 3, 4a and 4b). We also found a motif corresponding to box A in the leader region 5′ to the 16S rRNA of *Synechococcus* PCC 6301, *Synechocystis* PCC 6803 and the cyanelle, the only organisms for which the sequence of this region is available. As for the spacer box A, that of the leader region is similar in sequence (5′-GA/GACCUAGACAA-3′) in the free-living cyanobacteria but differs in the cyanelle (Fig. 4a, b). The box B of the leader was deduced uniquely from its structure and position relative to box A since, as in other bacteria (Berg *et al*., 1989), there was no homology in sequence between this domain in the spacer and leader regions.

Following transcription, the pre-mature rRNAs and mature tRNAs are released by cleavage involving enzymes RNase III and RNase P, respectively (King *et al*., 1986). The cleavage sites of RNase III are known in *E. coli* (Fig. 4a) and we have positioned the hypothetical sites in the cyanobacterial transcripts where possible. This was feasible for *Synechococcus* PCC 6301 for both 16S rRNA and 23S rRNA (Fig. 4a); for the cyanelle an apparent site was found only for the 23S rRNA (Fig. 4b). No potential RNase III sites could be identified in the other cyanobacterial sequences. The presumptive cleavage sites of RNase P are indicated for all sequences in Fig. 4(a, b).

DISCUSSION

Two different ITS regions (ITS-L and ITS-S) were amplified from *Nostoc* PCC 7120, which is known to contain four *rrn* operons (Ligon *et al*., 1991). Assuming no primer mismatch within two of the operons, the equal stoichiometry of ITS-L and ITS-S observed by gel electrophoresis shows that two of the four contain a longer ITS and two carry a shorter one. Minor differences within the ITS of each pair of operons cannot be excluded, but ambiguous positions were not observed following sequencing of three mixed clones of each spacer, suggesting that they may have the same length and nucleotide sequence.

ITS-L and ITS-S are virtually identical in sequence along their entire length (Fig. 3) and in secondary structure (not shown), except in the regions encoding $tRNA^{Ile}$ and $tRNA^{Ala}$ which are absent from the shorter ITS-S (Fig. 3). Multiple *rrn* operons containing ITSs of different sizes have also been observed in other bacteria,

and the ITSs of the different operons within a single organism commonly differ in their content of tRNA genes (Gürtler & Stanisich, 1996). However, in contrast to *Nostoc* PCC 7120, genome sequencing has revealed only two copies of the *rrn* operon in the *C. paradoxa* cyanelle (GenBank accession number U30821) and in *Synechocystis* PCC 6803 (D64000 and D90916). The ITSs within each copy are identical in length and in sequence, and the two copies therefore encode the same tRNA gene(s). Similarly, the two ITS regions of *Synechococcus* PCC 6301 show only six differences in the 395 nucleotides (representing 72% of the sequence and including both tRNA genes) available for comparison (Williamson & Doolittle, 1983; Tomioka & Sugiura, 1984).

It is not clear whether the two tRNA genes of *Nostoc* PCC 7120 have been acquired by ITS-L in the course of evolution or lost from ITS-S. In either case, however, two independent insertion or deletion events must have occurred, since both types of ITS share a part of the V2 structure (positions 401–440 of the alignment, Fig. 3). The formation of ITS-S from ITS-L by developmentally regulated excision events, such as occur in the *nif* operon during heterocyst differentiation in response to nitrogen starvation (Golden *et al*., 1985, 1988), can be excluded on two grounds. Firstly, both ITS-L and ITS-S were present in the same stoichiometry in cells of *Nostoc* PCC 7120 cultivated in both the presence and absence of combined nitrogen; secondly, the 401–440 region of ITS-S shows only 59% identity to that of ITS-L, suggesting a relatively ancient divergence, whereas excision events on either side of this region would not result in extensive sequence variation. The different ITS regions, although widespread in heterocystous cyanobacteria (Iteman *et al*., 1999), are thus unlikely to have been formed as a consequence of cellular differentiation. Provided that the transcripts are sufficiently stable, it will be of interest to determine whether expression of the two types of ITS is differentially regulated in response to nutritional or developmental stimuli.

