Genomic survey of cAMP and cGMP signalling components in the cyanobacterium *Synechocystis* PCC 6803

Jesús A. G. Ochoa de Alda and Jean Houmard

Author for correspondence: Jean Houmard. Tel: +33 144323519. Fax: +33 144323935. e-mail: jhoumard@biologie.ens.fr

Cyanobacteria modulate intracellular levels of cAMP and cGMP in response to environmental conditions (light, nutrients and pH). In an attempt to identify components of the cAMP and cGMP signalling pathways in *Synechocystis* PCC 6803, the authors screened its complete genome sequence by using bioinformatic tools and data from sequence–function studies performed on both eukaryotic and prokaryotic cAMP/cGMP-dependent proteins. Sll1624 and Slr2100 were tentatively assigned as being two putative cyclic nucleotide phosphodiesterases. Five proteins were identified as having all the determinants required to be cyclic nucleotide receptors, two of them being probably more specific for cGMP (an element of two-component regulatory systems – Slr2104 – and a putative cyclic-nucleotide-gated cation channel – Slr1575), the three others being probably more specific for cAMP: (i) a protein of unidentified function (Slr0842); (ii) a putative cyclic-nucleotide-modulated permease (Slr0593), previously annotated as a kinase A regulatory subunit; and (iii) a putative transcription factor (CRP-*Syn* – Sll1371), which possesses cAMP- and DNA-binding determinants homologous to those of the cAMP receptor protein of *Escherichia coli* (CRP-Ec). This homology, together with the presence in *Synechocystis* of CRP-Ec-like binding sites upstream of *crp*, *cya1*, *slr1575*, and several genes encoding enzymes involved in transport and metabolism, strongly suggests that CRP-*Syn* is a global regulator.

**Keywords:** CRP, cyclic-nucleotide gated channel

INTRODUCTION

Living organisms adapt to a variety of different environmental conditions. After the sensing of environmental stimuli, a largely unknown network of signal transduction pathways allows cells to rapidly respond by changing the activity of target enzymes and/or by modulating the expression of target genes, thus producing an integrated response. Cyanobacteria are photosynthetic prokaryotes, the survival of which requires sensing of light, oxygen and nutrient availability. Under standard growth conditions, cyanobacteria contain similar intracellular levels of cyclic AMP (cAMP) and cyclic GMP (cGMP) (Herdman & Elmorjani, 1988). Intracellular levels of both nucleotides vary in response to changes in environmental conditions: light, nutrients and oxygen (Herdman & Elmorjani, 1988; Ohmori, 1989; Sakamoto et al., 1991). Therefore, these two commonly encountered second messengers are likely to play an important role in cyanobacterial signalling pathways.

Intracellular levels of cAMP and cGMP depend on the relative rates at which they are synthesized by the adenylyl- (EC 4.6.1.1) or guanylyl-cyclases (EC 4.6.1.2), degraded by the cAMP- (EC 3.1.4.17) or cGMP- (EC 3.1.4.35) phosphodiesterases, and eventually excreted. The association of cAMP and cGMP with protein receptors in turn determines the physiological response(s). At present, only the cAMP and cGMP biosynthetic enzymes have been identified and characterized in cyanobacteria (Kasahara & Ohmori, 1999; Katayama & Ohmori, 1997; Ochoa de Alda et al., 2000; Terauchi & Ohmori, 1999).

Cyclic-nucleotide phosphodiesterases play a pivotal role in cAMP and cGMP signal transduction by regulating...
the intracellular levels of cyclic nucleotides. The two known classes of phosphodiesterases comprise the seven families of eukaryotic cyclic-nucleotide phosphodiesterases (Beavo, 1995) and the proteobacterial cAMP-phosphodiesterases (Macfadyen et al., 1998). Although cAMP-phosphodiesterase activities have been described in cyanobacteria (Sakamoto et al., 1991), no homologue of the proteobacterial enzymes has been recognized in the genome of Synechocystis PCC 6803 (Macfadyen et al., 1998). Five hypothetical coding regions in the Synechocystis PCC 6803 genome contain, however, a phosphohydrolase domain, named HD, characteristic of the N-terminal half region of the eukaryotic cAMP-phosphodiesterase catalytic centres (Aravind & Koonin, 1998). Mutational analysis of the HD domain of a cGMP-specific phosphodiesterase showed that the more conserved residues within HD domains are involved in catalysis (Turko et al., 1998).

