

Sequence analysis

Cyanobacterial response regulator PatA contains a conserved N-terminal domain (PATAN) with an alpha-helical insertionKira S. Makarova¹, Eugene V. Koonin¹, Robert Haselkorn² and Michael Y. Galperin^{1,*}¹National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA and ²Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, IL 60637, USA

Received on February 18, 2006; revised on March 9, 2006; accepted on March 10, 2006

Advance Access publication March 16, 2006

Associate Editor: Alex Bateman

ABSTRACT

The cyanobacterium *Anabaena* (*Nostoc*) PCC 7120 responds to starvation for nitrogen compounds by differentiating approximately every 10th cell in the filament into nitrogen-fixing cells called heterocysts. Heterocyst formation is subject to complex regulation, which involves an unusual response regulator PatA that contains a CheY-like phosphoacceptor (receiver, REC) domain at its C-terminus. PatA-like response regulators are widespread in cyanobacteria; one of them regulates phototaxis in *Synechocystis* PCC 6803. Sequence analysis of PatA revealed, in addition to the REC domain, a previously undetected, conserved domain, which we named PATAN (after PatA N-terminus), and a potential helix–turn–helix (HTH) domain. PATAN domains are encoded in a variety of environmental bacteria and archaea, often in several copies per genome, and are typically associated with REC, Roadblock and other signal transduction domains, or with DNA-binding HTH domains. Many PATAN domains contain insertions of a small additional domain, termed α -clip, which is predicted to form a four-helix bundle. PATAN domains appear to participate in protein–protein interactions that regulate gliding motility and processes of cell development and differentiation in cyanobacteria and some proteobacteria, such as *Myxococcus xanthus* and *Geobacter sulfurreducens*.

Contact: galperin@ncbi.nlm.nih.gov**Supplementary information:** http://www.ncbi.nlm.nih.gov/Complete_Genomes/SigCensus/PATAN.html**INTRODUCTION**

The ability of certain bacteria to form multicellular aggregates or filaments is a fascinating and poorly understood phenomenon that has the potential to provide insight into eukaryotic cell development (Haselkorn, 1998; Kaplan, 2003; Sogaard-Andersen, 2004). The filamentous cyanobacterium *Anabaena* PCC 7120 is a convenient model organism for studies of bacterial cell differentiation (Wolk, 1996; Kaneko *et al.*, 2001; Ehira *et al.*, 2003; Zhang *et al.*, 2006). In response to starvation for fixed nitrogen, certain

vegetative cells of *Anabaena* differentiate into nitrogen-fixing cells called heterocysts. Heterocysts do not conduct oxygenic photosynthesis and are surrounded by thick glycolipid cell walls, which allow them to maintain an anaerobic environment required for the nitrogenase activity. Heterocysts are regularly spaced, comprising approximately every 10th cell in the filament (Haselkorn, 1998). Heterocyst formation is subject to a complex regulation which includes the products of more than 10 genes, including *patA*, *patB*, *patN*, *patS*, *hetN*, *hetR*, *hglB* (*hetM*) and some others (Wolk, 1996; Adams, 2000; Meeks and Elhai, 2002; Golden and Yoon, 2003).

One of the components of the heterocyst formation regulatory system is an unusual response regulator PatA (Liang *et al.*, 1992) that contains a CheY-like phosphoacceptor (receiver, REC) domain (listed as domain PF00072 in the Pfam database, Bateman *et al.*, 2004) at its C-terminus. PatA mutants develop heterocysts mostly at the ends of filaments and grow poorly under nitrogen-fixing conditions (Liang *et al.*, 1992). A recent census of response regulators in sequenced prokaryotic genomes revealed multiple copies of PatA-type response regulators encoded in various cyanobacterial genomes (Galperin, 2006). In addition, sequences related to the N-terminal domain of PatA were detected in a variety of diverse bacteria. We report here a detailed sequence analysis of PatA, which revealed three novel conserved domains, and discuss their potential roles in prokaryotic signal transduction.