Comparison of the cyanobacterial ITS sequences shows that they are very variable in size (283–545 nucleotides). However, their length does not correlate with the presence or absence of the tRNA genes. For example, the ITS of *Nodularia* BCNOD 9427 lacks both tRNA genes but is similar in size (354 nucleotides) to that of '*Mastigocladus* HTF' strain PCC 7518 (382 nucleotides) that contains both genes and is larger than the ITS-S of *Nostoc* PCC 7120 (283 nucleotides) that lacks both genes. These discrepancies are the result of large differences in the length of the variable regions, particularly of region V3. The cyanelle of *C. paradoxa* is considered as an intermediary stage in the evolution of chloroplasts from an endosymbiotic cyanobacterium (Herdman & Stanier, 1977; Aitken & Stanier, 1979). Although the ITS of the cyanelle contains only 287 nucleotides, both tRNA genes are present; however, the stem–loop structure following the 16S rRNA gene is short and regions V2, V3 and box B of the spacer are absent (Fig. 4b). A similar organization is found in the chloroplast of the red alga *Porphyra* (data not shown), whose ITS has a length of 279 nucleotides. The cyanelle and chloroplast therefore probably represent the shortest cyanobacterial-type ITS that contains both tRNA genes. The absence of tRNA genes from the ITS-S of *Nostoc* PCC 7120 and from the ITSs of several other cyanobacteria is clearly not deleterious to the organisms, since a similar situation is often found in other bacteria (Gürtler & Stanisich, 1996). In addition, genome sequencing (Kaneko *et al*., 1996) has shown that the $tRNA^{Ile}$ gene of *Synechocystis* PCC 6803 is present not only in both *rrn* operons but also as a third copy elsewhere in the genome, and that the absence of the $tRNA^{Ala}$ gene from both operons is compensated by the presence of three copies at other locations.

Irrespective of the length of the ITS region, numerous short domains are conserved in all cyanobacteria examined, and most of them are homologous to those observed in *E. coli*. Regions D1, D2, D4 and D5 are required for the correct folding of the rRNA transcript (Fig. 4a, b), the D2 and D4 motifs being of particular importance in contributing to the formation of the double-stranded 16S rRNA and 23S rRNA processing stems, respectively. Correct secondary structure is indispensable for the pre-maturation of the 16S rRNA and 23S rRNA molecules by RNase III (Srivastava & Schlessinger, 1990). Although we identified putative RNase III cleavage sites in the 16S rRNA processing stem of *Synechococcus* PCC 6301 and in region D5 of the latter and of the cyanelle of *C. paradoxa*, we were unable to locate such sites in the sequences of *Nostoc* PCC 7120 and *Synechocystis* PCC 6803. Since a gene encoding RNase III exists in the latter organism (Kaneko *et al*., 1996), our sequence data indicate that, as in other bacteria (see Nicholson, 1999), this enzyme is most probably a sequence-nonspecific nuclease. In other bacteria, the antiterminator box B–box A regions are involved in maintaining the correct stoichiometry of transcription of the rRNA genes (Berg *et al*., 1989; Condon *et al*., 1995). Antiterminator sequences were previously thought to be absent from *Synechococcus* PCC 6301 (Berg *et al*., 1989), although Wilmotte *et al*. (1994) suggested that a conserved region identified in the ITS of *Synechococcus* PCC 6301, *Trichodesmium* NIBB 1067 and seven other cyanobacteria (genera and strain numbers were not provided) may be involved in processing of the rRNA precursor. We have shown in this communication that the latter motif corresponds to the antiterminator box A and, together with box B, is present in the unpaired leader and spacer regions of all cyanobacterial sequences examined. As in *E. coli*, the two box A sequences are located adjacent to, and not within, the 16S rRNA and 23S rRNA processing stems and are preceded by a box B stem structure (Fig. 4a, b). However, box B is absent from the spacer of the cyanelle of *C. paradoxa*, and may thus not be essential for the control of transcription. This is consistent with studies involving site-directed mutagenesis of the antiterminator region in *E. coli*, which suggested that only box A has an essential role (Gourse *et al*., 1986). The