In vertebrates, cAMP and cGMP couple visual and olfactory signals to electrical excitation and Ca\(^{2+}\) signalling by modulating cyclic-nucleotide-gated (CNG) channels (Zagotta & Siegelbaum, 1996). Both nucleotides can also activate protein kinases, which in turn regulate enzymes and proteins involved in intermediary metabolism as well as in transcription (Daniel et al., 1998). In Escherichia coli, the real occurrence of cGMP \([<10^{-8}\text{ mol}\cdot\text{mg protein}^{-1}]\) is debated (Hermdan & Elmorjani, 1988; Vogler & Lengeler, 1987). In this bacterium, cAMP functions as a cofactor of the cAMP-receptor protein (CRP) rather than as an activator of a protein kinase. The cAMP-receptor protein CRP, also known as CAP (Catabolite gene Activator Protein), is an allosteric DNA-binding protein that modulates the transcription of several genes (Gralla & Collado-Vides, 1996; Kolb et al., 1993). At present, all the cAMP-dependent responses in bacteria appear to be mediated through the binding of cAMP to its receptor, which in turn regulates directly or indirectly genes involved in pH regulation, sugar metabolism and taxis (Botsford & Harman, 1992).

The allosteric effect promoted by cyclic nucleotides on CRP, kinases and the aforementioned signalling channels is exerted through a similar cyclic nucleotide monophosphate (cNMP)-binding domain (Zagotta & Siegelbaum, 1996). The homodimeric three-dimensional structure of the E. coli cAMP-receptor protein (CRP-Ec) has been determined (Weber & Steitz, 1987). The structural data combined with site-directed mutagenesis studies showed the importance of the residues that are conserved among the cyclic-nucleotide-binding domains of CRPs, kinases and cyclic-nucleotide-gated channels (Varnum et al., 1993; Woodford et al., 1989; Zagotta & Siegelbaum, 1996). All the well-characterized cyclic-nucleotide-binding domains can accommodate both cGMP and cAMP, but the degree of activation that they confer upon the output domain depends on the bound nucleotide. Both cAMP and cGMP bind to the E. coli cAMP-receptor protein but cGMP does not activate transcription (Ebright et al., 1985). The olfactory cyclic-nucleotide-gated channel is fully activated by cGMP and cAMP. In contrast, for the photoreceptor channels, cAMP acts as a partial agonist, producing only a fraction of the current induced by cGMP (Zagotta & Siegelbaum, 1996).

The protein GAF domains (encountered in cGMP-specific phosphodiesterases, Adenylyl cyclases and Formate hydrogen lyases) which may allosterically regulate catalytic activities via ligand binding represent another type of cyclic nucleotide receptor (Aravind & Ponting, 1997). GAF domains participate in the architecture of many signalling proteins (phytochromes, ethylene receptors and members of two-component regulatory systems). The GAF domain of cGMP-stimulated phosphodiesterases binds cGMP. Structure–function studies of this GAF domain performed by alanine mutagenesis have identified a motif \([N(K/R)][XnD]\) necessary for full cGMP-binding activity (McAllister-Lucas et al., 1995; Turko et al., 1996).

Apart from the adenylyl cyclase Cya1 (Slr1991), which has just been shown to be involved in the regulation of cell motility in Synechocystis PCC 6803 (Terauchi & Ohnori, 1999), cGMP and cAMP signalling pathways in cyanobacteria remain sketchy. Here we report on the identification of putative components of the cGMP and cAMP cyanobacterial signalling pathways by making use of: (i) the complete genome sequence of Synechocystis PCC 6803 (Terauchi & Ohnori, 1999); (ii) recently available programs and databases (Table 1); and (iii) the data obtained by sequence–function studies for cAMP/cGMP-dependent proteins. Our study leads in particular to the presumptive identification of a prokaryotic cyclic-nucleotide-gated channel protein. Hypotheses are presented that could serve as good starting-points for functional genomic studies.

