ABSTRACT

THAYER, REBECCA ELIZABETH. Characterization of two genes up-regulated during heterocyst development in the cyanobacterium *Anabaena* sp. strain PCC 7120. (Under the direction of Stephanie E. Curtis.)

Anabaena sp. strain PCC 7120 is a cyanobacterium that carries out photosynthesis in a manner similar to plants and is capable of nitrogen fixation. This organism has developed a necessary spatial separation of the incompatible processes of photosynthesis and nitrogen fixation, as nitrogen fixation is sensitive to oxygen that is produced during photosynthesis. A differentiated cell type, called a heterocyst, is formed when *Anabaena* is in an environment lacking nitrogen, and these cells are the sites of nitrogen fixation. Heterocyst formation occurs about every tenth cell along a filament of photosynthetic vegetative cells after 24-36 hours of nitrogen starvation. A screen for sequences up-regulated at the transcript level during heterocyst development in *Anabaena* identified adjacent loci *alr4311* and *all4312*. The sequence of *alr4311* suggests it encodes the ATP-binding protein of an ABC transporter complex, while that of *all4312* suggests it encodes the response regulator of a two-component regulatory system. Phylogenetic analysis of the predicted protein sequences of *alr4311* and *all4312* indicated that both of these proteins have orthologs in *Nostoc punctiforme* and *Anabaena variabilis*, two filamentous, diazotrophic cyanobacteria. Additionally, *alr4311* appears to be most similar to ABC transporters involved in the import of cobalt, while *all4312* was most similar to uncharacterized response regulators.

The transcripts of *alr4311* and *all4312* are expressed at low levels in vegetative cells, and increase in abundance after nitrogen starvation and the induction of heterocyst development. Northern analysis and real-time RT-PCR showed that expression of *alr4311* and *all4312* are induced as early as 3 hours after initiation of differentiation, and expression

levels of both genes remain elevated through the first 24 hours of development. Expression of both of these genes was blocked in an *ntcA* mutant, and significantly decreased in a *hetR* mutant.

alr4311 was shown to be part of an operon with three unknown genes immediately upstream, while *all4312* is not organized in a cluster of genes. The promoters of *all4312* and the *alr4311* operon were mapped by primer extension analysis, and a single transcript was detected for each, mapping downstream from a –10 consensus sequence characteristic of a sigma 70 promoter.

Mutational analyses in which *alr4311* and *all4312* were individually inactivated suggested that *alr4311* is required for the response to nitrogen starvation, while *all4312* appears to be important in this response, but not essential. Both the *alr4311* and *all4312* mutants appeared to differentiate heterocysts, suggesting that the role of each of these genes may be related to heterocyst function, rather than development of the heterocyst structure.

CHARACTERIZATION OF TWO GENES UP-REGULATED DURING HETEROCYST DEVELOPMENT IN THE CYANOBACTERIUM *ANABAENA* SP. STRAIN PCC 7120

by
REBECCA ELIZABETH THAYER

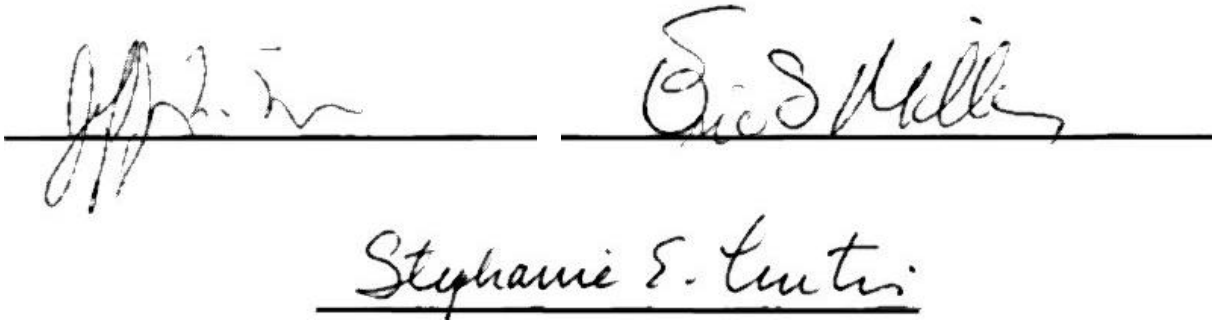
A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

FUNCTIONAL GENOMICS

Raleigh

2006

APPROVED BY:



Chair of Advisory Committee

BIOGRAPHY

Rebecca Elizabeth Thayer was born on January 13th, 1977 in Duluth, MN. She lived in the small town of Superior, WI until the age of 9, at which point Rebecca and her family relocated to Hendersonville in the mountains of western North Carolina. After graduating in the top four of her high school senior class, Rebecca began her undergraduate studies at the University of Massachusetts in Amherst, where she conducted an independent research project investigating chromosome segregation in the budding yeast *Saccharomyces cerevisiae*. Rebecca graduated *cum laude* with a B.S. in Biochemistry and a minor in French in the year 2000. She was accepted to the Functional Genomics Graduate Program as an IGERT fellow at North Carolina State University in the fall of 2002, where she began her graduate research studies under the direction of Dr. Stephanie Curtis.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor, Dr. Stephanie Curtis, for her guidance and support for the duration of my graduate studies, and Dr. Martha Ramirez for her help and encouragement in the lab. Additionally, I would like to thank my committee members, Dr. Eric Miller and Dr. Jeff Thorne, as well as those who helped me with specific portions of this work, including Dr. Laura Mathies for microscopy, and Dr. William Atchley and Dr. Jeff Thorne for help with the phylogenetic analyses. I would like to thank my husband Charles and my two wonderful children, Isabella and Lucas, for being my inspiration.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 2

CHAPTER 3

CHAPTER 1

LITERATURE REVIEW

1.1 INTRODUCTION

Cyanobacteria are autotrophic prokaryotes that carry out photosynthesis in a manner similar to plants. As the first organisms to evolve oxygenic photosynthesis, cyanobacteria generated oxygen in the Earth's atmosphere over 3 billion years ago, providing an oxic environment for the evolution of new life forms. Today, cyanobacteria are found in diverse habitats throughout the world, including both aquatic and terrestrial environments, from both arctic to tropical climates (21).

Cyanobacteria are highly adaptive and exist in several different forms (3). Some strains of cyanobacteria are unicellular, such as the marine cyanobacteria *Synechococcus* and *Prochlorococcus*, while some are filamentous, such as the freshwater strains *Anabaena* and *Nostoc*. The filamentous cyanobacteria can exist in single filaments or in colonies, and filaments may be branched or unbranched. Additionally, cyanobacteria produce several differentiated cell types: these vary according to species and include akinetes, spore-like cells produced in response to nutrient limitation; hormogonia, relatively short, motile filaments; and heterocysts, specialized cells for nitrogen fixation.

As a result of their abundant growth, cyanobacteria are recognized for their role in the global carbon cycle (17). Marine cyanobacteria, such as *Prochlorococcus* and *Synechococcus*, are abundant in the oceans and remove 10 billion tons of carbon from the air per year.

Some cyanobacteria are capable of nitrogen fixation and are important in the global nitrogen cycle. Nitrogen fixation in cyanobacteria can occur in heterocysts (spatial separation), or in the dark (temporal separation). For example, the nitrogen-fixing *Nostoc* form endosymbiotic relationships with plants and fungi, providing fixed nitrogen for the

plant (16). These cyanobacteria fix nitrogen in heterocysts, using spatial separation of photosynthesis and nitrogen fixation. Other cyanobacteria, such as the marine cyanobacterium *Crocosphaera watsonii*, use temporal separation, performing photosynthesis during the light of day and nitrogen fixation during the dark of night.

The advent of cyanobacterial genomic sequence data, combined with the development of molecular genetic advancements, has facilitated research in these organisms in areas such as photosynthesis, nitrogen fixation, and cellular differentiation. Our lab is interested in the regulation of gene expression during heterocyst development in the cyanobacterium, *Anabaena* sp. strain PCC 7120, simply referred to herein as *Anabaena*. The 7.2 Mb genome of *Anabaena* was sequenced in 2001 and consists of a single chromosome and six plasmids with a total of 6132 potential protein-encoding genes (13). The complete sequence and annotation data is available online at CyanoBase, the genome database for cyanobacterial genomes.

1.2 HETEROCYST DEVELOPMENT

Anabaena is a filamentous cyanobacterium that is capable of nitrogen fixation. This organism produces two types of cells to spatially separate the incompatible processes of photosynthesis and nitrogen fixation. In the presence of fixed nitrogen, *Anabaena* exists as filaments of photosynthetic, vegetative cells. Under nitrogen-limiting conditions, filaments of *Anabaena* differentiate to form nitrogen-fixing heterocysts in a semi-regular pattern, occurring approximately every tenth cell along a filament of vegetative cells (23). The development of a heterocyst from a vegetative cell represents one of the simplest examples of

cell differentiation and pattern formation found in nature, making *Anabaena* a prime model organism for the study of gene expression regulating these processes.

The heterocyst is the exclusive site of nitrogen fixation in *Anabaena*, and it possesses several characteristics to protect the enzyme nitrogenase, required for nitrogen fixation, from inactivation by oxygen. The cell wall of the heterocyst is surrounded by a thick envelope composed of an inner glycolipid layer and an outer polysaccharide layer. This envelope layer acts as a barrier to reduce the entry of oxygen into the cell, while simultaneously allowing a sufficient level of dinitrogen to enter. The low levels of oxygen that do enter the cell are rapidly eliminated by an increase in rate of respiration. Additionally, heterocysts lack Photosystem II, the reaction center for the oxidation of water in photosynthesis, to further keep nitrogenase from exposure to oxygen (22).

The energy-intensive process of nitrogen fixation requires high levels of ATP and reductant. Heterocysts retain Photosystem I activity, which generates ATP to meet this requirement. The reductant required for nitrogen fixation is obtained by the breakdown of carbohydrates from neighboring vegetative cells. In exchange, fixed nitrogen is transported to the vegetative cells from heterocysts (9). Therefore, heterocysts and vegetative cells are mutually dependent and function cooperatively to ensure survival under nitrogen-limiting conditions.

The regulation of gene expression during heterocyst development involves a network of genes that are expressed in response to nitrogen starvation (25). Genes regulated during heterocyst development include those required for synthesis of the heterocyst structure and those that are required for function of the heterocyst. The questions of particular interest include how it is determined which cells will differentiate in response to nitrogen starvation,

and, once differentiation is initiated, how the pattern of heterocysts is established and maintained along a filament. Much progress has been made in elucidating the genes involved in these processes (Table 1.2.1); however, how the expression of these genes is regulated to control the differentiation and development of heterocysts still remains unclear.

The presence of a source of combined nitrogen represses the formation of heterocysts in *Anabaena*. The preferential source of nitrogen is ammonium, but nitrate will be utilized in cases when ammonium is absent. When deprived of combined nitrogen, *Anabaena* rapidly activates genes involved in nitrate metabolism to scavenge for combined nitrogen before committing to heterocyst formation (10).

Several signals have been identified in effecting a response to nitrogen starvation. Accumulation of 2-oxoglutarate levels, along with an increase in free calcium ion concentration, acts as an early signal of nitrogen starvation (15, 26). Signaling pathways directing heterocyst differentiation are subsequently activated. The transcription factor NtcA, a global nitrogen regulator, responds to these early indicators of nitrogen starvation and activates genes that are involved in utilizing alternative nitrogen sources, if available. If not, NtcA activates expression of genes involved in heterocyst differentiation, including the gene *hetR*, which is essential for heterocyst development (11).

NtcA is a transcriptional regulator and a member of the cAMP receptor protein (CRP) family of proteins. NtcA controls an array of genes involved in heterocyst development by binding to promoter regions and activating expression. A consensus sequence for NtcA binding sites ($GTAN_8TAC$) has been determined experimentally (10). NtcA binding sites have also been predicted computationally for several cyanobacterial genomes, with the goal of identifying further genes regulated by NtcA (20).

HetR is considered the central processor of early signals in heterocyst development. This autoregulatory protein is a serine-type protease and a DNA-binding protein that activates expression of genes involved in heterocyst spacing, such as *patS*, and genes involved in synthesis of the polysaccharide layer, such as *hepA* (12). HetR is expressed within 2-3 hours after nitrogen starvation. Mutational studies in which *hetR* was inactivated have shown that heterocyst formation is blocked, and overexpression of *hetR* results in multiple contiguous heterocysts in the absence of nitrogen (4).

The *patS* gene encodes a small diffusible peptide of 13 or 17 amino acids that inhibits heterocyst differentiation. *patS* expression occurs in proheterocysts and plays an important role in establishment and maintenance of heterocyst pattern formation. *patS* is thought to control pattern formation by lateral inhibition, thus resulting in the formation of regularly-spaced single heterocysts along a filament of vegetative cells (24). Another gene, *hetN*, was shown to play a role in the maintenance of heterocyst pattern. *hetN* encodes a putative ketoacyl reductase. Mutational studies in which *hetN* and *patS* were simultaneously inactivated resulted in differentiation of every cell into a heterocyst (2).

hetC is a gene that encodes an ABC protein exporter that is required in the early stages of heterocyst differentiation. Expression of *hetC* is dependent on NtcA but independent of HetR. HetC is required for expression of *hepA*, a gene involved in the synthesis of the polysaccharide layer. Additionally, HetC is required for *hglD* and *hglE* expression, two genes involved in the formation of the glycolipid layer (14).

Following the activation of early genes required for heterocyst development, genes involved in structure and function of the heterocyst are activated. The *devBCA* operon encodes an ABC exporter that is essential for heterocyst differentiation. The transcription of

the *devBCA* operon is positively regulated by NtcA and plays a role in the formation of the glycolipid layer of the heterocyst envelope (7). The gene *devH* is required for regulation of the glycolipid layer during heterocyst differentiation (18). DevH belongs to the CRP family of transcriptional regulators and exhibits sequence similarity to NtcA.

Synthesis of the polysaccharide layer is regulated (in part) by a pair of genes, *devR* and *hepK*, encoding a response regulatory protein and a histidine kinase sensor protein, respectively, of a two-component system (27). HepA, HepB, and HepC are also involved in the synthesis of the polysaccharide layer (1).

The last stage of heterocyst development is marked by the activation the *nifHDK* operon, which encodes the structural genes for nitrogenase. The *nifHDK* operon is part of a cluster of 15 nitrogen fixation (*nif*) or nitrogen fixation-related genes that are organized in at least six transcriptional units. This cluster includes the genes (in order) *nifB*, *fdxN*, *nifS*, *nifU*, *nifH*, *nifD*, *nifK*, *nifE*, *nifN*, *nifX*, ORF2, *nifW*, *hesA*, *hesB*, and *fdxH* (10). A second cluster of *nif* genes, the *nifVZT* operon, is located in another region of the chromosome (19). These clusters of genes for nitrogen fixation are expressed once the morphogenesis of heterocysts reaches completion.

Three programmed DNA rearrangements are known to occur late in heterocyst differentiation, coinciding with the time of expression of the nitrogenase genes at approximately 18 hours after heterocyst differentiation is induced (8). The excised DNA elements are named for the genes they disrupt. The *nifD* element encodes the site-specific recombinase XisA, while the *fdxN* element encodes the site-specific recombinase XisF. It was recently shown that the excision of the *hupL* element requires *xisC*, encoding another site-specific recombinase (5). Each of these recombinases is required for excision of their

respective elements during heterocyst differentiation. Upon excision of these elements, the structures of these essential genes are restored in the chromosomes of heterocysts.

Our research is primarily concerned with the identification of new genes that may be involved in heterocyst development, and our goal is to understand how the expression of these genes is regulated during differentiation of a heterocyst from a vegetative cell. The specific focus of this project is the AD206.2 sequence, which was identified in a screen for sequences up-regulated in response to nitrogen starvation (6). This work describes the characterization of the genes that comprise AD206.2, *alr4311* and *all4312*, and their roles in heterocyst development.

1.3 REFERENCES

1. Adams, D. G. (2000) Heterocyst formation in cyanobacteria. Curr. Opin. Microbiol. 3:618-624.

2. Borthakur, P. B., Orozco, C. C, Young-Robbins, S. S, Haselkorn, R., and Callahan, S. M. (2005) Inactivation of *patS* and *hetN* causes lethal levels of heterocyst differentiation in the filamentous cyanobacterium *Anabaena* sp. PCC 7120. Mol. Microbiol. 57:111-123.

3. Bryant, D. A. (1994) The molecular biology of cyanobacteria. Kluwer, Dordrecht, The Netherlands.

4. Buikema, W. J. and Haselkorn, R. (1991) Characterization of a gene controlling heterocyst differentiation in the cyanobacterium *Anabaena* 7120. Genes Dev. 5:321-330.

5. Carrasco, C. D., Holliday, S. D., Hansel, A., Lindblad, P., Golden, J. W. (2005) Heterocyst-specific excision of the *Anabaena* sp. strain PCC 7120 *hupL* element requires *xisC*. J. Bacteriol. 187:6031-6038.

6. Curtis, S. E., Hebbar, P. B. (2001) A screen for sequences up-regulated during heterocyst development in *Anabaena* sp. strain PCC 7120. Arch. Microbiol. 175:313-22.

