BIOINFORMATICS DISCOVERY NOTE Vol. 22 no. 1 2006, pages 3–6
doi:10.1093/bioinformatics/bti739

*Sequence analysis*

# PilZ domain is part of the bacterial c-di-GMP binding protein

Dorit Amikam[1,2] and Michael Y. Galperin[3,*]

[1]Department of Biotechnology and Environmental Sciences, Tel-Hai Academic College, Tel-Hai, Israel, [2]Sharett Institute of Oncology, Hadassah University Medical Center, Ein-Kerem, Jerusalem, Israel, and [3]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

Received on October 17, 2005; revised on October 20, 2005; accepted on October 21, 2005
Advance Access publication October 25, 2005
Associate Editor: Alex Bateman

**ABSTRACT**

Recent studies identified c-di-GMP as a universal bacterial secondary messenger regulating biofilm formation, motility, production of extracellular polysaccharide and multicellular behavior in diverse bacteria. However, except for cellulose synthase, no protein has been shown to bind c-di-GMP and the targets for c-di-GMP action remain unknown. Here we report identification of the PilZ ('pills') domain (Pfam domain PF07238) in the sequences of bacterial cellulose synthases, alginate biosynthesis protein Alg44, proteins of enterobacterial YcgR and firmicute YpfA families, and other proteins encoded in bacterial genomes and present evidence indicating that this domain is (part of) the long-sought c-di-GMP-binding protein. Association of the PilZ domain with a variety of other domains, including likely components of bacterial multidrug secretion system, could provide clues to multiple functions of the c-di-GMP in bacterial pathogenesis and cell development.

**Contact**: galperin@ncbi.nlm.nih.gov

**Supplementary information:** http://www.ncbi.nlm.nih.gov/Complete_Genomes/SigCensus/PilZ.html

## 1 INTRODUCTION

The recent identification of bis-(3′-5′)-cyclic dimeric guanosine monophosphate, c-di-GMP, as a universal secondary messenger in bacteria was a key advance in microbiology, made possible, in part, by comparative genome analysis (Galperin *et al*., 2001; D'Argenio and Miller, 2004; Galperin, 2004; Jenal, 2004; Römling *et al*., 2005). The GGDEF (formerly DUF1) and EAL (formerly DUF2) domains, whose involvement in c-di-GMP turnover was discovered in the groundbreaking work by Moshe Benziman and co-workers (Tal *et al*., 1998), were found to be among the most abundant domains encoded in bacterial genomes, suggesting that c-di-GMP-dependent regulation was widespread in the bacterial world (Galperin *et al*., 2001; Galperin, 2005). Indeed, just in the past two years, c-di-GMP was implicated in regulating transition between motility and sessility in *Escherichia coli* and *Salmonella typhimurium*, twitching motility in *Pseudomonas aeruginosa*, biofilm formation in *Vibrio cholerae* and *Yersinia pestis*, and photosynthesis gene expression in *Synechococcus elongatus* (Huang *et al*., 2003; Kirillina *et al*., 2004; Simm *et al*., 2004; Thomas *et al*., 2004; Tischler and Camilli, 2004). An important advance was the recent demonstration that the diguanylate cyclase (c-di-GMP synthetase) activity resides in the GGDEF domain (Paul *et al*., 2004; Ryjenkov *et al*., 2005), whereas the EAL domain functions as c-di-GMP-specific phosphodiesterase, hydrolyzing c-di-GMP to linear diguanylate GpGp (Bobrov *et al*., 2005; Christen *et al*., 2005; Schmidt *et al*., 2005). Still, mechanisms of c-di-GMP-dependent signaling remain unknown, owing to the scarcity of data on the targets of c-di-GMP action.