### RESULTS AND DISCUSSION

#### Identification of potential cAMP- and cGMP-phosphodiesterases

Using sequence similarity search programs (BLAST and PSI-BLAST) no proteobacterial cAMP-phosphodiesterase homologues could be recognized in the Synechocystis PCC 6803 genome. Such homologues could however exist in other cyanobacteria because the fusion of the two overlapping ORFs, PID g139932 and g139969, of the unicellular cyanobacterium Synechococcus PCC 6301 (Kumano et al., 1983) would result in an ORF 30% identical (47% similar) to E. coli and Haemophilus influenzae phosphodiesterases. Recently, Aravind & Koonin (1998) reported that Slr1885, Slr2100, Sl1624 and Slr0104 contained a phosphohydrolase domain named HD, which is related to the catalytic region of eukaryotic cyclic-nucleotide phosphodiesterases. We observed that the HD-domains of ORFs Slr2100 and Sl1624 were similar (30% identity, 40% similarity), and that each HD domain was fused to a different N-
was clustered with elements of two-component systems. The analysis of the cyanobacterium *Synechocystis* PCC 6803 genome performed by Mizuno *et al.* (1996) indicated that *slr2100* was transcribed as a single unit. The corresponding polypeptides contain both the transmitter (sensory histidine kinase) and the receiver (response regulator) domains, likely forming a specific multi-step phospho-relay system (Appleby *et al.*, 1996; Kotani & Tabata, 1998; Mizuno *et al.*, 1996). By using a program aimed at detecting already identified protein domains (SMART), we observed that ORFs *slr2098* and *slr2104* contain PAS signalling motifs in the sensor region (Fig. 1), motifs known to regulate light- and O$_2$-stimulated signalling pathways in a variety of organisms (Zhulin & Taylor, 1998). PAS domains have also been related to the perception of blue light (Pellequer *et al.*, 1998; Zhulin *et al.*, 1997). Interestingly, in *Synechocystis* PCC 6803 blue light induces changes in the intracellular levels of cAMP (Terauchi & Ohmori, 1998). Altogether, data indicate that *slr2100* might participate in the blue light signal transduction pathway.

### Table 1. Web sites and programs used to identify components of the cAMP and cGMP signalling pathways in *Synechocystis* PCC 6803

<table>
<thead>
<tr>
<th>Tool or content</th>
<th>Website</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Annotated analysis of the <em>Synechocystis</em> PCC 6803 genome</strong></td>
<td><a href="http://www.kazusa.or.jp:8080/cyano/">http://www.kazusa.or.jp:8080/cyano/</a></td>
<td>Nakamura <em>et al.</em> (1998)</td>
</tr>
<tr>
<td><strong>Sequence similarity search tool</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Protein structure prediction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Prediction of transmembrane helices and topology of proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Protein structure data bank</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Domain architecture</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Alignment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLUSTALW</td>
<td></td>
<td><a href="http://www2.ebi.ac.uk/clustalw/">http://www2.ebi.ac.uk/clustalw/</a></td>
</tr>
</tbody>
</table>

* Matrix, Blosum 62; gap existence cost, 11; per residue gap cost, 1; lambda ratio, 0.85.
† Matrix, MTIDK; weighting, no; window width, all.
‡ Matrix, MTIDK; weighting, no; window width, all.
§ Size of pseudocount vector, 1000; length of helices, from 17 to 25; lengths of tails, from 1 to 5; lengths of loops, > 1; iterations, 1.
|| Gap opening 10; gap extension 0.005; gap distance 0.005.

*terminal response regulator domain through a coiled-coil structural domain (Fig. 1). Interestingly, this modular organization is that of the cAMP-specific phosphodiesterase of *Dictostelium discoideum* (RegA), for which the receiver domain has been shown to modulate the phosphodiesterase activity (Thomasson *et al.*, 1998). These similarities in sequence and domain organization suggest that Slr2100 and Slr1624 could be cyanobacterial cGMP- and cAMP-phosphodiesterases and that the cyclic nucleotide levels are likely controlled through two-component systems. In the filamentous cyanobacterium *Spirulina platensis*, such a regulation indeed controls the adenylyl cyclase CyaC and, thus, the intracellular levels of cAMP (Kasahara & Ohmori, 1999).

The analysis of the *Synechocystis* PCC 6803 genome (slr2098, slr2099 and slr2104), all these genes possibly being transcribed as a single unit. The corresponding polypeptides contain both the transmitter (sensory histidine kinase) and the receiver (response regulator) domains, likely forming a specific multi-step phospho-relay system (Appleby *et al.*, 1996; Kotani & Tabata, 1998; Mizuno *et al.*, 1996). By using a program aimed at detecting already identified protein domains (SMART), we observed that ORFs Slr2098 and Slr2104 contain PAS signalling motifs in the sensor region (Fig. 1), motifs known to regulate light- and O$_2$-stimulated signalling pathways in a variety of organisms (Zhulin & Taylor, 1998). PAS domains have also been related to the perception of blue light (Pellequer *et al.*, 1998; Zhulin *et al.*, 1997). Interestingly, in *Synechocystis* PCC 6803 blue light induces changes in the intracellular levels of cAMP (Terauchi & Ohmori, 1998). Altogether, data indicate that Slr2100 might participate in the blue light signal transduction pathway.
Cyclic nucleotide receptor proteins