7. Fiedler, G., Arnold, M., Hannus, S., and Maldener, I. (1998) The DevBCA exporter is essential for envelope formation in heterocysts of the cyanobacterium *Anabaena* sp. strain PCC 7120. Mol. Microbiol. 27:1193-1202.

8. Golden, J. W. (1997) Programmed DNA rearrangements in cyanobacteria, p. 162-173. In F. J. de Bruijn, J. R. Lupski, and G. Weinstock (ed.), Bacterial genomes: physical structure and analysis. Chapman & Hall, New York, N.Y.

9. Golden, J. W. and Yoon, H. (2003) Heterocyst development in *Anabaena*. Curr. Opin. Microbiol. 6:557-563.

10. Herrero, A., Muro-Pastor, A. M., and Flores, E. (2001) Nitrogen control in cyanobacteria. J. Bacteriol. 183:411-425.

11. Herrero, A., Muro-Pastor, A. M., Valladares, A., Flores, E. (2004) Cellular differentiation and the NtcA transcription factor in filamentous cyanobacteria. FEMS Microbiol Rev. 28:469-487.

12. Huang, X., Dong, Y., Zhao, J. (2004) HetR homodimer is a DNA-binding protein required for heterocyst differentiation, and the DNA-binding activity is inhibited by PatS. Proc. Natl. Acad. Sci. U S A. 101:4848-4853.

13. Kaneko, T., Nakamura, Y., Wolk, C. P., Kuritz, T., Sasamoto, S., Watanabe, A., Iriguchi, M., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kohara, M., Matsumoto, M., Matsuno, A., Muraki, A., Nakazaki, N., Shimpo, S., Sugimoto, M., Takazawa, M., Yamada, M., Yasuda, M., and Tabata, S. (2001) Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. DNA Res. 8:205-213.

14. Khudyakov, I. and Wolk, C. P. (1997) *hetC*, a gene coding for a protein similar to bacterial ABC protein exporters, is involved in early regulation of heterocyst differentiation in *Anabaena* sp. strain PCC 7120. J. Bacteriol. 179:6971-6978.

15. Laurent, S., Chen, H., Bedu, S., Ziarelli, F., Peng, L., Zhang, C. C. (2005) Nonmetabolizable analogue of 2-oxoglutarate elicits heterocyst differentiation under repressive conditions in *Anabaena* sp. PCC 7120. Proc. Natl. Acad. Sci. U S A. 102:9907-9912.

16. 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. R. 66:94–121.

17. Partensky, F., Hess, W. R., and Vaulot, D. (1999) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. Microbiol. Mol. Biol. Rev. 63:106–127.

18. Ramirez, M. E., Hebbar, P. B., Zhou, R., Wolk, C. P., and Curtis, S. E. (2005) *Anabaena* sp. strain PCC 7120 gene *devH* is required for synthesis of the heterocyst glycolipid layer. J. Bacteriol. 187:2326-2331.

19. Stricker, O., Masepohl, B., Klipp, W., Bohme, H. (1997) Identification and characterization of the *nifV-nifZ-nifT* gene region from the filamentous cyanobacterium *Anabaena* sp. strain PCC 7120. J. Bacteriol. 179:2930-2937.

20. Su, Z., Olman, V., Mao, F., Xu, Y. (2005) Comparative genomics analysis of NtcA regulons in cyanobacteria: regulation of nitrogen assimilation and its coupling to photosynthesis. Nucleic Acids Res. 33:5156-51571.

21. Whitton, B. A. and Potts, M. (2002) The ecology of cyanobacteria: their diversity in time and Space. Kluwer, New York.

22. Wolk, C. P. (1996) Heterocyst formation. Annu. Rev. Genet. 30:59-78.

23. Wolk, C. P., Ernst, A., and Elhai, J. (1994) Heterocyst metabolism and development, p.769-823. In D. A. Bryant (ed.), The molecular biology of cyanobacteria. Kluwer, Dordrecht, The Netherlands.

24. Yoon, H. S. and Golden, J. W. (2001) PatS and products of nitrogen fixation control heterocyst pattern. J. Bacteriol. 183:2605-2613.

25. Zhang, C., Laurent, S., Sakr, S., Peng, L., and Bedu, S. (2006) Heterocyst differentiation and pattern formation in cyanobacteria: a chorus of signals. Mol. Micro. 59:367-375.

26. Zhao, Y., Shi, Y., Zhao, W., Huang, X., Wang, D., Brown, N., Brand, J., and Zhao, J. (2005) CcbP, a calcium-binding protein from *Anabaena* sp. PCC 7120, provides evidence that calcium ions regulate heterocyst differentiation. Proc. Natl. Acad. Sci. U S A. 102:5744-5748.

27. 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.

TABLE 1.2.1

Genes Involved in Heterocyst Differentiation		
Event	**Early Genes**	**Function**
Response to Nitrogen Starvation: **0-30 minutes**	*ntcA*	Nitrogen-response regulatory protein
	nirA-nrtABCD-narB operon	Nitrate assimilation genes
Heterocyst **Differentiation:** **1-4 hours**	*hetR*	Central regulator of heterocyst differentiation
	hetP, hetC, hetF, hetL	Heterocyst differentiation proteins
Event	**Middle Genes**	**Function**
Heterocyst Spacing: **3-12 hours**	*patA, patB, patS* *hetM, hetN, hetI*	Establishment and maintenance of heterocyst pattern
Formation of the heterocyst envelope: **4-18 hours**	*hepA, hepB, hepC, hepK*	Synthesis of the polysaccharide layer
	hglC, hglD, hglE, hglK; devH, devABC operon	Synthesis of the glycolipid layer
Event	**Late Genes**	**Function**
Activation of nitrogen fixation: **18-24 hours**	*nif* gene clusters including *nifHDK* and *nifVZT* operons	Nitrogenase Activity
	nifD, *fdxN*, *hupL* elements	Genome rearrangements

CHAPTER 2

CHARACTERIZATION OF *alr4311*

2.1 INTRODUCTION

ABC transporters are proteins that utilize the energy from ATP hydrolysis to transport solutes across a cell membrane (15). ABC transporters facilitate the unidirectional import or export of their specific substrate in prokaryotes, whereas all eukaryotic ABC transporters appear to be exporters. Some ABC systems are not involved in transport, but rather in such cellular processes as DNA repair and mRNA translation (3). Import systems in prokaryotes employ ABC transporters in the uptake of nutrients such as sugars, phosphate, and sulfate. In both eukaryotes and prokaryotes, export systems transport macromolecules such as lipopolysaccharides and toxin polypeptides out of the cytoplasm (18). Although the substrates that are handled by ABC transporters vary widely, ABC transporters are highly conserved and make up one of the largest superfamilies of proteins (28). In general, there are several families within the ABC superfamily, and the specificity of the substrate for transport usually correlates with the family (7, 30).

The first ABC transporter sequence to be published was the histidine transporter of *Salmonella typhimurium*, a periplasmic transporter, in 1982 (16). The sequence of a component of the maltose transporter (MalK) of *Escherichia coli* soon followed, along with the sequence of an oligopeptide transporter from *S. typhimurium* (12). In 1986, the first eukaryotic transporter was identified, the human multidrug resistance P-glycoprotein (13). It was these studies that provided the basic fundamental principles regarding the structure and function of ABC systems.

ABC transporters, also known as traffic ATPases, are named for their highly conserved ATP-binding cassette. The members of the ABC family of proteins share characteristic, conserved motifs. These include the Walker A and Walker B motifs, as well as

a highly conserved signature sequence (consensus LSGGQ), commonly referred to as the linker peptide (28). The Walker A and Walker B motifs have been shown to be important for nucleotide-binding and ATPase activity, while mutations within the linker peptide sequence have been shown to block the hydrolysis of ATP.

ABC transporters typically consist of four structural domains, including two hydrophobic membrane-spanning domains, and two hydrophilic ATP-binding domains (29). Prokaryotic importers tend to encode these domains on separate polypeptides, while most exporters in both prokaryotes and eukaryotes fuse these domains in various combinations (10). Additional proteins are necessary to facilitate transport in some systems. Importers require a substrate-binding protein for transport, and many of the export systems require accessory proteins that are encoded by genes that are linked to the ABC protein-encoding gene. The two membrane-spanning domains are believed to span the membrane as 6 α-helices per domain, 12 α-helices in total for a transporter. These domains serve two functions, as they act as the pathway for transport of the solute across the membrane, and they confer specificity of the transporter (19).

The number of ABC transporters in a particular genome varies widely. In general, eukaryotic genomes contain fewer ABC transporters than those of prokaryotes, with the highest number of ABC transporters found in genomes of highly adaptable bacteria that are able to thrive in diverse habitats. For example, the human genome contains 48 known genes classified as ABC transporters (8), while the *E. coli* genome contains 79 ABC transporters (22).

ABC systems are widespread in cyanobacteria, as shown by the growing number of annotated sequences of cyanobacterial genomes. According to CyanoBase, there are 84

genes classified as ATP-binding proteins of ABC transporters within the genome of *Anabaena*. Many of these are predicted to be organized in operons; however, there are several putative ATP-binding proteins throughout the genome that do not appear to be located within a gene cluster.

Several ABC systems have been experimentally characterized in cyanobacteria. In *Anabaena*, the DevBCA exporter (Table 1.2.1) has been shown to be required for the formation of the envelope layer in heterocysts (11). Additionally, the vanadate transporter in *Anabaena variabilis* (25), the potassium transporter in *Anabaena* L-31 (2), and the manganese transporter in *Synechocystis* PCC 6803 (27) are currently being investigated.

The study of ABC transport systems has significant importance in the field of medicine. In humans, diseases such as cystic fibrosis, Dubin-Johnson syndrome, and Tangier disease are caused by mutations in genes encoding ABC transporters. Additionally, ABC transporters play a role in multidrug resistance of cancer (19). ABC transporters in prokaryotes serve as model systems to further advance our knowledge and understanding of these important systems. It is hoped that continued research in this field will lead to a better understanding of the molecular mechanisms and physiological functions of ABC transport systems. This chapter describes the characterization of the *alr4311* gene of *Anabaena*, which was chosen for further study based on the observation that this gene is up-regulated in response to nitrogen starvation. *alr4311* is predicted to encode an ATP-binding protein of an ABC transporter.

2.2 MATERIALS AND METHODS

Strains and culture conditions. The strains and plasmids used in this study are listed in Table 2.2.1. Liquid cultures of cyanobacterial strains were grown in BG-11 or in

modified Kratz and Meyers as described previously (14). *alr4311* mutant strains were maintained in cultures containing neomycin sulfate (80 μg/ml).

Escherichia coli strains were grown in 2YT or LB media supplemented with appropriate antibiotics.

Cyanobacterial developmental time courses (nitrogen step-down) and growth rate experiments were carried out as previously described (14). 1 ml samples of each culture were removed for $O.D._{750}$ measurements every 24 hours. Light microscopy at 60X magnification was used to observe filaments for growth and morphological characteristics.

Isolation and analysis of nucleic acids. Genomic DNA was isolated as described by Wu *et al.* (31). Total RNA was isolated using RNA Wiz (Ambion) according to the manufacturer's instructions, and RNA samples were purified using the RNeasy Mini Kit with the RNase-Free DNase set from Qiagen.

Southern blot analyses were performed using 10 μg genomic DNA. Northern blot analyses were performed using 20 μg total RNA isolated from *Anabaena* over a developmental time course. Probes for both Southern and Northern analyses were created by PCR amplification of *Anabaena* genomic DNA. Primers 4311-A and 4311-C were used to create a 722 bp probe (4311 AC) for Southern and Northern analyses.

Probes were labeled using the AlkPhos Direct Labeling kit (Amersham Biosciences) and hybridizations were performed following the manufacturer's recommendations. Detection of hybridization signals was achieved by using ECF substrate and a Storm scanner (Molecular Dynamics). Quantitation was performed using ImageQuant software.

RT-PCR was used to detect gene expression after nitrogen step-down. First-strand cDNA was synthesized using ImProm-II reverse transcriptase (Promega) with 1 μg total

RNA from a developmental time course. Reactions were carried out as recommended by the manufacturer. Induction of gene expression was detected by PCR using 1-2 μl cDNA in 25 μl reaction volumes.

Primer extension was carried out using the Primer Extension System-AMV Reverse Transcriptase (Promega) according to the manufacturer's instructions with 30 μg of total RNA isolated from cells from a nitrogen step-down experiment after 24 hours. Primer 4308PE-1 was used for this assay.

Sequence ladders were generated using Promega's fmol DNA Cycle Sequencing System following the manufacturer's instructions for end-labeling primers. The same primer (4308PE-1) was used in the sequencing reactions as for primer extension. A 458 bp PCR product containing the upstream region of *alr4308* plus a portion of the ORF was created by amplification with primers 4308-F and 4308-R, and this product was used as template in the sequencing reaction.

Real-time RT-PCR was performed using the SYBR-Green PCR Core Reagents from Applied Biosystems according to the manufacturer's recommendations with *Anabaena* total RNA isolated from a developmental time course. Primer set 4311-1 and 4311-2, which amplified a 119 bp product, was used to measure expression of *alr4311*, which was normalized against RNaseP transcript levels from amplification with primer set RNaseP-1 and RNaseP-2. *hepA* expression was measured as a positive control of heterocyst induction using primer set HepA-F3 and HepA-R3.

Inactivation of *alr4311* by single recombination. *alr4311* was initially inactivated by a single recombination insertion. PCR was used to create an insert containing a fragment of the *alr4311* gene for cloning. Primers 4311-A and 4311-B were used to create insert 4311

AB (505 bp). The insert was cloned into the *Eco*RV site of pBN2, and this construct was subsequently transferred to *Anabaena* using standard procedures for conjugation (9). Southern blot analysis was used to confirm single recombination in the mutant (A73) using *npt*- and *alr4311*-specific probes.

Inactivation of *alr4311* by double recombination. *alr4311* was inactivated by using the *sacB* method of positive selection for double recombinant mutants (4). A 736 bp fragment of the *alr4311* gene was created by PCR of *Anabaena* genomic DNA with primers 4311-D and 4311-E, which generated *Hind*III sites at the 5' and 3' ends. A 1.1 kbp *Sma*I fragment containing the *npt* gene was isolated from pRL648 (9) and cloned into the *Eco*RV site of this gene fragment in forward orientation relative to the direction of transcription. The disrupted gene fragment was cloned into the *Nae*I site of pRL271, a suicide vector containing the *sacB* gene, and transferred into *Anabaena* using standard procedures for conjugation. Ex-conjugants were selected by resistance to neomycin, followed by selection with 8% sucrose to identify double recombinant mutants. Southern blot analysis was used to confirm double recombination in the mutant (A75) using *npt*- and *alr4311*-specific probes.

Phylogenetic analysis of *alr4311*. BLASTP (1) deduced protein searches of the non-redundant GenBank protein database at NCBI and the cyanobacterial Swiss-Prot database (Swiss_cyanoP) at Cyanobase were performed using the translated amino acid sequence of *alr4311* (Alr4311), obtained from the genome database for cyanobacteria, CyanoBase (http://www.kazusa.or.jp/cyanobase/). The top 19 GenBank sequences (E-values of 5e-23 or less; 30-95% identities) and the top 20 Swiss_cyanoP sequences (E-values of 1e-17 or less; approximately 30% identities) were collected for multiple sequence alignment using the default parameters of CLUSTAL W (17). The alignments were subsequently imported into

GeneDoc version 2.6 (24) for visual inspection. Phylogenetic trees were constructed using the neighbor-joining method of the software program MEGA version 2.1 (21). The trees were evaluated using the p-distance amino acid option and the bootstrap test of phylogeny based on 1000 replications. A phylogenetic analysis employing a similar approach to the one described here was used in a previous study of ABC-transporters (28).

2.3 RESULTS AND DISCUSSION

***alr4311* encodes an ATP-binding protein of an ABC transporter.** To identify potential genes involved in heterocyst development, a screen for sequences up-regulated at the transcript level in the absence of fixed nitrogen was performed (5). From the collection of sequences identified, AD206.2 was selected for further study, as transcripts were shown to increase in abundance during heterocyst development. AD206.2 spans the *alr4311* and *all4312* genes (Figure 2.3.1).

alr4311 is predicted to encode a 251 amino acid ATP-binding protein of an ABC transporter. A sequence alignment of Alr4311 with similar sequences from two closely related nitrogen-fixing cyanobacteria, *Anabaena variabilis* (Ava_1262) and *Nostoc Punctiforme* (NpF3906), was performed to confirm the presence of the conserved motifs characteristic of ATP-binding proteins. Alr4311 was found to contain the Walker A motif, Walker B motif, and linker peptide sequence, confirming its classification as an ATP-binding protein and its similarity to Ava_1262 and NpF3906 (Figure 2.3.2).