In the original study of the regulation of the cellulose synthase in *Acetobacter xylinum* (currently *Gluconacetobacter xylinus*) and other bacteria, Benziman and co-workers detected c-di-GMP binding to the cellulose synthase with most label bound to its $\beta$-subunit, BcsB (Amikam and Benziman, 1989; Mayer *et al*., 1991). These data suggested that BcsB was the c-di-GMP binding protein, which was reflected in its SwissProt annotation. Subsequent studies, however, revealed that c-di-GMP was actually binding to a 200 kD membrane-bound protein complex (Weinhouse *et al*., 1997), which has not been further characterized, but could correspond to the dimer of the $\alpha$-subunit or to the second form of cellulose synthase whose single polypeptide chain contained both subunits (Saxena and Brown, 1995). Hence, it remained unclear which part of cellulose synthase, if any, would bind c-di-GMP and what were its other cellular targets. Our previous attempts to identify the c-di-GMP-binding adaptor protein by computational means have been unsuccessful, as no known protein exhibited the same phyletic distribution as the GGDEF and EAL domains (Römling *et al*., 2005). Here we report identification of the PilZ ('pills') domain (Pfam domain PF07238, Bateman *et al*., 2004) in the sequence of bacterial cellulose synthases and present evidence indicating that this domain is the long-sought c-di-GMP-binding protein.

## 2 RESULTS AND DISCUSSION

Development of twitching motility in *P.aeruginosa* is governed by about 40 genes (Mattick, 2002). Functions of most of them are known or could be predicted based on the available experimental data. PilZ, encoded by *P.aeruginosa* PA2960 gene, is a 118 amino acid protein (Fig. 1) that remains one of the very few without an

*To whom correspondence should be addressed.