RESULTS**Domain organization of the PatA protein**

Sequence analysis of the PatA protein from *Anabaena* (*Nostoc*) PCC 7120 (All0521, UniProt accession no. P39048) revealed a complex domain organization, in accordance with an earlier conclusion (Liang *et al.*, 1992) that PatA consists of three distinct domains. In addition to the C-terminal REC domain of PatA, we detected an N-terminal conserved domain, hereafter called PATAN ('pattern', after PatA N-terminus) domain, and a disrupted helix–turn–helix (HTH) domain in the middle of the protein (see below).

*To whom correspondence should be addressed.

PatA <i>Anabaena</i>	P39048	16-178	OPLSLLKIKITGKTI---GCLQVFS--4-WSIYVEEGKLIY-CYS-84-GSYEFIPPE-10-YLNVRLLVQCQQHGRVPE
TaxP1 <i>Synechococcus</i>	Q55448	20-177	RQASFFDGLKQPRFS---GQLILSS--5-WSFYLYMGRIMY-TGG-76-QSMEVACD-12-LIDADQMIABESQQQLWSKWQ
3109D10 <i>Fremyella</i>	Q6H049	9-181	DLRSILQLIELGQRIT---GQLLVKT-27-WFVFFLNQIILY-QEG-73-GSFIHQSS--8-TWEIAPLVSKIAKQLQNWQ
All12165 <i>Anabaena</i>	Q8YV16	16-178	HPLSLLAQLTSRHAT---GCLRVT-4-WTINLEASKLTY-SYS-84-GSYEFFPD-10-YLDRLLVYECQQLRQRQ
All2821 <i>Anabaena</i>	Q8YTA3	11-173	SPOGLLRHLSSCFDIT---TCLQAFS--4-WSIYLEDGTTTY-THS-86-GTYELSNS--8-RLDVTVKVMERCQIRLQNWQ
Ava 3775 <i>Avaria</i>	Q3M6K6	5-159	NVLNEFKTCTQLQYN---GQLIINS--5-WTFYRRLGRIVW-TGG-72-SNFVSISS-13-PTSADVSMQMLBESWKNWS
T110572 <i>Thermos</i>	Q8DLC4	45-202	OPEGLLRQLSTSGGT---GCFKVVV--4-WFLYFDRSDIY-THT-82-GQHQLVPY--7-RFDAASLIEKCKSRIQOWL
Tlr1992 <i>Thermos</i>	Q8DHG7	11-170	SLGEIFRFLEQGQKT---GCLSKP-15-YYIFFRLLQIIVA-TTS-69-GWFKFDAN-11-SASPMDISLAGLRLRDWT
ChP <i>Myxococcus</i>	Q84FE1	13-161	LQLVMPPLFSAPGVE---GVLSVER--4-RCFHVKDCYLVG-SSN-70-GEVGFTPG-10-KLSLSTLHRDAVSRQREWS
GSU0316 <i>Geobacter</i>	Q74GD0	22-114	SLADLIQLKGTNRFS---GCISVEF--4-GVIFFDGGEIVH-EKG-16-GRFTAYPH--8-RKNRQHLLDADRAMDERH
GSU1472 <i>Geobacter</i>	Q74D47	17-111	SKVKIPDMFEKLKSSSFTCYLSFT--4-AIFFFEAGKLIS-MLE-23-GALSVYKL---SKDLTMCLEHMLHGDVLYQ
GSU2224 <i>Geobacter</i>	Q74AX5	11-171	GLGEILQIVSLSRKS---GVLRVNS--4-GEIAFRQCKVIR-SSS-72-GTFDFELQ-20-GLNPQFLAMEGARLIDEQR
Pcar 0683 <i>Pelobacter</i>	Q3A6R5	56-203	AMVDVLSFLNMFRTK---GALVVSLL-4-KEVSLRGEIVVA-ASS-69-GSFFFLEK-10-AMTTQNMIMEGLRRVDERA
PproDRAFT_1493	Q3G2B4	12-160	SVMDDIQWADNSKRS---GTLILFQ--4-KKIYFQDCQIIF-WSD-70-GLFEFIDE-10-RLSSTQLLLESLQKFDETH
Adeh_3688 <i>Anaerobaculum</i>	Q4NTE4	146-240	GVVDLVQTMELMGKKS---GALHVKS--5-AVCYFKDGRILD-ELG-17-GEFAIEFK-9--VSTQGLLMEGMRRIDEWG
DP0804 <i>Desulfotomaculum</i>	Q6AQ40	281-375	AAAELLQIFHMNQKT---GMLFLAL-4-ASVFFLESSVVR-LYK-16-GRYRFTTG-10-IGDFMSILLEGIQQVDEAE
Aq_1845 <i>Aquificoccus</i>	Q67699	11-131	SFADILQVLHHDKKS---GVLIVIEW--4-VAYYIKDCELVLRP-42-GIFSFTQG-10-AFBAEELIMEAARHLTLEE
DR0150 <i>Deinococcus</i>	Q9RVH9	10-101	PLLPVLQMLLVSGR---GVFTVDH--3-GELWFHEHGVLRG-16-GTFFPEPG--8-SLKQDAALHWLLGETDAWA
TC0814 <i>Thermus</i>	Q72JF8	9-97	GPIDLLQLLAQGKKT---GAFRVEG---GEVYLERGRPVH-AYG-16-GRFRFPFG--8-EGPLEAYLLEAVRRLGEGV
AF1501 <i>Archaeoglobus</i>	Q28771	16-109	NFRDILKELSATGFT---GYLEVSY--7-AKVLFSNGKIVA-GIK-22-CVVDVYAL---DEGKVAKALEWNRNAVVEH
MJ0164 <i>Methanohalobium</i>	Q57628	21-111	NLEEIINEIDT-----GYILLV-8-GYIFVEDSKIVG-YYT-22-KVIDYKYK---NKDKINLMLKWLYPEIFACK
MJ0164 <i>Methanocaldococcus</i>	Q57628	161-243	KYIILNAYRKD-----GKFN-----GYILYKQTPIA-AYE-23-TVIDVYEY---NEKTHVILVLEYPQMKILD
PSI-PRED			HHHHHHHHH EEEEE EEEEE EEEEE EEEEE EEEEE HHHHHHHHHHH
Consensus/90%			t...h.t.ht...t G.h.h... h.h.h...sphh.s... .h.h... ..t...hh.....