Protein domain analysis of the *Synechocystis* genome, using the SMART program, indicated that single or multiple GAF domains are present in 28 ORFs of the *Synechocystis* genome (12 of which also containing a PAS domain). SMART server provides a clustal w alignment of GAF domains that permits comparison of the *Synechocystis* GAF domains with those of some eukaryotic phosphodiesterases. The GAF domain of ORF Slr2104 contains all the determinants (Asn258, Lys259, Asp267) for cGMP-binding. That some of these residues are missing in other GAF domains of *Synechocystis* may indicate: (i) that these sites do not participate in cGMP binding; (ii) that the substituted residues can functionally replace those found in GAF domains of eukaryotic phosphodiesterases; or (iii) that this GAF domain serves some function other than cGMP binding.

Based on protein domain sequence analyses (SMART), 12 ORFs of the *Synechocystis* genome could have cNMP-binding domains (Fig. 2). All of them contain a highly conserved stretch of approximately 120 amino acids that is homologous to the cNMP-binding domains of other proteins, including the cAMP- and cGMP-dependent protein kinases, the cyclic-nucleotide-gated channels and the cAMP receptor protein (CRP). In five of these ORFs the cNMP-binding domain is joined to putative transmembrane helices (Slr0593, Slr1529, Slr1575, Slr0510 and Sll1180), whereas in another five it is attached to a C-terminal domain which contains helix–turn–helix (HTH) DNA-binding motifs (Sll1371, Sll1924, Sll1169, Srr0449 and Slr1423).

Protein secondary structure predictions showed that the overall structure of the cNMP-binding domain of CRP-Ec was conserved in the putative corresponding domains of the 12 ORFs (Fig. 2b), suggesting a similar tertiary structure. In CRP-Ec, the aC helix (conserved in the 12 putative cNMP-binding domains) is required for the induction, via the binding of the nucleotide, of the conformational change that subsequently permits sequence-specific DNA-binding (Heyduk et al., 1992). In cyclic-nucleotide-gated channels this helix stabilizes the ligand in the open-state form of the channel (Tibbs et al., 1998). The strongest interactions between the ligand and CRP-Ec are achieved through non-covalent bonds: (i) between Arg82, Ser83 and the cyclic phosphate; (ii) between Gly71, Glu72 and the ribose; and (iii) between the N-6 amino group of the cAMP purine ring, Thr127 of one subunit and Ser128 of the other subunit of the dimer (Weber & Steitz, 1987). The residues that bridge
the cyclic phosphate and the ribose are conserved in cNMP-dependent protein kinases and in cyclic-nucleotide-gated channels. Site-directed mutagenesis of these conserved residues showed that they are indeed essential for the binding of cyclic nucleotides (Moore et al., 1992; Tibbs et al., 1998; Woodford et al., 1989). Our alignment showed that these residues were conserved in Sll1371 (CRP-Syn), Slr1575, Slr0593 and Slr0842 (Fig. 2). Therefore, among the 12 selected ORFs containing putative cNMP-binding domains, at least 4 polypeptides likely possess a functional cNMP-binding domain.

In general, ligand discrimination of cNMP-binding domains is achieved by the residues analogous to Ser83 and Thr127 of CRP-Ec. Since in CRP-Syn, Slr1575, Slr0593 and Slr0842 the position analogous to Ser83 of CRP-Ec is occupied by either a Ser or a Thr residue, this position probably does not allow discrimination between cAMP and cGMP. In CRP-Ec, residues Thr127 and Ser128 determine nucleotide selectivity and participate in the intersubunit communication. Amino acid substitutions that introduce a hydrophobic amino acid side chain at positions 127 or 128 decrease CRP-Ec discrimination between cAMP and cGMP (Lee et al., 1994). Mutation T127L results in an activation of transcription upon the binding of cGMP instead of cAMP (Moore et al., 1996). According to

the cyclic phosphate and the ribose are conserved in cNMP-dependent protein kinases and in cyclic-nucleotide-gated channels. Site-directed mutagenesis of these conserved residues showed that they are indeed essential for the binding of cyclic nucleotides (Moore et al., 1992; Tibbs et al., 1998; Woodford et al., 1989). Our alignment showed that these residues were conserved in Sll1371 (CRP-Syn), Slr1575, Slr0593 and Slr0842 (Fig. 2). Therefore, among the 12 selected ORFs containing putative cNMP-binding domains, at least 4 polypeptides likely possess a functional cNMP-binding domain.