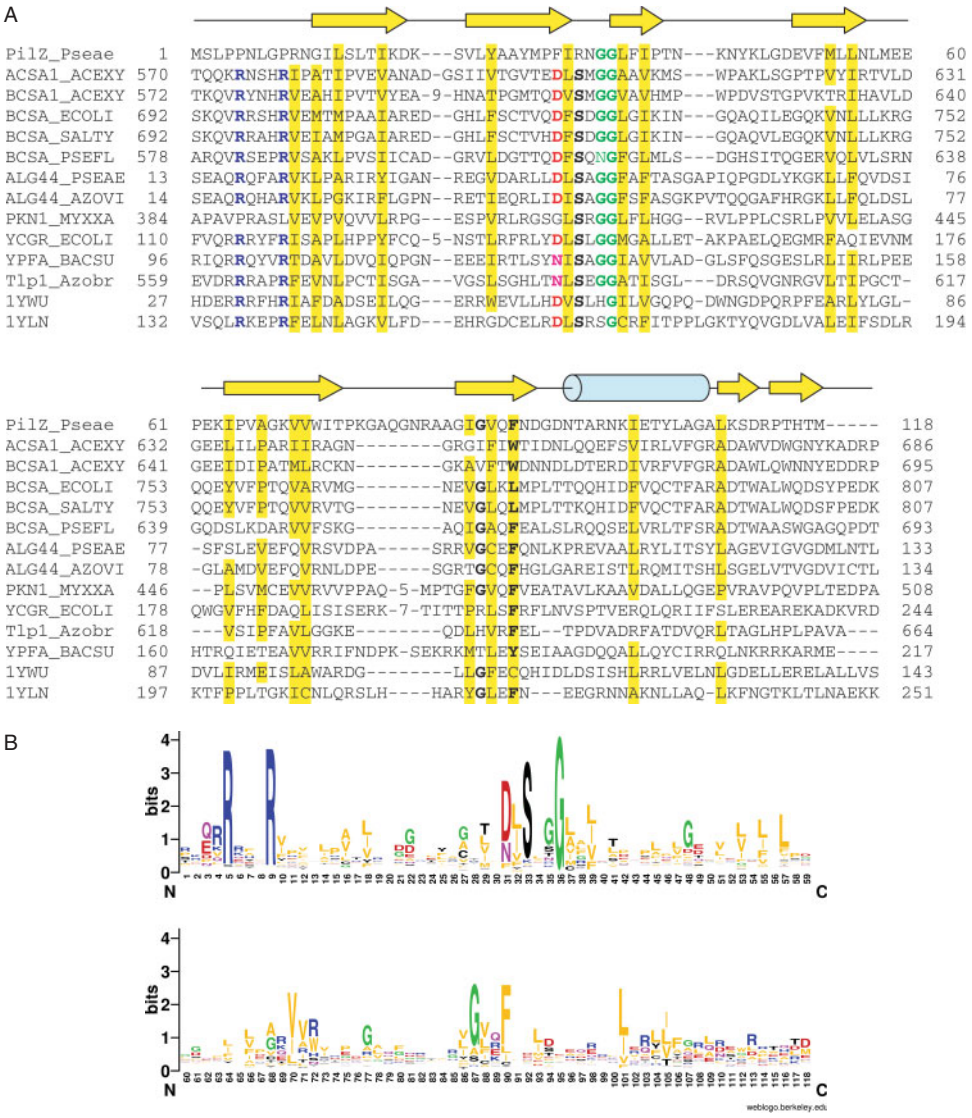


**Fig. 1.** Multiple alignment of the PilZ domain. (**A**) Experimentally studied proteins. Conserved hydrophobic residues are shaded, other conserved residues are shown in bold. Proteins are listed by their gene names, UniProt entry names, or PDB codes. GenBank/EMBL protein accession numbers for PilZ_Pseae and Tlp1_Azobr are AAA93519 and AAT76671, respectively. Secondary structure assignments are from the VCA0042 structure (PDB: 1YLN); arrows indicate $\beta$-strands, the cylinder indicates an $\alpha$-helix. (**B**) Sequence logo of 472 non-redundant sequences of the PilZ domain, generated by the WebLogo tool (http://weblogo.berkeley.edu, Crooks *et al*., 2004) from a master–slave alignment, produced by PSI-BLAST using PilZ-Pseae as the query. Residue numbering is as in PilZ_Pseae.

assigned function. *pilZ* mutants produce normal amounts of pilin but are unable to assemble functional pili (Alm *et al*., 1996). Close homologs of PilZ are encoded in many beta- and gamma-proteobacteria with distant homologs showing up in a variety of bacteria (Table S1 in the Supplementary Material). Sequence analysis of the PilZ protein was prompted by the discovery in *Geobacter sulfurreducens* of a response regulator GSU3263 with REC-PilZ domain architecture (Fig. 2), which suggested that PilZ might have a regulatory role. This analysis greatly benefited from the crystal structure of the *V.cholerae* protein VCA0042 (Protein Data Bank entry 1YLN), which contains C-terminal PilZ domain, solved recently by the Midwest Center for Structural Genomics (R.Zhang, M.Zhou, S.Moyi, F.Collart and A.Joachimiak, to be published). In addition, an NMR structure of stand-alone PilZ domain (PDB entry 1YWU) has been solved by the Northeast Structural Genomics Consortium (T.A.Ramelot, A.A.Yee, A.Semesi, C.H.Arrowsmith and M.A.Kennedy, to be published).

Iterative PSI-BLAST searches of the NCBI protein database started from the C-terminal 120 amino acid fragment of GSU3263 and retrieved more than 600 sequences from a variety of bacteria. Importantly, this search identified the PilZ domain near the C-terminus of the $\alpha$-subunit of cellulose synthase from *G.xylinus*, *E.coli*, and other bacteria (Fig. 2). In contrast, no PilZ domain was detected in otherwise closely related eukaryotic cellulose synthases from the slime mold *Dictyostelium discoideum* and marine urochordate *Ciona savignyi* (Blanton *et al*., 2000; Matthysse *et al*., 2004). Eukaryotes do not seem to encode GGDEF domains and presumably do not produce c-di-GMP.

