# UC San Diego

## UC San Diego Electronic Theses and Dissertations

**Title**

Transcriptional regulation of heterocyst development in Anabaena sp. strain PCC 7120

**Permalink**

https://escholarship.org/uc/item/1mq4x763

**Author**

Flaherty, Britt Lee

**Publication Date**

2012

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

# Transcriptional Regulation of Heterocyst Development in *Anabaena* sp. strain PCC 7120

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Biology

by

Britt Lee Flaherty

Committee in Charge:

Professor James Golden, Chair
Professor Eric Allen
Professor Steve Briggs
Professor Steve Mayfield
Professor Brian Palenik

2012

Copyright

Britt Lee Flaherty, 2012

All rights reserved.

The Dissertation of Britt Lee Flaherty is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2012

# TABLE OF CONTENTS

SIGNATURE PAGE ..... iii
TABLE OF CONTENTS ..... iv
LIST OF TABLES ..... vi
LIST OF FIGURES ..... vii
LIST OF ABBREVIATIONS ..... viii
ACKNOWLEDGEMENTS ..... ix
VITA ..... xii
ABSTRACT OF THE DISSERTATION ..... xiii
Chapter 1. Introduction ..... 1
1.1 CYANOBACTERIA ..... 1
1.1.1 A Brief History ..... 1
1.1.2 Cyanobacterial Genetics and their Role in Green Energy ..... 2
1.2 NITROGEN FIXATION ..... 3
1.2.1 Nitrogen Fixation in *Anabaena* ..... 3
1.2.2 A Transcriptional Cascade Regulates Heterocyst Development ..... 7
1.2.3 HetR is a Key Regulator of Heterocyst Development ..... 8
1.2.4 Pattern Formation: PatS, HetN, PatA, and PatN ..... 11
1.2.5 Degradation of the Phycobilisomes ..... 12
1.2.6 The Role of DevH in *hgl* gene expression ..... 13
1.3. BACTERIAL TRANSCRIPTOMES ..... 14
1.3.1 Deep Sequencing ..... 14
1.3.2 Transcriptional Studies in Cyanobacteria ..... 16
1.3.3 Bacterial transcriptomes encode abundant antisense transcription ..... 17
1.3.4 ChIP in Bacteria ..... 18
1.4 SUMMARY OF THIS WORK ..... 21
Chapter 2: Gene Expression after Nitrogen Deprivation ..... 23
2.1 INTRODUCTION ..... 23
2.2 MATERIALS AND METHODS ..... 25
2.2.1 Preparation of RNA samples ..... 25
2.2.2 Preparation of a cDNA library ..... 26
2.2.3 Illumina sequencing and analysis ..... 28
2.3 RESULTS AND DISCUSSION ..... 29
2.3.1 Analysis of the transcriptional response to nitrogen deprivation ..... 29
2.3.2 Identification of unstudied genes regulated in response to nitrogen deprivation ..... 35
2.3.3 Transcript mapping ..... 38

2.4 CONCLUSIONS ....43
Chapter 3: Antisense Transcription in *Anabaena* ....45
3.1 INTRODUCTION ....45
3.2 MATERIALS AND METHODS ....47
3.2.1 Antisense RNA mapping ....47
3.2.2 *gfp*-fusion to antisense promoters ....47
3.3 RESULTS AND DISCUSSION ....48
3.3.1 Identification of antisense RNAs ....48
3.3.2 Analysis of *anti-nblA* ....52
3.4 CONCLUSIONS ....56
Chapter 4: ChIP of HetR ....58
4.1 INTRODUCTION ....58
4.2 MATERIALS AND METHODS ....61
4.2.1 Cell Growth Conditions and Nitrogen Deprivation ....61
4.2.2 Chromatin Immunoprecipitation ....62
4.2.3 DNA Library Preparation and Sequencing ....63
4.2.4 Sequence Alignment and Peak Finding ....64
4.2.5 *gfp*-reporter Fusion Construction ....64
4.2.6 Imaging GPF promoter fusions ....68
4.3 RESULTS AND DISCUSSION ....68
4.3.1 ChIP Peaks ....68
4.3.2 Identification of a Refined Putative HetR Binding Site ....69
4.3.3 GO Term Enrichment in HetR Targets ....73
4.3.4 *gfp*-Promoter Fusions for HetR Targets ....76
4.4 CONCLUSIONS ....96
Chapter 5: ChIP of DevH ....97
5.1 INTRODUCTION ....97
5.2 MATERIALS AND METHODS ....98
5.2.1 Cell Growth Conditions and Nitrogen Deprivation ....98
5.2.2 Chromatin Immunoprecipitation ....99
5.2.3 DNA Library Preparation and Sequencing ....100
5.2.4 Sequence Alignment and Peak Finding ....101
5.3 RESULTS AND DISCUSSION ....101
5.4 CONCLUSIONS ....103
Chapter 6: Conclusions ....107
References ....112