In general, ligand discrimination of cNMP-binding domains is achieved by the residues analogous to Ser83 and Thr127 of CRP-Ec (Altenhofen et al., 1991; Lee et al., 1994; Shabb et al., 1990; Varnum et al., 1995). The Ser83 residue (or a conservative replacement of it such as Thr) allows the binding of both cGMP and cAMP. Since in CRP-Syn, Slr1575, Slr0593 and Slr0842 the position analogous to Ser83 of CRP-Ec is occupied by either a Ser or a Thr residue, this position probably does not allow discrimination between cAMP and cGMP. In CRP-Ec, residues Thr127 and Ser128 determine nucleotide selectivity and participate in the intersubunit communication. Amino acid substitutions that introduce a hydrophobic amino acid side chain at positions 127 or 128 decrease CRP-Ec discrimination between cAMP and cGMP (Lee et al., 1994). Mutation T127L results in an activation of transcription upon the binding of cGMP instead of cAMP (Moore et al., 1996). According to
these data, CRP-Syn, Slr1575 and Slr0593 would be activated by cGMP, since the positions equivalent to Thr127 in CRP-Ec are occupied by either Leu or Ala residues (Fig. 2). However, reconstruction of the putative cNMP-binding domain using the CRP-Ec structure as a template (Fig. 3) showed that: (i) the substitution of Ser128 by Asn (in CRP-Syn and Slr0842) preserved the interactions with positions N-6 and N-7 of the adenine; and (ii) the loss of specificity that would result from the substitution of Thr127 by Leu or Ala (as in CRP-Syn and Slr0593) could be overcome by the substitution of Ser62 by an Asn, which could create a new hydrogen bond with the N1 of the adenine (Fig. 3). Thus we propose that the cNMP-binding domains of CRP-Syn, Slr0593 and Slr0842 would be activated preferentially by cAMP (Fig. 3). This deductive method cannot be used to infer the specificity of Slr1575 because the aforementioned critical positions are occupied by non-polar amino acids.

Characterization of the proteins that could bind cyclic nucleotides

Slr2104 is a member of two-component regulatory systems. Searches for characteristic architecture domains by the SMART program indicated that Slr2104 contains a sensory histidine kinase N-terminal domain joined to a response regulator domain followed by a phosphoacceptor domain of histidine kinases (Fig. 1). Thus, Slr2104 integrates a multi-step phospho-relay. The sensor domain (input domain) of the histidine kinase is formed by a putative membrane domain, a cGMP-binding GAF domain and a PAS domain. Strikingly, this ORF is clustered with the putative phosphodiesterase Slr2100. Hence, two-component regulatory systems are likely involved in the control, amplification and integration of cyclic nucleotide signals.

Slr1575 is a putative cyclic-nucleotide-gated channel (CNG). ORF Slr1575 possesses structural elements clearly related to voltage-gated and cyclic-nucleotide-gated channels (Fig. 4). In such channels, the major determinants in channel gating have been mapped in the cytoplasmic linker (C-linker) that connects the last transmembrane segments of the channel and the cNMP-binding domain (Zong et al., 1998). The alignment of the cyclic-nucleotide-gated bovine cone photoreceptor channel (CNG3) with Slr1575 (Fig. 4a) showed that two (I439 and D481) of the three amino acids in the C-linker that determine the high specificity of the cone photoreceptor channel for cGMP were conserved, the Asp494 of CNG3 being replaced by a Gln residue. This alignment strongly suggests that Slr1575 would be activated preferentially by cGMP.

ORF Slr1575 contains a sequence similar to the pore-forming region of ion channels and a stretch of hydrophobic and basic amino acids similar to the transmembrane voltage-sensing S4 domain of voltage-activated and cyclic-nucleotide-gated channels (Fig. 4b, c). Putative voltage-sensing and pore-forming sequences form the N-terminus of Slr1575. We deduced its putative topology by comparison with the well-known structure of the bacterial K+ channel (KcsA) (Doyle et al., 1998). In the pore-forming sequence, the GXG motif boxed in Fig. 4c is of particular interest since it determines cation selectivity in ion channels (Heginbotham et al., 1992; Kerr & Sansom, 1995). The pore of Slr1575 would resemble the non-selective cation pore of the P2X2 purine receptor. Our comparative analysis leads us to conclude that Slr1575 would be a non-selective cation channel gated by cyclic nucleotides that would belong to the S4 ion channel superfamily (Heginbotham et al., 1992).

