

# Characterization of Genes Encoding Multi-domain Proteins in the Genome of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena* sp. Strain PCC 7120

Masayuki OHMORI,<sup>1</sup> Masahiko IKEUCHI,<sup>1</sup> Naoki SATO,<sup>2</sup> Peter WOLK,<sup>3</sup> Takakazu KANEKO,<sup>4</sup> Teruo OGAWA,<sup>5</sup> Minoru KANEHISA,<sup>6</sup> Susumu GOTO,<sup>6</sup> Shuichi KAWASHIMA,<sup>6</sup> Shinobu OKAMOTO,<sup>1</sup> Hidehisa YOSHIMURA,<sup>1</sup> Hiroshi KATO,<sup>1</sup> Takatomo FUJISAWA,<sup>1</sup> Shigeki EHIRA,<sup>2</sup> Ayako KAMEI,<sup>1</sup> Shizue YOSHIHARA,<sup>1</sup> Rei NARIKAWA,<sup>1</sup> and Satoshi TABATA<sup>4,\*</sup>

*Department of Life Sciences (Biology), The University of Tokyo, Komaba 3-8-1, Meguro, Tokyo 153-8902, Japan,<sup>1</sup> Department of Biochemistry and Molecular Biology, Faculty of Science, Saitama University, 255 Shimo-Ohkubo, Saitama City, Saitama 338-8570, Japan,<sup>2</sup> MSU-DOE Plant Research Laboratory, Michigan State University, E. Lansing, MI 48824, U.S.A.,<sup>3</sup> Kazusa DNA Research Institute, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan,<sup>4</sup> Bioscience Center, Nagoya University, Chikusa, Nagoya 464-8601, Japan,<sup>5</sup> and Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan<sup>6</sup>*

(Received 24 September 2001; revised 6 December 2001)

## Abstract

Computational analysis of gene structures in the genome of *Anabaena* sp. PCC 7120 revealed the presence of a large number of genes encoding proteins with multiple functional domains. This was most evident in the genes for signal transduction pathway and the related systems. Comparison of the putative amino acid sequences of the gene products with those in the Pfam database indicated that GAF and PAS domains which may be involved in signal recognition were extremely abundant in *Anabaena*: 87 GAF domains in 62 ORFs and 140 PAS domains in 59 ORFs. As for the two-component signal transduction system, 73, 53, and 77 genes for simple sensory His kinases, hybrid His kinases and simple response regulators, respectively, many of which contained additional domains of diverse functions, were presumptively assigned. A total of 52 ORFs encoding putative Hanks-type Ser/Thr protein kinases with various domains such as WD-repeat, GAF and His kinase domains, as well as genes for presumptive protein phosphatases, were also identified. In addition, genes for putative transcription factors and for proteins in the cAMP signal transduction system harbored complex gene structures with multiple domains.

**Key words:** GAF domain; multi-domain protein; PAS domain; photoreceptor domain; protein kinase; two-component system

## 1. Introduction

Cyanobacteria constitute a single but large taxonomic and phylogenetic group within the domain Eubacteria. The distinctive characteristic of cyanobacteria is their oxygenic photosynthesis ability. By the use of two photosystems, PSII and PSI, they are capable of utilizing H<sub>2</sub>O as a reductant source in photosynthesis, thereby producing molecular oxygen, as eukaryotic algae and plants do. Cyanobacteria are either unicellular or filamentous. Filamentous species may be subdivided into those with, and those without, heterocysts, differentiated cells that are specialized for fixation of N<sub>2</sub> in an oxygen-containing en-

vironment. Heterocysts, which appear in very early fossil records,<sup>1,2</sup> are thought to have evolved when the concentration of cyanobacteria-produced O<sub>2</sub> in the Earth's atmosphere reached levels capable of inactivating the N<sub>2</sub>-fixing enzyme, nitrogenase.

Sequencing of a genome of a unicellular cyanobacterium *Synechocystis* sp. PCC 6803 (*Synechocystis*) was completed in 1996.<sup>3</sup> By utilizing the sequence information, many studies concerning the structure, function and expression of *Synechocystis* genes have been carried out to date. Consequently, a large body of knowledge has been accumulated on the functions of 3000 genes in the 3.57-Mb genome. However, *Synechocystis* is merely one genus among the diverse cyanobacterial species, and extensive studies of genes carried by various cyanobacteria are needed to understand the genetic systems in

Communicated by Mituru Takanami

\* To whom correspondence should be addressed. Tel. +81-438-52-3933, Fax. +81-438-52-3934, E-mail: tabata@kazusa.or.jp

cyanobacteria. Very recently, Kaneko et al.<sup>4</sup> reported the complete sequence of the entire genome of a filamentous cyanobacterium *Anabaena* sp. strain PCC 7120 (*Anabaena*). This strain has been used extensively for genetic studies on nitrogen fixation and on cell differentiation during the heterocyst formation, and thus, combined with the known information of the entire genome structure, provides an excellent system to study common and unique features of genes among cyanobacteria.

The purpose of this study was to extend our understanding of the genetic system of *Anabaena* by clustering the predicted proteins at the genome-wide level and by analyzing their domains and domain organization. On the other hand, the genome project of *Nostoc punctiforme* PCC 73102, another filamentous cyanobacterium, is in progress and draft sequence data are available on the Web ([http://www.jgi.doe.gov/JGI\\_microbial/html/nostoc/nostoc\\_homepage.html](http://www.jgi.doe.gov/JGI_microbial/html/nostoc/nostoc_homepage.html)). However, it is premature to analyze each protein-coding sequence in detail because the data are still preliminary. Here, we first clustered the amino acid sequences of most of the putative gene products in the *Anabaena* genome and generated a list of genes in each cluster. Careful inspection of the largest cluster, consisting of 470 open reading frames (ORFs), revealed that the cluster was generated due to multiple functional domain structures in each of the putative protein sequences, and that majority of the genes in this cluster are those for signal transduction pathways. Accordingly, in this study, we focused on the analysis of genes involved in signal transduction and related pathways, and the characteristic features of the functional domains of the *Anabaena* genes are described.

## 2. Materials and Methods

### 2.1. Definition and designation of strains

In taxonomic studies that predated axenic cultures, species of *Nostoc* were distinguished from species of *Anabaena* by the disposition of the former to generate firm macroscopic assemblages.<sup>5</sup> Rippka and coauthors first designated PCC 7120 as *Anabaena* sp.<sup>6</sup> but later (see ref<sup>7</sup>) as *Nostoc* sp.<sup>8</sup> Where *Nostoc* cluster 1 includes a bona fide *Nostoc*, *Nostoc punctiforme* strain PCC 73102, and cluster 3 includes PCC 7120, those authors<sup>8</sup> state that "it is highly likely that members of 'Nostoc' clusters 1, 2, and 3 may in fact represent different genera rather than distantly related species of the same genus." Also, Wilmotte and Herdman<sup>9</sup> have pointed out the need for more detailed studies of the genotypic relationships of heterocyst-forming strains. Pending such studies, we have in this paper continued with the earlier designation, *Anabaena* sp. strain PCC 7120.

For readability, *Anabaena*, *Synechocystis*, and *Nostoc* are used in the present paper for *Anabaena* sp. PCC 7120, *Synechocystis* sp. PCC 6803, and *Nostoc punctiforme*

PCC 73102 (= ATCC 29133), respectively.

### 2.2. Computer analyses

The complete set of predicted protein sequences of *Anabaena* sp. PCC 7120<sup>4</sup> were extracted from the main chromosome (*all0001–all5371*) and six plasmids (*all7001–alr9505*).

Gene clustering was performed using the SISEQ package<sup>10</sup> and other original software developed in the laboratory. Homology analysis was performed using the BLAST (NCBI-BLAST, versions 2.1–2.2) and PSI-BLAST (version 2.1.1) programs<sup>11</sup> running locally or on the Web (non-redundant GenBank, SwissProt in NCBI and GenomeNet, Cyanobase, and various cyanobacterial draft sequence databases in JGI).