only major difference in secondary structure of the cyanobacterial and *E. coli* transcripts involves region D1, which is entirely paired with the leader sequence of the transcript in *E. coli*, but pairs with region D1′ of the ITS to form a stem–loop structure in the cyanobacteria examined (Fig. 4a, b). The similarity of essential regulatory and structural motifs in organisms as divergent as cyanobacteria and Proteobacteria suggests that mechanisms involved in the coordinated transcription and processing of the rRNA genes have been conserved throughout the course of evolution.

An unusual feature in the ITS-L of *Nostoc* PCC 7120 is that the V2 structure is composed of complementary paired STRR elements. STRR sequences appear to be common in heterocystous cyanobacteria, where they occur primarily, though not exclusively, in intergenic regions (Jackman & Mulligan, 1995; Mazel *et al.*, 1990; Vioque, 1997), but they are rarely paired with a complementary sequence and have not been reported previously in the *rrn* operon. At least nine different STRR families are known (Jackman & Mulligan, 1995). The *Nostoc* PCC 7120 STRR sequences within the ITS are similar to STRR type 6 sequences that form a similar structure (Vioque, 1997) within the Rnase P transcript of another *Nostoc* strain, PCC 7937 (*Anabaena* ATCC 29413), but differ from those found (Mazel *et al.*, 1990) elsewhere in the *Nostoc* PCC 7120 genome. Since STRR sequences are absent from the ITSs of all other cyanobacteria examined, and stem V2 is missing from many, it would appear that this region serves no essential function. Site-directed mutagenesis of the similar region in the gene encoding RNase P led to the same conclusion (Vioque, 1997).

$tRNA^{Ile}$ and $tRNA^{Ala}$ are the only tRNA genes identified to date in cyanobacterial *rrn* operons. The percentage of identity of each of these genes in the strains analysed is very high (Fig. 3). With the exception of the $tRNA^{Ala}$ gene of *Synechococcus* PCC 6301, the cyanobacteria and the cyanelle do not encode the 3′-terminal CCA extension found in many bacteria (including *E. coli*; Fig. 3), Archaea and eukaryotes (see Dirheimer *et al.*, 1995). However, they contain a characteristic subterminal CCA sequence at the 3′ end of the stem that may be easily mistaken for this extension. When not encoded, the 3′-terminal CCA extension, which is also absent from chloroplast tRNA genes (Martin, 1995), is most likely added post-transcriptionally, as in the latter, by ATP(CTP):tRNA nucleotidyltransferase (Martin, 1995) to form the mature tRNA.

The size and number of ITS bands recovered by PCR amplification will be valuable for the genetic classification of cyanobacterial strains when a sufficiently large database has been established. However, we have shown unequivocally (Fig. 2) that, if a strain contains several different ITS regions, bands corresponding to heteroduplexes may be produced. Such heteroduplexes have also been described in other bacteria, their formation being influenced by the PCR protocols employed (Jensen *et al.*, 1993). The interpretation of multiple products should therefore be made with caution, and they are best verified by denaturing gel electrophoresis. The successful alignment of cyanobacterial ITS regions (Fig. 3) may be exploited for the rapid incorporation of new sequences, the choice of restriction sites for RFLP analysis of unknown strains, and the design of probes and primers suitable for *in situ* hybridization and for PCR amplification. However, target regions resembling STRR sequences should be avoided, since these have been found at different locations in the genome (Mazel *et al.*, 1990; Vioque, 1997) and do not appear to be specific to a given organism. Cyanobacterial-specific probes can be developed from the conserved motifs of the ITS, whereas several of the variable regions appear to be useful for discrimination at higher taxonomic resolution.

