The existence of eukaryotic ribonucleoprotein consensus sequence-type RNA-binding proteins in a prokaryote, *Synechococcus* 6301

Mamoru Sugita and Masahiro Sugiura*

Center for Gene Research, Nagoya University, Nagoya 464-01, Japan

Received October 25, 1993; Revised and Accepted December 6, 1993

DDBJ accession nos D17358 and D17359

ABSTRACT

A group of proteins containing a conserved ribonucleoprotein consensus sequence (RNP-CS)-type RNA-binding domain (CS-RBD) of ~80 amino acids is present in eukaryotic cells and binds specifically to a wide variety of RNA molecules. We have isolated 12 kDa single-stranded DNA binding proteins from the unicellular cyanobacterium *Synechococcus* 6301. The amino-terminal sequence was determined and two distinct genomic clones were isolated from a *Synechococcus* 6301 genomic library. Sequence analysis revealed that two closely related proteins contain a single CS-RBD of 82 amino acids and are named as 12RNP1 and 12RNP2. Both of the CS-RBDs share the highest amino acid identity with those of chloroplast ribonucleoproteins (40–51%). The 12RNP proteins were expressed in *Escherichia coli* bearing plasmids encoding glutathione S-transferase/12RNP fusion proteins and subjected to *in vitro* nucleic acid-binding assay. Both 12RNP1 and 12RNP2 bind to RNA homopolymers poly(U) and poly(G), indicating that they might be RNA-binding proteins. This is the first example of such proteins in prokaryotes. The 12RNP1 and 12RNP2 genes are transcribed as monocistronic mRNAs and the steady-state mRNA level of 12RNP1 is over 20-fold than that of 12RNP2. Due to the easiness of genetic manipulations the cyanobacterium will provide an excellent system to analyze the function of not only cyanobacterial but also plant RNA-binding proteins.

INTRODUCTION

Various kinds of proteins with one or more consensus sequence type-RNA binding domains (CS-RBD) or RNA recognition motifs (RRM) of about 80 amino acids are localized in the different subcellular compartments, nucleus, cytoplasm, mitochondria and plastids in eukaryotic cells and play a central role in RNA metabolism: capping, processing, polyadenylation and splicing of pre-RNAs, and nucleocytoplasmic transport of RNAs (1–4). CS-RBD includes two highly conserved motifs;

an octamer sequence termed ribonucleoprotein consensus sequence (RNP-CS) and a less conserved hexamer (RNP-2).

Chloroplasts, that are specialized plastids for photosynthesis, contain at least five related nuclear-encoded ribonucleoproteins (RNPs) of 28 to 33 kDa (cp28, cp29A, cp29B, cp31 and cp33) in tobacco (5, 6). The chloroplast proteins consist of an acidic amino (N)-terminal domain and two CS-RBDs of 83 amino acids. Their acidic N-terminal domains are unique to chloroplasts, distinct from any of the known RNA-binding proteins, including heterogeneous nuclear ribonucleoprotein (hnRNP), small nuclear ribonucleoprotein particle (snRNP) proteins, mRNA-binding proteins, splicing factors, or helix-destabilizing proteins of mammals, insects and yeast (2). Chloroplast RNPs are thought to be involved in mRNA processing or stabilization (5, 7). Likewise, glycine-rich proteins containing an N-terminal CS-RBD have been found in several plants (8, 9) but not in non-plant sources. Thus, at least two types of protein families containing CS-RBD are unique to plant cells. Chloroplasts are thought to descend from ancestral cyanobacteria (10, 11). The fact that the chloroplast RNPs are encoded in the nuclear genome raises the question whether the chloroplast proteins are of a cyanobacterial or of primordial nuclear origin. Here we report isolation of two 12 kDa CS-RBD containing proteins from *Synechococcus* 6301. The CS-RBD region has been found homologous to its chloroplast counterparts, thereby suggesting the involvement of these proteins in RNA binding. This is the first evidence for the existence of such proteins in prokaryotes. Moreover, we describe a possible evolutionary relation of the cyanobacterial proteins to chloroplast and other RNPs.

MATERIALS AND METHODS

Growth of cyanobacteria

Synechococcus sp. 6301 (formerly designated *Anacystis nidulans*) cells were grown in 500 ml of C medium of Kratz and Myers (12) at 30°C with shaking under 16 h light (2,500 lux) and 8 h dark period. Cells ($A_{730} = 0.8$) were harvested and stored at –70°C. For preparation of RNA from dark-grown cells, small scale cultures (50 ml) were wrapped in aluminium foil and incubated further for 12 h and then exposed to light for 5 h.