Fig. 1. Multiple alignments of the PATAN domain. Conserved small residues are shown with green shading, conserved hydrophobic residues are shaded yellow, those favoring a turn are shaded blue, polar residues are in red and other conserved residues are shown in bold. Proteins are listed by their gene and organism names and UniProt accession codes; names of experimentally studied proteins are shown in bold. The two bottom lines show secondary structure prediction by PSI-PRED (Jones, 1999) with α -helices and β -strands indicated by H and E, respectively, and the sequence consensus, where h indicates hydrophobic amino acid residues, s indicates small, p indicates polar, and t indicates favoring a turn. The red lines show position of the α -clip insertion.

PSI-BLAST (Altschul *et al.*, 1997) searches with inclusion cut-off E -value of 0.05 started with the 190 amino acid N-terminal fragment of PatA retrieved \sim 120 sequences from various prokaryotes. Most of the hits were from cyanobacteria and δ -proteobacteria, but putative PATAN domains were also detected in certain representatives of Actinobacteria, Aquificae, Chloroflexi, Deinococcus/Thermus group and in several archaea (see Table 1 in the Supplementary Materials). While this domain had not been listed in public domain databases, N-terminal regions of 10 cyanobacterial PatA proteins have been included in the ProDom database (Bru *et al.*, 2005) as family PD017487 and in Pfam_B as domain PB009947. A list of PATAN-containing proteins with several convenient query sequences and corresponding PSSMs for PSI-BLAST searches are available in the Supplementary Material.