ACKNOWLEDGEMENTS

This work was supported by contract BIO4-CT96-0256 (BASIC) of the European programme BIOTECH (Life Sciences and Technologies, Biotechnology Programme, 1994–1998), the Institut Pasteur and the Centre National de la Recherche Scientifique (CNRS URA 1129). We thank Gert Van der Auwera and Annick Wilmotte for permission to include the sequence of ‘*Mastigocladus* HTF’ strain PCC 7518.

REFERENCES

Aitken, A. & Stanier, R. Y. (1979). Characterization of peptidoglycan from the cyanelles of *Cyanophora paradoxa*. *J Gen Microbiol* **112**, 219–223.

Bensaadi-Merchermek, N., Salvado, J.-C., Cagnon, C., Karama, S. & Mouchès, C. (1995). Characterization of the unlinked 16S rDNA and 23S–5S rRNA operon of *Wolbachia pipientis*, a prokaryotic parasite of insect gonads. *Gene* **165**, 81–86.

Berg, K. L., Squires, C. & Squires, C. L. (1989). Ribosomal RNA operon anti-termination. Function of leader and spacer region box B–box A sequences and their conservation in diverse micro-organisms. *J Mol Biol* **209**, 345–358.

Birnboim, H. C. & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7**, 1513–1523.

Cai, Y. & Wolk, C. P. (1990). Use of a conditionally lethal gene in *Anabaena* sp. strain PCC 7120 to select for double recombinants and to entrap insertion sequences. *J Bacteriol* **172**, 3138–3145.

Castenholz, R. W. & Waterbury, J. B. (1989). Group 1. Cyanobacteria. In *Bergey's Manual of Systematic Bacteriology*, pp. 1710–1727. Edited by J. T. Staley, M. P. Bryant, N. Pfennig & J. G. Holt. Baltimore: Williams & Wilkins.

Chen, X. & Widger, W. R. (1993). Physical genome map of the unicellular cyanobacterium *Synechococcus* sp. strain PCC 7002. *J Bacteriol* **175**, 5106–5116.

Condon, C., Squires, C. & Squires, C. L. (1995). Control of rRNA transcription in *Escherichia coli*. *Microbiol Rev* **59**, 623–645.

De Rijk, P. & De Wachter, R. (1997). RnaViz, a program for the vizualisation of RNA secondary structure. *Nucleic Acids Res* **25**, 4679–4684.

Dirheimer, G., Keith, G., Dumas, P. & Westhof, E. (1995). Primary, secondary and tertiary structures of tRNAs. In *tRNA: Structure, Biosynthesis, and Function*, pp. 93–126. Edited by D. Söll & U.

RajBhandary. Washington, DC: American Society for Microbiology.

Fukunaga, M. & Mifuchi, I. (1989). Unique organization of *Leptospira interrogans* rRNA genes. *J Bacteriol* **171**, 5763–5767.

Giovannoni, S. J., Turner, S., Olsen, G. J., Barns, S., Lane, D. J. & Pace, N. R. (1988). Evolutionary relationships among cyanobacteria and green chloroplasts. *J Bacteriol* **170**, 3584–3592.

Golden, J. W., Robinson, S. J. & Haselkorn, R. (1985). Rearrangement of nitrogen fixation genes during heterocyst differentiation in the cyanobacterium *Anabaena*. *Nature* **314**, 419–423.

Golden, J. W., Carrasco, C. D., Mulligan, M. E., Schneider, G. J. & Haselkorn, R. (1988). Deletion of a 55-kilobase-pair DNA element from the chromosome during heterocyst differentiation of *Anabaena* sp. strain PCC 7120. *J Bacteriol* **170**, 5034–5041.

Gourse, R. L., deBoer, H. A. & Nomura, M. (1986). DNA determinants of rRNA synthesis in *E. coli*: growth rate dependent regulation, feedback inhibition, upstream activation, antitermination. *Cell* **44**, 197–205.

Gürtler, V. & Stanisich, V. A. (1996). New approaches to typing and identification of bacteria using the 16S–23S rDNA spacer region. *Microbiology* **142**, 3–16.

Hayes, P. K. & Barker, G. L. A. (1997). Genetic diversity within Baltic Sea populations of *Nodularia* (Cyanobacteria). *J Phycol* **33**, 919–923.