Motif analysis was performed by Pfam (versions 6.0–6.4)<sup>12</sup> and SMART (version 3.1)<sup>13</sup> searches performed locally or on the Web (<http://www.sanger.ac.uk/Pfam/>, <http://smart.embl-heidelberg.de/>). It is known that the three-dimensional (3D) structure of various PAS domains is conserved, whereas their primary sequences are rather diverse. A PAS domain consists of a PAS-core motif (Pfam PF00989), a helical connector and a PAC motif (Pfam PF00785) but all motifs were not always identified by Pfam search. Therefore, when only a portion of a PAS domain was predicted, the entire PAS domain was determined manually. A His kinase (or transmitter) domain generally combines an ATPase subdomain (HATPase\_c; Pfam PF02518) and a signal subdomain (Pfam PF00512) or an HPT subdomain (Pfam PF01627), which contains the autophosphorylatable His residue, whereas the response regulator proteins bear receiver domains (Pfam PF00072) that carry a phospho-acceptor Asp residue. Cluster analysis of GAF and PAS domains was done by automatic sequence alignment and classification with the neighbor-joining algorithm using the ClustalX program.<sup>14</sup> The clusters were visualized as a non-rooted tree.

When the total number of genes and Pfam-defined domain structures of *Anabaena* ORFs were compared with those of other bacteria including *Synechocystis*, the complete list was obtained from the GTOPI web site at <http://spock.genes.nig.ac.jp/~genome/gtop.html>.

## 3. Results and Discussion

### 3.1. Clustering of *Anabaena* proteins

Similarity among the 6004 protein-coding genes in the *Anabaena* genome, excepting those for putative transposases, was analyzed by using the BLASTP scores with a cutoff value of  $1 \times 10^{-8}$ . This cutoff value, selected after several trials, gave reasonable clustering results. Accordingly, proteins with multiple domains were clustered by similarity of at least one of the domains. The largest cluster consisting of 470 ORFs was obtained for pro-

teins having the following Pfam motifs: GAF and PAS domains, transcription factors, response regulators, and protein kinases. The size of this cluster was much larger in *Anabaena* than those of *Synechocystis* (133 ORFs) and *Pseudomonas aeruginosa* (213 ORFs). Such a difference can be partly explained by the fact that *Anabaena* is exceptionally abundant in multi-domain protein sequences. Examples are multifunctional adenylate cyclase, CyaC (All4963)<sup>15</sup> and hybrid His kinases (All0729 and Alr2428) (Fig. 1C). Since many of them are assumed to be involved in signal transduction, the large size seems to be correlated with the complex physiology of *Anabaena* as mentioned earlier. In this study, we focused on the structure of genes with multi-domains in the largest cluster, especially those involved in the signal transduction cascade.

### 3.2. Proteins with GAF domains

The GAF domain is a distinctive but rather diverged motif consisting of approximately 150 amino acid residues, which was first recognized in cGMP-specific phosphodiesterases, *Anabaena* adenylate cyclase and *Escherichia coli* FhlA protein.<sup>16</sup> The survey of the *Anabaena* genome revealed the presence of 87 GAF domains encoded by 62 ORFs, 60 of which are on the chromosome (Fig. 2), although their functional roles remain mostly unknown. This is much larger than in *Synechocystis* (33 GAF domains encoded by 28 ORFs), or other bacteria, for example, 6 GAF domains encoded by 6 ORFs in *P. aeruginosa*, and 8 GAF domains encoded by 7 ORFs in *E. coli*. Proteins containing three or four GAF domains were found exclusively in cyanobacteria: There were 9 and 2 ORFs for such proteins in *Anabaena* (see also Fig. 1A) and *Synechocystis*, respectively, whereas only one or two GAF domains were detected in single proteins in other organisms including eukaryotes.

The 87 GAF domains could be categorized into more than ten subclasses based on the cluster analysis (Fig. 2). Four of them were recognized as follows: (I) chromophore-binding domains of bacteriophytochrome, (II) cGMP-binding domains, (III) a group of domains in Ser/Thr-type protein kinases, and (IV) a group of domains in the LuxR family of transcription factors.

*Anabaena* had 16 ORFs encoding 24 putative chromophore-binding GAF domains, which are related to plant phytochromes and bacteriophytochromes that act as photoreceptors for various light signals. This is unique to *Anabaena*, since other heterotrophic bacteria have none or only a few such domains. The plant phytochromes and bacteriophytochromes covalently bind a linear tetrapyrrole molecule, such as phytychromobilin or phycocyanobilin, via a Cys residue (CH-type and CX-type), a His residue (XH-type) or an unknown residue (XX-type) in the subclass I GAF domains.<sup>17</sup>

Of the 24 chromophore-binding GAF domains in the *Anabaena* genome, 12 were classified as CH-type, 6 as CX-type, 4 as XH-type and 2 as XX-type.

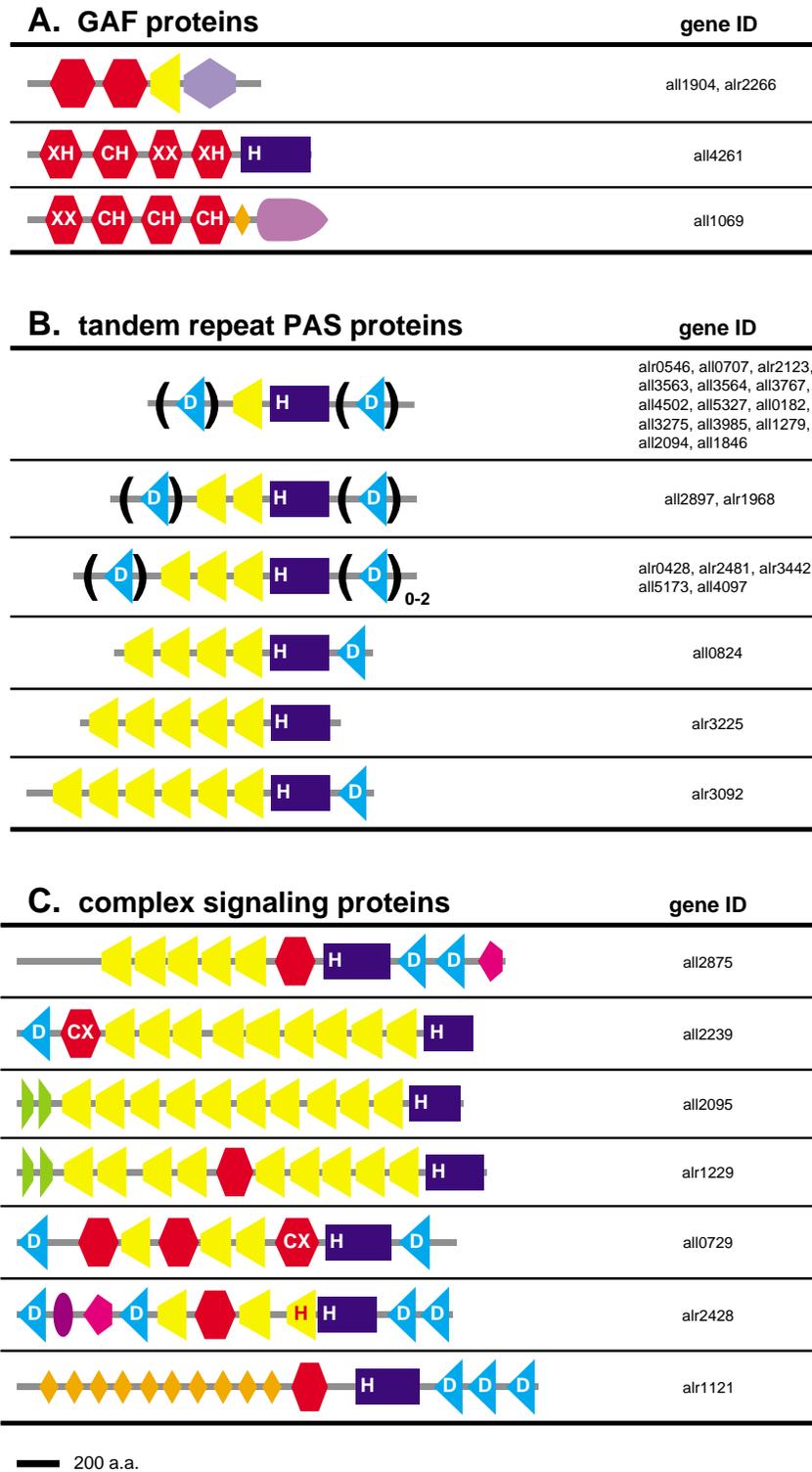
Two *Anabaena* phytochrome genes, *aphA* (*alr3157*) and *aphB* (*all2899*), were highly homologous to each other, although *aphA* encodes a CH motif, whereas *aphB* encodes an XH motif. Other genes of known function in this subclass were *all1069*, a *pixJ1* homolog for positive phototaxis,<sup>18</sup> and *all1688*, a *cikA* homolog for circadian adjustment.<sup>19</sup> Fourteen of 16 ORFs encoded a His kinase domain as an output domain.