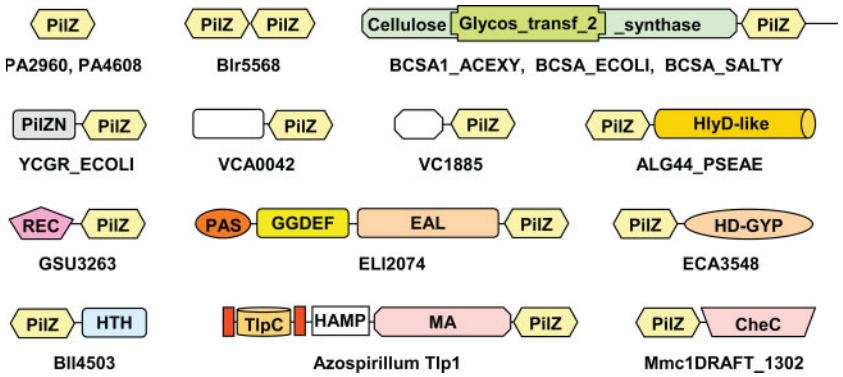


**Fig. 2.** Domain architectures of selected PilZ-containing proteins. The first two rows show the most common domain structures, the last two rows show rare domain combinations that indicate involvement of PilZ in signaling. Proteins are listed by UniProt entry names or their gene names. Source organisms and GenBank/EMBL protein accession numbers are as follows: PA2960 and PA4608, *Pseudomonas aeruginosa* AAG06348 and AAG07996, respectively; Blr5568 and Bll4503, *Bradyrhizobium japonicum* BAC50833 and BAC49768; VCA0042 and VC1885, *Vibrio cholerae* AAF95956 and AAF95033; GSU3263, *Geobacter sulfurreducens* AAR36654; ELI2074, *Erythrobacter litoralis* EAL74814; ECA3548, *Erwinia carotovora* CAG76446; Mmc1draft_1302, *Magnetococcus* sp. MC-1 EAN28119. Domain designations are as follows: Cellulose synthase, Pfam domain PF03552; Glycos_transf_2, PF00535; PilZN and PilZNR, new domains that will be described elsewhere; HlyD-like, a divergent variant of the PF00529 domain; REC, PF00072; PAS, PF0989; GGDEF, PF00990; EAL, PF00563; HD-GYP, COG2206, a variant of PF01966 (Galperin *et al*., 1999); HTH_3, PF01381; TlpC, a periplasmic signaling domain (Greer-Phillips *et al*., 2004); HAMP, PF00672; MCPsignal, PF00015; CheC, COG1776, a protein phosphatase domain (Park *et al*., 2004).

In addition, PilZ domain was found in the Alg44 proteins that are required for alginate biosynthesis in *P.aeruginosa* and *Azotobacter vinelandii* (Maharaj *et al*., 1993; Mejia-Ruiz *et al*., 1997). Unlike cellulose, alginate is a polymer of mannuronic acid that is produced from GDP-mannuronate in a different biosynthetic pathway, which indicates that PilZ plays a regulatory, rather than enzymatic, role in controlling alginate formation and pili biogenesis in *P.aeruginosa*. Given that the phenotype of *pilZ* mutation (see above) is similar to that resulting from a deletion of the GGDEF-EAL domain protein FimX (Huang *et al*., 2003), these observations suggested that PilZ might serve as the c-di-GMP-binding domain of cellulose synthase and as the c-di-GMP-binding protein in other instances.

Binding of c-di-GMP by PilZ explains an earlier observation that an *E.coli* motility defect caused by an *hns* mutation could be reversed either by increased expression of the *yhjH* gene or by inactivation of the *ycgR* gene (Ko and Park, 2000). Since the *yhjH* gene encodes a stand-alone EAL domain, its overexpression results in a decreased cellular level of c-di-GMP, which indeed stimulates motility (Simm *et al*., 2004). The mutation reversal without decreasing the cellular level of c-di-GMP could have been caused by the loss of the c-di-GMP receptor. This suggests that the product of the *ycgR* gene, which contains a PilZ domain (Figs 1 and 2), serves as the cellular receptor for c-di-GMP. Indeed, YcgR and cellulose synthase are the only PilZ-containing proteins encoded in *E.coli* (see Table S1 in the Supplementary Material).