A phylogenetic analysis was utilized in the characterization of *alr4311* to infer the function of the protein it encodes based on similarity to known proteins. This approach is often used as part of initial characterization of a gene, as it provides a basis for experimental

design. A BLASTP search of the non-redundant GenBank database was performed using the translated amino acid sequence of *alr4311*. The top 19 most similar sequences were collected and aligned: these sequences included putative bacterial ATP-binding proteins, the majority of which are classified as a result of prediction (based on sequence similarity) rather than experimental validation. A phylogeny of these sequences was constructed based on the neighbor-joining method (Figure 2.3.3). The most closely related sequences were found to be uncharacterized, putative ATP-binding proteins from *Anabaena variabilis* (E-value = 3e-120, 95% identity) and *Nostoc punctiforme* (E-value = 2e-87, 74% identity). Other similar sequences were putative ATP-binding proteins from *Trichodesmium erythraeum* IMS101 (E-value = 1e-60, 50% identity), *Synechocystis* sp. PCC 6803 (E-value = 1e-39, 44% identity), and *Crocosphaera watsonii* WH 8501 (E-value = 3e-48, 51% identity).

The results of this phylogenetic analysis indicate that Alr4311 shares a high sequence similarity with ATP-binding proteins from two nitrogen-fixing, filamentous, freshwater cyanobacteria, *A. variabilis* and *N. punctiforme*. Interestingly, Alr4311 also shares a 50% sequence identity to an ATP-binding protein from the marine cyanobacterium, *T. erythraeum*, which is a non-heterocystous nitrogen-fixer. Furthermore, Alr4311 shares sequence similarity with ATP-binding proteins from the unicellular marine diazotroph, *C. watsonii*, and the unicellular, non-nitrogen fixing cyanobacterium *Synechocystis.* These results suggest that *alr4311* may play a role related to metabolic processes during differentiation, rather than functioning directly in development of the heterocyst structure.

To further investigate the function of *alr4311*, a search of the cyanobacterial Swiss-Prot database (Swiss_cyanoP) using the translated sequence of *alr4311* was performed at Cyanobase. This database was chosen for its high level of annotation to determine if

Alr4311 would cluster with a known, characterized group of ABC transporters. A neighbor-joining tree of the top 20 most similar sequences was constructed (Figure 2.3.4). The majority of the sequences in this tree are ABC import proteins, and they formed several distinct clusters, representing proteins for nitrate, phosphonate, cobalt, sulfate and phosphate import. Alr4311 was found to cluster near the ATP-binding proteins of cobalt import systems with a bootstrap value of 63.

The majority of the most closely related proteins to Alr4311 in Figure 2.3.4 are classified as ABC importers, and each of these proteins fell into distinct clusters based on the type of import they facilitate. Only one protein in the tree, HepA from *Anabaena* sp. strain PCC 7120, did not cluster with one of these groups. HepA is an ATP-binding protein of 607 amino acids that is essential for synthesis of the heterocyst polysaccharide layer during differentiation. The length of this protein suggests that HepA is an ABC exporter. Given these results, this phylogenetic analysis suggests that *alr4311* may be involved in a type of ABC import, rather than export, and may possibly function in the import of cobalt or another nutrient/molecule needed for metabolism.

***alr4311* expression is up-regulated in response to nitrogen starvation.** To characterize the expression of *alr4311* during heterocyst development, Northern blot analysis and real-time RT-PCR were performed. Total RNA was isolated from samples collected at 0, 3, 12, and 24 hours after cultures of *Anabaena* were starved for fixed nitrogen. Northern analysis was used as a preliminary experiment to qualitatively determine the expression pattern of *alr4311* during a developmental time course. The results of this analysis showed that *alr4311* is expressed in response to nitrogen starvation as early as 3 hours after step-down, and *alr4311* continues to be expressed through 24 hours of development (data not

shown). Subsequently, real-time RT-PCR was performed to determine the relative ratio of increase in expression during differentiation (Table 2.3.1). Three independent cultures of *Anabaena* were subjected to nitrogen step-down for RNA isolation at the same time points as collected for Northern analysis. These biological triplicates were run in duplicate to account for pipetting/technical errors, and expression was normalized against RNaseP transcript levels. The 0-hour time point was designated as the "calibrator" sample, and the ratio of transcript increase for later time points is expressed relative to this time point. Expression of *hepA* was used as a positive control for induction of expression. *alr4311* expression was found to increase approximately 6-fold after 3 hours of nitrogen starvation, and this steady state level was observed through the 24-hour time point (as for Northern analysis). These results suggest that *alr4311* is required early in the differentiation process. Whether expression of *alr4311* is localized in heterocysts or expressed along a filament remains unknown, as total RNA is isolated from whole filaments containing both vegetative cells and heterocysts.

Expression of *alr4311* was analyzed in two developmental mutants, *ntcA* and *hetR*. Real-time RT-PCR was performed with samples obtained from 0 and 24 hours after nitrogen step-down. No induction of *alr4311* transcription was observed in the *ntcA* mutant, and induction in the *hetR* mutant was significantly decreased (data not shown). As both of these genes are required early in differentiation, this result suggests that *alr4311* may be directly or indirectly regulated by NtcA and/or HetR.

***alr4311* sequence and upstream region is conserved in other cyanobacteria.** The organization of the genome upstream of *alr4311* suggests that this gene may be part of an operon. The genes *alr4308*, *alr4309*, and *alr4310* are located within close proximity to

alr4311, spanning a region covering about 2.5 kb, including the length of *alr4311*. Each gene is separated by a short distance of 149 bp, 80 bp, and 84 bp, respectively. Additionally, the gene order of this region appears to be conserved in *Nostoc punctiforme* and *Anabaena variabilis ATCC 29413*. The functions of the genes comprising this cluster are unknown, as they are annotated as hypothetical proteins in CyanoBase, but may encode other components of the ABC transport system in which *alr4311* is involved.

To investigate the possibility of *alr4311* being in an operon, RT-PCR was used to detect expression from this region. Primers designed to amplify overlapping fragments of the ~2.5 kb region were used in RT-PCR to determine if this region is transcribed as a single message. Overlapping transcripts from both the 5´ and 3´ ends were observed with this method, suggesting that *alr4311* is part of an operon (Figure 2.3.5).

Primer extension analysis was used to map the transcription start point (tsp) of the operon encoding genes *alr4308*, *alr4309*, *alr4310*, and *alr4311*. RNA isolated from cells at 24 hours after nitrogen step-down was used in the primer extension reaction with primer 4308PE-1 (Figure 2.3.6 and Figure 2.3.7). A single transcript mapping at 140 bp from the 5´ end of the *alr4308* ORF was detected. The 5´ terminus of the transcript mapped 3´ to a region conforming to the –10 consensus sequence for s^{70} promoters. This –10 sequence (TATAAT) is conserved in *A. variabilis* and *N. punctiforme* (Figure 2.3.7). Many cyanobacterial genes lack elements conforming to the *E. coli* –35 promoter element (6); however, a sequence resembling the –35 consensus sequence ($TTGN_3$) is present in *Anabaena* sp. strain PCC 7120 and *A. variabilis*, but apparently absent in *N. punctiforme*.

The role of *alr4311* in heterocyst development/function. Preliminary experiments to inactivate *alr4311* by single recombination were performed to investigate the role of this

gene during heterocyst development. A fragment of *alr4311* was cloned into a vector bearing a neomycin-resistant cassette (*npt*) and transferred to *Anabaena* by conjugation. A map showing the location of the gene fragment used for single recombination is shown in Figure 2.3.8 (A). Single recombinant mutants were selected for neomycin resistance. The single recombinant *alr4311* mutant was found to be inviable in the absence of fixed nitrogen when grown in nitrogen-free media supplemented with neomycin (data not shown). The nature of the single recombinant mutant is such that the plasmid insertion disrupting the gene of interest is unstable and can be lost by the organism in the absence of antibiotic selection. Subsequently, experiments aimed at inactivating *alr4311* by double recombination were carried out to generate a more stable mutant to further investigate the role of *alr4311* in heterocyst development.

The wild-type *alr4311* gene of *Anabaena* was inactivated by the *sacB* method of double-recombination (4). A fragment of *alr4311* was disrupted by the *npt* cassette, cloned into a vector bearing the *sacB* gene, and transferred to *Anabaena* via conjugation. A map of the gene and the location of the fragment used for disruption are shown in Figure 2.3.8 (B). Double recombinant mutants were identified by selection on neomycin and the ability to grow on sucrose-containing media. Genomic DNA from the mutant strain was analyzed by Southern blotting to confirm the presence of the neomycin-resistance cassette in the wild-type *alr4311* gene (data not shown).

Growth rate analysis of the *alr4311* double recombinant mutant (designated strain A75) was carried out in media with and without fixed nitrogen over a period of 20 days (Figure 2.3.9). In the presence of nitrate, strain A75 exhibited growth characteristics similar to wild type throughout the 20-day time period, and both cultures appeared dark green with

long filaments, as observed by microscopy. In the absence of nitrate, wild type cultures grew (exponentially) and developed their characteristic blue-green color after approximately 5-7 days. In contrast, cultures of the mutant strain A75 were unable to grow without nitrate, as the cultures turned yellow and cell growth ceased after 3 days without nitrate. This mutant phenotype was observed in nitrogen-free media in the presence and absence of antibiotic supplementation.

Morphology of the mutant strain A75 was observed by light microscopy to examine the mutant for heterocyst formation and pattern (Figure 2.3.10). Samples of the wild type and mutant cultures were removed at 48 hours after nitrogen step-down and stained with Alcian blue, a dye that stains the polysaccharide layer of heterocysts. Wild type filaments were observed to contain regularly spaced heterocysts with Alcian blue staining. The heterocysts were easily discerned by their somewhat larger, smoother appearance in comparison to vegetative cells. Filaments of the A75 mutant appeared to contain heterocysts with normal spacing, based on the presence of Alcian blue staining.

The observation that the *alr4311* double recombinant mutant A75 appears to differentiate heterocysts but is unable to grow in the absence of a source of combined nitrogen suggests that *alr4311* is required for the response to nitrogen-limiting conditions. The apparent formation of the heterocyst structure indicates that *alr4311* may not be required for structural development, consequently suggesting a role in heterocyst function instead, such as a role in metabolic activities that occur during heterocyst differentiation. However, although there are no observed abnormalities of the heterocyst structure in the *alr4311* mutant, a structural defect cannot be entirely ruled out. As an example, the *devH* mutant

appears to differentiate normal heterocysts, but the glycolipid layer of the heterocyst envelope is absent (26).

Conclusion. In summary, sequence analysis of *alr4311* supports the prediction of this gene encoding an ATP-binding protein of an ABC transporter. Evidence that *alr4311* is part of an operon comprised of four genes supports the idea that *alr4311* is part of an ABC transporter complex. Furthermore, phylogenetic analysis of this gene indicates that *alr4311* is related to the family of ABC importers, suggesting a role in a type of import process. The expression profile of *alr4311* shows that *alr4311* is induced as early as 3 hours after nitrogen step-down, placing it in the category of "early" genes required for heterocyst development. These observations, in conjunction with the observed phenotype of the mutant in the absence of fixed nitrogen, imply a role of *alr4311* in the import of an essential nutrient during heterocyst differentiation, and that this role may be conserved in other closely related cyanobacteria, particularly *Anabaena variabilis* and *Nostoc punctiforme.*

2.4 REFERENCES

1. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D.J. (1990) Basic local alignment search tool. J. Mol. Biol. 215:403-410.

2. Ballal, A., Apte, S. K. (2005) Differential expression of the two kdp operons in the nitrogen-fixing cyanobacterium *Anabaena* sp. strain L-31. Appl. Environ. Microbiol. 71:5297-303.

3. Bouige, P., Laurent, D, Piloyan, L., and Dassa, E. (2002) Phylogenetic and functional classification of ATP-binding cassette (ABC) systems. Curr. Protein and Pept. Sc. 3:541-559.

4. Cai, Y. and Wolk, C. P. (1990) Use of a conditionally lethal gene in *Anabaena* sp. strain PCC 7120 to select for double recombinants and to entrap insertion sequences. J. Bacteriol. 172:3138-3145.

5. Curtis, S. E. and Hebbar, P. B. (2001) A screen for sequences up-regulated during heterocyst development in *Anabaena* sp. strain PCC 7120. Arch. Microbiol. 175:313-322.

6. Curtis, S. E. and Martin, J. A. (1994) The transcription apparatus and the regulation of transcription initiation, p. 613-639. In D. A. Bryant (ed.), The molecular biology of cyanobacteria. Kluwer, Dordrecht, The Netherlands.

7. Dassa, E. and Bouige, P. (2001) The ABC of ABCs: a phylogenetic and functional classification of ABC systems in living organisms. Res. Microbiol. 152:211-229.

8. Dean, M., Rzhetsky, A., Allikmets, R. (2001) The human ATP-binding cassette (ABC) transporter superfamily. Genome Res. 11:1156-1166.

9. Elhai, J., and Wolk, C.P. (1988) Conjugal transfer of DNA to cyanobacteria. Methods Enzymol. 167:747-754.

10. Fath, M. J. and Kolter, R. (1993) ABC Transporters: Bacterial Exporters. Microbiol. Rev. 57:995-1017.

11. Fiedler, G., Arnold, M., Hannus, S., and Maldener, I. (1998) The DevBCA exporter is essential for envelope formation in heterocysts of the cyanobacterium *Anabaena* sp. strain PCC 7120. Mol. Microbiol. 27:1193-1202.

12. Gilson, E., Higgins, C. F., Hofnung, M., Ames, G. F. L., Nikaido, H. (1982) Extensive homology between membrane-associated components of histidine and maltose transport systems of *Salmonella typhimurium* and *Escherichia coli*. J. Biol. Chem. 257:9915-9918.

13. Gros, P., Croop, J., Housman, D. (1986) Mammalian multidrug resistance gene: complete cDNA sequence indicating strong homology to bacterial transport proteins. Cell 47:371–380.

14. Hebbar, P. B. and Curtis, S. E. (2000) Characterization of *devH*, a gene encoding a putative DNA binding protein required for heterocyst function in *Anabaena* sp. strain PCC 7120. J. Bacteriol. 182:3572-3581.

15. Higgins, C. F. (1992) ABC transporters: from microorganisms to man. Annu. Rev. Cell Biol. 8:67-113.

16. Higgins, C. F., Haag, P. D., Nikaido, K., Ardeshir, F., Garcia, G., Ferro-Luzzi, A. G. (1982) Complete nucleotide sequence and identification of membrane components of the histidine transport operon of *S. typhimurium*. Nature. 298:723-727.

17. Higgins, D., Thompson, J., Gibson, T., Thompson, J.D., Higgins, D.G., Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence

alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673-4680.

18. Holland, I. B. and Blight, M. A. (1999) ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. J. Mol. Biol. 293:381-399.

19. Holland, I. B., Cole, S. P. C.; Kuchler, K., and Higgins, C. F. (2003) ABC proteins: from bacteria to man. London. Academic Press.

20. Huang, G., Fan, Q., Lechno-Yossef, S., Wojciuch, E., Wolk, C. P., Kaneko, T., Tabata, S. (2005) Clustered genes required for the synthesis of heterocyst envelope polysaccharide in *Anabaena* sp. strain PCC 7120. J. Bacteriol. 187:1114-1123.

21. Kumar, S., Tamura, K., Jakobsen, I. B., and Nei, M. (2001) MEGA2: Molecular Evolutionary Genetics Analysis software, Arizona State University, Tempe, Arizona, USA.

22. Linton, K. J. and Higgins, C. F. (1998) The *Escherichia coli* ATP-binding cassette (ABC) proteins. Molec. Microbiol. 28:5-13.

23. Marsh, J., Enfle, M., and Wykes, E. (1984) The pIC plasmid and phage vectors with versatile cloning sites for recombinant selection by insertional inactivation. Gene 32:481-485.

24. Nicholas, K. B. and Nicholas, H. B., Jr. (1997) GeneDoc: A tool for editing and annotating multiple sequence alignments. Distributed by the author.

25. Pratte, B. S. and Thiel, T. (2006) High-affinity vanadate transport system in the cyanobacterium *Anabaena variabilis* ATCC 29413. J Bacteriol. 188:464-468.

26. Ramirez, M. E., Hebbar, P. B., Zhou, R., Wolk, C. P., and Curtis, S. E. (2005) *Anabaena* sp. strain PCC 7120 gene *devH* is required for synthesis of the heterocyst glycolipid layer. J. Bacteriol. 187:2326-2331.

27. Rukhman, V., Anati, R., Melamed-Frank, M., Adir, N. (2005) The MntC crystal structure suggests that import of Mn2+ in cyanobacteria is redox controlled. J Mol Biol. 348:961-969.

28. Saurin, W., Hofnung, M., Dassa, E. (1999) Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. J. Mol Evol. 48:22–41.

29. Schneider, E. and Hunke, S. (1998) ATP-binding-cassette (ABC) transport systems: functional and structural aspects of the ATP-hydrolyzing subunits/domains. FEMS Microbiol. Rev. 22:1–20.

30. Tomii, K. and Kanehisa, M. (1998) A comparative analysis of ABC transporters in complete microbial genomes. Genome Res. 8:1048-1059.

31. Wu, X., Zarka, A.; and Boussiba, S. (2000) A simplified protocol for preparing DNA from filamentous cyanobacteria. Plant Mol. Biol. Rep. 18:385-392.