Subclass II (putative cGMP-binding GAF domains) was found in adenylate cyclases CyaB1 (All1904) and CyaB2 (Alr2266), suggesting that binding of cGMP may regulate the activity of adenylate cyclase (Fig. 1A). Domains in subclass III were unique to 13 ORFs of Ser/Thr-type protein kinases (subclass HstK, see section 3.6) and an additional ORF (*alr2137*), while domains in subclass IV were unique to 3 ORFs of the LuxR family of transcription factors, as mentioned in section 3.4. GAF domains of subclass III and IV were present in some cyanobacteria but not in heterotrophic bacteria.

### 3.3. Proteins with PAS domains

PAS domains are important signaling modules that monitor changes in light, redox potential, oxygen, small ligands, and the overall energy level of a cell.<sup>20</sup> A total of 140 PAS domains was detected in 59 ORFs on the *Anabaena* chromosome (Fig. 3). This is a much larger figure than is found in other organisms: There were 30 PAS domains encoded by 21 ORFs in *Synechocystis*, 30 domains by 24 ORFs in *P. aeruginosa*, 35 domains by 21 ORFs in *Archaeoglobus fulgidus*. Multiple PAS domains were often found in a single protein, and *Anabaena* was unique in this regard. The largest ORF (*all2095*) encodes 10 PAS domains, and 10 additional ORFs encode more than 4 PAS domains in each putative product (Fig. 1C), whereas other bacterial proteins contain at most 3 PAS domains. Some multiple PAS domains may have originated from tandem duplication in *Anabaena*. Interestingly, increasing PAS domains from one to six were found in a series of His kinases and/or hybrid His kinases (Fig. 1B).

The 140 PAS domains could be grouped into approximately 65 subclasses (Fig. 3) based on sequence alignment. However, putative function could be predicted for only four PAS domains. Three of them, encoded by *alr1229*, *all2875*, and *alr3170*, were significantly homologous to the FMN-binding pocket of blue light receptors NPH1, NPL1, and PHY3 in plants.<sup>21</sup> The Cys residue in the pocket that forms an adduct with flavin during photocycling was conserved in All2875 and Alr3170 but not in Alr1229. This Cys residue is not essential for flavin binding, as shown by site-directed mutagenesis in NPH1.<sup>22</sup>



**Figure 1.** Domain structure of typical or extreme examples of deduced signaling proteins in the Anabaena genome. Each domain is defined in the text. The domain structure from N-terminus to C-terminus is shown from left to right. Domains are depicted mostly as the number of amino acid residues (see scale) except when many proteins in one category are represented schematically. Potentially chromophore-binding GAF domains are subdivided according to their conserved Cys/His residues (section 3.1) and indicated with C/H/X within domain boxes, whereas the other GAF domains are shown with simple domain boxes. Parentheses indicate domains that are not present in all of the listed proteins. Subscript 0–2 means that the number of the specified domain ranges from 0 to 2.

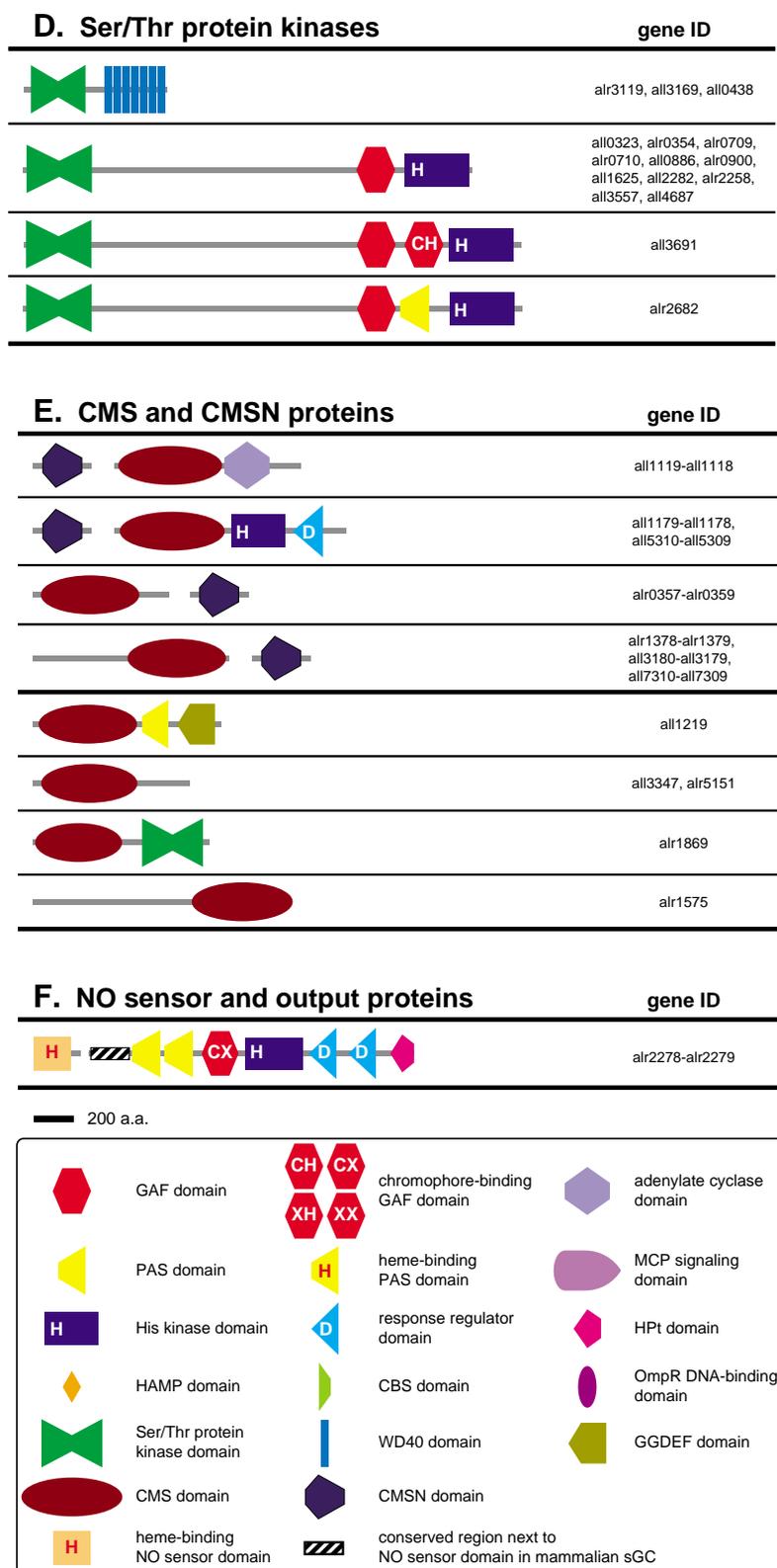
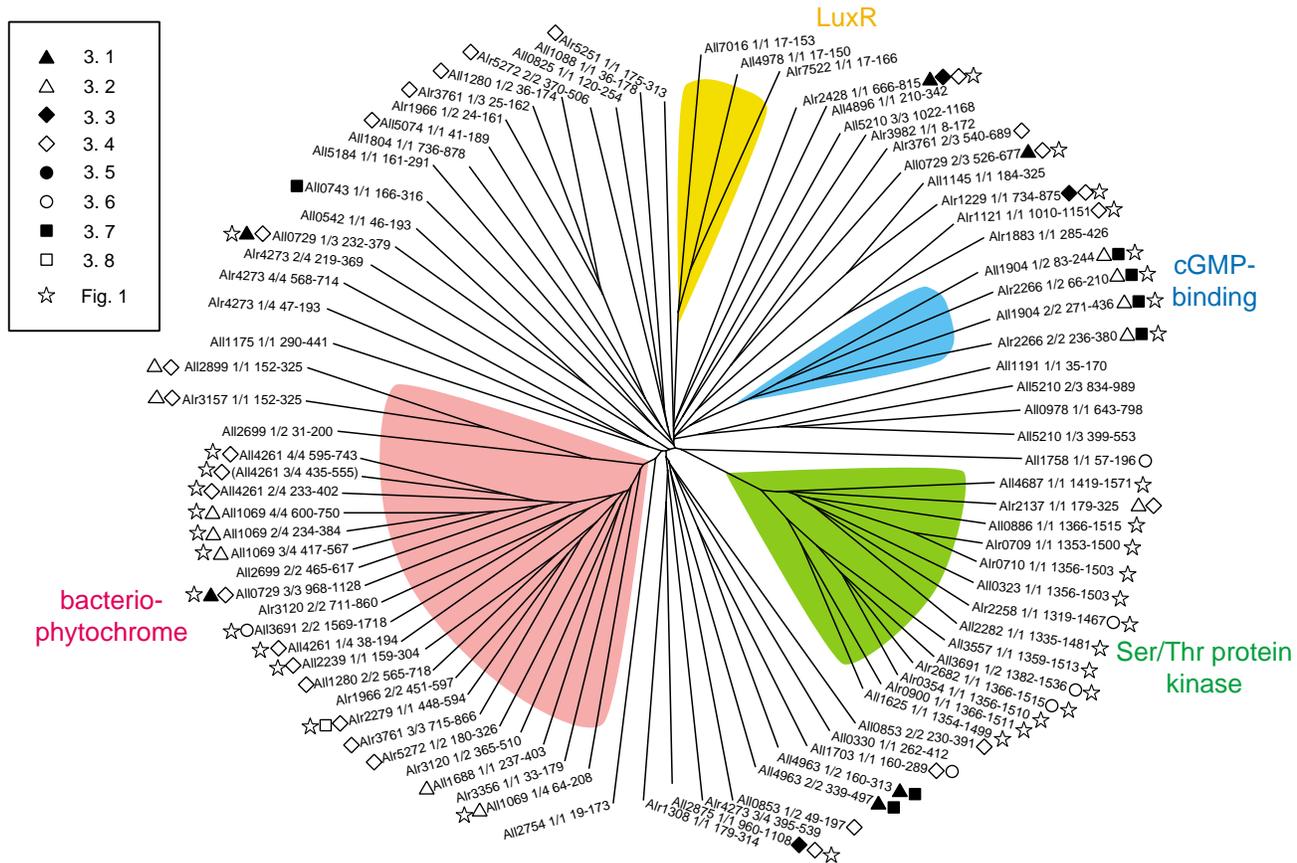