Several other PilZ-containing proteins have been experimentally characterized, revealing phenotypes that are consistent with its role as the c-di-GMP adaptor protein. An *Azospirillum brasilense* chemotaxis receptor Tlp1, which contains a C-terminal PilZ domain (Fig. 2), in addition to energy taxis, was found to be required for colonization of plant roots (Greer-Phillips *et al*., 2004). In *Myxococcus xanthus*, a protein kinase combining the Ser/Thr kinase and PilZ domains was shown to control onset of cell differentiation; its deletion caused premature differentiation, resulting in poor spore production (Muñoz-Dorado *et al*., 1991).

Judging from the structure of PilZ domain (PDB: 1YWU), c-di-GMP binding might require its oligomerization or interaction with additional protein domains. Several alpha-proteobacterial proteins, such as *Bradyrhizobium japonicum* proteins Bll4394 and Blr5568, contain tandem duplications of the PilZ domain, whereas in VCA0042 and related proteins (PDB: 1YLN), PilZ is bound to a separate N-terminal domain, PilZNR (Fig. 2). The same two-domain organization, albeit with an apparently unrelated N-terminal domain (PilZN), is seen in the proteins of the YcgR family. Domain architectures of other PilZ-containg proteins (Fig. 2) include its fusions with signaling domains, such as the CheY-like receiver domain, GGDEF, EAL and HD-GYP domains, and are consistent with the notion that PilZ binds c-di-GMP. In Alg44 family proteins, the PilZ domain is fused to a domain, very similar to HlyD, the membrane component of a multidrug secretion system (Lewis, 2001; Holland *et al*., 2005). Such association could explain the observed role of c-di-GMP in regulating protein secretion and production of extracellular polysaccharide. In addition, PilZ forms a number of clade-specific fusions with uncharacterized domains that are found only in proteins from beta- (e.g. CV2716), gamma- (e.g. VC2344, PA2989) or delta- (e.g. GSU0137, GSU0943) proteobacteria (data not shown).

As would be expected of a c-di-GMP adaptor protein, the phyletic distribution of PilZ domain is generally similar with those of the GGDEF and EAL domains. Like GGDEF and EAL domains, PilZ domain is encoded in many bacterial genomes, including those of early-branching bacteria *Thermotoga maritima* and *Aquifex aeolicus*, but not in any archaeal or eukaryotic genome. Some genomes encode multiple copies of the PilZ domain, up to 15 such genes in *Bdellovibrio bacteriovorus*. Among proteobacteria belonging to beta, gamma and delta subdivisions, chlamydia, spirochetes and several other lineages, there is absolute correlation between presence or absence of consensus GGDEF and EAL domains and presence or absence of the PilZ domain (see Table S2 in the Supplementary Material). In alpha-proteobacteria, however,

this correlation is not absolute. Intracellular bacterial parasites and symbionts, representing genera *Bartonella*, *Brucella*, *Rickettsia*, *Ehrlichia* and *Wolbachia*, appear to encode functional GGDEF domains but do not encode discernible PilZ domains. No PilZ domains have been found in certain GGDEF-encoding actinobacteria, cyanobacteria and firmicutes, including *Staphylococcus aureus*, which has been experimentally shown to respond to exogenous c-di-GMP (Karaolis *et al*., 2005). These organisms might harbor c-di-GMP adaptors other than PilZ or just PilZ-related domains that have diverged beyond recognition by sequence comparison alone.