TABLE 2.2.1 Bacterial strains, plasmids, and oligonucleotides

Strain, plasmid, or oligonucleotide	Characteristics	Source or reference
Strains		
Anabaena spp.		
PCC 7120	Wild type	R. Haselkorn
A73	PCC 7120 *alr4311*::pBN2-4311#1, Nm^R	This study
A75	PCC 7120 *alr4311*::*npt*, Nm^R	This study
E. coli		
EP-Max 10B	Electro-competent cells	Bio-Rad
Plasmids		
pBN2	Cloning vector, Ap^R	S. Curtis
pRL271	Suicide vector, Em^RCm^R	C. P. Wolk
pIC20H	Cloning vector, Ap^R	23
pRL648	Contains neomycin resistance cassette, Nm^R	9
pRL1383a	RSF1010-based cloning vector, Sp^RSm^R	20
Oligonucleotides		
4311-A	TAATCTGTTTGC	This study
4311-B	GGCTATACCGATATCTAGATG	This study
4311-C	TTCCATTCTTCATCGG	This study
4311-D	GGAAAAGCTTATTTGCTACCG	This study
4311-E	GCTACAAGCTTTCTCTGTCCAG	This study
4311-1	AGTCCAACAAGCCTTAGC	This study
4311-2	TGCGTCCTAACCAATCAC	This study
RNaseP-1	CGAGGGCGATTATCTATCTG	This study
RNaseP-2	CAAGGCCGAAGGAACTATG	This study
HepA-F3	TATCGCCTACGGTACATCTG	This study
HepA-R3	AACCCTTCGGGCATTTCTTC	This study
4308-F	GCCCTAAGAAGGGTGAAAC	This study
4308-R	AAACACGAGGCGTTGTC	This study
4308PE-1	GAGCCATAACTGCTGAGGAGCTTGTCG	This study

Table 2.3.1. Results of *alr4311* expression study. RNA was isolated from triplicate cultures of *Anabaena* at time 0, 3, 12, and 24 h after nitrogen step-down and used for real-time RT-PCR. Ratios are given relative to expression levels of *alr4311* at time 0 h.

alr4311								
Relative Expression Ratios	**3 h**			**12 h**			**24 h**	
	Plate 1	**Plate 2**		**Plate 1**	**Plate 2**		**Plate 1**	**Plate 2**
Culture 1	23.91	27.30		14.66	20.35		30.85	18.11
Culture 2	51.98	35.25		21.32	23.73		73.41	NA
Culture 3	83.81	153.78		18.35	26.53		86.84	NA
Mean (Plate)	53.24	72.11		18.11	23.54		63.70	18.11
Mean (Time)	62.67			20.82			40.90	
Log base 2 ratio	5.97			4.38			5.35	

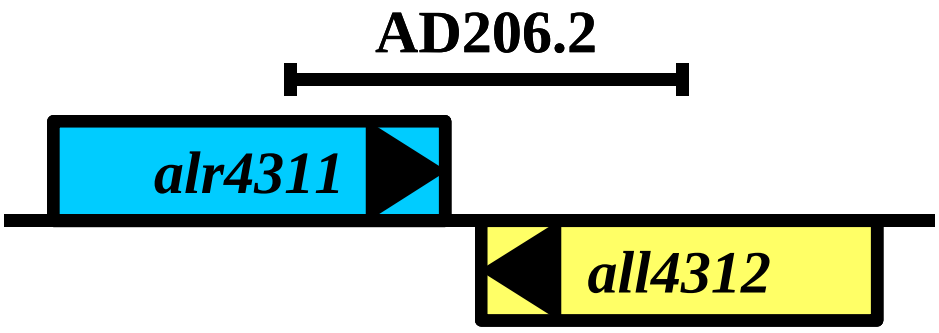


Figure 2.3.1. Diagram of the AD206.2 region which includes two convergently transcribed genes, *alr4311* and *all4312*.

Figure 2.3.2. Alr4311 contains characteristic, conserved motifs of ATP-binding proteins of ABC transporters. The Walker A motif, Walker B motif, and the linker peptide sequence of Alr4311 and similar sequences from *Anabaena variabilis* (Ava1262) and *Nostoc punctiforme* (NpF3906) are indicated in boxes. Green shading, 100% sequence similarity; yellow shading, 80% sequence similarity.

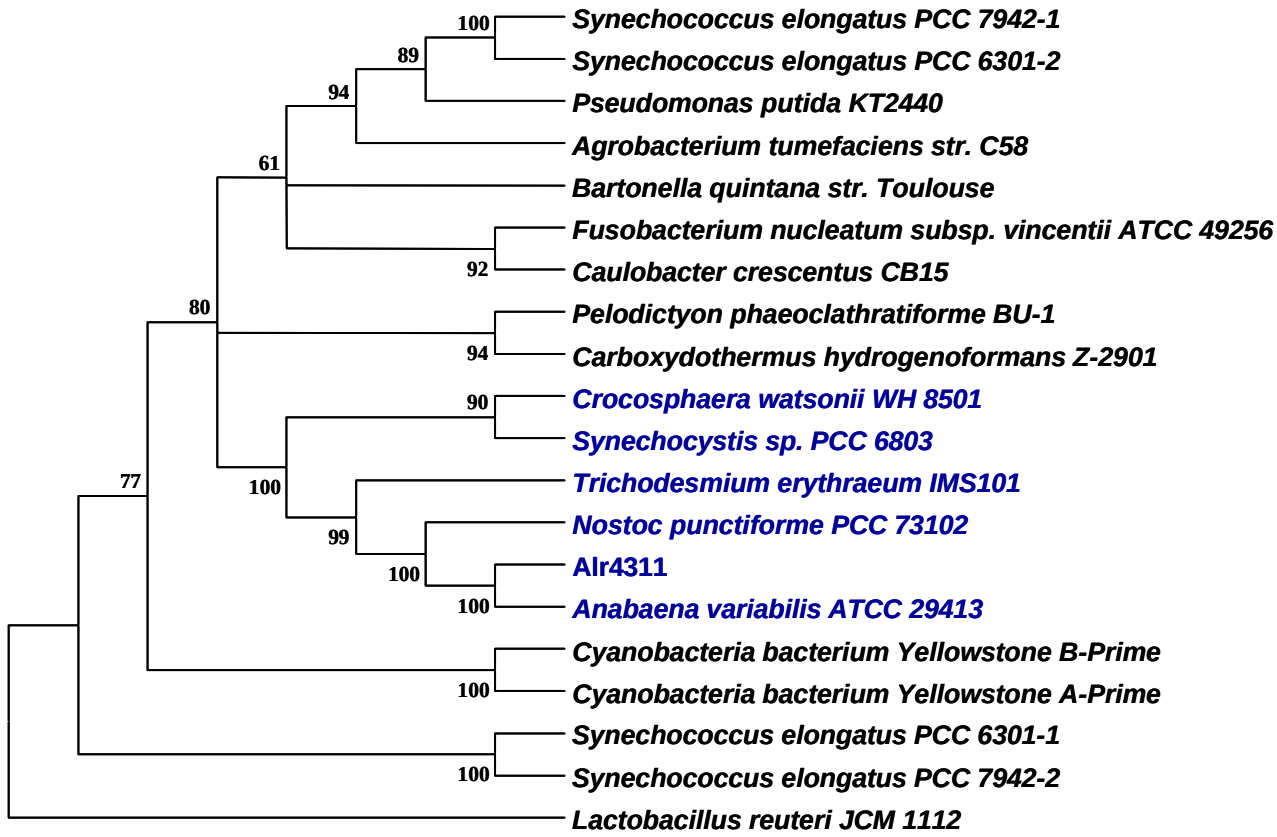


Figure 2.3.3. Neighbor-joining tree of bacterial ATP-binding proteins found in a BLASTP search of the GenBank database using the translated sequence of *alr4311*. Bootstrap values (>50) based on 1000 replications are shown at each node.

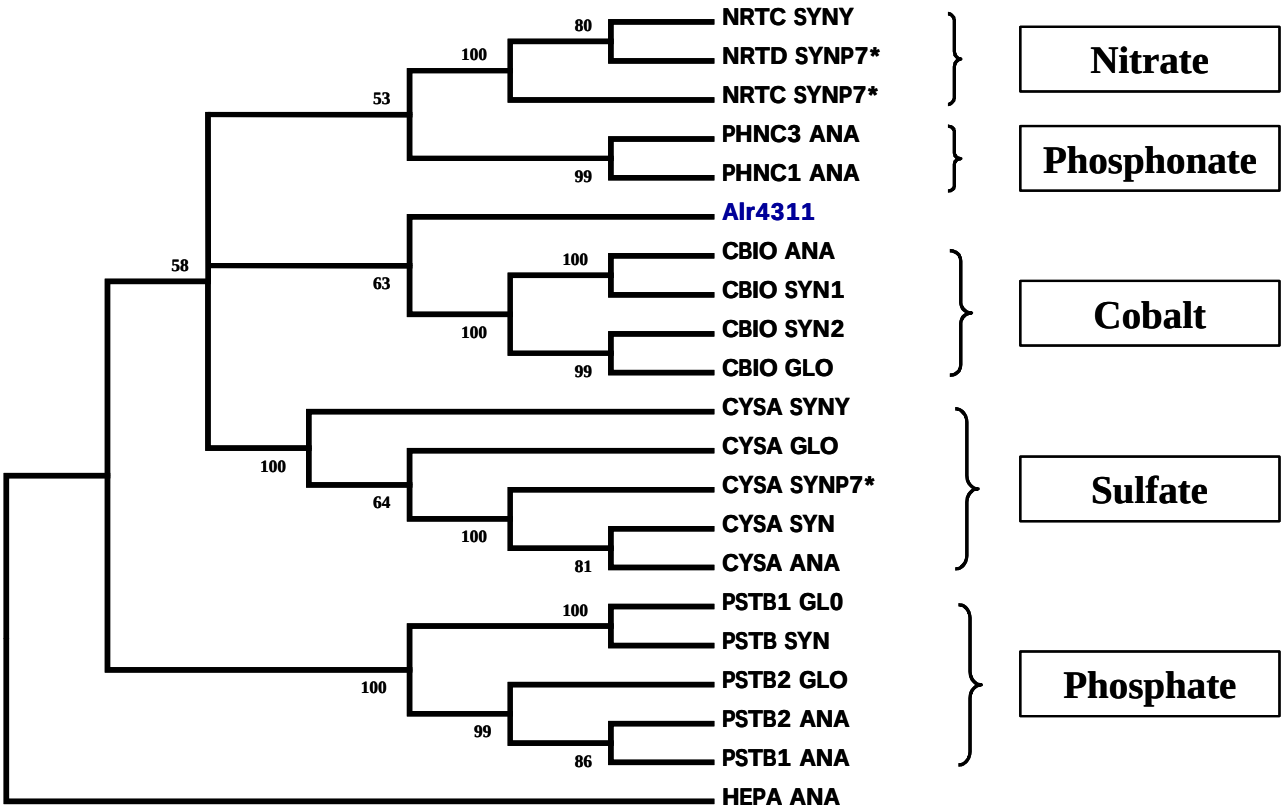
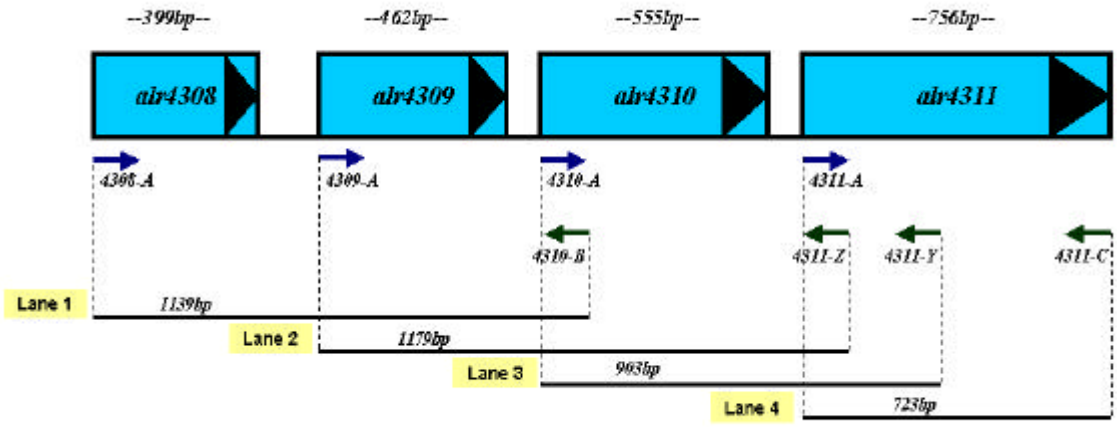


Figure 2.3.4. Neighbor-joining tree of ABC proteins found in a BLASTP search of the SwissProt cyanobacterial database using the translated sequence of *alr4311*. Bootstrap values (>50) based on 1000 replications are shown at each node. *Indicates experimentally characterized proteins.

A



B

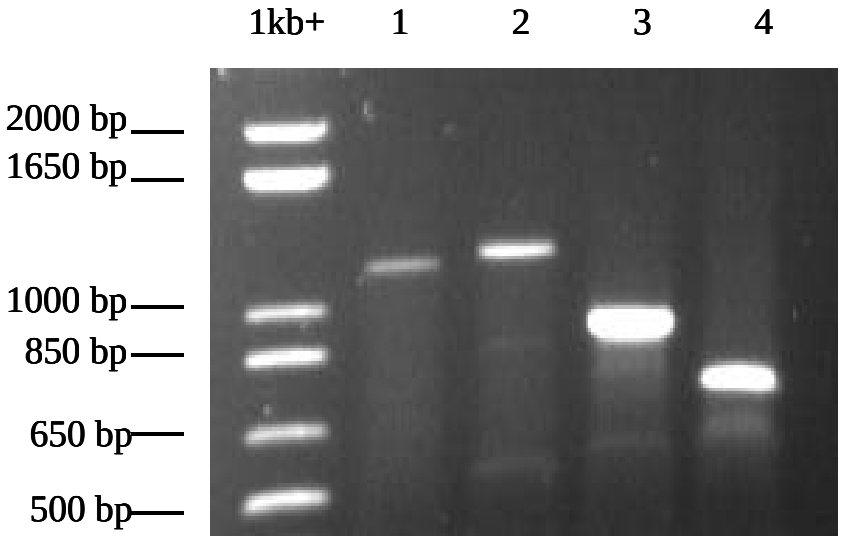


Figure 2.3.5. A) Map of the region upstream and including *alr4311*. Primers are indicated by arrows. B) RT-PCR was used to determine if the *alr4308-alr4311* gene cluster is transcribed as an operon. RT-PCR products were visualized on an agarose gel. Primer pairs: lane 1, 4308-A + 4310-B; lane 2, 4309-A + 4311-Z; lane 3, 4310-A + 4311-Y; lane 4, 4311-A + 4311-C.

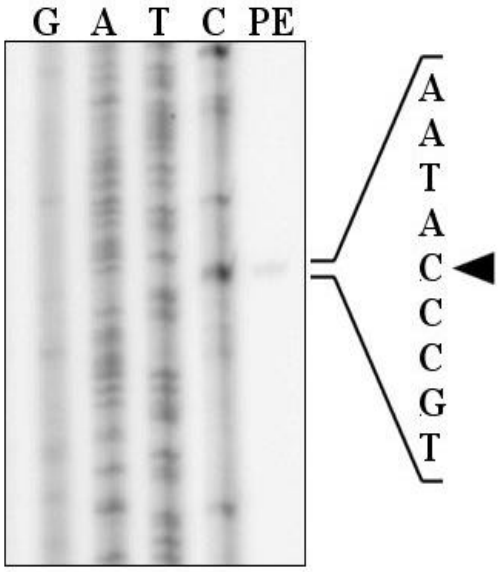


Figure 2.3.6. Identification of the 5' terminus of *alr4308* transcripts. Total RNA (30µg) from cells grown for 24 hours in the absence of fixed nitrogen was used in the primer extension reaction with ^{32}P-end-labeled primer 4308PE-1. The sequence ladder was generated using the same primer. The arrowhead indicates the transcription start point of the gene.