# LIST OF TABLES

Table 2.1: Temporal response of the *nifHDK* operon and selected heterocyst glycolipid (*hgl*) and polysaccharide (*hep*) genes to nitrogen deprivation 33

Table 2.2: Unstudied regulatory genes with 5-fold or greater increase in expression by 6, 12, or 21 hours after nitrogen deprivation[1] .....36

Table 2.3: Transposase gene families with a 2-fold or greater increase in expression by 6 hours after nitrogen deprivation[1] .....37

Table 3.1: Antisense RNAs transcribed within the ORF or 5' UTR of genes involved in heterocyst differentiation .....50

Table 4.3: High-Confidence ChIP Peaks.....71

Table 4.4: HetR Targets Chosen for Further Analysis .....77

Table 5.1: DevH ChIP Hits .....105

Table 5.2: High Confidence DevH ChIP Hits.....106

# LIST OF FIGURES

Figure 1.1: Morphology of Differentiating Cells .....5
Figure 1.3: The transcriptional cascade that regulates heterocyst development.....10
Figure 2.1: Gene clusters upregulated during nitrogen deprivation. .....34
Figure 2.2: Transcript 5' ends of *psbB*, *petF*, *nrrA*, *psbAI*, *atp1*, and *rbcL*. ....40
Figure 2.3: Analysis of the *nifB-fdxN-nifS-nifU* operon structure, 5' end, and expression levels in response to nitrogen deprivation. .....41
Figure 3.1: Examples of antisense RNAs.....51
Figure 3.2: *Anti-nblA* is expressed in vegetative cells .....55
Figure 4.1: Refined HetR Consensus binding site as identified by ChIP peaks. .....72
Figure 4.2: GO Term enrichment in HetR target genes identified by ChIP. ...75
Figure 4.2: Promoter-*gfp* fusions and RNA-seq expression of *asr1469*.....81
Figure 4.3: Transient expression of a *asr1469-gfp reporter fusion* in heterocysts.....82
Figure 4.4: Promoter-*gfp* fusions and RNA-seq expression of *alr3758*.....84
Figure 4.5: Specific expression of *alr3758,* an anti-sigma factor antagonist, in heterocysts.....85
Figure 4.6: Promoter-*gfp* fusions and RNA-seq expression of *asl2028*. .....88
Figure 4.7: *gfp* fusions show vegetative cell-specific expression of *asl2028*. 89
Figure 4.8: Promoter-*gfp* fusions and RNA-seq expression of *alr2242*.....92
Figure 4.9: *gfp* fusions show suppression of *alr2242* in Wild Type cells. .....93
Figure 4.10: Promoter-*gfp* fusions and RNA-seq expression of *all5131*. .....95