While *patA* is so far the only experimentally characterized PATAN-encoding gene, two genes of this family have been identified in mutation screens. One of the six *patA*-like genes from *Synechocystis* sp. PCC 6803 (*slr0038*) is involved in phototaxis and has been named *taxP1* gene (Yoshihara *et al.*, 2000; Bhaya *et al.*, 2001). In *Myxococcus xanthus*, a PATAN-containing protein is required for adventurous gliding motility (Yoderian *et al.*, 2003). Finally, a *patA*-containing contig from *Fremyella diplosiphon* (also called *Calothrix* sp. PCC 7601) showed an apparent increase in expression when illuminated with red light as compared with green light (Stowe-Evans *et al.*, 2004).

PSI-BLAST-generated multiple sequence alignment of various PATAN domains, combined with secondary structure prediction using several different algorithms (Cuff *et al.*, 1998; Jones, 1999; Kelley *et al.*, 2000), produced a fairly consistent picture of four β -strands flanked by two α -helices (Fig. 1). However, the region between β -strands 3 and 4 varied in length from 16 to 84 amino acid residues and contained from 1 to 4 predicted α -helices,

indicating the presence of an insertion domain (Fig. 2). Using a 65 amino acid fragment of *Geobacter sulfurreducens* protein GSU0213 (UniProt entry Q74GN3), roughly corresponding to this insertion (amino acid residues 55–119), as a query in a PSI-BLAST search with inclusion threshold E -value of 0.1 retrieved a large number of proteins, of which many belong to COG2804 in the COG database (Tatusov *et al.*, 2000) and have been annotated as type II secretory pathway ATPase PufE and/or type IV pilus assembly pathway ATPase PilB. Other hits included enterobacterial bacteriophage N4 receptor protein NfrB, δ -proteobacterial chemotaxis histidine kinase CheA and a variety of uncharacterized multidomain proteins (see Fig. 3 in the Supplementary Materials). Some of these proteins produced two distinct hits, suggesting that this small insertion domain was tandemly duplicated. Indeed, a \sim 30 amino acid-long overlap in these hits was detected, indicating the presence of a repeat unit of \sim 30 amino acid residues. Furthermore, a weak (E -value \sim 0.1–0.5) but reproducible similarity with TPR repeats was observed in many searches starting from different query sequences of this insert domain. In proteins of the PufE/PilB and NfrB families, the query sequence mapped to the region preceding the GspII_E_N domain (PF05157), previously observed in the N-terminal region of the XpsE protein (PDB: 2d27) and found to form a distinct four-helical bundle domain (Chen *et al.*, 2005). Taken together, these observations suggested that the insertion in the PATAN domain is another distinct, small domain (hereinafter, α -clip domain) that contains two repetitive units, each formed by two antiparallel α -helices (in some PATAN domains, these four α -helices are reduced to two or just one). Given that such helical units are mostly found in multidomain proteins that form regulatory protein complexes (Possot *et al.*, 2000; Ulrich and Zhulin, 2005), the α -clip domain is likely to have the same function as TPR repeats, namely, facilitating protein–protein interactions (Groves and Barford, 1999). Several cyanobacteria encode a family