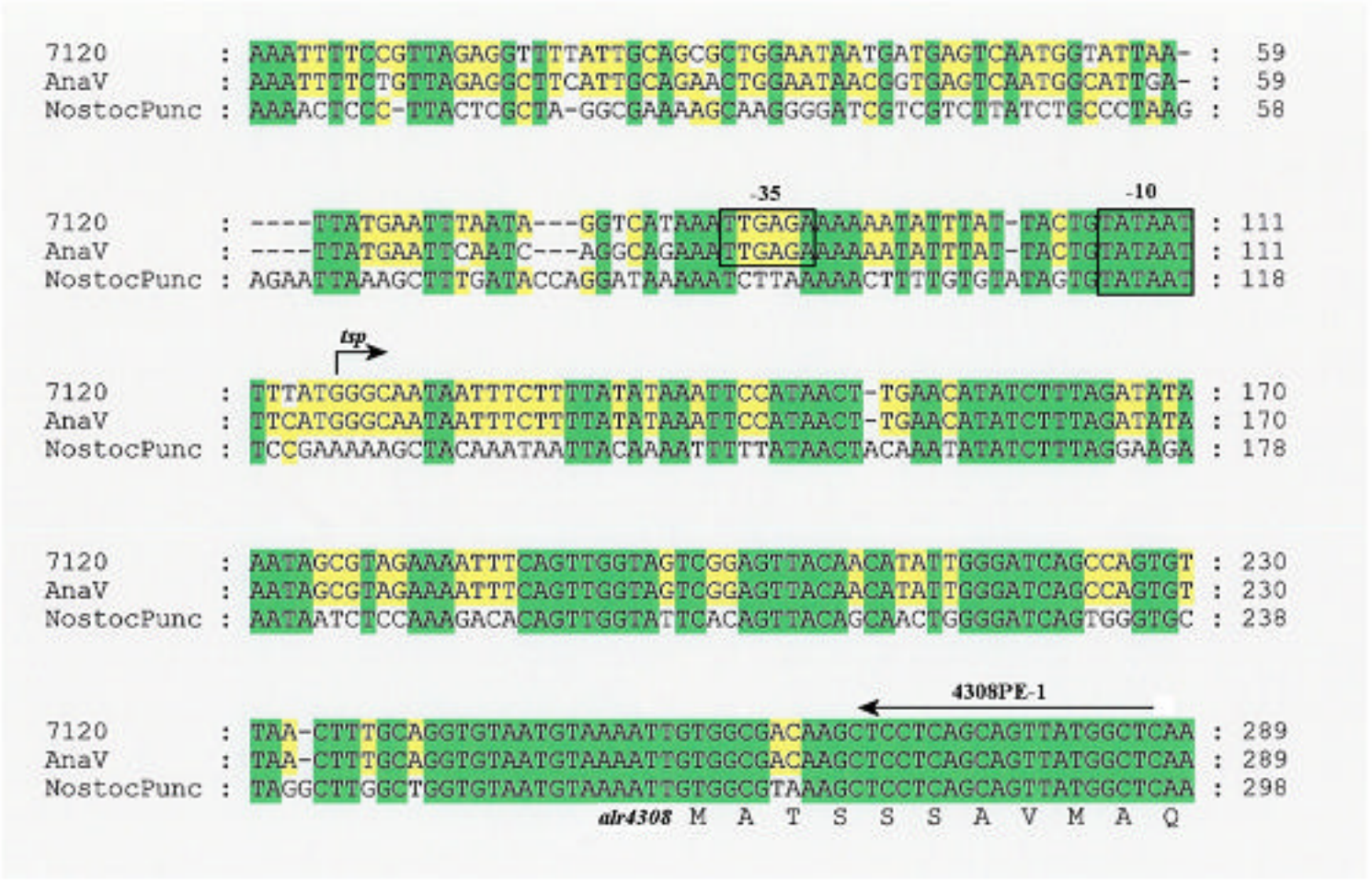


Figure 2.3.7. Multiple sequence alignment of *alr4308* upstream region from *Anabaena* sp. strain PCC 7120 (7120) with similar sequences from *Anabaena variabilis* (AnaV) and *Nostoc punctiforme* (NostocPunc). The primer used for the primer extension assay (4308PE-1) is indicated by a long arrow. The transcription start point (*tsp*) of *alr4308* is indicated by a bent arrow. The putative –10 and –35 consensus sequences are boxed. The translation product of the open reading frame is shown below the alignment in single letter amino acid code. Identical sequences are shaded in green.

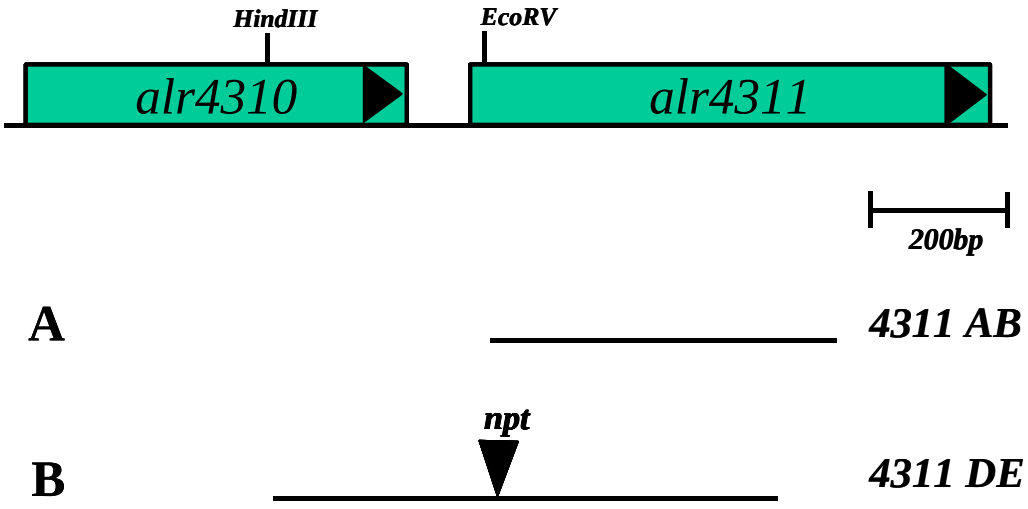
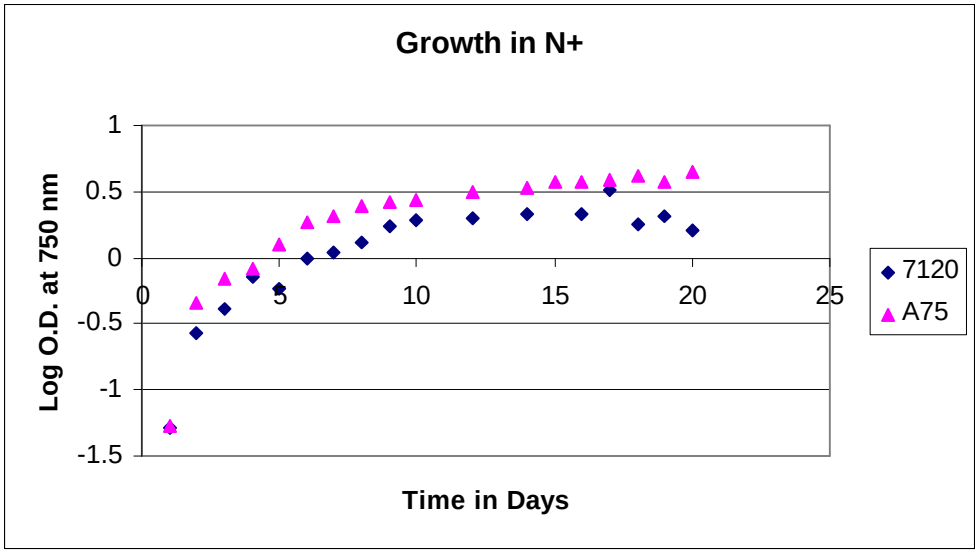


Figure 2.3.8. Organization of the *alr4311* gene and corresponding gene fragments for mutagenesis. A) Gene fragment used for single recombination mutagenesis. B) Gene fragment used for double recombination mutagenesis. Arrowhead indicates *npt* cassette insertion.

A



B

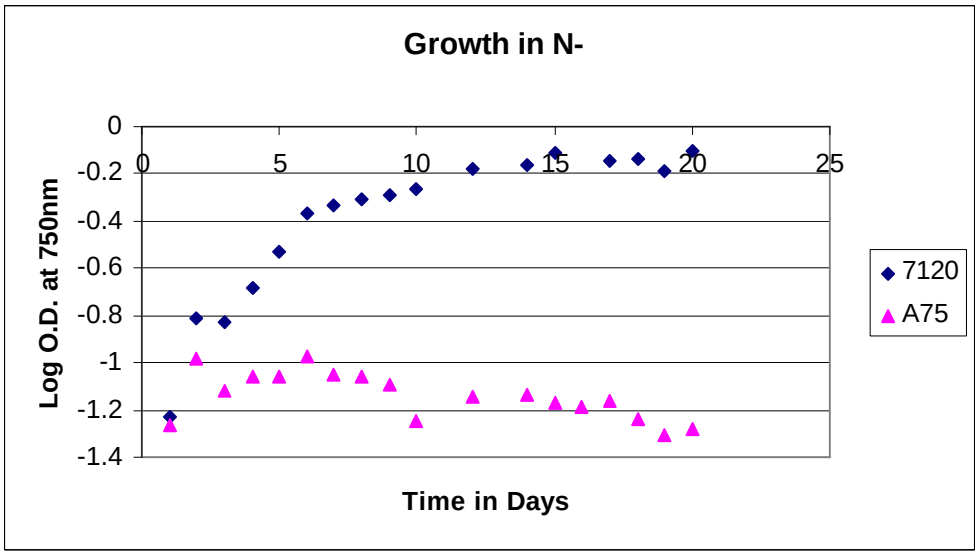
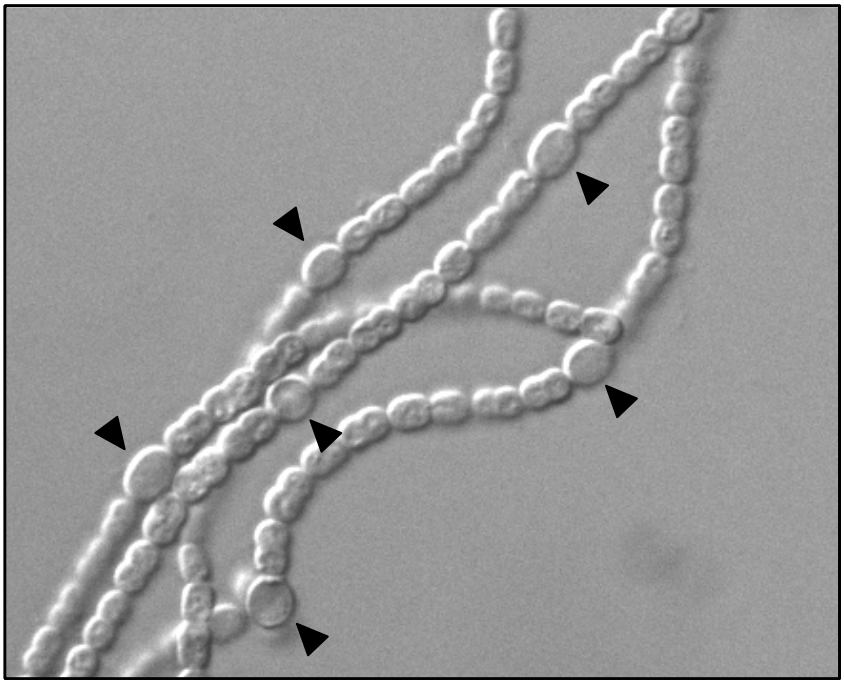


Figure 2.3.9. Growth analysis of wild type and mutant strain A75 in BG-11 with nitrate (A) and without nitrate (B).

A



B

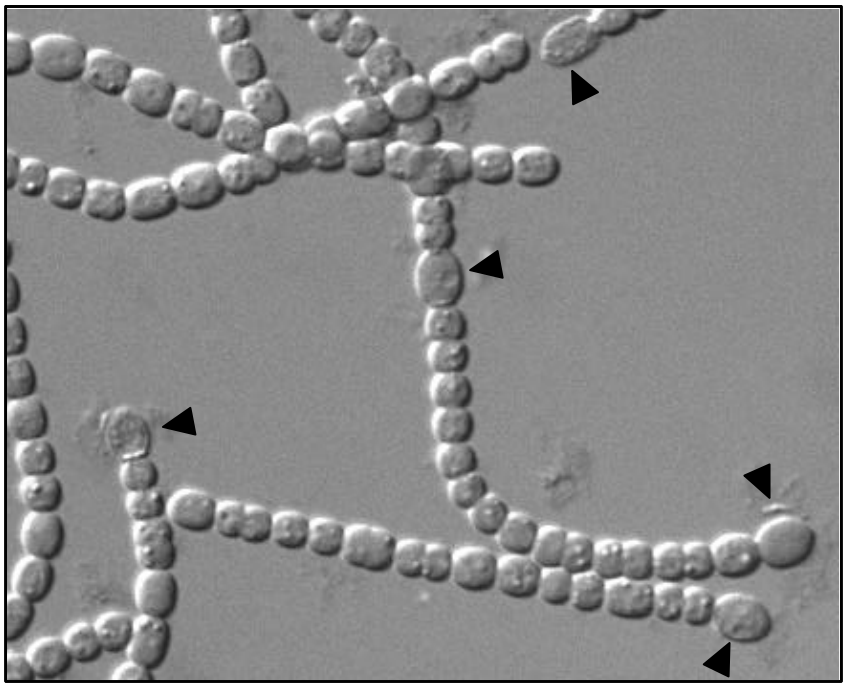


Figure 2.3.10. Light microscopy (60X magnification) of wild type (A) and A75 mutant (B) strains starved for fixed nitrogen. Cultures were grown in BG-11 without nitrate for 48 hours. Arrowheads indicate heterocysts.

CHAPTER 3

CHARACTERIZATION OF *all4312*

3.1 INTRODUCTION

Two-component regulatory systems are composed of two partner proteins that regulate a diverse array of metabolic processes in response to environmental changes. These regulatory systems play a significant role in the adaptation of an organism to its surroundings to allow survival in a changing environment. Through a histidine protein kinase (HPK) sensor protein and a response regulator (RR) protein, the two-component system is able to sense and rapidly respond to such environmental stresses as changes in osmolarity, temperature, acidity, and nutrient and toxin levels (24). Although the majority of these systems are found in prokaryotes, examples of two-component systems in eukaryotes, namely fungi, amoebae, and plants, have been recently discovered. No two-component systems have been identified in the animal kingdom to date (26).

The first two-component system to be described was the nitrate assimilation system of *E. coli* (22). This work focused on elucidating the biochemical events underlying the sensory-response mechanism of the nitrogen regulation system to regulate gene expression of the *glnALG* operon according to nitrogen availability. Shortly thereafter, it became clear that proteins of two-component systems share conserved domains with other regulatory proteins, and that they probably function via similar mechanisms (23).

Much of the knowledge of signal transduction pathways formed by two-component systems comes from studies of model organisms in bacteria. Well-characterized two-component systems include those for chemotaxis in *E. coli* and *S. typhimurium* (Che proteins), porin regulation in *E. coli* (Omp proteins), and sporulation in *Bacillus subtilis* (Spo proteins).

The HPK sensor consists of a variable N-terminal domain that is commonly referred to as the periplasmic sensing domain (or input domain) and a conserved C-terminal transmitter domain, also known as the cytoplasmic kinase domain (26, 10). The HPK senses an environmental stimulus through the sensor domain in the form of an input signal, which then triggers an autokinase activity of the transmitter domain. The HPK protein autophosphorylates at a conserved histidine residue in the transmitter domain by catalyzing the transfer of the gamma-phosphate of ATP (3, 24). The RR protein subsequently catalyzes transfer of the phosphoryl group from the histidine residue to an aspartate residue in the conserved N-terminal receiver domain of the RR (25). The variable C-terminal effector domain (or output domain) of the RR is then activated to elicit the specific response. In many cases, this response is transcriptional activation of genes required for responding to the particular environmental stress.

The level of phosphorylation of the response regulator acts as a regulatory mechanism for controlling the output response (26). The phosphorylation level can be regulated by the HPK, as some HPKs have phosphatase activity, or by autophosphatase activity of the RR itself. Auxiliary phosphatases have also been observed to enhance the autophosphatase activity of the RR.

The number of two-component regulatory systems in a given genome varies among organisms. Two-component systems appear to be more prevalent in prokaryotes than in eukaryotes, with more complex phosphorelay systems and multi-component regulatory systems being predominant in eukaryotic organisms. Among the prokaryotes, some bacterial genomes have multiple two-component signal transduction systems, while others appear to have no signal transduction systems of this type at all. For example, the genome of *E. coli*

has 62 proteins classified as two-component system proteins (Mizuno 1997), *Synechocystis* has 80 (20), *Prochlorococcus* MED4 has 11 (18) and *Mycoplasma pneumoniae* appears to have none (2). It is presumed that the number of HPKs in a genome directly correlates to the number of environmental stimuli to which an organism can respond.

The genome of *Anabaena* has a large family of proteins belonging to two-component signal transduction systems. There are 71 genes encoding HPKs, 71 genes encoding response regulators, and 53 genes encoding HPK-RR hybrid proteins, for a grand total of 195 genes predicted to encode components of these regulatory systems (13). More than half of these genes are present individually in the genome, while 85 of these genes are organized in 36 gene clusters of two to five genes. The large number of two-component genes is reflective of the highly adaptive nature of cyanobacteria in general, suggesting that *Anabaena* is capable of handling a wide variety of environmental stimuli.

There have been a few experimentally characterized genes encoding proteins of two-component systems in *Anabaena*. *patA* (Table 1.2.1) is a gene encoding a response regulator protein with similarity to CheY that plays a role in pattern formation of heterocysts and has no apparent DNA-binding domain (16). The HepK-DevA two-component system (Table 1.2.1) is involved in the synthesis of the heterocyst polysaccharide layer during differentiation (28). The ManR response regulator has been studied for its regulation of genes in a Mn(2+)-sensing two-component system (12).