* To whom correspondence should be addressed

Preparation of ssDNA-binding proteins

Frozen cells (ca. 10 g wet weight) were homogenized with 8 g of quartz sand and 5 ml of solution A (50 mM Tris-HCl, pH 7.9, 2 mM EDTA, 2 mM DTT and 1 mM PMSF) by a mortar and a pestle. The homogenate was centrifuged at 8,000 rpm for 10 min and solid $(\text{NH}_4)_2\text{SO}_4$ was added to the supernatant to 80% saturation. The blue green pellet was dissolved in 10 ml of solution B (10% glycerol, 10 mM Tris-HCl, pH 8.0, 0.1 M NaCl, 0.1 mM DTT and 1 mM PMSF) and passed through a dsDNA cellulose column (1 ml). Its flow-through fraction was applied to an ssDNA cellulose column (1 ml) as described (5). Fractions eluted between 0.3 M and 0.6 M NaCl in solution B were pooled, concentrated by acetone precipitation and separated by 15% polyacrylamide-0.1% SDS gel electrophoresis (15% SDS-PAGE). The proteins were then transferred to a PVDF membrane, stained with Coomassie brilliant blue and sequenced using an Applied Biosystems 470A gas-phase sequencer (13). An oligonucleotide of 41 nucleotides (nt), 5'-ATCTACGTGGIAAC(C/T)TTCITT(C/T)GA(A/G)GCIAC(C/G)GA(A/G)GCI-IGA-3', which is deduced from IYVGNLSFEATEAD from amino acid residues 2-15 (Fig. 1B), was prepared using an Applied Biosystems 381A DNA synthesizer.

Isolation and sequencing of the 12RNP genes

A *Synechococcus* 6301 genomic DNA library (10^4 phage plaques) previously constructed in λ EMBL4 (14) was screened by plaque hybridization with the above mentioned ^{32}P -labeled oligonucleotide or with a 249 bp DNA fragment (Fig. 3A, positions 786-1034) corresponding to the 12RNP1 CS-RBD of 82 amino acids as probes. The 249 bp DNA fragment was prepared by PCR using its genomic clone and primers as below (non-genomic sequence parts are underlined):

1N = 5' AAGGATCCATTTACGTTGGTAACCTGTCCT 3'
1RBD = 5' AAGAATTCTCCTCACGGGGCTTTGCTT 3'

Southern blot hybridization was performed under low stringent condition at 60°C in solution C (5×SSPE, 5×Denhardt's solution, 0.5% SDS and 0.1 mg/ml denatured salmon sperm DNA) as described (15). Plaque hybridization was done in solution C at 45°C (for the oligonucleotide) or at 65°C (for the DNA fragment) overnight. The membranes were washed with 2×SSC at room temperature and then at 45°C for 5 min each (for the oligonucleotide) or 5×SSPE-0.1% SDS for 30 min and then 2×SSC-0.1% SDS for 20 min at 55°C (for the DNA fragment). The inserts in genomic clones AN7 and AN21 were digested with *Bam*HI and *Hind*III, and the DNA fragments hybridized to the probe were subcloned into Bluescript KS(+) plasmid and both strands were sequenced by dideoxy chain termination method with Sequenase (USB) as described (16).

Expression of 12RNPs in *E. coli*

The DNA fragment encoding 12RNP1 or 12RNP2 was prepared by PCR using its genomic clone as a template and primers as below:

1N = the same as above
1C = 5' AAGAATTCTAAGCAGGCAGCTTAGTAGTT 3'
2N = 5' AAGGATCCATTTACGTTGGTAATCTTTCCTT 3'
2C = 5' AAGAATTCACACAGATTTAACTTAGA 3'

The amplified fragment was cloned into the expression vector pGEX2T using *Bam*HI and *Eco*RI restriction sites. The DNA

sequences encoding 12RNP proteins were verified by DNA sequencing. *E. coli* cells bearing the expression vector encoding glutathione *S*-transferase (GST)-12RNP were cultured and the induction of fusion protein by addition of IPTG were carried out as described (17). The fusion protein was adsorbed to glutathione-Sepharose 4B (Pharmacia) and was cleaved by digestion with thrombin. Approximately 100 μg of 12RNP proteins were recovered from 100 ml cultures. The N-terminal sequence of purified 12RNP1 and 12RNP2 is GSIYVGNL- and verified by protein micro-sequencing.

Nucleic acid-binding assay

The assay was carried out essentially as described (9). The 12RNP protein prepared from an *E. coli* lysate (3 μg protein) was mixed with 20 μl each of ssDNA- and dsDNA-cellulose (Sigma), poly(U)- and poly(A)-Sepharose 4B (Pharmacia), and poly(C)- and poly(C)-polyacrylydrazido agarose (Sigma) in 1 ml buffer D (10 mM Tris-HCl, pH 7.5, 2.5 mM MgCl_2 , 0.5% Triton X-100, 1 $\mu\text{g}/\text{ml}$ leupeptin (Boehringer), 1 mM PMSF and 0.2 M NaCl). The mixture was mixed at 4°C for 10 min by rotation. The nucleic acid-resins were washed successively once with buffer D containing 2 mg/ml heparin, twice with buffer D and twice with distilled water. Bound proteins were eluted with 100 μl of SDS-PAGE loading buffer, and 20 μl of released proteins were separated by 17% SDS-PAGE and visualized by silver-staining.