Herdman, M. & Stanier, R. Y. (1977). The cyanelle: chloroplast or endosymbiotic prokaryote? *FEMS Microbiol Lett* **1**, 7–12.

Herdman, M., Janvier, M., Rippka, R. & Stanier, R. Y. (1979a). Genome size of cyanobacteria. *J Gen Microbiol* **111**, 73–85.

Herdman, M., Janvier, M., Waterbury, J. B., Rippka, R., Stanier, R. Y. & Mandel, M. (1979b). Deoxyribonucleic acid base composition of cyanobacteria. *J Gen Microbiol* **111**, 63–71.

Iteman, I., Rippka, R., Tandeau de Marsac, N. & Herdman, M. (1999). Use of molecular tools for the study of genetic relationships of heterocystous cyanobacteria. In *Marine Cyanobacteria* (*Bulletin de l'Institut Océanographique*, *Monaco*, special issue 19), pp. 13–20. Edited by L. Charpy & A. Larkum. Monaco: Institut Océanographique.

Jackman, D. M. & Mulligan, M. E. (1995). Characterization of a nitrogen-fixation (*nif*) gene cluster from *Anabaena azollae* 1a shows that closely related cyanobacteria have highly variable but structured intergenic regions. *Microbiology* **141**, 2235–2244.

Janssen, I., Mucke, H., Löffelhardt, W. & Bohnert, H. J. (1987). The central part of the cyanelle rDNA unit of *Cyanophora paradoxa*: sequence comparison with chloroplasts and cyanobacteria. *Plant Mol Biol* **9**, 479–484.

Jensen, M. A., Webster, J. A. & Straus, N. (1993). Effect of PCR conditions on the formation of the heteroduplex and single-stranded DNA products in the amplification of bacterial ribosomal DNA spacers regions. *PCR Methods Appl* **3**, 186–194.

Kaneko, T., Sato, S., Kotani, H. & 21 other authors (1996). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res* **3**, 109–136.

King, T. C., Sirdeskmukh, R. & Schlessinger, D. (1986). Nucleolytic processing of ribonucleic acid transcripts in procaryotes. *Microbiol Rev* **50**, 428–451.

Lachance, M. A. (1981). Genetic relatedness of heterocystous cyanobacteria by deoxyribonucleic acid–deoxyribonucleic acid reassociation. *Int J Syst Bacteriol* **31**, 139–147.

Ligon, P. J. B., Meyer, K. G., Martin, J. A. & Curtis, S. E. (1991). Nucleotide sequence of a 16S rRNA gene from *Anabaena* sp. strain PCC 7120. *Nucleic Acids Res* **19**, 4553.

Lu, W., Evans, H. E., McColl, M. & Saunders, V. A. (1997). Identification of cyanobacteria by polymorphisms of PCR-amplified ribosomal DNA spacer region. *FEMS Microbiol Lett* **153**, 141–149.

Martin, N. (1995). Organellar tRNAs: biosynthesis and function. In *tRNA: Structure*, *Biosynthesis*, *and Function*, pp. 127–140. Edited by D. Söll & U. RajBhandary. Washington, DC: American Society for Microbiology.

Matzura, O. & Wennborg, A. (1996). RNAdraw: an integrated program for RNA secondary structure calculation and analysis under 32-bits Microsoft Windows. *CABIOS* **12**, 247–249.

Mazel, D., Houmard, J., Castets, A.-M. & Tandeau de Marsac, N. (1990). Highly repetitive DNA sequences in cyanobacterial genomes. *J Bacteriol* **172**, 2755–2761.

Naïmi, A., Beck, G. & Branlant, C. (1997). Primary and secondary structures of rRNA spacer regions in enterococci. *Microbiology* **143**, 823–834.

Neilan, B. A., Stuart, J. L., Goodman, A. E., Cox, P. T. & Hawkins, P. R. (1997). Specific amplification and restriction polymorphisms of the cyanobacterial rRNA operon spacer region. *Syst Appl Microbiol* **20**, 612–621.