# LIST OF ABBREVIATIONS

| | |
|---|---|
| BG-11 | Blue Green 11 medium |
| $BG\text{-}11_0$ | Blue Green 11 medium without sodium nitrate |
| bp | base pairs |
| CDS | Coding Sequence |
| ChIP | Chromatin Immunoprecipitation |
| DIC | Differential Interference Contrast |
| GFP | Green Fluorescent Protein |
| GO | Gene Ontology |
| h | hours |
| *hep* | heterocyst-specific exopolysaccharide genes |
| *hgl* | heterocyst-specific glycolipid genes |
| MACS | Model Based Analysis of ChIP Sequences |
| N- | Conditions without a source of fixed nitrogen |
| N+ | Conditions with fixed nitrogen present |
| *nif* | nitrogen fixation genes |
| OD | Optical Density |
| RACE | Rapid Amplification of cDNA Ends |
| RPKM | Reads per Kb gene model per million reads in a sample |
| sRNAs | small RNAs |
| TRITC | Tetramethyl Rhodamine Iso-Thiocyanate microscope filter |
| TSS | Transcriptional Start Site |
| UTR | Untranslated Region |

# ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. James Golden for taking me on as a young graduate student and for constantly supporting this work throughout its journey. His consistent commitment to science and teaching as well as incredible patience and enthusiasm have proven invaluable throughout the course of this work, particularly during the drafting of this manuscript.

I would also like to acknowledge the J. Golden and S. Golden labs, as well as Dr. Susan Golden, for providing scientific guidance, feedback on presentations and writing, and support over the years. In particular, I would like to thank Dr. Rodrigo Mella-Herrera for help with RNA purification, Dr. Arnaud Taton for bioinformatics help, and Dr. David Johnson for help with RNA-seq and for being a great officemate. Furthermore, I would like to acknowledge my truly dedicated undergraduate researcher, Neil Raina, whose work was vital both to this thesis and to lab moral over the years, as well as Natalie Ortiz, whose recent support was essential as I finished up my work. I would also like to acknowledge great post docs in the Golden Labs, in particular Dr. Susan Cohen and Dr. Ryan Simkovsky, for constant scientific and personal support.

In addition, I would like to acknowledge my graduate committee, Dr. Eric Allen for help with experimental design, software, and use of his facilities,

as well as Dr. Steve Mayfield, Dr. Brian Palenik, and Dr. Steve Briggs for their experimental guidance.

Personally, I would like to acknowledge my family and friends for their unwavering support during my graduate career. I would like to thank my family of academics, Dr. John Flaherty, Pam Flaherty, M.S.W., Dr. Heather Flaherty, Dr. Brett Pearson, and Jon Flaherty M.B.A., as well as the young Nora Flaherty, for always understanding this path. From my friends, I would particularly like to acknowledge the men and women who have inspired balance and health during these last five years: Dr. Amanda Herman and Marcy Erb for the hundreds of miles of inspiration; Dr. Shannon Seidel for being an incredibly supportive shackmate; Elizabeth Clark, Michaelanne Munoz, Rachel Patterson, Dr. Jesse Vargas, Pagkapol Pongsawakul, and Emily Grossman for their encouragement and company over the years; Dr. Julie Bordowitz for starting a new group of runners and for many miles to come; Alexia Cervantes for helping me create a strong mind and body; Megan Becker for support and roommateship; and Drs. Dawn Adin and John Buchner for sharing their warmth and family throughout the years. Finally, I would like to thank Rachael Mills, Esq., first for becoming a pseudo-expert on cyanobacterial genetics, but most importantly for making my life and home encouraging and full of love as I finished this work.

This work was supported by National Science Foundation Grant 0925126 (JWG). BLF was supported by National Institutes of Health Training

Grant T32GM007240 as well as the Tyler Dylan Fund at the University of California at San Diego.

Chapters 2 and 3 are, in part, a reprint of the material as it appears in BMC Genomics, *2011* B. Flaherty, F. Van Nieuwerburgh, S.R. Head, J.W. Golden. The dissertation author was the primary author of this publication.

# VITA

## EDUCATION

**Ph.D.** **UCSD**, 2012
Biology

**B.A.** **Grinnell College**, 2007
Biological Chemistry

## TEACHING

UCSD Undergraduate Labs, 2008-2011
UCSD Lab Mentor, 2008-present
Johns Hopkins Center for Talented Youth, 2007
Grinnell College Lab Assistant, 2006

## RESEARCH

**UCSD Division of Biological Sciences Ph.D. Program, 2007-present**
Supervisor: Dr. James Golden, Ph.D.
**New York State Department of Health, 2006**
Supervisor: Dr. Erasmus Schneider, Ph.D.
**Grinnell College, 2005-2007**
Supervisor: Dr. Leslie Gregg-Jolly, Ph.D.