Northern analysis and primer extension

Total RNA was isolated from the frozen cells essentially as described (18). The yield of total RNA per 50 ml of culture ($A_{730} = 0.6-0.8$) was 20-60 μg . RNA electrophoresis and northern blotting to a Hybond N membrane were as described (19). The 341 bp and 315 bp genomic DNA probes were produced by amplifications by using the primers 1N and 1C (for 12RNP1) and 2N and 2C (for 12RNP2), respectively. Hybridization and washing of the membranes were done at 65°C as described (19). The dried membranes were exposed to an imaging plate and analyzed by a Fuji Bioimaging Analyzer BAS2000.

Oligonucleotide primers used for primer extension were:

PE-1 (12RNP1) = 5'-CGTAAATAGACATGGAGTTGTCTCCGA-3'
PE-2 (12RNP2) = 5'-GACTGTTCCGGGAAAACGGAAAGCAA-3'

Annealing of total *Synechococcus* RNA (10 μg) and a [^{32}P]primer ($1-2 \times 10^6 \text{cpm}$) and primer extension were carried out as described (19).

Phylogenetic tree

Phylogenetic tree was constructed using unweighted pair-group method with arithmetic mean (20). The region from the 1st beta sheet ($\beta 1$) to the 4 residues after the 4th beta sheet ($\beta 4$) of CS-RBD was used.

RESULTS**12 kDa ssDNA-binding proteins contain RNP-2**

To investigate the possibility of a cyanobacterial origin of chloroplast CS-RBD proteins we isolated proteins by using an ssDNA column, which is the same procedure applied for the isolation of tobacco chloroplast RNPs (5). Of several ssDNA-binding proteins eluted between 0.3 M and 0.6 M NaCl solutions, an abundant protein of 12 kDa could be characterized (Fig. 1A).

Figure 1. A. ssDNA-binding proteins eluted with 0.3 to 0.6 M NaCl were resolved by 15–25% SDS-PAGE and silver-stained. An abundant 12 kDa protein is indicated by arrow (lane 1). Lane 2 shows low molecular mass markers (myoglobin digests, Pharmacia). **B.** The N-terminal amino acid sequence of the 12 kDa protein. The 1st, 9th and 11th residues are a mixture of two amino acids and Xs denote unidentified residues.

Figure 2. A. Restriction maps of the genomic clones AN7 and AN21 encoding 12RNP1 and 12RNP2, respectively. Thick lines represent the insert and open boxes indicate λ EMBL4 arms. Stippled bars (below the insert) show the subcloned fragment hybridizing with the probes. **B.** *Bam*HI; E, *Eco*RI; H, *Hind*III; S, *Sal*I sites. **B.** Genomic Southern blot analysis. *Synechococcus* 6301 DNA (0.5 mg) was digested with *Eco*RI/*Hind*III (lane E/H) or *Bam*HI/*Hind*III (lane B/H). The probe was the 249 bp DNA encoding the CS-RBD of 12RNP1. Hybridized fragments are indicated by bars with size (kb).

The determined N-terminal sequence includes a hexapeptide IYVGNL, which resembles the RNP-2 motif of chloroplast RNPs (Figs. 1B and 4), suggesting that the 12 kDa protein is an RNP-CS type protein. The mixtures of two amino acids observed at

Figure 3. Nucleotide sequences of the genes for 12RNP1 (A) and 12RNP2 (B) together with the predicted amino acid sequences. Initiation and termination codons are indicated by double-lines. Putative ribosome-binding sequences are enclosed. CS-RBDs and a glycine-rich region are boxed and indicated by dashed underlines, respectively. Thick arrows indicate the positions of the putative transcription start sites. The positions of the primer PE-1 and PE-2 are indicated by dashed arrows. Palindromic sequences are symbolized by inverted arrows. The tRNA^{Val} (UAC) gene is marked by thickened underlining. Numbers correspond to the sequences deposited in the database.

positions 1, 9 and 11 suggests that at least two closely related proteins exist in *Synechococcus* 6301.