Due to the widespread nature of two-component systems in bacteria and their absence from animals, two-component systems are actively being investigated for the development of antimicrobial drugs. This chapter describes the characterization of the gene *all4312* of *Anabaena*, which was chosen for further study based on the observation that this gene is up-

regulated in response to nitrogen starvation. *all4312* is predicted to encode a response regulator of a two-component signal transduction system.

3.2 MATERIALS AND METHODS

Strains and culture conditions. The strains and plasmids used in this study are listed in Table 3.2.1. Liquid cultures of cyanobacterial strains were grown in BG-11 or in modified Kratz and Meyers as described previously (7). *all4312* mutant strains were maintained in cultures containing neomycin sulfate (80 μg/ml). Complemented mutant and control strains were maintained in the same manner as the mutant strain with the addition of spectinomycin (10 μg/ml).

Escherichia coli strains were grown in 2YT or LB media supplemented with appropriate antibiotics.

Cyanobacterial developmental time courses (nitrogen step-down) and growth rate experiments were carried out as previously described (7). 1 ml samples of each culture were removed for $O.D._{750}$ measurements every 24 hours. Light microscopy at 60X magnification was used to observe filaments for growth and morphological characteristics.

Isolation and analysis of nucleic acids. Genomic DNA was isolated as described by Wu *et al.* (27). Total RNA was isolated using RNA Wiz (Ambion) according to the manufacturer's instructions, and RNA samples were purified using the RNeasy Mini Kit with the RNase-Free DNase set from Qiagen.

Southern blot analyses were performed using 10 μg genomic DNA. Northern blot analyses were performed using 20 μg total RNA isolated from *Anabaena* over a developmental time course. Probes for both Southern and Northern analyses were created by

PCR amplification of *Anabaena* genomic DNA. Primers 4312-A and 4312-D were used to create a 675 bp probe (4312 AD) for Southern analyses. Primers 4312-Bam2 and 4312-Sal were used to create a 1.3 kb probe (4312 B2S) for Northern analysis.

Probes were labeled using the AlkPhos Direct Labeling kit (Amersham Biosciences) and hybridizations were performed following the manufacturer's recommendations. Detection of hybridization signals was achieved by using ECF substrate and a Storm scanner (Molecular Dynamics). Quantitation was performed using ImageQuant software.

Primer extension was carried out using the Primer Extension System-AMV Reverse Transcriptase (Promega) according to the manufacturer's instructions with 30 μg of total RNA isolated from cells from a nitrogen step-down experiment after 24 hours. Primer 4312PE-2 was used for this assay.

Sequence ladders were generated using Promega's fmol DNA Cycle Sequencing System following the manufacturer's instructions for end-labeling primers. The same primer (4312PE-2) was used in the sequencing reactions as for primer extension. A 558 bp PCR product containing the upstream region of *all4312* plus a portion of the ORF was created by amplification with primers 4312-1 and 4312-300, and this product was used as template in the sequencing reaction.

Real-time RT-PCR was performed using the SYBR-Green PCR Core Reagents from Applied Biosystems according to the manufacturer's recommendations with *Anabaena* total RNA isolated from a developmental time course. Primer set 4312-1 and 4312-2, which amplified an 80 bp product, was used to measure expression of *all4312*, which was normalized against RNaseP transcript levels from amplification with primer set RNaseP-1

and RNaseP-2. *hepA* expression was measured as a positive control of heterocyst induction using primer set HepA-F3 and HepA-R3.

Inactivation of *all4312* by single recombination. *all4312* was initially inactivated by a single recombination insertion. PCR was used to create an insert containing a fragment of the *all4312* gene for cloning. Primers 4312-B and 4312-C were used to create insert 4312 BC (431 bp). The insert was cloned into the *Eco*RV site of pBN2, and this construct was subsequently transferred to *Anabaena* using standard procedures for conjugation (6). Southern blot analysis was used to confirm single recombination in the mutant (A76) using *npt*- and *all4312*-specific probes.

Inactivation of *all4312* by double recombination. *all4312* was inactivated by using the *sacB* method of positive selection for double recombinant mutants (4). A 768 bp fragment of the *all4312* gene was created by PCR of *Anabaena* genomic DNA with primers 4312-E and 4312-F, which generated *Hind*III sites at the 5' and 3' ends. A 1.1 kbp *Sma*I fragment containing the *npt* gene was isolated from pRL648 (6) and cloned into the *Eco*RV site of this gene fragment in the reverse orientation relative to the direction of transcription. The disrupted gene fragment was cloned into the *Nae*I site of pRL271, a suicide vector containing the *sacB* gene, and transferred into *Anabaena* using standard procedures for conjugation. Ex-conjugants were selected by resistance to neomycin, followed by selection with 8% sucrose to identify double recombinant mutants. Southern blot analysis was used to confirm double recombination in the mutant (A80) using *npt*- and *all4312*-specific probes.

Complementation. PCR amplification of *Anabaena* genomic DNA with primers 4312-Bam2 and 4312-Sal was used to create a fragment spanning the entire sequence of *all4312* plus ~500 bp of sequence upstream of the gene, with engineered *Bam*HI and *Sal*I

sites at the 5' and 3' ends, respectively. This fragment was cloned into the *Bam*HI/*Sal*I site of the promoterless RSF1010-based vector, pRL1383a (11), and transferred to the double recombinant *all4312* mutant strain (A80) using standard procedures for conjugation to create strain A81. pRL1383a without the *all4312* insert was transferred to A80 via conjugation, and the resulting strain (A82) was used as a control. Ex-conjugants were selected on media containing neomycin (80 μg/ml) and spectinomycin (10 μg/ml).

Phylogenetic analysis of *all4312*. A BLASTP (1) deduced protein search of the non-redundant GenBank protein database at NCBI was performed using the translated amino acid sequence of *all4312* (All4312), obtained from the genome database for cyanobacteria, CyanoBase (http://www.kazusa.or.jp/cyanobase/). The top 30 sequences (E-value equal to or less than 6e-35) were collected for multiple sequence alignment using the default parameters of CLUSTAL W (9). The alignment was subsequently imported into GeneDoc version 2.6 (21) for editing and visual inspection. A phylogenetic tree was constructed using the neighbor-joining method of the software program MEGA version 2.1 (14). The tree was evaluated using the p-distance amino acid option and the bootstrap test of phylogeny based on 1000 replications. Simple Modular Architecture Research Tool (SMART) analysis was used to identify domain locations of the translated sequence of *all4312* (15).

3.3 RESULTS AND DISCUSSION

***all4312* encodes a response regulator protein of a two-component system.** To identify potential genes involved in heterocyst development, a screen for sequences up-regulated at the transcript level in the absence of fixed nitrogen was performed (5). From the collection of sequences identified, AD206.2 was selected for further study, as transcripts

were shown to increase during heterocyst development. AD206.2 spans the *alr4311* and *all4312* genes (see Figure 2.3.1).

all4312 is predicted to encode a 256 amino acid response regulator of a two-component system. All4312 shows similarity to the OmpR subfamily of response regulators. Members of this protein family are characterized by the presence of two domains, namely an N-terminal receiver domain and a C-terminal winged helix DNA-binding domain. A multiple sequence alignment of All4312 with similar sequences from two closely related nitrogen-fixing cyanobacteria, *Anabaena variabilis* (Ava_1263) and *Nostoc punctiforme* (NpR3907) with the OmpR protein from *E. coli* was performed to identify conserved regions among these proteins (Figure 3.3.1). A conserved aspartate residue present in each amino acid sequence was identified at position 56, presumably corresponding to the Asp-55 phosphorylation site of OmpR. The receiver domain and DNA-binding domain were identified by SMART analysis at positions 4-114 and 148-223, respectively.

A phylogenetic analysis similar to that performed for *alr4311* was utilized in the characterization of *all4312* to infer the function of the protein it encodes based on similarity to known proteins. A BLASTP search of the non-redundant GenBank database was performed using the translated amino acid sequence of *all4312.* The top 30 most similar sequences were collected and aligned: these sequences included putative bacterial response regulator proteins, the majority of which are classified as a result of prediction (based on sequence similarity) rather than experimental validation. A phylogeny of these sequences was constructed based on the neighbor-joining method (Figure 3.3.2). The most closely related sequences were found to be uncharacterized, putative response regulators from *Anabaena variabilis* (E-value = 1e-143, 100% identity) and *Nostoc punctiforme* (E-value =

3e-122, 91% identity). Other similar sequences were putative response regulators from *Synechocystis* sp. PCC 6803 (E-value = 2e-91, 72% identity), *Crocosphaera watsonii* WH 8501 (E-value = 2e-94, 73% identity), and *Trichodesmium erythraeum* IMS101 (E-value = 6e-91, 71% identity).

The results of this phylogenetic analysis indicate that All4312 shares strong sequence similarity with response regulator proteins from two nitrogen-fixing, filamentous, freshwater cyanobacteria, *A. variabilis* and *N. punctiforme*. All4312 also shares high sequence similarity to a response regulator protein from the marine cyanobacterium, *T. erythraeum*, which is a non-heterocystous nitrogen-fixer, and to a response regulator from the unicellular, non-nitrogen fixing cyanobacterium *Synechocystis*. Furthermore, All4312 shares sequence similarity with a response regulator protein from the unicellular marine diazotroph, *C. watsonii*. The fact that All4312 shares high sequence similarity to response regulators from unicellular cyanobacteria and a diazotrophic cyanobacterium that does not form heterocysts indicates that All4312 may play a functional role in cellular processes during heterocyst differentiation, rather than a structural role.

***all4312* expression is up-regulated in response to nitrogen starvation.** To characterize the expression of *all4312* during heterocyst development, Northern blot analysis and real-time RT-PCR were performed. Total RNA was isolated from samples collected at 0, 3, 12, and 24 hours after cultures of *Anabaena* were starved for fixed nitrogen. Northern analysis was used as a preliminary experiment to qualitatively determine the expression pattern of *all4312* during a developmental time course. The results of this analysis showed that *all4312* is expressed in response to nitrogen starvation as early as 3 hours after step-down, and *all4312* continues to be expressed through 24 hours of development (data not

shown). Subsequently, real-time RT-PCR was performed to determine the relative ratio of increase in expression during differentiation (Table 3.3.1). Three independent cultures of *Anabaena* were subjected to nitrogen step-down for RNA isolation at the same time points as collected for Northern analysis. These biological triplicates were run in duplicate to account for pipetting/technical errors, and expression was normalized against RNaseP transcript levels. The 0-hour time point was designated as the "calibrator" sample, and the ratio of transcript increase for later time points is expressed relative to this time point. Expression of *hepA* was used as a positive control for induction of expression. *all4312* expression was found to increase approximately 7-fold after 3 hours of nitrogen starvation, and this steady state level was observed through the 24-hour time point (as for Northern analysis). These results suggest that *all4312* is required early in the differentiation process. Whether expression of *all4312* is localized in heterocysts or expressed along a filament remains unknown, as total RNA is isolated from whole filaments containing both vegetative cells and heterocysts.

Expression of *all4312* was analyzed in two developmental mutants, *ntcA* and *hetR*. Real-time RT-PCR was performed with samples obtained from 0 and 24 hours after nitrogen step-down. No induction of *all4312* transcription was observed in the *ntcA* mutant, and induction in the *hetR* mutant was significantly decreased (data not shown). As both of these genes are required early in differentiation, this result suggests that *all4312* may be directly or indirectly regulated by NtcA and/or HetR.

Identification of the 5´ terminus of *all4312*. Unlike *alr4311*, which is part of a cluster of four genes, *all4312* is situated as a single gene on the chromosome. This places *all4312* in the group of response regulators in *Anabaena* that are present individually, rather

than part of a cluster, and no potential HPK proteins appear to be encoded upstream of *all4312*. The closest gene is *asr4313*, a gene encoding a small hypothetical protein of 80 amino acids, and is situated 538 bp upstream of *all4312* in the opposite direction for transcription.

Primer extension analysis was used to map the transcription start point (tsp) of *all4312*. RNA isolated from cells at 24 hours after nitrogen step-down was used in the primer extension reaction with primer 4312PE-2 (Figure 3.3.3 and Figure 3.3.4). A single transcript mapping at 27 bp from the 5´ end of the *all4312* ORF was detected. The 5´ terminus of the transcript mapped 3´ to a region conforming to the –10 consensus sequence for s^{70} promoters. This –10 sequence (TAN_3T) is conserved in *A. variabilis* and *N. punctiforme* (Figure 3.3.4). Additionally, a conserved NtcA binding site ($GTAN_3TAC$) was identified 22 bp upstream of the –10 sequence in each of the genes. This spacing fits the requirements for regulation by NtcA, as it has previously been shown that NtcA binds promoter regions in which the –10 box is spaced 5-7 nucelotides from the tsp and 22-23 nucleotides from the $GTAN_3TAC$ consensus sequence (8). The presence of an NtcA binding site in the promoter region of *all4312* is consistent with the lack of *all4312* expression in an *ntcA* mutant, suggesting that *all4312* is regulated by NtcA. This is consistent with the conservation of the NtcA binding site in the sequences from *A. variabilis* and *N. punctiforme*.

The role of *all4312* in heteocyst development/function. Preliminary experiments to inactivate *all4312* by single recombination were performed to investigate the role of this gene during heterocyst development. A fragment of *all4312* was cloned into a vector bearing a neomycin-resistant cassette (*npt*) and transferred to *Anabaena* by conjugation. A map showing the location of the gene fragment used for single recombination is shown in Figure

3.3.5 (A). Single recombinant mutants were selected for neomycin resistance. The single recombinant *all4312* mutant appeared to be inviable in the absence of fixed nitrogen when grown in nitrogen-free media supplemented with neomycin, which was required to maintain the plasmid insertion. Experiments aimed at inactivating *all4312* by double recombination were subsequently carried out to generate a more stable mutant.

The wild-type *all4312* gene was inactivated by the *sacB* method of double-recombination (4). A fragment of *all4312* was disrupted by the *npt* cassette, cloned into a vector bearing the *sacB* gene, and transferred to *Anabaena* via conjugation. A map of the gene and the location of the fragment used for disruption are shown in Figure 3.3.5 (B). Double recombinant mutants were identified by selection on neomycin and the ability to grow on sucrose-containing media. Genomic DNA from the mutant strain was analyzed by Southern blotting to confirm the presence of the neomycin-resistance cassette in the wild-type *all4312* gene (data not shown).

Complementation of the *all4312* mutant strain A80 was performed by cloning the wild-type *all4312* gene and its promoter region into the replicative vector pRL1383a. This construct was transferred to A80 by conjugation to create strain A81. A complement control strain (A82) was created by transferring pRL1383a without the *all4312* gene insert to A80 by conjugation.

Growth rate analyses of strains A80, A81, and A82 were initially carried out in media with and without fixed nitrogen in the presence of antibiotics. None of these strains were able to grow in the absence of fixed nitrogen with antibiotics present in the media. It had previously been observed that *Anabaena* displays an increased sensitivity to certain antibiotics, such as spectinomycin, when grown without a source of combined nitrogen (11).

It is speculated that this may be due to higher sensitivity of protein synthesis in heterocysts compared to vegetative cells. As strains A81 and A82 were being maintained with spectinomycin, the growth rate analyses were performed in media with and without fixed nitrogen and antibiotics, under the premise that the double recombinant is a stable mutant (Figure 3.3.6). The removal of antibiotics during nitrogen step-down of complemented mutant strains in *Anabaena* has been done before (11). At the conclusion of the growth rate experiments, PCR was used to confirm that each strain had retained its original genetic identity.

In the presence of nitrate, all strains (wild type, A80, A81, and A82) grew exponentially and developed a dark green color with long filaments, as observed by microscopy. In the absence of nitrate, wild type cultures grew (exponentially) and developed their characteristic blue-green color after approximately 5-7 days. In contrast to the original growth characteristics observed for the *all4312* single and double recombinant mutants in response to nitrogen starvation, strain A80 was viable in the absence of fixed nitrogen when antibiotics were absent from the media. Additionally, strains A81 and A82 were viable under these conditions, as well.

The double recombinant *all4312* mutant shows a subtle difference in growth rate in comparison to that of wild type. Under our laboratory conditions, wild type cultures of *Anabaena* generally enter the logarithmic growth phase by the end of the first week after nitrogen step-down. However, strain A80 exhibited a slower growth response after nitrogen step-down, but reached a cell density comparable to that of wild type by the end of the 20-day growth period. Strain A81 grew only slightly better than strain A80, while strain A82 grew similarly to strain A80 by the end of the growth period (data not shown). The

observation that the *all4312* mutant is able to recover from an initially slow growth rate suggests that *all4312* is not essential for heterocyst development, but that this gene may play a role in the response to nitrogen starvation.

Morphology of the mutant strain A80 was observed by light microscopy to examine the mutant for heterocyst formation and pattern (Figure 3.3.7). Samples of the wild type and mutant cultures were removed at 48 hours after nitrogen step-down and stained with Alcian blue, a dye that stains the polysaccharide layer of heterocysts. Wild type filaments were observed to contain regularly spaced heterocysts with Alcian blue staining. The heterocysts were easily discerned by their somewhat larger, smoother appearance in comparison to vegetative cells. Filaments of the A80 mutant appeared to contain heterocysts based on the presence of Alcian blue staining; however, heterocysts did not appear to be present at the wild type level of frequency. Additionally, all of the cells along a given filament were of similar size, making the discernment of heterocysts difficult based on relative cell size and appearance alone.