## OUTREACH AND SCIENCE WRITING

San Diego Center for Algal Biotechnology, 2011-present
Small Things Considered Blog, 2011
Biology Graduate Peer Mentoring Program, 2009-2011
BioEASI, 2009 (Bio Education through Art and Science Initiative)
San Diego Science Festival, 2009
UCSD Graduate Recruitment, 2007-2009

## FELLOWSHIPS AND AWARDS

ARCS Fellowship, 2011-2012
Tyler Dylan Award, 2010-2011
NSF Graduate Research Fellowship Honorable Mention. 2009
UCSD Cellular and Molecular Genetics Fellow, 2008-2011
Grinnell College Trustee Honor Scholar, 2003-2007

## PUBLICATION

**Flaherty, B., et al**. (2011). "Directional RNA deep sequencing sheds new light on the transcriptional response of Anabaena sp. strain PCC 7120 to combined-nitrogen deprivation." BMC Genomics **12**(1): 332. **Highly Accessed**

# ABSTRACT OF THE DISSERTATION

Transcriptional Regulation of Heterocyst Development in *Anabaena* sp. strain PCC 7120

by

Britt Flaherty

Doctor of Philosophy in Biology

University of California, San Diego, 2012

Professor James Golden, Chair

Fixed nitrogen is a limited resource for growth in the environment and the fixation of atmospheric nitrogen is vital to nutrient cycling and growth. Cyanobacteria are a group of photosynthetic bacteria that evolved 4.5 billion years ago to harvest sunlight as energy. Cyanobacteria have since evolved a wide array of metabolic capabilities over time, including the ability to fix atmospheric nitrogen. *Anabaena* sp. strain PCC 7120, hereafter *Anabaena*, is a species of cyanobacteria that fixes atmospheric $N_2$ into ammonia by forming specialized nitrogen-fixing cells called heterocysts. Heterocysts form only in

the absence of a source of fixed nitrogen and are evenly spaced along a filament of *Anabaena* cells. We studied the gene expression networks that regulate heterocyst development through deep sequencing, employing both RNA-seq on a nitrogen-deprived culture as well as ChIP-seq on key transcription factors involved in heterocyst development. Deep sequencing gave us a global view of gene expression in response to nitrogen deprivation in *Anabaena.* Our RNA-seq work identified new genes involved in heterocyst development, mapped operon structure and transcript length, and discovered abundant antisense transcription in the genome. In particular, we identified antisense transcription in the coding region of the gene *nblA*, which codes for a small peptide that triggers the proteolysis of the photosynthetic machinery in response to nutrient stress. Furthermore, we used ChIP-seq to identify the regulon of two transcription factors, HetR and DevH, in response to nitrogen deprivation. Our work on HetR, a transcription factor with known roles in regulating heterocyst development, identified many new HetR targets, including genes involved in HetR's role during nitrogen deprivation and during vegetative cell growth. Our work on DevH, a transcription factor required for forming the heterocyst-specific cell wall, also identified new DevH targets, including many genes involved in cell wall formation and transcriptional regulators. This work adds to our understanding of transcriptional networks that regulate heterocyst development in *Anabaena*. Furthermore, our study

provides insight into gene structure and transcriptional regulation in cyanobacteria as a whole

# Chapter 1. Introduction

## 1.1 CYANOBACTERIA

### 1.1.1 A Brief History

The Earth's early atmosphere supported mostly anaerobic single-celled organisms in the oceans and on land. However, 2.5 billion years ago, cyanobacteria evolved the ability to harvest sunlight for energy in an oxygenic process we now know as photosynthesis (51). Over millions of years, these organisms thrived in salt and fresh water, producing enough oxygen to rust iron in our oceans and, eventually, oxygenate the earth's atmosphere. Cyanobacteria shaped the earth into an environment that supported aerobic microbes, thus laying the groundwork for evolution of plants and animals. During that time, cyanobacteria colonized niche environments and evolved a diverse array of metabolic processes, including the ability to harvest atmospheric $N_2$ and fix this nitrogen it ammonia for incorporation into amino acids.