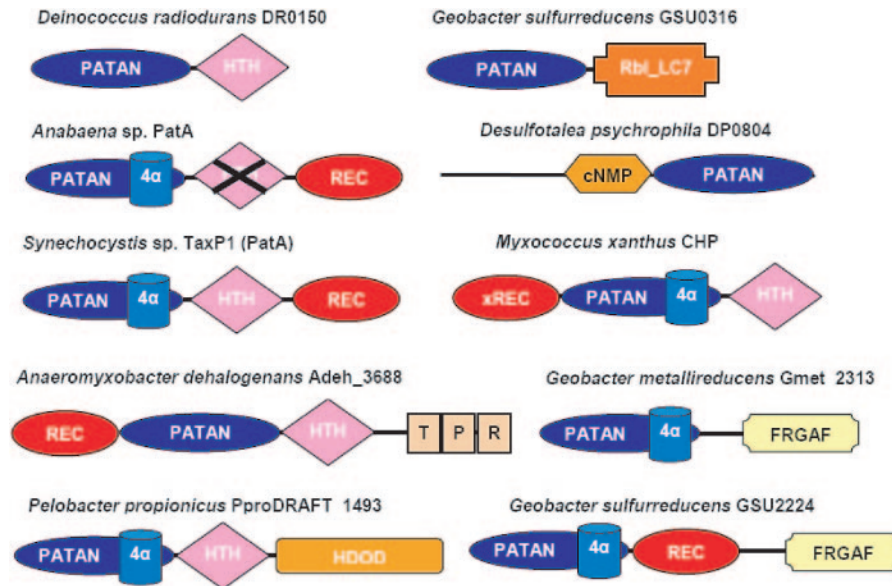


Fig. 2. Domain architectures of proteins containing the PATAN domain. Protein accession numbers are as in Figure 1. Domain designations are as follows: HTH, ArsR-like HTH domain (Fig. 4); Rbl_LC7, Roadblock (Pfam domain PF03259); 4 α , α -clip insert domain (Fig. 3); REC, CheY-like receiver domain, PF00072; cNMP, PF00027; TPR, PF00515; HDOD, PF08668 or COG1639; FRGAF, GAF-related domain of the *M.xanthus* FrgA protein (Fig. 5). Figures 3–5 are available in the Supplementary Materials.

of proteins, exemplified by *Thermosynechococcus elongatus* protein Tlr1669, which consist entirely of α -clip domains and might also participate in protein–protein interactions.

Threading the predicted structure of the PATAN domain against a library of known structural folds using GenThreader and 3D-PSSM (Jones, 1999; Kelley *et al.*, 2000) returned no reliable hits, suggesting that PATAN might have a novel structure. However, the predicted organization of the PATAN domain with its four-strand antiparallel β -sheet and two flanking α -helices is suggestive of the profilin fold, which is found in a variety of ligand-binding signaling domains, including PAS, GAF and Roadblock/LC7 (Taylor and Zhulin, 1999; Ho *et al.*, 2000; Lunin *et al.*, 2004).

Sequence analysis of the 60 amino acid middle region of the PatA family proteins from cyanobacteria revealed a conserved HTH motif similar to the ones in the transcriptional regulators of the ArsR family (see Fig. 4 in the Supplementary Materials). Thus, cyanobacterial PatA family proteins can be predicted to bind DNA and regulate transcription. However, in the original PatA protein from *Anabaena* PCC 7120, the HTH domain appears to be disrupted, suggesting that this protein has lost DNA-binding properties.

Domain architectures of PATAN-containing proteins

In cyanobacteria, PATAN domains are almost invariably found in association with HTH and REC domains, which conforms to the canonical domain architecture of response regulators of PatA type (Fig. 2). The number of *patA* paralogs per genome generally correlates with the genome size but not with the ability of cyanobacteria to form heterocysts (Table 1). Indeed, genomes of non-heterocystous cyanobacteria *Crocospaera watsonii* and *Synechocystis* sp. PCC 6803 encode more PatA-like proteins than heterocystous *Anabaena* PCC 7120 and *Anabaena variabilis*. However, the relatively large genome of the early-branching

cyanobacterium *Gloeobacter violaceus* encodes only a single PatA-type regulator with a truncated PATAN domain, which might be related to the fact that *G.violaceus* does not form thylakoids. The notion that the functions of PatA-type response regulators go beyond regulation of heterocyst formation is also supported by experimental data on their involvement in phototaxis and response to low temperature in *Synechocystis* sp. PCC 6803 (Suzuki *et al.*, 2000; Yoshihara *et al.*, 2000; Bhaya *et al.*, 2001).