Isolation and characterization of genes encoding 12 kDa proteins

To isolate the genes encoding the 12 kDa ssDNA-binding proteins we screened a *Synechococcus* 6301 genomic DNA library by plaque hybridization with a degenerate oligonucleotide probe based on the N-terminal amino acid sequence described above. Of six positive overlapping clones, the clone AN7 which contains a 17 kb genomic DNA fragment was analyzed (Fig. 2A). The DNA sequence of the 1203 bp *Bam*HI–*Hind*III fragment which hybridizes to the oligonucleotide probe was determined. It contains a single uninterrupted reading frame of 111 codons

Figure 4. Sequence alignment of CS-RBDs from cyanobacterial 12RNP proteins, tobacco chloroplast RNPs, cp28, cp29A and cp31 (5, 6), maize chloroplast NBP (22), RNA-binding glycine-rich proteins RGP-1b, RGP-1c (9), maize AAIP (8), potato U2-B" (23), human U1-A (24), human U1-70K (25), human hnRNP A1 (26), human ASF (27), sex-lethal (28), tra-2, *Drosophila* sex-determination protein (29), yeast PABP, poly(A)-binding protein (30), yeast RNA15 (31) and *E. coli* rho (32). The regions of two conserved motifs, RNP-2 and RNP-CS, are shown in open boxes. # means one of more than two CS-RBDs. The amino acids conserved among more than four proteins including either of 12RNP1 and 12RNP2 are in bold type. The position in the primary sequence of the first amino acid on every line is shown. The regions of β -sheets and α -helices are referred to a review by Kenan *et al.* (3).

starting from an ATG codon (Fig. 3A). A sequence GGAG which is complementary to the 3' end of *Synechococcus* 6301 16S rRNA (21) is located 8 bp apart from the ATG codon. The N-terminal region of its deduced amino acid sequence showed good matches with the determined amino acid sequence (Fig. 1B), confirming the gene encoding the 12 kDa protein. The predicted protein contains a single CS-RBD of 82 amino acids and therefore we designated this protein as 12RNP1.

To estimate the copy number of the 12RNP1 gene, *Synechococcus* 6301 genomic DNA was digested to completion with restriction enzymes, and subjected to Southern blot hybridization with the 249 bp DNA fragment corresponding to the CS-RBD of 12RNP1. Digestion with *EcoRI*/*HindIII* and *BamHI*/*HindIII* resulted in single bands corresponding to sizes of 4.3 kb and 1.2 kb, respectively, as expected which suggests that 12RNP1 is encoded by a single gene. In addition to these strong bands, weak signals (3.7 kb *EcoRI*/*HindIII* and 2.7 kb *BamHI*/*HindIII* fragments) were also detected (Fig. 2B). This result suggests that an additional gene homologous to the 12RNP1 gene is present in the *Synechococcus* 6301 genome. The same genomic DNA library was screened again by using the 249 bp DN as a probe and several positive clones giving weaker signals were selected. All positive clones overlap with one another and encompass a single locus. The 2.7 kb *HindIII*/*BamHI* subfragment in the genomic clone AN21 (12 kb insert) was sequenced. A nucleotide sequence of the 755 bp *EcoRI*/*BamHI* region is shown in Fig. 3B. A reading frame of 100 codons starting from GTG and terminating by TAA was found. A putative ribosome-binding sequence, AGGA, is present 8 to 11 bp prior to the GTG codon. Its N-terminal amino acid sequence shows high homology with the determined protein sequence (Fig. 1B), hence we named this protein as 12RNP2.

Structure of 12RNP1 and 12RNP2

Protein and DNA sequencing (Figs. 1B and 3) revealed that the 12RNP1 is 110 amino acids long with a calculated mol. mass of 11,342 Da and an isoelectric point of 8.2 and the 12RNP2 consists of 99 amino acids (11,100 Da) with an isoelectric point

of 7.9. The two 12RNP proteins contain a single CS-RBD of 82 amino acids, which shows 64% identity with each other. 12RNP1 has a short glycine-rich region of 18 amino acids (78% Gly) that is appended to the CS-RBD, while 12RNP2 does not have such a glycine-rich sequence but its C-terminal sequence, TPTTKQLTV, is homologous to that of 12RNP1 (TATTK-LPA).

The two highly conserved motifs of 12RNP1 are IYVGNL and RGFGFVEM, respectively, and resemble the chloroplast consensus sequence motifs, (I/L)(Y/F)VGNL and RGFGFVTM (Fig. 4). The RNP-CS (RGFAFIEM) of 12RNP2 is less homologous to those of chloroplast RNPs. The overall CS-RBD of 12RNP1 shows 40–51% amino acid identities, 78–94% similarities to those of tobacco chloroplast RNPs cp28, cp29A and cp31 (5, 6) and maize chloroplast NBP (22). Likewise, identities to CS-RBDs of the RNA-binding glycine-rich proteins (8, 9) are approximately 36 to 40%, while their homology to those of the potato U2-snRNP B' is only 22% (23). They show 18–34% homology with those of poly(A)-binding proteins, nuclear hnRNP proteins and snRNP proteins from non-plant sources (24–31). The 12RNP2 protein is the smallest of the CS-RBD-containing proteins identified so far. A phylogenetic tree constructed from 52 CS-RBDs supports the supposition that cyanobacterial CS-RBDs are most closely related to the chloroplast RNPs and plant RNA-binding glycine-rich proteins (Fig. 5).