Cyanobacteria still contribute significantly to carbon fixation on earth, accounting for 20-30% of the planet’s photosynthetic activity, and one species of cyanobacteria, *Prochlorococcus*, is thought to make up as much as 50% of the carbon fixation that occurs in Earth’s oceans (117, 123). Despite their ability to photosynthesize, cyanobacteria are not closely related to plants. In fact, plants evolved the photosynthetic chloroplast via a symbiotic relationship with an early species of cyanobacteria over 600 million years ago (22, 26, 37,

80). However, in their own right, cyanobacteria continue to shape our planet through large contributions to global carbon and nitrogen cycling. Cyanobacteria can fix nitrogen both alone and in symbiotic relationships, such as with fungi or coral, and some nitrogen fixing species are even used today as a nitrogen source for rice paddies (64, 70, 110, 120).

### 1.1.2 Cyanobacterial Genetics and their Role in Green Energy

Cyanobacteria are still abundant on our planet's surface today, and they are the focus of molecular biologists and bacterial geneticists as both model organisms and in applied metabolic engineering. Recent work has focused on using cyanobacteria as producers for biofuels and renewable chemicals (78). Cyanobacteria have great potential as producers of renewable fuels due to their efficient harvesting of solar energy, simple cell structure (eliminating the pre-treatment necessary to harvest cellulosic biofuel from plants), low water usage, and diversity both in nutritional requirements and in metabolites produced (24). This, combined with decades of research on how to manipulate their genomes, makes cyanobacteria prime targets for green energy and chemical production.

A few studies have proven cyanobacteria as a platform for genetic engineering and as a future provider of biofuels. Ethanol, isoprene, glucose, and fructose have all been produced in cyanobacteria as preliminary examples of genetic engineering to produce industrial products (28).

Furthermore, yield estimates place algal biofuels as a whole among the least energy and space intensive candidates for producing biofuels (45). With an increased understanding of gene networks in cyanobacteria, gene regulation, and genetic diversity, we will be able to further engineer synthetic networks and select for better strains of cyanobacteria for commercial use.

This work will focus on one species of cyanobacteria, *Anabaena* sp. strain PCC 7120 (hereafter *Anabaena*) and its ability to fix atmospheric nitrogen into ammonia in the absence of a source of combined nitrogen. Fixed nitrogen is a limiting nutrient for crop and bacterial growth and increasing our understanding of the genes involved in nitrogen fixation may one day allow us to engineer better nitrogen-fixing organisms. In addition, this work delves into the regulation of transcription for *Anabaena*, thus increasing our understanding of how cyanobacterial gene regulation works as a whole.

## 1.2 NITROGEN FIXATION

### 1.2.1 Nitrogen Fixation in *Anabaena*

Fixed nitrogen is required for biological growth, yet nitrogen is often a limited resource in the environment. Therefore, many species have evolved ways to fix abundant atmospheric nitrogen ($N_2$ gas) into ammonia for metabolic use. *Anabaena* is a multicellular filamentous cyanobacterium that fixes atmospheric $N_2$ in nitrogen-limiting conditions. Since the nitrogenase enzyme, which breaks the $N_2$ bond, is oxygen sensitive, *Anabaena* must form

specialized microoxic cells called heterocysts for nitrogen fixation (Figure 1.1). Heterocysts develop in response to nitrogen deprivation in a semiregular pattern along the filament, about once ever 10-15 cells. As the filament continues to grow, heterocysts will continue to form as long as the environment remains low in nitrogen.