Figure 1. Continued.



**Figure 2.** Dendrogram of GAF domains in the Anabaena genome. The complete list of 87 GAF domains is represented as gene ID number plus domain number and sequence position. For example, “All1069 1/4 64–208” represents the first of the four GAF domains, which is located from position 64 to 208 in All1069 protein. Domains in proteins which were referred to in the text or in Fig. 1 are shown with symbols.

Thus, the potentially flavin-binding PAS domain encoded by *alr1229* may serve not as a photoreceptor but as a sensor for such other stimuli as redox changes. Another PAS domain whose putative function could be deduced was that for heme binding in Alr2428. The best characterized, similar domain is in the heme-based oxygen sensor FixL of rhizobia, which regulates the expression of nitrogen fixation genes.<sup>23</sup> The His residue that ligates the heme iron was also conserved in Alr2428. Considering that no such PAS domain was detected in the *Synechocystis* genome, the heme-binding PAS domain of Alr2428 may be involved in nitrogen fixation in the heterocyst of *Anabaena*.

### 3.4. Two-component regulatory systems

Two-component regulatory systems play central roles in regulation of a number of processes in prokaryotes. *Anabaena* had a remarkably large number of genes for two-component systems. Based on Pfam search and examination of sequence alignment,<sup>4</sup> we detected 73, 53, and 77 ORFs that may encode simple sensory His kinases,

hybrid sensory His kinases and simple response regulators, respectively, in the *Anabaena* genome including its plasmids (Table 1). These numbers are quite large in comparison with those in the genome of the heterotrophic bacterium *P. aeruginosa* (43, 19, and 72, respectively)<sup>24</sup> or the unicellular cyanobacterium *Synechocystis* (25, 16, and 41, respectively).<sup>3</sup> The highly developed genetic system for the two-component phosphotransfer may reflect the complicated physiology of *Anabaena* including its filamentous cell organization, heterocyst development and N<sub>2</sub> fixation, in addition to photosynthesis.

#### 3.4.1. Sensory His kinases

According to the domain architecture, the 73 predicted products were categorized into 6 subclasses: (I) GAF (some, plus PAS) followed by His kinase domains, (II) PAS (with GAF for some cases) followed by His kinase domains, (III) HAMP followed by His kinase domains, (IV) Ser/Thr protein kinase followed by GAF and His kinase domains, (V) a single His kinase domain, and (VI) a novel His kinase domain with multiple PAS domains.

Table 1. Lists of genes for two-component regulatory systems.

| [His kinase]         |            |              |              |              |               |
|----------------------|------------|--------------|--------------|--------------|---------------|
| I. GAF+HK            | II. PAS+HK | III. HAMP+HK | IV. PK+HK    | V. simple HK | VI. novel HK  |
| all0330              | alr0428    | alr1192      | all0323      | alr0117      | all1280       |
| all0542              | alr0546    | alr1551      | alr0354      | alr1171      | all2095       |
| all0825              | alr0642    | alr2739      | alr0709      | alr1665      |               |
| all0853              | all0707    | alr3511      | alr0710      | alr2572      |               |
| all1088              | all1145    | alr4105      | all0886      | all2772      |               |
| all1191              | alr1229    | all4496      | alr0900      | all2883      |               |
| alr1308              | alr2123    | alr4586      | all1625      | all2956      |               |
| all1688              | alr2481    | all4726      | alr2258      | alr3037      |               |
| alr1966              | alr3225    | alr5189      | all2282      | alr3155      |               |
| alr2137              | all3563    |              | alr2682      | all3167      |               |
| all2699              | alr3767    |              | alr3557      | all3359      |               |
| all2899              | all4502    |              | alr3691      | alr3547      |               |
| alr3157              | alr5272    |              | all4687      | all3587      |               |
| alr3982              | all5327    |              |              | all3600      |               |
| all4261              |            |              |              | alr4716      |               |
| all5074              |            |              |              | alr4882      |               |
|                      |            |              |              | alr4905      |               |
|                      |            |              |              | all7583      |               |
|                      |            |              |              | all7605      |               |
| [Hybrid His kinase]  |            |              |              |              |               |
| I. HK+RR             | II. HK+mRR | III. RR+HK   | IV. RR+HK+RR | V. cheA      | VI. composite |
| all0182              | alr1121    | all0637      | all0729      | all0926      | alr2428       |
| all0638              | alr1883    | all1177      | all1846      | all1068      |               |
| all0824              | alr2279    | alr1231      | alr1968      | all2161      |               |
| all0978              | all2875    | all1279      | all4097      |              |               |
| all1178              | alr3761    | all1388      | all4963      |              |               |
| alr1285              | all5173    | all2094      |              |              |               |
| all1389              |            | all2239#     |              |              |               |
| all1639              |            | alr2241      |              |              |               |
| all1716              |            | alr2307      |              |              |               |
| all1804              |            | all2897      |              |              |               |
| all2379              |            | alr3121      |              |              |               |
| alr3092              |            | alr3159      |              |              |               |
| alr3120              |            | alr3671      |              |              |               |
| all3275              |            | all3764      |              |              |               |
| alr3442              |            | all3765      |              |              |               |
| all3985              |            | all4096      |              |              |               |
| alr4878              |            | alr4879      |              |              |               |
| all5210              |            | alr4880      |              |              |               |
| all5309              |            | all5308      |              |              |               |
| [Response regulator] |            |              |              |              |               |
| I. OmpR              | II. NarL   | III. AraC    | IV. CheY     | V. PatA      | VI. Others    |
| alr0072              | alr0913    | alr0429      | alr0264      | all0521      | alr0356       |
| all0129              | all1704    | all3232      | alr0442      | all0930      | alr0960       |
| alr1170              | all1736    |              | alr0774      | all1072      | all1012       |
| alr1194              | alr2138    |              | all0823      | all2165      | alr1086       |
| all1964              | alr3156    |              | all0929      | all2821      | alr1230       |
| alr3260              | all3660    |              | all1071      |              | all1281       |
| all3788              | alr3768    |              | all1640      |              | all1703       |
| all3822              | all4635    |              | alr1967      |              | alr2280       |
| all4312              | all5069    |              | alr2049      |              | all2281       |
| all4503              | alr5150    |              | all2164      |              | alr2306       |
| all4727              | alr7219    |              | alr2240      |              | all3759       |
| all4750              | alr8531    |              | alr2429      |              | all2874       |
| all5174              | alr8535    |              | alr2726      |              | all3348       |
| alr5188              | alr9013    |              | all2898      |              | alr3599       |
| all5323              |            |              | all2955      |              | alr3920,      |
| all7584              |            |              | alr3158      |              | alr5251       |
| all7606              |            |              | all3239      |              |               |
|                      |            |              | alr3386      |              |               |
|                      |            |              | alr3594      |              |               |
|                      |            |              | all3766      |              |               |
|                      |            |              | all4927      |              |               |
|                      |            |              | all5172      |              |               |
|                      |            |              | alr5328      |              |               |
| [HPt]                |            |              |              |              |               |
| simple HPt           |            |              |              |              |               |
| alr4086              |            |              |              |              |               |
| all8565              |            |              |              |              |               |