Nelissen, B., Wilmotte, A., Neefs, J.-M. & De Wachter, R. (1994). Phylogenetic relationships among filamentous helical cyanobacteria investigated on the basis of 16S ribosomal RNA gene sequence analysis. *Syst Appl Microbiol* **17**, 206–210.

Nicholas, K. B. & Nicholas, H. B., Jr (1997). Genedoc: a tool for editing and annotating multiple sequence alignments. www.cris.com/~ketchup/genedoc.shtml.

Nicholson, A. W. (1999). Function, mechanism and regulation of bacterial ribonucleases. *FEMS Microbiol Rev* **23**, 371–390.

Ojaimi, C., Davidson, B. E., Saint Girons, I. & Old, I. G. (1994). Conservation of gene arrangement and an unusual organization of rRNA genes in the linear chromosomes of the Lyme disease spirochaetes *Borrelia burgdorferi*, *B. garinii* and *B. afzelii*. *Microbiology* **140**, 2931–2940.

Otsuka, S., Suda, S., Li, R. H., Watanabe, M., Oyaizu, H., Matsumoto, S. & Watanabe, M. M. (1999). Phylogenetic relationships between toxic and non-toxic strains of the genus *Microcystis* based on 16S to 23S internal transcribed spacer sequence. *FEMS Microbiol Lett* **172**, 15–21.

Rippka, R. & Herdman, M. (1992). *Catalogue of Strains*. *Pasteur Culture Collection of Cyanobacterial Strains in Axenic Culture*. Paris: Institut Pasteur.

Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M. & Stanier, R. Y. (1979). Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J Gen Microbiol* **111**, 1–61.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Srivastava, A. K. & Schlessinger, D. (1990). Mechanism and regulation of bacterial ribosomal RNA processing. *Annu Rev Microbiol* **44**, 105–129.

Tomioka, N. & Sugiura, M. (1984). Nucleotide sequence of the 16S–23S spacer region in the *rrn*A operon from a blue-green alga, *Anacystis nidulans*. *Mol Gen Genet* **193**, 427–430.

Tomioka, N., Shinozaki, K. & Sugiura, M. (1981). Molecular cloning and characterization of ribosomal RNA genes from a blue-green alga, *Anacystis nidulans*. *Mol Gen Genet* **184**, 359–363.

Vioque, A. (1997). The RNase P RNA from cyanobacteria: short

tandemly repeated repetitive (STRR) sequences are present within the RNase P RNA gene in heterocyst-forming cyanobacteria. *Nucleic Acids Res* **25**, 3471–3477.

Williamson, S. E. & Doolittle, W. F. (1983). Gene for $tRNA^{Ile}$ and $tRNA^{Ala}$ in the spacer between the 16S and 23S rRNA genes of a blue-green alga: strong homology to chloroplast tRNA genes and tRNA genes of *E. coli rrnD* cluster. *Nucleic Acids Res* **11**, 225–235.

Wilmotte, A. (1994). Molecular evolution and taxonomy of the cyanobacteria. In *The Molecular Biology of Cyanobacteria*, pp. 1–25. Edited by D. A. Bryant. The Netherlands: Kluwer Academic Publishers.

Wilmotte, A., Van der Auwera, G. & De Wachter, R. (1993). Structure of the 16S ribosomal RNA of the thermophilic cyanobacterium *Chlorogloeopsis* HTF ('*Mastigocladus laminosus* HTF') strain PCC7518, and phylogenetic analysis. *FEBS Lett* **317**, 96–100.

Wilmotte, A., Neefs, J. M. & De Wachter, R. (1994). Evolutionary affiliation of the marine nitrogen-fixing cyanobacterium *Trichodesmium* sp strain NIBB 1067, derived by 16S ribosomal RNA sequence analysis. *Microbiology* **140**, 2159–2164.

Wolk, C. P. (1996). Heterocyst formation. *Annu Rev Genet* **30**, 59–78.

Received 11 November 1999; revised 24 January 2000; accepted 13 March 2000.