To express active nitrogenase, heterocysts must make their cellular environment microoxic. First, they deposit an oxygen impermeable glycolipid layer and an additional thick polysaccharide layer outside of their cell wall. These extracellular layers keep out atmospheric oxygen (92). Heterocysts reorganize their thylakoid membranes into a honeycomb shape that is almost exclusively involved in increased respiration to scavenge free $O_2$, particularly at the cell poles (88). Heterocysts lack photosystem II activity (which normally produces $O_2$), but maintain ATP production by photosystem I (65). Heterocysts fix atmospheric $N_2$ and pass fixed nitrogen in the form of amino acids along the filament to neighboring vegetative cells (111). In return, vegetative cells continue to photosynthesize and pass the products of carbon fixation to heterocysts, thus creating a mutual dependence between these two cell types (Figure 1. 2).

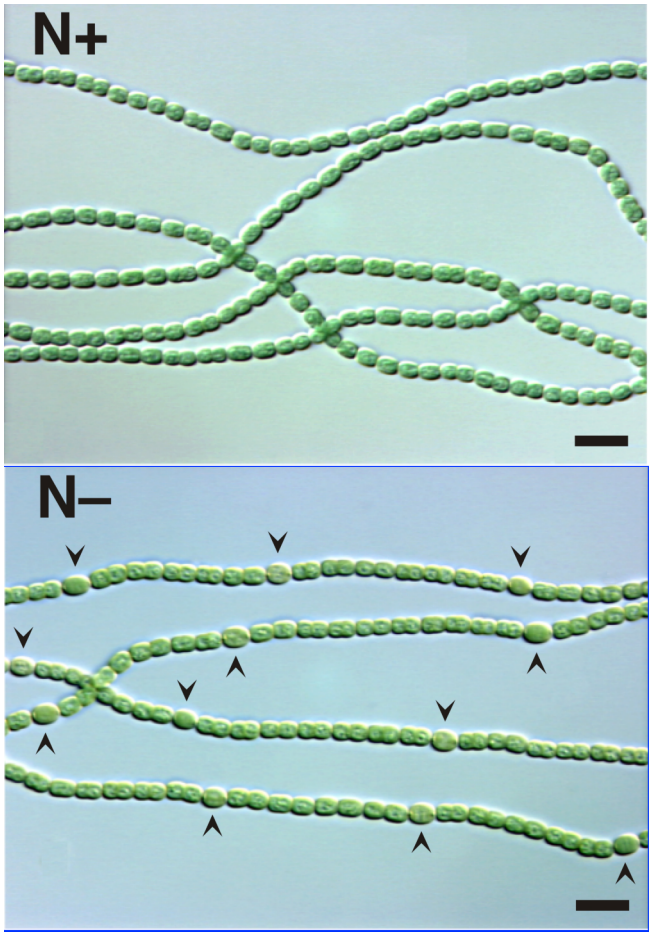



**Figure 1.1: Morphology of Differentiating Cells**
*Anabaena* differentiates into two distinct cell types upon removal of combined nitrogen from the medium. In the top panel, filaments are grown in the presence of nitrate (N+) and form only vegetative cells. In the bottom panel, cells are grown without nitrate (N-) and must fix atmospheric $N_2$ in heterocysts, which have been marked in this image by Λ. Heterocysts have thick polysaccharide and glycolipid layers that make them appear rounder and larger than their neighboring vegetative cells. Bar is 10 μm.