## 3 CONCLUSIONS

Sequence analysis shows that PilZ domain is encoded in a variety of bacterial genomes with a phyletic pattern similar to those of the diguanylate cyclase (GGDEF) and c-di-GMP-specific phosphodiesterase (EAL) domains. The notion that PilZ serves as c-di-GMP-binding adaptor protein is supported by its presence in bacterial cellulose synthases and other proteins and is consistent with the available experimental data. However, since most genetic data involve loss of function, they only demonstrate that the PilZ domain is necessary for c-di-GMP binding in many bacteria, but not whether it is sufficient for binding or requires additional protein domains. PilZ forms numerous domain associations, which could mediate the diverse signaling mechanisms by c-di-GMP. Many of these domain fusions involve uncharacterized protein domains, opening new avenues for further studies of c-di-GMP-mediated signal transduction.

## ACKNOWLEDGEMENTS

D.A. was supported by Tel-Hai Academic College, Tel-Hai, and Sharett Institute of Oncology, Hadassah University Medical Center, Israel. M.Y.G. was supported by the Intramural Research Program of the National Library of Medicine at the National Institutes of Health. Funding to pay the Open Access publication charges for this article was provided by the NIH Intramural Research Program.

*Conflict of Interest:* none declared.

## REFERENCES

Alm,R.A. *et al*. (1996) Identification of a novel gene, *pilZ*, essential for type 4 fimbrial biogenesis in *Pseudomonas aeruginosa*. *J. Bacteriol*., **178**, 46–53.

Amikam,D. and Benziman,M. (1989) Cyclic diguanylic acid and cellulose synthesis in *Agrobacterium tumefaciens*. *J. Bacteriol*., **171**, 6649–6655.

Bateman,A. *et al*. (2004) The Pfam protein families database. *Nucleic Acids Res*., **32**, D138–D141.

Blanton,R.L. *et al*. (2000) The cellulose synthase gene of *Dictyostelium*. *Proc. Natl Acad. Sci. USA*, **97**, 2391–2396.

Bobrov,A.G. *et al*. (2005) The phosphodiesterase activity of the HmsP EAL domain is required for negative regulation of biofilm formation in *Yersinia pestis*. *FEMS Microbiol. Lett*., **247**, 123–130.

Christen,M. *et al*. (2005) Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. *J. Biol. Chem*., **280**, 30829–30837.

Crooks,G.E. *et al*. (2004) WebLogo: a sequence logo generator. *Genome Res*., **14**, 1188–1190.

D'Argenio,D.A. and Miller,S.I. (2004) Cyclic di-GMP as a bacterial second messenger. *Microbiology*, **150**, 2497–2502.

Galperin,M.Y. *et al*. (1999) A specialized version of the HD hydrolase domain implicated in signal transduction. *J. Mol. Microbiol. Biotechnol*., **1**, 303–305.

Galperin,M.Y. *et al*. (2001) Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiol. Lett*., **203**, 11–21.

Galperin,M.Y. (2004) Bacterial signal transduction network in a genomic perspective. *Environ Microbiol*, **6**, 552–567.

Galperin,M.Y. (2005) A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. *BMC Microbiol*., **5**, 35.

Greer-Phillips,S.E. *et al*. (2004) An energy taxis transducer promotes root colonization by *Azospirillum brasilense*. *J. Bacteriol*., **186**, 6595–6604.

Holland,I.B. *et al*. (2005) Type 1 protein secretion in bacteria, the ABC-transporter dependent pathway. *Mol. Membr. Biol*., **22**, 29–39.

Huang,B. *et al*. (2003) FimX, a multidomain protein connecting environmental signals to twitching motility in *Pseudomonas aeruginosa*. *J. Bacteriol*., **185**, 7068–7076.

Jenal,U. (2004) Cyclic di-guanosine-monophosphate comes of age: a novel secondary messenger involved in modulating cell surface structures in bacteria? *Curr. Opin. Microbiol*., **7**, 185–191.

Karaolis,D.K. *et al*. (2005) c-di-GMP (3′-5′-cyclic diguanylic acid) inhibits *Staphylococcus aureus* cell–cell interactions and biofilm formation. *Antimicrob. Agents Chemother*., **49**, 1029–1038.