\* HK, His kinase domain; mPAS, multiple PAS domains; PK, Ser/Thr protein kinase domain; RR, receiver domain of response regulator; mRR, multiple receiver domains.

# includes novel His kinase domain.



brid kinase may encode a novel His kinase because a specific His residue was present at an unusual position instead of the typical signal subdomain (i.e. His at position 752 in All1280). A region flanking this His residue was conserved among the three predicted His kinase domains and also found in a few proteins of some bacteria and archaea. We propose that these ORFs encode a novel subclass of sensory His kinases. As a sensor domain, All1280 had a motif for bacteriophytochrome-like chromophore binding, while All2095 and All2239 had 10 and 9 PAS domains, respectively.

### 3.4.2. Hybrid sensory His kinases

The 53 ORFs that encode hybrid sensory kinases containing both His kinase (transmitter) domains and response regulator (receiver) domains were divided into six subclasses: (I) one transmitter domain followed by one receiver domain, (II) one transmitter domain followed by multiple receiver domains, (III) a receiver domain followed by a transmitter domain, (IV) a receiver domain followed by a transmitter and another receiver domain; (V) a CheA domain plus a receiver domain, and (VI) a composite type (Table 1).

Numerous multi-domain proteins unique to *Anabaena* were found in this category. The most complex one was Alr2428 that bears at least 6 different domains (receiver, OmpR-type DNA-binding, HPT [His-containing phosphotransmitter], PAS, GAF and transmitter domains) (Fig. 1C). As mentioned in section 3.4, one of the PAS domains was homologous to the FixL-type heme-binding motif. A few examples having at least four different domains are shown in Fig. 1C (All0729, Alr1121, Alr1229, All2239, and All2875).

HPT is a specific domain that receives a phosphate group from a receiver domain at a His residue and transmits it to another receiver in a multistep phosphotransfer processes.<sup>28</sup> The *Anabaena* genome harbored 7 ORFs whose predicted products carry an HPT domain. They included three ORFs of subclass II, three *cheA*-like ORFs of subclass V, one ORF (*alr2428*) of subclass VI, one response-regulator ORF (*all5174*) (next section) and two ORFs with a simple HPT domain (Table 1). The three ORFs in subclass II (*alr2279*, *all2875*, and *alr3761*) predicted the same arrangement of a transmitter plus two consecutive receivers followed by a C-terminal HPT domain, whereas their N-terminal sensor regions consisted of different combinations of PAS and GAF domains. The C-terminal HPT domain probably transfers a phosphate group to another response regulator as does ArcB in *E. coli*.<sup>28</sup> The two simple HPT proteins (Alr4086 and All8565) may mediate phosphotransfer as does Spo0F<sup>29</sup> because they do not have an extra domain.

### 3.4.3. Response regulators

Seventy-seven ORFs presumptively encoded simple response regulators that carry one or more receiver domains but no transmitter domain. Almost all receiver domains in *Anabaena* consisted of three signature sequences containing the specific Asp residue that is phosphorylated via phosphotransfer from the transmitter His residue of the sensory His kinase. According to potential output domains,<sup>28</sup> the predicted products of 77 ORFs could be classified into six subclasses: (I) the OmpR family of transcription factors, (II) the transcription factor NarL subfamily of the LuxR family, (III) the AraC family of transcription factors, (IV) the CheY subfamily of response regulators, (V) the PatA subfamily of response regulators and (VI) others.

All5323 of the OmpR family was unique since it has two receiver domains at the C-terminus in addition to the N-terminal receiver and DNA-binding domains. This protein appears to be homologous to RcaC of another filamentous cyanobacterium *Fremyella diplosiphon* that mediates signal transduction for chromatic adaptation.<sup>30</sup> Members of the CheY subfamily were small proteins that bear only a receiver domain with no output domain. Whereas *Anabaena* had 23 such predicted proteins, this subfamily is generally small in most bacteria (e.g. 7 such predicted proteins in *P. aeruginosa*). Five predicted proteins of the PatA subfamily had an N-terminal putative output domain and the C-terminal receiver domain. Such ORFs appear to be unique to cyanobacteria. The original *patA* (*all0521*) was identified as involved in the regulation of heterocyst pattern formation in *Anabaena*.<sup>31</sup>

The unclassified products of 16 ORFs in subclass VI were predicted to carry potential output domains at the C-termini. All1703 had a GAF domain followed by a Ser/Thr-protein phosphatase domain. Alr5251 also had a GAF domain, whereas All1012 had a PAS domain between the receiver and GGDEF (Pfam PF00990) domains. The presence of these domains may suggest that these predicted response regulators not only receive signals from His kinases but also recognize environmental stimuli directly. Alr3920 has an HD domain, which may confer metal-dependent phosphohydrolase activity.<sup>32</sup> The predicted products of four ORFs, *all1012*, *alr2306*, *all2874*, and *alr3599*, possessed a GGDEF domain which is often found in bacterial signaling molecules. Additionally, Alr2306 and Alr1230 had a widely conserved EAL domain of unknown function (Pfam PF00563). Alr0960 and its homolog (Sll1879) of *Synechocystis* had a novel conserved domain, which is homologous to Ycf55 of unknown function in the chloroplast genome of a red alga. The output function of these domains has not clearly been assigned yet.

**Table 2.** Lists of genes for transcription factors.

| I. CRP     | II. LysR       | III. ArsR  | IV. AraC    | V. LuxR    | VI. OmpR  | VII. GntR      |
|------------|----------------|------------|-------------|------------|-----------|----------------|
| alr0295    | alr0353        | alr0831    | alr0429     | alr0913    | alr0072   | all1076        |
| alr2325    | all0602        | alr1044    | all1651     | all1704    | all0129   | alr2226        |
| alr3952    | all0862        | all1103    | all2035     | all1736    | alr1170   | alr4738        |
| alr4392    | all3953        | alr1867    | all2149     | alr2138    | alr1194   |                |
| alr4454    | alr4099        | alr2766    | all2159     | asl2551    | all1964   |                |
| all4500    | all4986        | all3743    | alr2174     | alr3156    | alr2428   |                |
| alr4564    | asl8072        | all3903    | alr2178     | all3660    | alr3260   |                |
|            | all8086        | all5056    | alr2184     | alr3768    | all3788   |                |
|            |                | all7621    | alr2189     | all4635    | all3822   |                |
|            |                |            | alr2208     | all4978    | all4312   |                |
|            |                |            | all2237     | all5069    | all4503   |                |
|            |                |            | alr2575     | all7016    | all4727   |                |
|            |                |            | all2580     | alr7219    | all4750   |                |
|            |                |            | alr2587     | alr7522    | all5174   |                |
|            |                |            | alr2591     | alr7652    | alr5188   |                |
|            |                |            | alr2595     | alr8531    | all5323   |                |
|            |                |            | all2613     | alr8535    | all7584   |                |
|            |                |            | all2621     | alr9013    | all7606   |                |
|            |                |            | alr2625     |            |           |                |
|            |                |            | all3171     |            |           |                |
|            |                |            | all3232     |            |           |                |
|            |                |            | all4925     |            |           |                |
| VIII. MerR | IX. PemK       | X. TetR    | XI. MarR    | XII. FurR  | XIII. ROK | XIV. DegT-like |
| all0345    | all0337        | alr0810    | all0372     | alr0957    | alr1982   | all0498        |
| alr1629    | all3211        | alr1721    | alr4692     | all1691    | all5002   | alr3012        |
| alr3162    | alr4921        | alr1941    | alr7168     | all2473    |           |                |
| alr7637    |                | all2813    |             |            |           |                |
|            |                | alr3646    |             |            |           |                |
|            |                | all3728    |             |            |           |                |
|            |                | all4279    |             |            |           |                |
|            |                | alr4567    |             |            |           |                |
|            |                | alr5296    |             |            |           |                |
|            |                | all7523    |             |            |           |                |
|            |                | alr7526    |             |            |           |                |
| XV. LexA   | XVI. bZip-like | XVII. PhoU | XVIII. TenA | XIX. NifR3 |           |                |
| alr4908    | all4801        | all4501    | alr2769     | alr4898    |           |                |