The subunit composition of cyclic-nucleotide-gated channels, heteromeric or homomeric, modulates both cyclic nucleotide sensitivity (Bradley et al., 1994) and cation permeation (Dzeja et al., 1999). Although Slr0510 does not contain the determinants for cyclic nucleotide binding (Fig. 1), this hypothetical protein is homologous to Slr1575 in both sequence (33% identity and 51% similarity) and domain architecture (Fig. 2). By analogy with eukaryotic cyclic-nucleotide-gated channels, the specificity for cations and cyclic nucleotides of Slr1575 homomeric channels might be modulated by heteromerization with Slr0510.

Slr0593 is a putative cyclic-nucleotide-modulated permease. In both Cyanobase and Genequiz databases, Slr0593 is annotated as a CAMP protein kinase regulatory chain because the cNMP-binding domains of these proteins share a high degree of homology (30% identity and 48% similarity). Our sequence analysis indicates, however, that of the five domains that define such proteins (Taylor et al., 1990), two were not conserved, namely the dimer-interaction site and the peptide-inhibitory site. Two programs aimed at predicting transmembrane helices and protein topology (PHDhtm and ΠΜΠΤΩΠ) predict eight transmembrane helices for Slr0593, the cNMP-binding domain being located on the cytosolic side, between the third and fourth helices (Fig. 2). The three C-terminal membrane helices of Slr0593 show 27% identity (43% similarity) with three membrane helices of the amino acid permease ROCC of Bacillus subtilis (PID g730600) (Glasner et al., 1993). The five C-terminal membrane helices of Slr0593 show 26% identity (42% similarity) with the phenylacetic acid permease PhaJ of Pseudomonas putida (PID g3253206) (Olivera et al., 1998). Although the specificity of the permease cannot be inferred from sequence comparison, our results indicate that Slr0593 is likely to be a permease, the activity of which would be modulated by cyclic nucleotides.

Slr0842 is a putative signal transducer. In ORF Slr0842, the CAMP-binding domain is joined to a DUF2 domain (Fig. 2). This domain was first recognized in the N-terminal region of Acetobacter xylinum enzymes that control the turnover of c-di-GMP: diguanylate cyclase and phosphodiesterase A (Tal et al., 1998). It is possible that DUF2 domains mediate one of these activities. Using a program for detecting characteristic architecture domains (SMART), we have found DUF2 domains in 13 ORFs.
cAMP and cGMP signalling components in *Synechocystis*

Fig. 3. Models for the cAMP-binding sites of CRP-Syn, Slr1575, Slr0593 and Slr0842, by analogy with the cAMP receptor protein (CRP-Ec) structure. Key amino acids were deduced from the alignment shown in Fig. 1. Spatial structures were modelled by substituting the corresponding residues in the three-dimensional structure of CRP-Ec (PDB entry 3GAP). Residues contacting the ribose and cyclic phosphate are not shown because they are fully conserved.

of *Synechocystis* PCC 6803 (Slr0842, Slr0267, Slr1588, Slr1102, Slr0821, Slr1103, Slr0359, Slr1104, Slr1895, Slr1305, Slr1593, Slr2077 and Slr1692). They may be preceded by input sensory signalling domains like response regulators, GAF domains, FHA (Forkhead-associated domain) domains and/or PAS domains. Hence, DUF2 domains can probably be regulated by different signals. We conclude that ORF Slr0842 might be a cyclic-nucleotide-modulated signal transducer of a pathway yet to be determined.