While the presence of PatA-type response regulators appears to be a specifically cyanobacterial trait, other bacteria and archaea encode combinations of the PATAN domains with predicted HTH domains that are likely to function as transcriptional regulators. There are also PATAN-containing response regulators that have the characteristic domain organization with the REC domain at their N-termini (Fig. 2). In addition, the PATAN domain is often found in combination with other signaling domains, such as Roadblock/LC7, TPR, cNMP-binding domain and the recently described HD-superfamily response output domain, HDOD (PF08668; Galperin, 2004, 2006). In several δ -proteobacteria, the PATAN domain is fused to FRGAF (FrgA-GAF, see Fig. 5 in the Supplementary Materials), a distinct variant of the GAF domain (Aravind and Ponting, 1997) that is found in the *M.xanthus* protein FrgA (UniProt entry Q9RF11), which is involved in the regulation of *M.xanthus* fruiting body formation (Cho *et al.*, 2000). Domain architectures of PATAN-containing proteins support the hypothesis that the PATAN domain is a signaling domain that participates in prokaryotic cell development and differentiation. After the REC and HTH domains, the third domain most commonly associated with PATAN is the Roadblock/LC7 (PF03259) domain (Bowman *et al.*, 1999; Koonin and Aravind, 2000). This association could already be seen in the original description of the prokaryotic Roadblock/LC7 (MglB-family) domain, which showed the genomic context of the Roadblock/LC7 in *Aquifex aeolicus*: the PATAN-HTH-encoding

aq_1845 gene, sandwiched between *mglB*-like *aq_1847* and the *mglA*-like *aq_1844* gene, has been marked as an uncharacterized gene (Koonin and Aravind, 2000). Adjacent genes encoding PATAN and Roadblock/LC7 domains are also found in many other genomes, including bacteria *Deinococcus radiodurans*, *Chloroflexus aurantiacus*, each of the three PATAN-encoding actinobacterial genomes (*Nocardia farcinica*, *Streptomyces avermitilis*, *Streptomyces coelicolor*), as well as in archaea *Archaeoglobus fulgidus* and *Methanothermobacter thermoautotrophicus*. Most δ -proteobacteria encode at least one fusion protein that combines PATAN and Roadblock/LC7 domains.

DISCUSSION

Despite the abundance of PATAN-encoding genes in bacterial genomes, only three of them have known mutant phenotypes. These include the *Anabaena patA* gene that is required for heterocyst formation (Liang et al., 1992), the *taxP1* gene (*sl10038*) in *Synechocystis* sp., involved in phototaxis (Yoshihara et al., 2000; Bhaya et al., 2001) and response to low temperature (Suzuki et al., 2000) and a gene that is required for adventurous gliding motility in *M.xanthus* (Youderian et al., 2003). Microarray experiments identified two *patA*-like genes as the ones induced by red light (Stowe-Evans et al., 2004) and by low temperature (Ehira et al., 2005). These phenotypes, as well as domain combinations formed by PATAN (Fig. 2), indicate that the PATAN domain is a novel component of prokaryotic signal transduction machinery. In cyanobacteria and δ -proteobacteria, PATAN-containing proteins are involved in chemotaxis and in regulation of complex developmental processes, such as formation of heterocysts (*Anabaena*) or pili and fruiting bodies (*M.xanthus*). The exact function of the PATAN domain remains unclear; the absence of universally conserved amino acid residues (Fig. 1) makes it unlikely that this domain has an enzymatic activity. However, the predicted structural similarity with PAS and GAF domains suggests that, like these domains, PATAN binds specific ligands. Ligand-binding by the PATAN domain might modulate DNA-binding by the HTH domains of predicted transcriptional regulators with PATAN-HTH domain architecture (Fig. 2). In cyanobacterial PatA-type response regulators, which additionally contain the α -clip (Fig. 3 in Supplementary Material) and REC domains, DNA-binding by the predicted HTH domain (Fig. 4 in Supplementary Material) could be additionally modulated by phosphorylation of the REC domain by a two-component sensor kinase and/or protein-protein interactions mediated by the α -clip domain.