### 3.5. Transcription factors

The Anabaena genome contained at least 118 genes that presumptively encode transcriptional regulatory proteins (Table 2). The notable features of the genes in this category are as follows.

Compared with other bacteria, Anabaena had more ORFs for the CRP family (7 ORFs). One of them (Alr4564) harbored a PAS domain in the N-terminal region, which might be involved in signal sensing. Such PAS-type CRP has not been reported in other bacteria including Synechocystis.

Two ORFs in the AraC family, *all1651* and *all2035*, encoded unique sensor-like PAS domains in the N-terminal regions. Such PAS-AraC hybrid protein has not been detected in any known genomes of organisms besides those of cyanobacteria. Another characteristic of genes in the AraC family is a hybrid with a receiver domain of a response regulator, observed in *alr0429* and *all3232*. Such receiver-AraC hybrid proteins have been detected only in Synechocystis and *Bacillus* but not in other organisms.

Notably, the Anabaena genome harbored 13 ORFs homologous to *pchR*, which induces gene expression for biosynthesis of the nonribosomal peptide siderophore py-

ochelin and the ferripyochelin receptor in response to iron deprivation of *P. aeruginosa*.<sup>33</sup> The large number of presumptive regulatory genes appears to concord with the large number of transporter proteins for the acquisition of iron. In fact, 56 of 310 transporter-related genes are annotated for iron uptake.<sup>4</sup>

### 3.6. Ser/Thr-type protein kinases and protein phosphatases

Many bacteria and archaea bear a “Hanks-type” (eukaryotic-type) Ser/Thr protein kinase that has been extensively studied in eukaryotes.<sup>34,35</sup> In Anabaena, the main chromosome and pCC7120 $\alpha$  contained 51 and 1 ORF, respectively, that encode presumptive Ser/Thr protein kinases, whereas no presumptive Tyr protein kinase was detected (Table 3).<sup>4</sup> Although the presence of ORFs that encode Ser/Thr protein kinases has been reported in various bacterial genomes, the number in Anabaena is unusually high; for example, 13 ORFs in Synechocystis<sup>36</sup> and 11 ORFs in *Mycobacterium tuberculosis*.<sup>37</sup> The ORFs in this category could be divided into three subclasses: Pkn2, HstK, and ABC1 (Table 3).

**Table 3.** Lists of genes for Ser/Thr protein kinases and phosphatases.

| [protein kinase]        |                         |           | [protein phosphatase] |                         |          |
|-------------------------|-------------------------|-----------|-----------------------|-------------------------|----------|
| I. Pkn2                 | II. HstK                | III. ABC1 | I. Low MW PTP         | II. PPP                 | III. PPM |
| all0192                 | all0323                 | all4960   | alr5068               | alr4370                 | all1758  |
| alr0344                 | alr0354                 | all0592   | alr1105               | all0373                 | all1731  |
| all0438                 | alr0709                 | alr4515   | alr1067               | alr3731 ( <i>prpA</i> ) | all2470  |
| alr0548                 | alr0710                 | alr0189   | all3436               |                         | alr0547  |
| alr1311                 | all0886                 |           |                       |                         | alr4516  |
| alr1336                 | alr0900                 |           |                       |                         | all1703  |
| alr1869                 | all1625                 |           |                       |                         | all3759  |
| all1919                 | all2282                 |           |                       |                         | alr1086  |
| all2334                 | all3557                 |           |                       |                         | alr2280  |
| alr2259 ( <i>pknB</i> ) | all4687                 |           |                       |                         |          |
| alr2411                 | alr2258 ( <i>hstK</i> ) |           |                       |                         |          |
| alr2412 ( <i>spkE</i> ) | all3691                 |           |                       |                         |          |
| alr2502                 | alr2682                 |           |                       |                         |          |
| all2760                 |                         |           |                       |                         |          |
| alr3119                 |                         |           |                       |                         |          |
| all3169                 |                         |           |                       |                         |          |
| all3206                 |                         |           |                       |                         |          |
| alr3268 ( <i>spkB</i> ) |                         |           |                       |                         |          |
| alr3706                 |                         |           |                       |                         |          |
| all3207                 |                         |           |                       |                         |          |
| alr3732 ( <i>pknE</i> ) |                         |           |                       |                         |          |
| all3773                 |                         |           |                       |                         |          |
| alr3877                 |                         |           |                       |                         |          |
| alr3997                 |                         |           |                       |                         |          |
| alr4141                 |                         |           |                       |                         |          |
| alr4366 ( <i>pknA</i> ) |                         |           |                       |                         |          |
| alr4368 ( <i>pknD</i> ) |                         |           |                       |                         |          |
| all4518 ( <i>spkA</i> ) |                         |           |                       |                         |          |
| all4668                 |                         |           |                       |                         |          |
| all4813 ( <i>pknC</i> ) |                         |           |                       |                         |          |
| all4838                 |                         |           |                       |                         |          |
| alr4949                 |                         |           |                       |                         |          |
| alr4954                 |                         |           |                       |                         |          |
| all5278                 |                         |           |                       |                         |          |
| alr7232                 |                         |           |                       |                         |          |

Thirty-five ORFs encoded protein kinases of subclass Pkn2. The size of this subclass is much larger in *Anabaena* than in other bacteria. The genes in this subclass included *pknA* through *pknE*,<sup>38</sup> homologs of *Synechocystis* protein kinase genes (*spkA* and *spkB*) and other unique genes. Seven tandem domains of WD40 (WD repeat)<sup>39</sup> were found C-terminally in the predicted proteins of three genes, *all0438*, *alr3119*, and *all3169*, in this subclass (Fig. 1D). Generally, the WD repeat proteins have few other defined domains within single proteins as is observed in the *Synechocystis* WD proteins. Among prokaryotes and eukaryotes whose entire genomes have been sequenced, only two filamentous cyanobacteria, *Anabaena* and *Nostoc*, contained the presumptive WD-repeat proteins with a Ser/Thr protein kinase domain. The second subclass consisted of 13 ORFs that encode unique protein kinases carrying a GAF domain and a His kinase domain in a C-terminal part (Fig. 1D). These genes have been found only in *Anabaena* and *Nostoc* and here we designated them as subclass HstK according to the proposed gene name for *alr2258*.<sup>37</sup> The region between the N-terminal protein kinase and C-terminal GAF/His kinase domains did not

match any sequences in Pfam but showed high conservation among the subclass HstK. Additionally, All3691 and Alr2682 had the chromophore-binding GAF domain and PAS domain, respectively (Fig. 1D). The third subclass ABC1 contains four ORFs as shown in Table 3.