**The cAMP receptor protein (CRP-Syn, slr1371) is a transcription factor.** The homology of CRP-Syn with the *E. coli* cAMP receptor protein (CRP-Ec) extends to the helix–turn–helix DNA-binding domain (Fig. 5). The crystal structure of the CRP-DNA complex enabled the identification of the residues of CRP-Ec that interact with the DNA (Schultz et al., 1991), and the compilation of several CRP-DNA binding sites led to the identification of a 22 bp palindromic consensus site of sequence $\text{AAATGTGATCT}^\ast \text{AGATCAGATT}$ (the most conserved bases are underlined and the base pairs that interact with the protein are doubly underlined) (Berg & von Hippel, 1988). Interactions of CRP-Ec with the consensus DNA-binding sequence involve specific base contacts with Arg180, Glu181 and Arg185, the determinants of specificity (Kolb et al., 1993). These and other less selective contacts involved in the specific recognition of the consensus DNA binding site by *E. coli* cAMP receptor protein are conserved in the *Synechocystis* PCC 6803 protein (Fig. 5a). This suggests that CRP-Syn could recognize the same DNA sequence as CRP-Ec. Upstream of the CRP-Syn coding region we found DNA sequences showing a high degree of similarity with the promoter region of the *E. coli* crp, including the two CRP binding sites (Fig. 5b). These identities reinforce the hypothesis that CRP-Syn binds to the same DNA sequence as CRP-Ec, and suggest that *crp* is autoregulated as it is in *E. coli* (Hanamura & Aiba, 1991). While this paper was under review, Yoshimura et al. (2000) reported that slr1371 (referred to as CRP-Syn in this article) not only binds cAMP but...
Fig. 4. Features common to Slr1575 and cation channels. (a) PSI-BLAST sequence alignment of Slr1575 with the C-linker (doubled underlined) and cNMP-binding domain (underlined) regions of the bovine cone photoreceptor cyclic-nucleotide-gated channel (CNG3). Boxes contain the determinants for the specificity of CNG3 for cGMP (Zong et al., 1998). (b) Alignment of the voltage-sensing S4 region of the shaker K⁺ channel, CNG3 (Finn et al., 1996), and Slr1575. Basic residues are underlined. (c) Alignment of the pore-forming region of the shaker K⁺ channel, CNG3 (Zong et al., 1998), the P2X2 purinoreceptor (Kerr & Sansom, 1995) and Slr1575. The putative determinants for cation selectivity (Kerr & Sansom, 1995) are boxed. Numbers at the left indicate the position of the first residue in the corresponding sequence. Sequence identity and similarity are represented by (*) and (:), respectively. GenBank accession codes (PID) are shown to the right of the alignment.

also, when complexed with cAMP, can alter the electrophoretic mobility of an oligonucleotide that contains the E. coli CRP binding site. Slr0593 (referred to as a cyclic nucleotide permease in this article) also bound cAMP while sll1924 could not. These data are in agreement with the hypotheses proposed in this article.

We then searched for CRP-binding sites in the upstream region of the Synechocystis PCC 6803 genes that are homologues of E. coli CRP-regulated genes (Gralla & Collado-Vides, 1996). As in E. coli, we found putative CRP-binding sites upstream of the adenyllyl cyclase gene (cyt) and of the acetolactate synthase small subunit gene (ilvN slr0065). Further screening of the whole genome of Synechocystis using a program aimed at recognizing short highly similar DNA sequences (Find-patterns, GCG package) revealed putative CRP-binding sites upstream of the genes encoding: (i) the ribose-5-phosphate isomerase (rpiA slr0194) involved in the pentose phosphate pathway; (ii) the UDP-N-acetyl-muramoylalanyl-d-glutamyl-2,6-diaminopimelate-d-alanyl-d-alanine ligase (murF slr1351), implicated in the synthesis of the cell envelope; (iii) the above-mentioned putative cyclic-nucleotide-gated-channel slr1575; (iv) the high-affinity branched-chain amino acid transport permease protein (luvH slr1200); (v) a putative methyl-accepting chemotaxis protein I (cheD slr0041) involved in bacterial taxis; (vi) the AT103 protein (AT103 slr1874), a putative photchrome-regulated gene; (vii) subunit 3 of the NADH dehydrogenase of the respiratory electron transport (ndhC slr1279); (viii) the homologue of the response regulator NarL (slr1708) involved in the adaptation of E. coli to anaerobic respiration; (ix) the uncharacterized histidine kinase Slr1805; and (x) the permease protein PstA (psta or phoT) implicated in phosphate transport. The diversity of the putative CRP-Syn-regulated genes suggests the existence in Synechocystis PCC 6803 of a CRP-Syn-regulated modulon.

Concluding remarks

After standard search for identification by homology, about half of the putative ORFs of every fully sequenced genome are annotated as hypothetical proteins. More thorough analyses, coupled with the physiological and structure–function data that are available, make it possible to refine the assignments and/or propose
cAMP and cGMP signalling components in *Synechocystis*

A CLUSTAL W sequence alignment of the cAMP receptor protein (CRP-Syn and CRP-Ec) sequences involved in DNA-binding. Contacts involved in specific recognition of DNA by CRP-Ec (Schultz et al., 1991) are indicated above its sequence: specific base contacts (\(\text{y}\)), charged interactions with the DNA backbone phosphate (\(-\text{ac}\)), side chain H bonds (\(+\)) and backbone amide H bonds (\(\text{r}\)). Numbers at the left indicate the position of the first residue in the corresponding sequence. Sequence identity and similarity are represented by (*) and (:), respectively. Numbers in parentheses represent amino acids not shown but considered for the alignment.