Cyanobacterial *patA* genes are mostly found in chemotaxis operons (Bhaya et al., 2001), suggesting that their role involves transduction of environmental signals sensed by the chemotactic machinery. An additional function of the PatA protein in *Anabaena* is implied by the observation that sonicated filaments, so shortened that the terminal heterocysts make up >10% of the cells, nevertheless grow very slowly on N₂ as nitrogen source (Liang et al., 1992). This observation was interpreted to mean that PatA functions in N₂ reduction or in the transport of fixed nitrogen to neighboring vegetative cells. A phylogenetic tree of cyanobacterial PATAN domains (see Fig. 6 in the Supplementary Materials) shows four clear branches, which indicates that these groups have diverged early in cyanobacterial evolution and might reflect their functional divergence. It must be noted that PATAN domains are absent in marine

picocyanobacteria with their relatively small genome sizes (Table 1). They are also missing in such δ -proteobacteria as *Desulfovibrio vulgaris*, which otherwise has a fairly sophisticated signal transduction system (Galperin, 2005). In contrast, some other δ -proteobacteria encode multiple PATAN-containing proteins (Table 1). The reasons for such a phyletic distribution remain unknown and might reflect still poorly understood peculiarities of signal transduction in these bacteria.

In other organisms, frequent association of the PATAN domain with the Roadblock/LC7 domain and occasional association with the GSPII_E_N domain suggest that PATAN participates in multiprotein signaling complexes that could also include MglA-like GTPases, MasK-like Ser/Thr protein kinases (Thomasson et al., 2002) and a variety of other proteins. The identification of the PATAN domain should help in the identification of additional components of this complex signal transduction machinery and in understanding its role(s) in the regulation of prokaryotic cell motility, development, and differentiation.

ACKNOWLEDGEMENTS

The authors thank Patricia Hartzell for helpful discussion. This work was supported by the Intramural Research Program of the National Library of Medicine at the US National Institutes of Health. Funding to pay the Open Access publication charges was provided by the Intramural Research Program of the NIH, National Library of Medicine.

Conflict of Interest: none declared.

REFERENCES

- Adams,D.G. (2000) Heterocyst formation in cyanobacteria. *Curr. Opin. Microbiol.*, **3**, 618–624.
- Altschul,S.F. et al. (1997) Gapped BLAST and PSI-BLAST—a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Aravind,L. and Ponting,C.P. (1997) The GAF domain: an evolutionary link between diverse phototransducing proteins. *Trends Biochem. Sci.*, **22**, 458–459.
- Bateman,A. et al. (2004) The Pfam protein families database. *Nucleic Acids Res.*, **32**, D138–D141.
- Bhaya,D. et al. (2001) Light regulation of type IV pilus-dependent motility by chemosensor-like elements in *Synechocystis* PCC6803. *Proc. Natl Acad. Sci. USA*, **98**, 7540–7545.
- Bowman,A.B. et al. (1999) *Drosophila* roadblock and *Chlamydomonas* LC7: a conserved family of dynein-associated proteins involved in axonal transport, flagellar motility, and mitosis. *J. Cell Biol.*, **146**, 165–180.
- Bru,C. et al. (2005) The ProDom database of protein domain families: more emphasis on 3D. *Nucleic Acids Res.*, **33**, D212–D215.
- Chen,Y. et al. (2005) Structure and function of the XpsE N-terminal domain, an essential component of the *Xanthomonas campestris* type II secretion system. *J. Biol. Chem.*, **280**, 42356–42363.
- Cho,K. et al. (2000) Developmental aggregation of *Myxococcus xanthus* requires *frgA*, an *frz*-related gene. *J. Bacteriol.*, **182**, 6614–6621.
- Cuff,J.A. et al. (1998) JPred: a consensus secondary structure prediction server. *Bioinformatics*, **14**, 892–893.
- Ehira,S. et al. (2003) Genome-wide expression analysis of the responses to nitrogen deprivation in the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res.*, **10**, 97–113.
- Ehira,S. et al. (2005) Identification of low-temperature-regulated ORFs in the cyanobacterium *Anabaena* sp. strain PCC 7120: distinguishing the effects of low temperature from the effects of photosystem II excitation pressure. *Plant Cell Physiol.*, **46**, 1237–1245.
- 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.