Kirillina,O. *et al*. (2004) HmsP, a putative phosphodiesterase, and HmsT, a putative diguanylate cyclase, control Hms-dependent biofilm formation in *Yersinia pestis*. *Mol. Microbiol*., **54**, 75–88.

Ko,M. and Park,C. (2000) Two novel flagellar components and H-NS are involved in the motor function of *Escherichia coli*. *J. Mol. Biol*., **303**, 371–382.

Lewis,K. (2001) In search of natural substrates and inhibitors of MDR pumps. *J. Mol. Microbiol. Biotechnol*., **3**, 247–254.

Maharaj,R. *et al*. (1993) Sequence of the *alg8* and *alg44* genes involved in the synthesis of alginate by *Pseudomonas aeruginosa*. *Gene*, **136**, 267–269.

Matthysse,A.G. *et al*. (2004) A functional cellulose synthase from ascidian epidermis. *Proc. Natl Acad. Sci. USA*, **101**, 986–991.

Mattick,J.S. (2002) Type IV pili and twitching motility. *Annu. Rev. Microbiol*., **56**, 289–314.

Mayer,R. *et al*. (1991) Polypeptide composition of bacterial cyclic diguanylic acid-dependent cellulose synthase and the occurrence of immunologically crossreacting proteins in higher plants. *Proc. Natl Acad. Sci. USA*, **88**, 5472–5476.

Mejia-Ruiz,H. *et al*. (1997) The *Azotobacter vinelandii alg8* and *alg44* genes are essential for alginate synthesis and can be transcribed from an *algD*-independent promoter. *Gene*, **199**, 271–277.

Muñoz-Dorado,J. *et al*. (1991) A gene encoding a protein serine/threonine kinase is required for normal development of *M. xanthus*, a gram-negative bacterium. *Cell*, **67**, 995–1006.

Park,S.Y. *et al*. (2004) Structure and function of an unusual family of protein phosphatases: the bacterial chemotaxis proteins CheC and CheX. *Mol. Cell*, **16**, 563–574.

Paul,R. *et al*. (2004) Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev*., **18**, 715–727.

Römling,U. *et al*. (2005) C-di-GMP: The dawning of a novel bacterial signalling system. *Mol. Microbiol*., **57**, 629–639.

Ryjenkov,D.A. *et al*. (2005) Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: insights into biochemistry of the GGDEF protein domain. *J. Bacteriol*., **187**, 1792–1798.

Saxena,I.M. and Brown,R.M.,Jr (1995) Identification of a second cellulose synthase gene (*acsAII*) in *Acetobacter xylinum*. *J. Bacteriol*., **177**, 5276–5283.

Schmidt,A.J. *et al*. (2005) Ubiquitous protein domain EAL encodes cyclic diguanylate-specific phosphodiesterase: enzymatically active and inactive EAL domains. *J. Bacteriol*., **187**, 4774–4781.

Simm,R. *et al*. (2004) GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. *Mol. Microbiol*., **53**, 1123–1134.

Tal,R. *et al*. (1998) Three *cdg* operons control cellular turnover of cyclic di-GMP in *Acetobacter xylinum*: genetic organization and occurrence of conserved domains in isoenzymes. *J. Bacteriol*., **180**, 4416–4425.

Thomas,C. *et al*. (2004) PsfR, a factor that stimulates *psbAI* expression in the cyanobacterium *Synechococcus elongatus* PCC 7942. *Microbiology*, **150**, 1031–1040.

Tischler,A.D. and Camilli,A. (2004) Cyclic diguanylate (c-di-GMP) regulates *Vibrio cholerae* biofilm formation. *Mol. Microbiol*., **53**, 857–869.

Weinhouse,H. *et al*. (1997) c-di-GMP-binding protein, a new factor regulating cellulose synthesis in *Acetobacter xylinum*. *FEBS Lett*., **416**, 207–211.