*Anabaena* encoded three types of presumptive protein phosphatases as listed in Table 3. Four ORFs encoded low molecular weight phosphotyrosine protein phosphatase or acid phosphatase, which acts on Tyr phosphorylated proteins and acyl phosphates. Two of them, *alr1067* and *alr5068*, showed significant homology to genes for Tyr protein phosphatases, though no Tyr protein kinase was identified in the genome. Three ORFs, *all0373*, *alr3731* (*prpA*), and *alr4370*, encoded presumptive protein phosphatases 1/2A/2B (PPP subfamily). One of them, *alr3731*, is necessary for the development of N<sub>2</sub> fixation.<sup>40</sup> Nine ORFs encoded presumptive protein phosphatases 2C (PPM subfamily). The putative products of four ORFs in this subclass, *alr1086*, *all1703*, *alr2280*, and *all3759*, also contained a response-regulator receiver domain in the C-terminal part. *all1758* is a homolog of *Synechocystis slr2031*, which was shown to be involved in motility and sensitivity to high light.<sup>41</sup>

*alr4516* is a homolog of *Synechocystis sll1771*, which was recently shown to dephosphorylate P<sub>II</sub> protein.<sup>42</sup>

### 3.7. cAMP signal transduction

Anabaena has numerous adenylate cyclase genes (*cya*). In addition to five known genes, *cyaA* (*all1118*), *cyaB1* (*alr2266*), *cyaB2* (*all1904*), *cyaC* (*all4963*), and *cyaD* (*all0743*),<sup>15</sup> a sixth gene for adenylate cyclase, *cyaE* (*all0661*) was found in the complete genome.<sup>4</sup> The catalytic domain of CyaE is homologous to CyaG of *Spirulina platensis* and to mammalian adenylate cyclase.<sup>43</sup> The Cya catalytic domains were C-terminal, whereas various putative input domains were N-terminal in all Cya proteins of Anabaena. Four Cya proteins had GAF domain(s) and the four GAF domains of CyaB1 and CyaB2 (see section 3.2) were predicted to bind cGMP. A novel conserved domain in the N-terminal region of CyaA consisted of four transmembrane segments and a large periplasmic loop. Domains with similar sequence and configuration were found in 11 other ORFs in Anabaena. The putative products of two genes, All1178 and All5309, bear hybrid His kinase and response regulator domains at the C-terminus, respectively, while Alr1869 had a C-terminal Ser/Thr protein kinase domain (Fig. 1E). These findings suggest that the N-terminal conserved domain with a periplasmic loop is a sensor for some environmental stimuli; hence, we designated it as the CMS (cyanobacterial membrane sensor) domain. Interestingly, 7 of 12 CMS-encoding genes, including *cyaA*, were accompanied either upstream or the downstream on the genome by genes that encode mutually homologous hydrophilic proteins with another novel domain. This observation leads us to speculate that a CMS protein and its accompanying protein may form a complex that transmits signals from the periplasm to the cytoplasm. Accordingly, we propose to designate these proteins as CMSN (Cyanobacterial Membrane Sensor Neighbor) (Fig. 1E). A set of CMS and CMSN proteins has been found in *Synechocystis* and *Nostoc* but in no other organisms.

The following are additional notable features of the genes in this category.

1. Anabaena had 18 ORFs that have a coding capacity of cNMP-binding domains (PF00027). Six of them, *alr0295*, *alr2325*, *ntcA* (*alr4392*), *alr4454*, *all4500*, and *alr4564*, could encode helix-turn-helix DNA binding motifs of the CRP family. Five ORFs, *alr1614*, *alr1927*, *all2676*, *alr4239*, and *all7315*, could encode a putative, C-terminal ABC transporter domain in addition to an N-terminal cNMP-binding domain.
2. SYCRP1 and SYPK are cAMP-binding proteins in *Synechocystis*.<sup>44</sup> The Anabaena genome had SYCRP homologs (Alr0295, Alr2325) but no SYPK

homolog.

3. Only one copy of a *cpdA* gene (*alr5338*) coding for a presumptive cAMP phosphodiesterase was identified, whereas multiple copies of *cya* genes were assigned in the Anabaena genome.

### 3.8. Genes related to signal transduction

Anabaena had a novel gene, *alr3165*, that encodes a retinal-binding light-dependent proton-pump protein with 7 transmembrane helices. It was significantly homologous to a group of archaerhodopsins in halobacteria and fungi,<sup>45</sup> but such a gene has not yet been documented in other cyanobacteria or even in the eubacterial kingdom.

Another unique ORF, *alr2278*, encodes a heme protein significantly homologous to the nitric oxide (NO)-binding heme domain of mammalian soluble guanylate cyclases.<sup>46</sup> Such a putative NO-binding heme protein was not found in *Synechocystis* or in any heterotrophic bacteria. A domain next to the heme domain in the soluble guanylate cyclases was homologous to an N-terminal domain of a protein encoded by the downstream gene *alr2279*, suggestive of split genes in Anabaena (Fig. 1F). Moreover, Alr2279 bore a complex signaling region including chromophore-binding GAF, hybrid His kinase and HPT domains. A similar pair of genes was also found in *Nostoc*, although the signaling region was much simpler than Anabaena Alr2279.

One of the most characteristic features of the Anabaena genome is the presence of a large number of genes that may respond to various stress conditions or may be involved in stress tolerance. A two-component response regulator gene, *orrA* (*alr3768*) is known to be involved in osmotic responses.<sup>47</sup> Two *orrA* homologs (*alr8535* and *all1704*) were also present. Various osmoprotectant molecules accumulate in response to high salinity or desiccation.<sup>48</sup> The Anabaena genome had putative genes for biosynthesis and degradation of trehalose (*all0166*, *all0167*, and *all0168*) and sucrose (*all4985* and *all1059*) but not of glycine betaine or glucosylglycerol.

Although not known to form differentiated cells called akinetes, PCC 7120 has an ortholog of an akinete-specific gene (Zhou and Wolk, unpubl.), and is likely a mutant whose progenitors formed akinetes. Insertion sequence IS891, where present as bp 4801396–4802754 and 4944643–4946001 of the chromosome, and bp 245638–246996 of pCC7120 $\alpha$ , lies within presumptive protein kinase gene *all3985* and within the 5' continuation of the ORFs of presumptive protein kinase genes *alr4105* and *alr7232*, respectively. The possible consequences, including developmental, of deleting those copies of IS891 will be of much interest.

**Acknowledgements:** K. Jaeger helped to find regulatory-type genes that are interrupted by insertion sequences. This work was partly supported by a Research

Fellowship for Young Scientists from the Japan Society for the Promotion of Science (to S. O., A. K., and S. Y.), by a Grant-in-Aid for Scientific Research "Genome Biology" (to M. O., M. I., N. S., T. O., and M. K.) from the Ministry of Education and Science, by the Genome Frontier Project "Genetic and Molecular Networks" (to M. K.), by the Program for Promotion of Basic Research Activities for Innovative Biosciences of Japan (to M. O. and M. I.), by the Kazusa DNA Research Institute Foundation (to T. K. and S. T.) and by a Grant for Scientific Research from the Human Frontier Science Program (to M. I.). Contributions from the Wolk laboratory were supported by the U.S. Department of Energy under grant DOE-FG02-91ER20021.