- Galperin, M.Y. (2006) Structural classification of bacterial response regulators: diversity of output domains and domain combinations. *J. Bacteriol.*, in press.
- Golden, J.W. and Yoon, H.S. (2003) Heterocyst development in *Anabaena*. *Curr. Opin. Microbiol.*, **6**, 557–563.
- Groves, M.R. and Barford, D. (1999) Topological characteristics of helical repeat proteins. *Curr. Opin. Struct. Biol.*, **9**, 383–389.
- Haselkorn, R. (1998) How cyanobacteria count to 10. *Science*, **282**, 891–892.
- Ho, Y.S. *et al.* (2000) Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. *EMBO J.*, **19**, 5288–5299.
- Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 195–202.
- Kaneko, T. *et al.* (2001) Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res.*, **8**, 205–213.
- Kaplan, H.B. (2003) Multicellular development and gliding motility in *Myxococcus xanthus*. *Curr. Opin. Microbiol.*, **6**, 572–577.
- Kelley, L.A. *et al.* (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.*, **299**, 499–520.
- Koonin, E.V. and Aravind, L. (2000) Dynein light chains of the Roadblock/LC7 group belong to an ancient protein superfamily implicated in NTPase regulation. *Curr. Biol.*, **10**, R774–R776.
- Liang, J. *et al.* (1992) The *patA* gene product, which contains a region similar to CheY of *Escherichia coli*, controls heterocyst pattern formation in the cyanobacterium *Anabaena* 7120. *Proc. Natl Acad. Sci. USA*, **89**, 5655–5659.
- Lunin, V.V. *et al.* (2004) The structure of the MAPK scaffold, MP1, bound to its partner, p14. A complex with a critical role in endosomal map kinase signaling. *J. Biol. Chem.*, **279**, 23422–23430.
- Meeks, J.C. and Elhai, J. (2002) Regulation of cellular differentiation in filamentous cyanobacteria in free-living and plant-associated symbiotic growth states. *Microbiol. Mol. Biol. Rev.*, **66**, 94–121.
- Possot, O.M. *et al.* (2000) Multiple interactions between pullulanase secretion components involved in stabilization and cytoplasmic membrane association of PulE. *J. Bacteriol.*, **182**, 2142–2152.
- Sogaard-Andersen, L. (2004) Cell polarity, intercellular signalling and morphogenetic cell movements in *Myxococcus xanthus*. *Curr. Opin. Microbiol.*, **7**, 587–593.
- Stowe-Evans, E.L. *et al.* (2004) Genomic DNA microarray analysis: identification of new genes regulated by light color in the cyanobacterium *Fremyella diplosiphon*. *J. Bacteriol.*, **186**, 4338–4349.
- Suzuki, I. *et al.* (2000) The pathway for perception and transduction of low-temperature signals in *Synechocystis*. *EMBO J.*, **19**, 1327–1334.
- Tatusov, R.L. *et al.* (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.*, **28**, 33–36.
- Taylor, B.L. and Zhulin, I.B. (1999) PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol. Mol. Biol. Rev.*, **63**, 479–506.
- Thomasson, B. *et al.* (2002) MglA, a small GTPase, interacts with a tyrosine kinase to control type IV pili-mediated motility and development of *Myxococcus xanthus*. *Mol. Microbiol.*, **46**, 1399–1413.
- Ulrich, L.E. and Zhulin, I.B. (2005) Four-helix bundle: a ubiquitous sensory module in prokaryotic signal transduction. *Bioinformatics*, **21** (Suppl. 3), iii45–iii48.
- Wolk, C.P. (1996) Heterocyst formation. *Annu. Rev. Genet.*, **30**, 59–78.
- Yoshihara, S. *et al.* (2000) Novel putative photoreceptor and regulatory genes required for the positive phototactic movement of the unicellular motile cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Cell Physiol.*, **41**, 1299–1304.
- Youderian, P. *et al.* (2003) Identification of genes required for adventurous gliding motility in *Myxococcus xanthus* with the transposable element *mariner*. *Mol. Microbiol.*, **49**, 555–570.
- Zhang, C.C. *et al.* (2006) Heterocyst differentiation and pattern formation in cyanobacteria: a chorus of signals. *Mol. Microbiol.*, **59**, 367–375.