The apparent decrease in heterocyst frequency seen in the *all4312* mutant after 48 hours without fixed nitrogen supports the observation of a slow growth rate in the mutant compared to wild type. Taken together, these results suggest that *all4312* may play a role in the early response to nitrogen starvation, perhaps as part of a regulatory network with NtcA.

Conclusion. In summary, phylogenetic and sequence analyses of *all4312* support the prediction of this gene as a response regulator of a two-component system with similarity to the OmpR subfamily of response regulators. The expression profile of *all4312* shows that *all4312* is induced as early as 3 hours after nitrogen step-down, placing it in the category of "early" genes required for heterocyst development. The presence of a conserved NtcA-

binding site suggests that *all4312* is regulated by NtcA, which is further supported by the absence of *all4312* induction in the *ntcA* mutant. The observation that the *all4312* mutant displays a slow response to nitrogen starvation suggests that *all4312* is important in the early stages of heterocyst differentiation, but not required, and that perhaps there are alternate/redundant signaling pathways involved in the response to nitrogen availability. The fact that *all4312* shares strong sequence similarity to several other cyanobacterial response regulators suggests that the function of *all4312* is conserved in other cyanobacteria.

3.4 REFERENCES

1. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool. J. Mol. Biol. 215:403-410.

2. Ashby, M. K. (2004). Survey of the number of two-component response regulator genes in the complete and annotated genome sequences of prokaryotes. FEMS Microbiol. Lett. 231:277–281.

3. Bourret, R. B, Borkovich, K. A., and Simon, M. I. (1991) Signal transduction pathways involving protein phosphorylation in prokaryotes. Annu. Rev. Biochem. 60:401-441.

4. Cai, Y. and Wolk, C. Peter. (1990) Use of a conditionally lethal gene in *Anabaena* sp. strain PCC 7120 to select for double recombinants and to entrap insertion sequences. J. Bacteriol. 172:3138-3145.

5. Curtis, S. E. and Hebbar, P. B. (2001) A screen for sequences up-regulated during heterocyst development in *Anabaena* sp. strain PCC 7120. Arch Microbiol 175: 313-322.

6. Elhai, J., and Wolk, C. P. (1988) Conjugal transfer of DNA to cyanobacteria. Methods Enzymol. 167:747-754.

7. Hebbar, P. B. and Curtis, S. E. (2000) Characterization of *devH*, a gene encoding a putative DNA binding protein required for heterocyst function in *Anabaena* sp. strain PCC 7120. J. Bacteriol. 182:3572-3581.

8. Herrero, A; Muro-Pastor, A. M.; and Flores, E. (2001) Nitrogen control in cyanobacteria. J. Bacteriol. 183:411-425.

9. Higgins, D., Thompson, J., Gibson, T., Thompson, J. D., Higgins, D. G., Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673-4680.

10. Hoch, J. A. and Silhavy, T. J. (1995) Two-component signal transduction. ASM Press, Washington, DC.

11. Huang, G., Fan, Q., Lechno-Yossef, S., Wojciuch, E., Wolk, C. P., Kaneko, T., Tabata, S. (2005) Clustered genes required for the synthesis of heterocyst envelope polysaccharide in *Anabaena* sp. strain PCC 7120. J. Bact. 187:1114-1123.

12. Huang, W. and Wu, Q. Y. (2004) Identification of genes controlled by the manganese response regulator, ManR, in the cyanobacterium, *Anabaena* sp. PCC 7120. Biotechnol. Lett. 26:1397-1401.

13. Kaneko, T., Nakamura, Y., Wolk, C. P., Kuritz, T., Sasamoto, S., Watanabe, A., Iriguchi, M., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kohara, M., Matsumoto, M., Matsuno, A., Muraki, A., Nakazaki, N., Shimpo, S., Sugimoto, M., Takazawa, M., Yamada, M., Yasuda, M., Tabata, S. (2001) Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. DNA Res. 8:205-213.

14. Kumar, S., Tamura, K.; Jakobsen, I. B., and Nei, M. (2001) MEGA2: Molecular Evolutionary Genetics Analysis software, Arizona State University, Tempe, Arizona, USA.

15. Letunic, I., Copley, R. R., Pils, B., Pinkert, S., Schultz, J., and Bork, P. (2006) SMART 5: domains in the context of genomes and networks. Nucl. Acids Res. 34:D257-D260.

16. Liang, J., Scappino, L., Haselkorn, R. (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. U S A. 89:5655-5659.

17. Marsh, J., Enfle, M., and Wykes, E. (1984) The pIC plasmid and phage vectors with versatile cloning sites for recombinant selection by insertional inactivation. Gene 32:481-485.

18. Mary, I., Vaulot, D. (2003) Two-component systems in *Prochlorococcus* MED4: Genomic analysis and differential expression under stress. FEMS Microbiology Letters. 226:135-144.

19. Mizuno, T. (1997) Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of *E. coli*. DNA Res. 4:161-168.

20. Mizuno, T., Kaneko, T., Tabata, S. (1996) Compilation of all genes encoding bacterial two-component signal transducers in the genome of the cyanobacterium, *Synechocystis* sp. strain PCC 6803. DNA Res. 3:407-414.

21. Nicholas, K. B. and Nicholas, H. B., Jr. (1997) GeneDoc: A tool for editing and annotating multiple sequence alignments. Distributed by the author.

22. Ninfa, A. J. and Magasanik, B. (1986) Covalent modification of the *glnG* product, NRI, by the *glnL* product, NRII, regulates the transcription of the *glnALG* operon in *Escherichia coli*. PNAS. 83:5909-5913.

23. Nixon, B. T., Ronson, C. W., and Ausubel, F. M. (1986) Two-component regulatory systems responsive to environmental stimuli share strongly conserved domains with the nitrogen assimilation regulatory genes *ntrB* and *ntrC*. Proc. NatI. Acad. Sci. USA. 83:7850-7854.

24. Parkinson, J. S. and Kofoid, E. C. (1992) Communcication modules in bacterial signaling proteins. Annu. Rev. Genet. 26:71-112.

25. Stock, J. B., Stock, A. M., Mottonen, J. M. (1990) Signal transduction in bacteria. Nature 344:395–400.

26. West, A. H. and Stock, A. M. (2001) Histidine kinases and response regulator proteins in two-component signaling systems. Trends Biochem Sci. 26:369-376.

27. Wu, X., Zarka, A., and Boussiba, S. (2000) A simplified protocol for preparing DNA from filamentous cyanobacteria. Plant Mol. Biol. Rep. 18:385-392.

28. Zhou, R. and Wolk, C. P. (2003) A two-component system mediates developmental regulation of biosynthesis of a heterocyst polysaccharide. J. Biol. Chem. 278:19939-19946.

TABLE 3.2.1. Bacterial strains, plasmids, and oligonucleotides

Strain, plasmid, or oligonucleotide	Characteristics	Source or reference
Strains		
Anabaena spp.		
PCC 7120	Wild type	R. Haselkorn
A76	PCC 7120 *all4312*::pBN2-4312#6, Nm^R	This study
A80	PCC 7120 *all4312*::*npt*, Nm^R	This study
A81	PCC 7120 *all4312*::*npt/all4312*$^+$, Nm^RSp^R	This study
A82	PCC 7120 *all4312*::*npt/all4312*$^-$, Nm^RSp^R	This study
E. coli		
EP-Max 10B	Electro-competent cells	Bio-Rad
Plasmids		
pBN2	Cloning vector, Ap^R	S. Curtis
pRL271	Suicide vector, Em^RCm^R	C. P. Wolk
pIC20H	Cloning vector, Ap^R	17
pRL648	Contains neomycin resistance cassette, Nm^R	6
pRL1383a	RSF1010-based cloning vector, Sp^RSm^R	11
Oligonucleotides		
4312-A	ATACCGTGTGCATCAAG	This study
4312-B	CTGCTCGGACTAATGAG	This study
4312-C	GTTAAATCGATATCCAACAT	This study
4312-D	CTTAGAGTTGAGCGTTG	This study
4312-E	GTTTGTGGAAGCTTAGCATTCAA	This study
4312-F	CTACAAAAGCTTGGGCATGGG	This study
4312-1	GTCCGAGCAGATAACATGAG	This study
4312-2	CCAGATGGTGACGGTATTG	This study
RNaseP-1	CGAGGGCGATTATCTATCTG	This study
RNaseP-2	CAAGGCCGAAGGAACTATG	This study
HepA-F3	TATCGCCTACGGTACATCTG	This study
HepA-R3	AACCCTTCGGGCATTTCTTC	This study
4312-Bam2	AAGGATCCAGGCTGAAATG	This study
4312-Sal	TAGTCGACAGTCATTAGTGCTGAG	This study
4312PE-2	GCTTGATGCACACGGTATTCC	This study

Table 3.3.1. Results of *all4312* expression study. RNA was isolated from triplicate cultures of *Anabaena* at time 0, 3, 12, and 24 h after nitrogen step-down and used for real-time RT-PCR. Ratios are given relative to expression levels of *all4312* at time 0 h.

all4312								
Relative Expression Ratios	**3 h**			**12 h**			**24 h**	
	Plate 1	**Plate 2**		**Plate 1**	**Plate 2**		**Plate 1**	**Plate 2**
Culture 1	225.46	114.37		187.94	85.16		101.03	59.20
Culture 2	116.87	219.13		87.09	122.99		465.88	NA
Culture 3	311.29	236.65		103.59	75.12		128.80	NA
Mean (Plate)	217.87	190.05		126.21	94.42		231.90	59.20
Mean (Time)	203.96			110.31			145.55	
Log base 2 ratio	7.67			6.79			7.19	

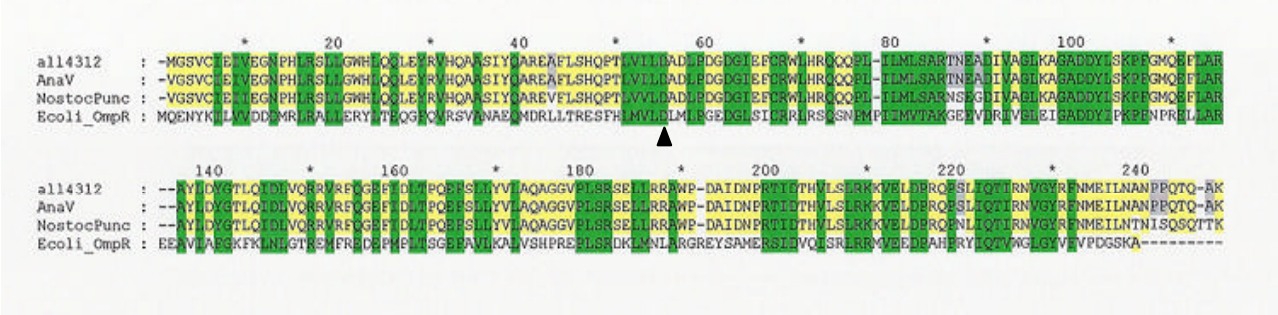


Figure 3.3.1. Multiple sequence alignment of All4312 with response regulators from *A. variabilis* (AnaV) and *N. punctiforme* (NostocPunc) and the OmpR protein from *E. coli*. Shading represents similarity levels as follows: green, 100%; yellow 80%; gray 60%. N- terminal receiver domain located at positions ~4-114; C-terminal DNA-binding domain located at positions ~148-223. Arrowhead indicates conserved Asp residue in the receiver domain.

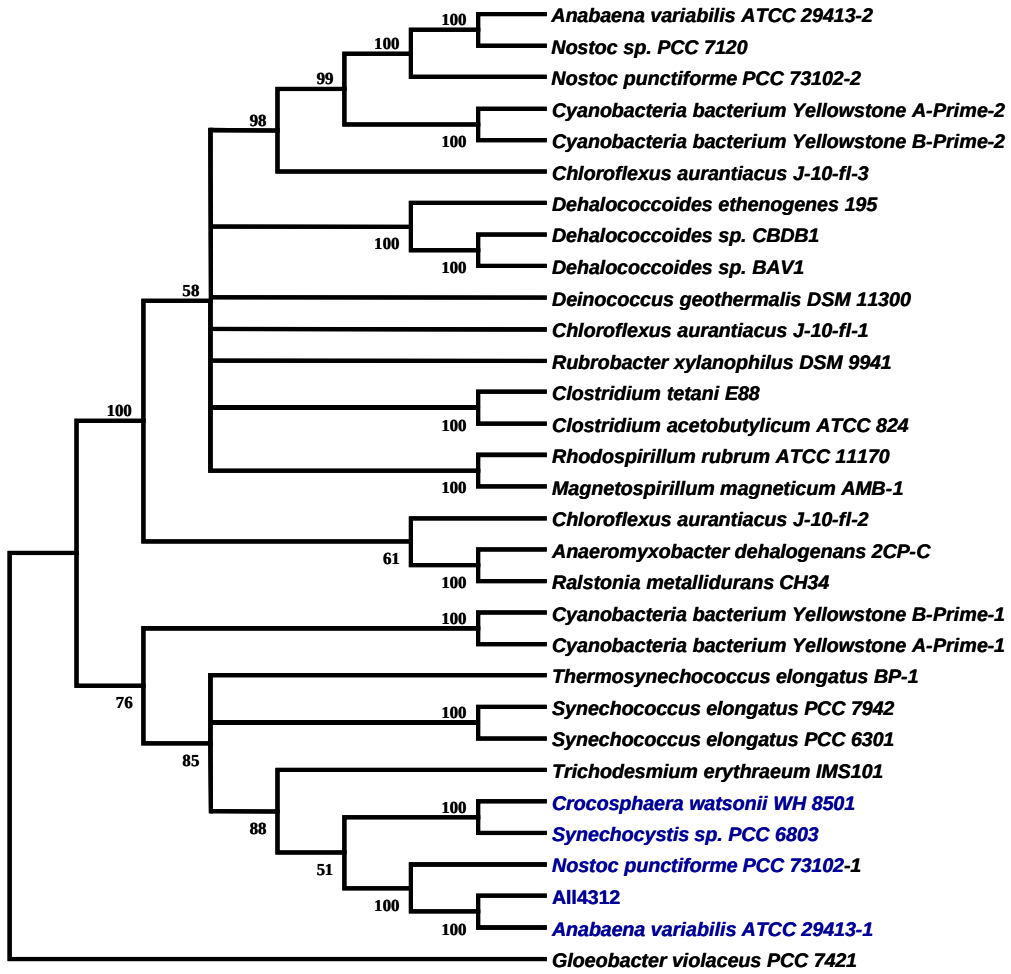


Figure 3.3.2. Neighbor-joining tree of bacterial response regulator proteins found in a BLASTP search of GenBank using the translated sequence of *all4312*. Bootstrap values (>50) based on 1000 replications are shown at each node.

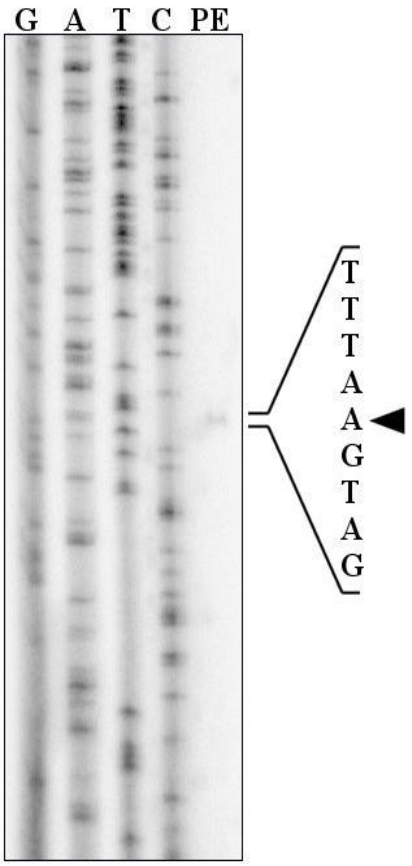


Figure 3.3.3. Identification of the 5' terminus of *all4312* transcripts. Total RNA (30µg) from cells grown for 24 hours in the absence of fixed nitrogen was used in the primer extension reaction with ^{32}P-end-labeled primer 4312PE-2. The sequence ladder was generated using the same primer. The arrowhead indicates the transcription start point of the gene.