A comparison of the *E. coli* crp promoter region P\(_{crp}\) with a region located upstream of the gene encoding CRP-Syn. The CRP-Ec binding sites I (repressing site) and II (activating site), f\(_{35}\) and f\(_{10}\) sequences and the ATG are underlined (Hanamura & Aiba, 1991). The CRP-dependent transcription start point is boxed. Sequences of *Synechocystis* PCC 6803 matching the CRP-Ec consensus DNA-binding site are double underlined. Identity scores (at the right) correspond to the number of nucleotides matching those of the residues involved in specific base contacts (\(\text{y}\)) with respect to the number of residues matching the 10 most conserved nucleotides of the palindrome (boxed). Residues identical to the consensus are doubly underlined.

A working hypotheses (Pallen, 1999). Using this approach in conjunction with targeted gene inactivation, we have recently shown that *Synechocystis* Cya2, a protein previously identified as an adenylyl cyclase (Kaneko et al., 1996), is responsible for cGMP synthesis in this organism, thus identifying the first prokaryotic guanylyl cyclase (Ochoa de Alda et al., 2000). Although phosphodiesterase activities have been measured in cyanobacteria (Herdman & Elmorjani, 1988), including *Synechocystis* PCC 6803, no obvious homologues of proteobacterial enzymes have been previously found; we tentatively propose Sll1624 and Slr2100 as being cyanobacterial phosphodiesterases.

Among 40 ORFs showing a high overall homology with cyclic nucleotide receptors, we identified 5 that contain the determinants required for ligand binding. One important conclusion derived from the presence of cGMP and cAMP at similar levels in *Synechocystis* PCC 6803 (Herdman & Elmorjani, 1988), and from the presence of the aforementioned cNMP-binding domains, is that one cyclic nucleotide could function as an inhibitor or as an agonist of the activity induced by the other, both of them being able to bind to the cNMP-binding domain but only one of the two inducing full activation.

Our sequence analysis study of the genome of *Synechocystis* PCC 6803 revealed the previously unrecognized presence of eukaryotic-type cGMP/cAMP signal-
ling components in a prokaryote, such as the cyclic-
nucleotide-gated channels and the cGMP-binding GAF
domain. The presence of the genes encoding eukaryotic-
type cGMP/cAMP signalling components that are
detected, together with that of a putative CRP modulon,
suggests that the complexity of the cAMP/cGMP
signalling pathways in cyanobacteria is greater than in
other bacteria, in agreement with the striking mor-
phological and physiological diversity found in the
cyanobacteria. This complexity could be even higher in
some filamentous cyanobacteria able to carry out
external differentiation, because *Anabaena* PCC 7120
for example has five different nucleotide cyclases (Katayama &
Ohmori, 1997).

Targeted inactivation of the ORFs that we have char-
acterized would be a first approach to validate our
hypotheses. In parallel, site-directed mutagenesis of the
hypothetical residues essential for the function of the
identified domains would give more information about
the signal transduction pathways in which they par-
ticipate. Substitution of the gene encoding *slr2100* by a
mutated gene, for example, should remove the phos-
dhodiesterase activity and affect the phosphorelay(s) in
which cyclic nucleotides are involved.

Acknowledgements

We are grateful to M.-I. Tapia for careful reading of the
manuscript and to J.-P. Roux for computer assistance. We
wish to thank A.-L. Etienne for her encouraging support. This
work was supported by grants from the Centre National de la
Recherche Scientifique to the UMR8543 and to J. A. G. Ochoa
de Alda, and from the Ecole Normale superieure (Paris).

References

Altenhofen, W., Ludwig, J., Eismann, E., Kraus, W., Bonigk, W. &
nucleotide-gated channels from rod photoreceptors and olfactory

Altschul, S. F. & Koonin, E. V. (1998). Iterated profile searches with
*psi-Blas*: a tool for discovery in protein databases. *Trends

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z.,
Miller, W. & Lipman, D. J. (1997). Gapped Blas and *psi-Blas*: a new
generation of protein database search programs. *Nucleic
Acids Res* 25, 3389–3402.

transduction via the multi-step phosphorylase: not necessarily a
road less travelled. *Cell* 86, 843–848.

Aravind, L. & Koonin, E. V. (1998). The HD domain defines a
new superfamily of metal-dependent phosphohydrolases. *Trends
Biochem Sci* 23, 469–472.

evolutionary link between diverse phototransducing proteins.

Beavo, J. A. (1995). Cyclic nucleotide phosphodiesterases: func-
tional implications of multiple isoforms. *Physiol Rev* 75, 725–748.

Benson, D. A., Boguski, M. S., Lipman, D. J., Ostell, J., Ouellette,
Acids Res* 27, 12–17.


Received 7 February 2000; revised 14 August 2000; accepted 30 August 2000.