Nucleic acid-binding property of 12RNP proteins

To demonstrate that the cyanobacterial 12RNP proteins are real RNA-binding proteins, *in vitro* nucleic acid-binding assay was carried out using the 12RNP proteins expressed in *E. coli* bearing a construct of GST-12RNP fusion protein. As shown in Fig. 6, both of the 12RNP proteins bind to ssDNA, poly(G) and poly(U) but not to dsDNA, poly(A) nor poly(C). Their binding property resembles in part those of chloroplast RNPs (36, 37), plant RNA-binding glycine-rich proteins (9, 38) and some of the HeLa hnRNP proteins (4). This suggests that the 12RNP proteins are likely RNA-binding proteins.

Figure 5. Phylogenetic tree constructed from the sequence of 52 CS-RBDs using the unweighted pair-group method with arithmetic mean (20). Only the topology of the tree is shown. The references for chloroplast RNPs, RGPs, AAIP, U2-snRNP B", U1-snRNP A, U1-snRNP70K, hnRNP A1, ASF, sex-lethal, tra-2, PABP, RNA15 and rho are shown in Fig. 4 and the others are in (3). PRP24, splicing factor; SSB, ssDNA-binding protein; bicoid, homeotic protein; PPTB, polypyrimidine tract-binding protein; eIF-4B, eukaryotic initiation factor; RNA15, proteins involved in mRNA stability or poly(A)-tail length; X16, proteins for RNA processing; elav, embryonic-lethal, abnormal-visual; SRP55, serine-arginine protein 55 (33); NAM8P, mitochondrial splicing factor (34); T4 GP32, bacteriophage T4 gene product 32 (35). SC, *Saccharomyces cerevisiae*; DM, *D. melanogaster*; HS, human; MM, mouse; XL, *Xenopus laevis*.

Expression of the 12RNP genes

Northern blot analysis of total RNA with the DNA fragments encoding the 12RNP proteins showed that 12RNP1 and 12RNP2 mRNAs are approximately 530 nt long and the 12RNP1 mRNA level is at least 20-fold higher than that of 12RNP2 (Fig. 7A). Primer extension experiments showed that transcription of the 12RNP1 and 12RNP2 gene starts at 112 and 205 bp upstream

Figure 6. Nucleic acid-binding properties of the 12RNP proteins. The proteins expressed in *E. coli* were mixed with various nucleic acid-beads, calf thymus ssDNA (ss) and dsDNA (ds), and RNA homopolymers poly(G), poly(A), poly(U) and poly(C) in 0.2 M NaCl. Bound proteins were analyzed by 17% SDS-PAGE. Lanes I are control to show the amount of proteins applied. Molecular mass markers are trypsin inhibitor (21.5 kDa) and lysozyme (14 kDa).

Figure 7. A. Northern blot analysis of transcripts from the 12RNP genes in *Synechococcus* 6301 total RNA (5 and 10 mg). Probes used were 341 bp DNA for 12RNP1 (positions 786–1126, Fig. 3A) and 315 bp DNA for 12RNP2 (positions 338–652, Fig. 3B). Size markers are rRNAs (23S and 16S) and RNA ladder (280 to 1280 nt). The film was exposed for 40 h at -70°C . **B.** Modulation of 12RNP transcript levels in light/dark—and temperature-shifted cells. *Synechococcus* 6301 cells grown under 16 h light/8 h dark cycle were transferred to darkness for 12 h. Cells were collected 0, 1, 3, 6, and 12 h after transfer of the cells to the dark and 5 h after transfer back to the light. 6 h light-grown cells (lane L6) were also collected as a control of 6 h dark. For cold-shock treatment, the cells were collected 3 h after shift from 30°C to 20°C . Total RNAs (3 μg) were hybridized with the DNA probes for 12RNP1 and 12RNP2 and for *rbcLS* (1866 bp DNA encoding the entire coding regions, 40, 41) genes. The films were exposed for 20 h (12RNP1, *rbcLS*) and 48 h (12RNP2) at -70°C . The lowest panel shows ethidium bromide staining patterns of total RNAs.

from the translation initiation codon, respectively (data not shown). However, canonical prokaryotic promoter motifs, '–10' and '–35' regions, are not found in the upstream regions of the two genes (Fig. 3). Thus the question whether the 5' ends determined by primer extension reflect genuine transcriptional starting sites or only processing sites remains open.

A gene encoding tRNA^{Val} (UAC), which corresponds to a group II intron-containing chloroplast counterpart and shows 68% sequence homology (39), is located 277 bp further upstream, on

the same strand of the 12RNP1 gene. Like chloroplast tRNA genes, the cyanobacterial tRNA^{Val} gene does not encode the 3' CCA end. A 685 bp genomic DNA fragment containing the tRNA^{Val} gene hybridized only to tRNA-sized transcripts (data not shown).