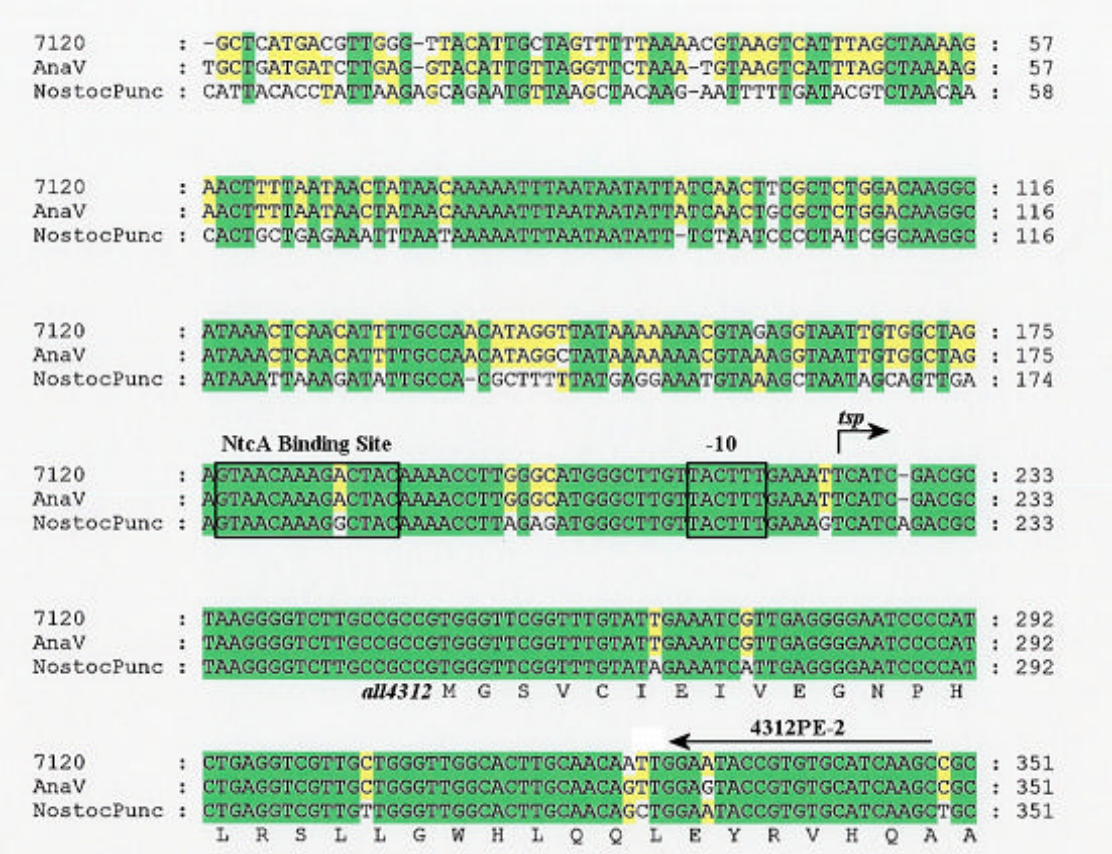


Figure 3.3.4. Multiple sequence alignment of *all4312* upstream region from *Anabaena* sp. strain PCC 7120 (7120) with similar sequences from *Anabaena variabilis* (AnaV) and *Nostoc punctiforme* (NostocPunc). The primer used for the primer extension assay (4312PE-2) is indicated by a long arrow. The transcription start point (*tsp*) of *all4312* is indicated by a bent arrow. The putative –10 consensus sequence and NtcA binding site are boxed. The translation product of the open reading frame is shown below the alignment in single letter amino acid code. Identical sequences are shaded in green.

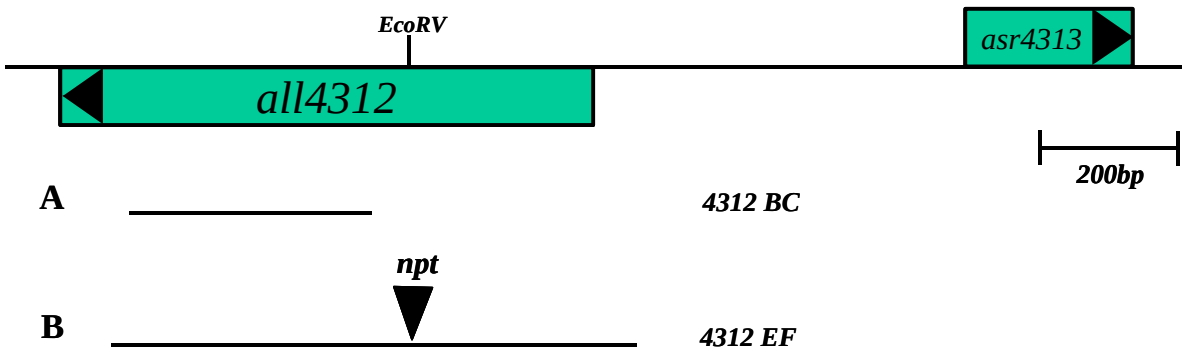
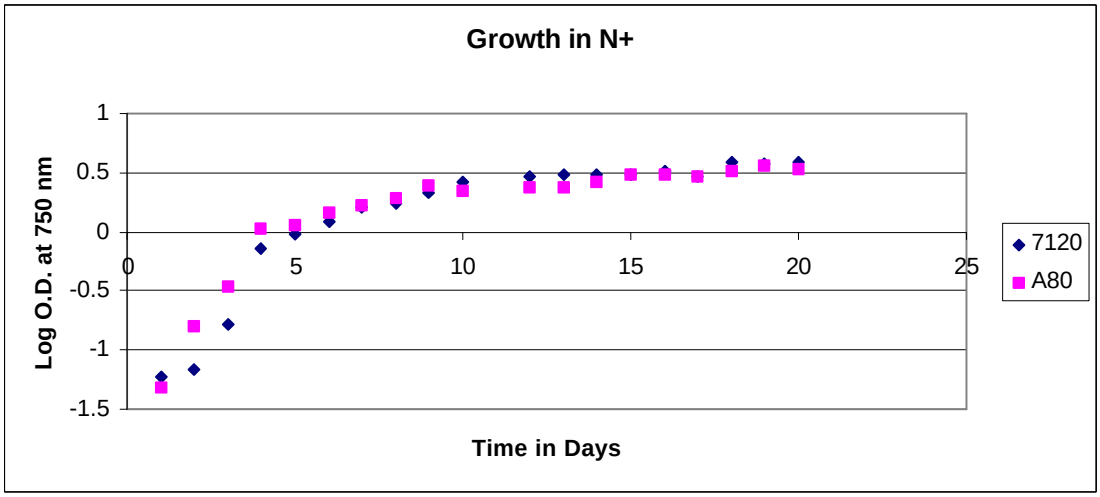


Figure 3.3.5. Organization of the *all4312* gene and corresponding gene fragments for mutagenesis. A) Gene fragment used for single recombination mutagenesis. B) Gene fragment used for double recombination mutagenesis. Arrowhead indicates *npt* cassette insertion.

A



B

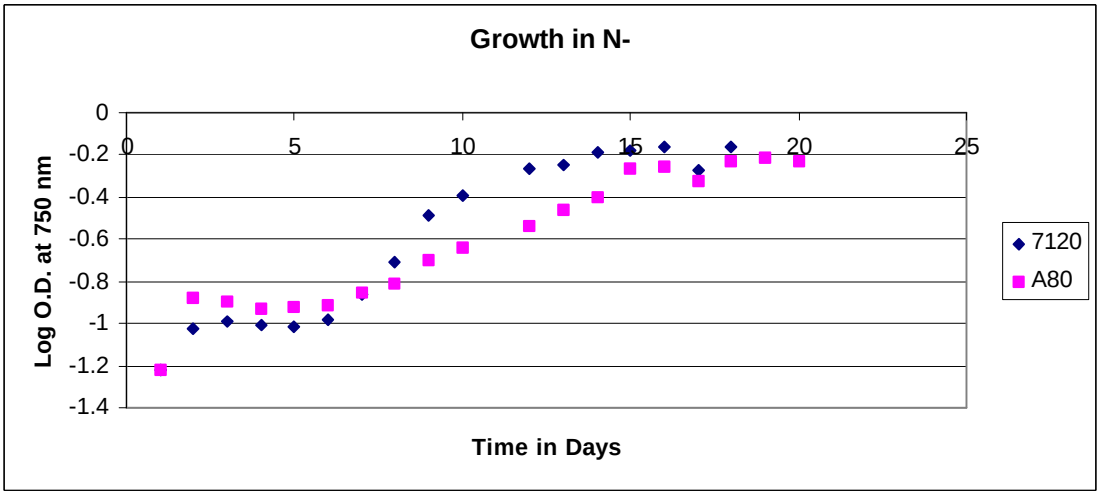
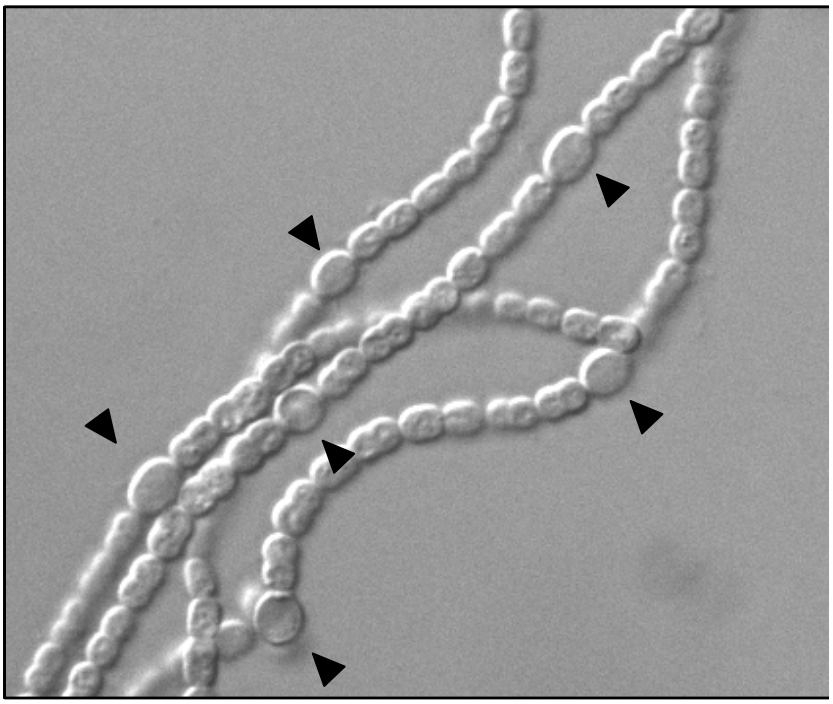


Figure 3.3.6. Growth analysis of wild type and mutant strain A80 in BG-11 with nitrate (A) and without nitrate (B).

A



B

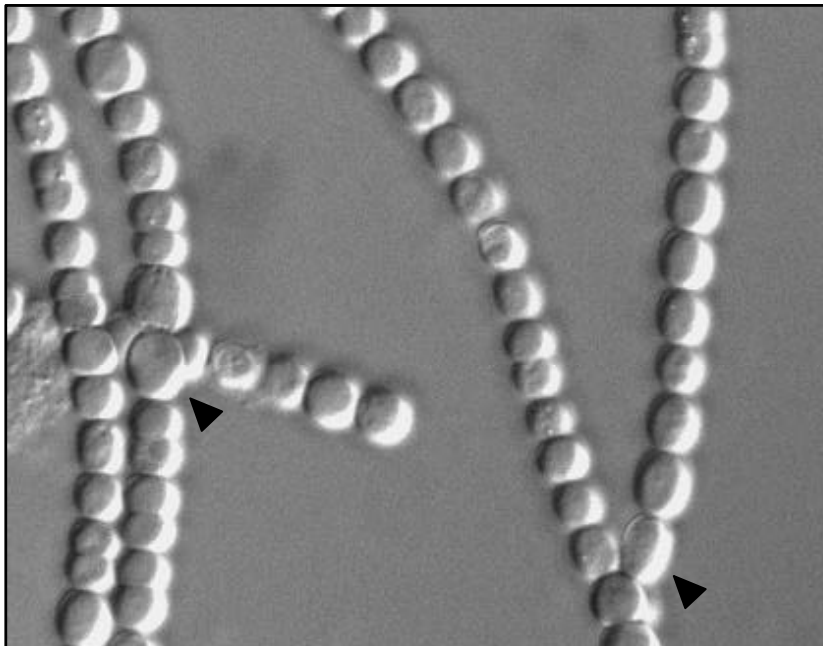


Figure 3.3.7. Light microscopy (60X magnification) of wild type (A) and A80 mutant (B) strains starved for fixed nitrogen. Cultures were grown in BG-11 without nitrate for 48 hours. Arrowheads indicate heterocysts.

CHAPTER 4

CONCLUSIONS AND FUTURE DIRECTIONS

4.1 CONCLUSIONS AND FUTURE DIRECTIONS

Based on the fact that *alr4311* and *all4312* are convergently transcribed and encode proteins involved in two different types of cellular processes, it appears that *alr4311* and *all4312* are unrelated. However, the experiments in characterization of these genes revealed several similarities between these genes. Through phylogenetic analysis, it is clear that the most closely related sequences to both *alr4311* and *all4312* come from *A. variabilis* and *N. punctiforme* and that their functions are conserved in filamentous, nitrogen-fixing freshwater cyanobacteria that differentiate heterocysts. Both *alr4311* and *all4312* are expressed (as single transcripts) early in heterocyst differentiation, suggesting that they play a role related to the response to nitrogen starvation and/or the early stages of heterocyst development. From mutational analyses, it was shown that *alr4311* is required for this response, while *all4312* is important for this response but not essential. Both mutants appeared to differentiate normal heterocysts; however, unlike the *alr4311* mutant, the *all4312* mutant was able to recover and grow like wild type after an initially slow response to nitrogen starvation, essentially eliminating the possibility of a heterocyst structural defect (thus indicating a functional role for *all4312* during heterocyst development). That the *alr4311* mutant was inviable in the absence of combined nitrogen may be due to either an impaired function of the heterocyst or to a problem in the synthesis of completion of the heterocyst structure. The observation of wild type-like heterocysts in the *alr4311* mutant is more suggestive of a functional role for *alr4311* during development, such as import of an essential nutrient for metabolism.

While the presence of a binding site for NtcA in the promoter region of *all4312* provides evidence that *all4312* is regulated by NtcA, there does not appear to be such a site

for *alr4311*. This raises the question of which genes might control expression of the *alr4308-alr4311* operon. In bacteria, there are several examples in which a response regulator directly activates genes of an ABC transporter. In *Anabaena*, the ManR response regulator controls expression of an ATP-binding protein in a manganese-sensing system (2). In *Bacillus subtilis*, the BceR response regulator of a two-component system for bacitracin resistance activates expression of the ABC transporter genes BceA and BceB (4). However, the genes in these examples are organized in clusters and transcribed in the same direction.

The fact that both *alr4311* and *all4312* are induced early during differentiation and expression is blocked in both the *ntcA* mutant and the *hetR* mutant suggests that these genes could be part of the same pathway. Further experiments are needed to investigate several questions. For further characterization of *alr4311*, experiments are needed to identify the other components of the ABC transporter, as well as the molecule for transport, and how the expression of these genes are induced in response to nitrogen starvation. For *all4312*, experiments are needed to identify the partner HPK sensor protein, as well as the specific stimulus perceived by the two-component system, and the genes that are regulated (activated or repressed) by the *all4312* response regulator. Additionally, experiments are needed to investigate whether the functions of *alr4311* and *all4312* are related after all.

While working on the characterization of *alr4311*, another lab had identified this gene as part of a transposon mutagenesis screen for mutants unable to grow in the absence of fixed nitrogen (Wolk, unpublished data). Mutational analysis of *alr4311*, including complementation of the mutant strain, were performed by this group, so complementation of the *alr4311* mutant presented in this work was not completed. We have an agreement to combine our results in a publication.

At the conclusion of the experimental work for *all4312*, two publications describing the characterization of this gene were published (3, 1). Sequence analyses, expression studies, and mutational analyses of *all4312* were consistent with our results, as was identification of the 5´ terminus of *all4312*. Regulation of *all4312* by NtcA was demonstrated by band-shift assays (3). GFP-fusion studies showed that expression of *all4312* is increased along whole filaments of cells within 3 hours of nitrogen starvation, with expression becoming localized to heterocysts by 24 hours (1). Heterocyst development was shown to be delayed in an *all4312* deletion mutant, and it was shown that induction of the *nifHDK* operon did not occur after 24 hours of growth without fixed nitrogen (1). Analysis of *hetR* expression in the *all4312* mutant showed that induction was blocked, and it was proposed that NtcA upregulates *hetR* through *all4312* (1).

4.2 REFERENCES

1. Ehira, S. and Ohmori, M. (2006) NrrA, a nitrogen-responsive response regulator facilitates heterocyst development in the cyanobacterium *Anabaena* sp. strain PCC 7120. Molec. Micro. 59:1692–1703.

2. Huang, W. and Wu, Q. Y. (2004) Identification of genes controlled by the manganese response regulator, ManR, in the cyanobacterium, *Anabaena* sp. PCC 7120. Biotechnol Lett. 26:1397-1401.

3. Muro-Pastor, A. M.; Olmedo-Verd, E.; and Flores, E. (2006) All4312, an NtcA-regulated two-component response regulator in *Anabaena* sp. strain PCC7120. FEMS Microbiol Lett. 256:171–177.

4. Ohki, R., Giyanto, T. K., Masuyama, W., Moriya, S., Kobayashi, K., and Ogasawara, N. (2003) The BceRS two-component regulatory system induces expression of the bacitracin transporter, BceAB, in *Bacillus subtilis*. Mol Microbiol. 49:1135-1144.