## References

- Croft, W. N. and George, E. A. 1959, Blue-green algae from the middle Devonian of Rhynie Aberdeenshire, *Bull. Br. Mus. Nat. Hist. Geology*, **3**, 339–353.
- Schopf, J. W. 1970, Precambrian micro-organisms and evolutionary events prior to the origin of vascular plants, *Biol. Rev.*, **45**, 319–352.
- Kaneko, T., Sato, S., Kotani, H. et al. 1996, Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions, *DNA Res.*, **3**, 109–136.
- Kaneko, T., Nakamura, Y., Wolk, C. P. et al. 2001, Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120, *DNA Res.*, **8**, 205–213.
- Geitler, L. 1932, Cyanophyceae, Akademische Verlagsgesellschaft, Leipzig.
- Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M., and Stanier, R. Y. 1979, Generic assignments, strain histories and properties of pure cultures of cyanobacteria, *J. Gen. Microbiol.*, **111**, 1–61.
- Lachance, M. A. 1981, Genetic relatedness of heterocystous cyanobacteria by DNA-DNA reassociation, *Int. J. Syst. Bacteriol.*, **31**, 139–147.
- Herdman, M., Castenholz, R. W., and Rippka, R. 2001, In: Boone, D. R. and Castenholz, R. W. (eds) *Bergey's Manual of Systematic Bacteriology*, 2nd Ed., Vol. 1, Springer-Verlag, New York, pp. 575–580.
- Wilmotte, A. and Herdman, M. 2001, In: Boone, D. R. and Castenholz, R. W. (eds) *Bergey's Manual of Systematic Bacteriology*, 2nd Ed., Vol. 1, Springer-Verlag, New York, pp. 487–493.
- Sato, N. 2000, SISEQ: Manipulation of multiple sequence and large database files for common platforms, *Bioinformatics*, **16**, 180–181.
- Altschul, S. F., Madden, T. L., Schaffer, A. A. et al. 1997, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.*, **25**, 3389–3402.
- Croft, L., Beatson, S. A., Whitchurch, C. B., Huang, B., Blakeley, R. L., and Mattick, J. S. 2000, An interactive web-based *Pseudomonas aeruginosa* genome database: discovery of new genes, pathways and structures, *Microbiology*, **146**, 2351–2364.
- Ponting, C. P., Schultz, J., Milpetz, F., and Bork, P. 1999, SMART: identification and annotation of domains from signalling and extracellular protein sequences, *Nucleic Acids Res.*, **27**, 229–232.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. 1997, The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, *Nucleic Acids Res.*, **24**, 4876–4882.
- Katayama, M. and Ohmori, M. 1997, Isolation and characterization of multiple adenylate cyclase genes from the cyanobacterium *Anabaena* sp. strain PCC 7120, *J. Bacteriol.*, **179**, 3588–3593.
- Aravind, L. and Ponting, C. P. 1997, The GAF domain: an evolutionary link between diverse phototransducing proteins, *Trends Biochem. Sci.*, **22**, 458–459.
- Vierstra, R. D. and Davis, S. J. 2000, Bacteriophytochromes: new tools for understanding phytochrome signal transduction, *Semin. Cell Dev. Biol.*, **11**, 511–521.
- Yoshihara, S., Suzuki, F., Fujita, H., Geng, X. X., and Ikeuchi, M. 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.
- Schmitz, O., Katayama, M., Williams, S. B., Kondo, T., and Golden, S. S. 2000, CikA, a bacteriophytochrome that resets the cyanobacterial circadian clock, *Science*, **289**, 765–768.
- 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.
- Crosson, S. and Moffat, K. 2001, Structure of a flavin-binding plant photoreceptor domain: Insights into light-mediated signal transduction, *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 2995–3000.
- Salomon, M., Christie, J. M., Knieb, E., Lempert, U., and Briggs, W. R. 2000, Photochemical and mutational analysis of the FMN-binding domains of the plant blue light receptor, phototropin, *Biochemistry*, **39**, 9401–9410.
- Miyatake, H., Mukai, M., Park, S. Y. et al. 2000, Sensory mechanism of oxygen sensor FixL from *Rhizobium meliloti*: crystallographic, mutagenesis and resonance Raman spectroscopic studies, *J. Mol. Biol.*, **301**, 415–431.
- Stover, C. K., Pham, X. Q., Erwin, A. L. et al. 2000, Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen, *Nature*, **406**, 959–964.
- Zhu, J., Kong, R., and Wolk, C. P. 1998, Regulation of *hepA* of *Anabaena* sp. strain PCC 7120 by elements 5' from the gene and by *hepK*, *J. Bacteriol.*, **180**, 4233–4242.
- Aravind, L. and Ponting, C. P. 1999, The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins, *FEMS Microbiol. Lett.*, **176**, 111–116.
- Dutta, R., Qin, L., and Inouye, M. 1999, Histidine kinases: diversity of domain organization, *Mol. Microbiol.*,

- 34, 633–640.
28. Mizuno, T. 1998, His-Asp phosphotransfer signal transduction, *J. Biochem. (Tokyo)*, **123**, 555–563.
  29. Perraud, A. L., Rippe, K., Bantscheff, M. et al. 2000, Dimerization of signalling modules of the EvgAS and BvgAS phosphorelay systems, *Biochim. Biophys. Acta*, **1478**, 341–354.
  30. Chiang, G. G., Schaefer, M. R., and Grossman, A. R. 1992, Complementation of a red-light-indifferent cyanobacterial mutant, *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 9415–9419.
  31. Liang, J., Scappino, L., and Haselkorn, R. 1992, The *pataA* 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. U.S.A.*, **89**, 5655–5659.
  32. Aravind, L. and Koonin, E. V. 1998, The HD domain defines a new superfamily of metal-dependent phosphohydrolases, *Trends Biochem. Sci.*, **23**, 469–472.
  33. Heinrichs, D. E. and Poole, K. 1993, Cloning and sequence analysis of a gene (*pchR*) encoding an AraC family activator of pyochelin and ferripyochelin receptor synthesis in *Pseudomonas aeruginosa*, *J. Bacteriol.*, **175**, 5882–5889.
  34. Shi, L., Potts, M., and Kennelly, P. J. 1998, The serine, threonine, and/or tyrosine-specific protein kinases and protein phosphatases of prokaryotic organisms. A family portrait, *FEMS Microbiol. Rev.*, **22**, 229–253.
  35. Kamei, A., Yuasa, T., Orikawa, K., Geng, X. X., and Ikeuchi, M. 2001, A eukaryotic-type protein kinase, SpkA, is required for normal motility of the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803, *J. Bacteriol.*, **183**, 1505–1510.
  36. Leonard, C. J., Aravind, L., and Koonin, E. V. 1998, Novel families of putative protein kinases in bacteria and archaea: evolution of the “eukaryotic” protein kinase superfamily, *Genome Res.*, **8**, 1038–1047.
  37. Av-Gay, Y. and Everett, M. 2000, The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*, *Trends Microbiol.*, **8**, 238–244.
  38. Zhang, C. C. 1996, Bacterial signalling involving eukaryotic-type protein kinases, *Mol. Microbiol.*, **20**, 9–15.
  39. Neer, E. J., Schmidt, C. J., Nambudripad, R. et al. 1994, The ancient regulatory-protein family of WD-repeat proteins, *Nature*, **371**, 297–300.
  40. Zhang, C. C., Friry, A., and Peng, L. 1998, Molecular and genetic analysis of two closely linked genes that encode, respectively, a protein phosphatase 1/2A/2B homolog and a protein kinase homolog in the cyanobacterium *Anabaena* sp. strain PCC 7120, *J. Bacteriol.*, **180**, 2616–2622.
  41. Kamei, A., Ogawa, T., and Ikeuchi, M. 1998, In: Garab, G. (ed) Photosynthesis: Mechanisms and Effects, Kluwer Academic Publishers, Dordrecht, pp. 2901–2905.
  42. Irmeler, A. and Forchhammer, K. 2001, A PP2C-type phosphatase dephosphorylates the PII signaling protein in the cyanobacterium *Synechocystis* PCC 6803, *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 12978–12983.
  43. Kasahara, M., Unno, T., Yashiro, K., and Ohmori, M. 2001, CyaG, a novel cyanobacterial adenylyl cyclase and a possible ancestor of mammalian guanylyl cyclases, *J. Biol. Chem.*, **276**, 10564–10569.
  44. Yoshimura, H., Hisabori, T., Yanagisawa, S., and Ohmori, M. 2000, Identification and characterization of a novel cAMP receptor protein in the cyanobacterium *Synechocystis* sp. PCC 6803, *J. Biol. Chem.*, **275**, 6241–6245.
  45. Blanck, A., Oesterhelt, D., Ferrando, E., Schegk, E. S., and Lottspeich, F. 1989, Primary structure of sensory rhodopsin I, a prokaryotic photoreceptor, *EMBO J.*, **8**, 3963–3971.
  46. Zhao, Y., Schelvis, J. P. M., Babcock, G. T., and Marletta, M. A. 1998, Identification of histidine 105 in the  $\beta$ 1 subunit guanylate cyclase as the heme proximal ligand, *Biochemistry*, **37**, 4502–4509.
  47. Schwartz, S. H., Black, T. A., Jager, K., Panoff, J. M., and Wolk, C. P. 1998, Regulation of an osmoticum-responsive gene in *Anabaena* sp. strain PCC 7120, *J. Bacteriol.*, **180**, 6332–6337.
  48. Reed, R. H., Borowitzka, L. J., Mackay, M. A. et al. 1986, Organic solute accumulation in osmotically stressed cyanobacteria, *FEMS Microbiol. Rev.*, **39**, 51–56.