In the 3' flanking regions of the two 12RNP genes short inverted repeat sequences are observed, which can be folded into stable stem-loop structures (Fig. 3). If these repeats are assumed to be transcriptional termination signals, their mRNAs can be calculated to be 520–550 nt in length, consistent with the result of northern analysis (Fig. 7A). Taken together, these results demonstrate that both of the 12RNP genes are transcribed as monocistronic mRNAs.

Steady-state mRNA levels of 12RNP1 and 12RNP2 decreased when *Synechococcus* 6301 cells grown under light was transferred to dark (Fig. 7B). The mRNA level of 12RNP1 declined gradually and maintains 30% of the control level even after 12 h of culture in dark. The 12RNP2 mRNA level in contrast decreased rapidly within 3 h after transfer to dark and remained undetectable after 6 h dark period. Upon transfer of the cells into light, the transcripts for 12RNP1 and 12RNP2 as well as for ribulose-1,5-bisphosphate carboxylase/oxygenase showed rapid accumulation. When the cells grown at 30°C were exposed for 3 h to low temperature (at 20°C), the 12RNP1 mRNA level increased 2-fold, while the transcript level of 12RNP2 gene showed 3-fold decrease (Fig. 7B). From this we conclude that the expression of these two genes is regulated differentially by environmental conditions.

DISCUSSION

We describe here the first identification of RNP-CS containing proteins, 12RNP1 and 12RNP2, from a cyanobacterium. Nucleic acid-binding assays indicate their probable involvement in RNA binding. Their CS-RBD structures show significant similarity to those of chloroplast RNPs, although chloroplast RNPs have two CS-RBDs. It is intriguing that the 12RNP1 contains a short glycine-rich region in the C-terminal portion and that its overall protein structure is similar to those of plant RNA-binding glycine-rich proteins (9, 38). Those proteins have a single CS-RBD and a glycine-rich domain at the N- and C-terminal halves, respectively. Chloroplast RNPs, hnRNP A1, U1-snRNP A, U2-snRNP B' and fly sex-lethal proteins contain two CS-RBDs, fly *elav* and polypyrimidine tract-binding proteins contain three, and poly(A)-binding protein and nucleolin possess four such domains (2–4). In contrast, several other RNPs contain a single CS-RBD together with one or more auxiliary domains, e.g. arginine-, aspartic acid- or serine-rich domain (U1-snRNP 70K and fly *tra2*), ATP-binding domain (La, *E. coli* rho, hnRNP C1/C2), and zinc-finger motif (Ro-60K) (2). Therefore, we conclude that the cyanobacterial 12RNP proteins represent the simplest versions among CS-RBD containing proteins identified so far.

E. coli termination factor, rho, contains an atypical RNA-binding domain, which is less similar to the consensus than most other members of the CS-RBD family found in eukaryotes (25). The *E. coli* domain shares only 12 and 9% amino acid identity with 12RNP1 and 12RNP2, respectively (Fig. 4). In cyanobacteria, rho-dependent transcription termination is unknown, though its presence can not be excluded. Recent immunological study with human anti-RNP suggests that components similar to eukaryotic snRNP proteins and snRNAs

exist in the cyanobacterium *Synechococcus leopliensis* and *B. subtilis* (42). This implies that snRNP-like proteins or other RNP-CS type proteins exist in the cyanobacterium and some eubacteria.

Some of the HeLa cell hnRNP proteins exhibit ssDNA-binding activity *in vitro* (4, 43), and subsequently have been purified by ssDNA affinity chromatography (44). To our knowledge, in cyanobacteria, no ssDNA-binding proteins have been reported. This study shows that ssDNA affinity chromatography is an efficient procedure to isolate novel RNP-CS type proteins.

Although the mRNA levels of 12RNP1 and 12RNP2 differ at least 20-fold (Fig. 7A), their expression is regulated in a light-dependent manner as has been reported for photosynthesis related genes (45, 46). This observation suggests that 12RNP proteins are involved in the biosynthesis of the photosynthetic apparatus. Interestingly, on exposure of *Synechococcus* 6301 cells to lower temperature, the 12RNP1 mRNA level increases while the transcript level of 12RNP2 gene remarkably declines (Fig. 7B). A similar increase has also been observed in a filamentous cyanobacterial (*Anabaena variabilis*) gene encoding a protein similar to 12RNP1 (N.Sato, personal communication). *E. coli* and *B. subtilis* cold-shock proteins, CS7.4 and CspB, have been shown to include the RNP-CS motif in the cold-shock domain (47, 48). Recent crystallographic study of the CspB reveals that the cold-shock domain may also interact with RNA as well as DNA (49). However, the cyanobacterial 12RNP proteins differ in overall structure from such proteins. Because of the different mRNA level, the distinct response to low temperature and the presence or absence of a glycine-rich domain, the two 12RNP proteins may have distinct function; 12RNP1 may be essential to maintain cell growth under normal conditions whereas 12RNP2 may play a regulatory role. A glycine-rich region has been shown to confer cooperative RNA-binding and mediate protein–protein interactions (50, 51).

The 12RNP1 and 12RNP2 proteins show high binding specificity to poly(G) and poly(U) (Fig. 6). This suggests that they have high affinity for G/U-rich regions in RNA molecules. A *Synechococcus* 7942 tRNA^{Leu} (UAA) gene contains a 239 bp intron (52). In *Synechococcus* 6301, an intron in the tRNA^{Leu} (UAA) gene has also been demonstrated (unpublished). Since snRNP-like components are suggested to exist in cyanobacteria (42), the 12RNP proteins may also be components in snRNP-like particles which may be involved in splicing of intron-containing pre-tRNAs. Identification of RNA sequences recognized by these 12RNP proteins and the disruption of both or either of the two 12RNP genes will provide clues to the elucidation of their function in the cyanobacterium. In addition, a cyanobacterial transformation system will also provide an excellent tool to analyze plant RNA-binding proteins.

Chloroplasts are thought to originate from bacteria-like endosymbionts, whose closest organisms are cyanobacteria (11). In accordance to this the present study shows that chloroplast RNPs are closely related to their homologs from ancient cyanobacteria, thus providing additional evidence for the endosymbiotic theory (10). In addition to this the similarity in structure of plant RNA-binding glycine-rich proteins with the 12RNPs raises the possibility that they also are of a cyanobacterial origin. It appears likely that the cyanobacteria-like endosymbionts contained a single CS-RBD-type protein encoded in their own genomes, and that this gene was subsequently transferred to the nucleus and that it was then duplicated and fused to other genes already present in the nuclear DNA. Consequently, genes

encoding either one domain- or a two domain-type protein were produced. The former evolved further to plant RNA-binding glycine-rich protein genes and the latter to chloroplast RNP genes after acquiring a DNA region encoding a transit-peptide. It is intriguing that cp29A and cp29B have a glycine-rich spacer region between the two CS-RBDs (6).

ACKNOWLEDGEMENTS

We thank Dr T. Wakasugi and Dr A. Vera for valuable discussions and Dr H. Kössel and Dr S. Kapoor for critical reading of this manuscript. We also thank Dr T. Matsubayashi for technical advice on protein work, Ms S. Kimura and Ms A. Mase for protein sequencing, and Ms C. Sugita for subcloning and DNA sequencing. We are grateful to Dr N. Sato, Tokyo Gakugei University for the unpublished sequence of *A. variabilis* gene. This work was supported by a Grant-in-Aid from the Ministry of Education, Science and Culture (Japan) and by the Special Coordination Funds of the Science and Technology Agency (Japan).

REFERENCES

- Mattaj, I. W. (1989) *Cell*, **57**, 1–3.
- Keene, J. D. and Query, C. C. (1991) *Prog. Nucleic Acid Res.*, **41**, 179–202.
- Kenan, D. J., Query, C. C. and Keene, J. D. (1991) *Trends Biochem. Sci.*, **16**, 214–220.
- Dreyfuss, G., Matunis, M. J., Piñol-Roma, S. and Burd, C. G. (1993) *Annu. Rev. Biochem.*, **62**, 289–321.
- Li, Y. and Sugiura, M. (1990) *EMBO J.*, **9**, 3059–3066.
- Ye, L., Li, Y., Fukami-Kobayashi, K., Go, M., Konishi, T., Watanabe, A. and Sugiura, M. (1991) *Nucleic Acids Res.*, **19**, 6485–6490.
- Schuster, G. and Grüsssem, W. (1991) *EMBO J.*, **10**, 1493–1502.
- Gómetz, J., Sánchez-Martínez, D., Stiefel, V., Rigau, J., Puigdomènech, P. and Pagès, M. (1988) *Nature*, **334**, 262–264.
- Hirose, T., Sugita, M. and Sugiura, M. (1993) *Nucleic Acids Res.*, **21**, 3981–3987.
- Margulis, L. (1981) *Symbiosis in Cell Evolution*, W. H. Freeman and Co., San Francisco.
- Gray, M. W. (1989) *Trends Genet.*, **5**, 294–299.
- Kratz, W. A. and Myers, J. (1955) *Amer. J. Bot.*, **42**, 282–287.
- Matsudaira, P. (1987) *J. Biol. Chem.*, **262**, 10035–10038.
- Meng, B. Y., Shinozaki, K. and Sugiura, M. (1989) *Mol. Gen. Genet.*, **216**, 25–30.
- Murayama, Y., Matsubayashi, T., Sugita, M. and Sugiura, M. (1993) *Plant Mol. Biol.*, **22**, 767–774.
- Li, Y., Nagayoshi, S., Sugita, M. and Sugiura, M. (1993) *Mol. Gen. Genet.*, **239**, 304–309.
- Smith, D. B. and Johnson, K. S. (1988) *Gene*, **67**, 31–40.
- Mohamed, A. and Jansson, C. (1989) *Plant Mol. Biol.*, **13**, 693–700.
- Sugita, M., Murayama, Y. and Sugiura, M. (1993) *Curr. Genet.* in press.
- Nei, M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press, NY, pp. 293–298.
- Tomioka, N. and Sugiura, M. (1983) *Mol. Gen. Genet.*, **191**, 46–50.
- Cook, W. B. and Walker, J. C. (1992) *Nucleic Acids Res.*, **20**, 359–364.
- Simpson, G. G., Vaux, P., Clark, G., Waugh, R., Beggs, J. D. and Brown, J. W. S. (1991) *Nucleic Acids Res.*, **19**, 5213–5217.
- Sillekens, P. T. G., Habets, W. J., Beijer, R. P. and van Venrooij, W. J. (1987) *EMBO J.*, **6**, 3841–3848.
- Query, C. C., Bentley, R. C. and Keene, J. D. (1989) *Cell*, **57**, 89–101.
- Riva, S., Morandi, C., Tsoulfas, P., Pandolfo, M., Biamonti, G., Merrill, B., Williams, K. R., Multhaup, G., Beyreuther, K., Werr, H., Hendrich, B. and Schäfer, K. P. (1986) *EMBO J.*, **5**, 2267–2273.
- Ge, H., Zuo, P. and Manley, J. L. (1991) *Cell*, **66**, 373–382.
- Bell, L. R., Maine, E. M., Schedl, P. and Cline, T. W. (1988) *Cell*, **55**, 1037–1046.
- Amrein, H., Gorman, M. and Nöthiger, R. (1988) *Cell*, **55**, 1025–1035.
- Sachs, A. B., Bond, M. W. and Kornberg, R. D. (1986) *Cell*, **45**, 827–835.
- Minvielle-Sebastia, L., Winsor, B., Bonneaud, N. and Lacroute, F. (1991) *Mol. Cell. Biol.*, **11**, 3075–3087.
- Pinkham, J. L. and Platt, T. (1983) *Nucleic Acids Res.*, **11**, 3531–3545.
- Roth, M. B., Zahler, A. M. and Stolk, J. A. (1991) *J. Cell. Biol.*, **115**, 587–596.
- Ekwall, K., Kermorgant, M., Dujardin, G., Groudinsky, O. and Slonimski, P. P. (1992) *Mol. Gen. Genet.*, **233**, 136–144.
- Krisch, H. M. and Allet, B. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4937–4941.
- Li, Y. and Sugiura, M. (1991) *Nucleic Acids Res.*, **19**, 2893–2896.
- Ye, L. and Sugiura, M. (1992) *Nucleic Acids Res.*, **20**, 6275–6279.
- Ludevid, M. D., Freire, M. A., Gómez, J., Burd, C. G., Albericio, F., Giralt, E., Dreyfuss, G. and Pagès, M. (1992) *Plant J.*, **2**, 999–1003.
- Deno, H., Kato, A., Shinozaki, K. and Sugiura, M. (1982) *Nucleic Acids Res.*, **10**, 7511–7520.
- Shinozaki, K. and Sugiura, M. (1983) *Nucleic Acids Res.*, **11**, 6957–6964.
- Shinozaki, K., Yamada, C., Takahata, N. and Sugiura, M. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 4050–4054.
- Kovacs, S. A., O'Neil, J., Watcharapijarn, J., Moe-Kirvan, C., Vijay, S. and Silva, V. (1993) *J. Bacteriol.*, **175**, 1871–1878.
- Kumar, A., Williams, K. R. and Szer, W. (1986) *J. Biol. Chem.*, **261**, 11266–11273.
- Pandolfo, M., Valentini, O., Biamonti, G., Rossi, P. and Riva, S. (1987) *Eur. J. Biochem.*, **162**, 213–220.
- Brand, S. N., Tan, X. and Widger, W. R. (1992) *Plant Mol. Biol.*, **20**, 481–491.
- Smart, L. B. and McIntosh, L. (1991) *Plant Mol. Biol.*, **17**, 959–971.
- Goldstein, J., Pollitt, N. S. and Inouye, M. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 283–287.
- Willmsky, G., Bang, H., Fischer, G. and Marahiel, M. A. (1992) *J. Bacteriol.*, **174**, 6326–6335.
- Schindelin, H., Marahiel, M. A. and Heinemann, U. (1993) *Nature*, **364**, 164–168.
- Cobianchi, F., Karpel, R. L., Williams, K. R., Notario, V. and Wilson, S. H. (1988) *J. Biol. Chem.*, **263**, 1063–1071.
- Nadler, S. G., Merrill, B. M., Roberts, W. J., Keating, K. M., Lisbin, M. J., Barnett, S. F., Wilson, S. H. and Williams, K. R. (1991) *Biochemistry*, **30**, 2968–2976.
- Kuhse, M. G., Strickland, R. and Palmer, J. D. (1990) *Science*, **250**, 1570–1573.