*Images courtesy of JW Golden*

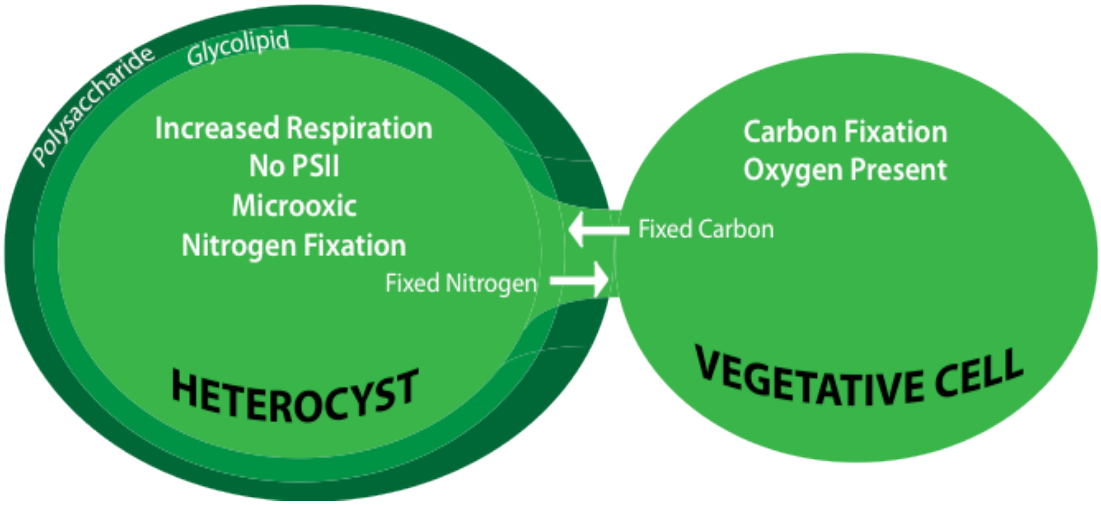



**Figure 1.2: *Anabaena* heterocysts exchange fixed nitrogen with vegetative cells.**

### 1.2.2 A Transcriptional Cascade Regulates Heterocyst Development

Combined nitrogen deprivation is sensed by the accumulation of 2-oxoglutarate (2-OG), a byproduct of the Krebs pathway in cyanobacteria. 2-OG is normally used as the backbone for nitrogen assimilation (69, 132, 137). When 2-OG accumulates, it binds directly to the transcription factor NtcA, enhancing NtcA's DNA binding activity (118, 137). In turn, NtcA activates the **N**itrogen **R**esponse **R**egulator **A** (NrrA) and NrrA activates HetR, a key regulator of heterocyst development. In addition, a number of feedback loops allow these transcription factors to regulate themselves and each other in response to nitrogen deprivation (Figure 1.3). For instance, NtcA binds its own promoter and HetR binds both its own promoter and that of NtcA as a positive feedback loop.

After HetR activation, a series of downstream transcriptional cascades results in expression of *patS*, which encodes a peptide involved in pattern formation, the *hgl* and *hep* genes responsible for deposition of the exopolysaccharide and glycolipid layers, genes responsible for shutting down photosynthesis, and finally the nitrogen fixation or *nif* genes, which form the active nitrogenase complex. The mechanisms that regulate this cascade of transcription have been parsed out through molecular biology and genetic techniques. DNA microarrays have been used to analyze global transcription during heterocyst development (32, 33, 60, 104). However, both gene-specific studies and DNA microarrays have significant drawbacks: Gene specific

studies have a high false negative rate and often miss subtle or redundant effects caused by overexpression or knock out specific gene, while microarrays performed on *Anabaena* thus far have looked at specific genes or clusters of genes, resulting in low resolution or targeted results that do not encompass the entire genome. In this work, we applied whole transcriptome sequencing to give a less biased view of the transcriptional networks that regulate heterocyst development.

### 1.2.3 HetR is a Key Regulator of Heterocyst Development

Heterocyst differentiation is regulated by a homodimeric transcription factor, HetR. HetR is expressed at a basal level in all cells but increases in expression in heterocysts by about 6 hours after nitrogen deprivation (Figure 3) (13). HetR is required for nitrogen fixation and diazotrophic growth and overexpression of HetR leads to multiple contiguous heterocysts (an MCH phenotype) (10, 56). Because HetR overexpression appears to override other cellular queues governing heterocyst development and spacing, HetR is a key regulator of heterocyst development, however it is becoming clear that additional regulators such as HetP and HetZ play important regulatory roles (53, 134).

*in vitro w*ork shows that HetR is capable of binding DNA fragments upstream of a handful of genes upregulated in heterocysts, including *hetZ, hetP, hepA, patS,* and *hetR* itself (27, 54, 56). This work also suggests that

HetR binds a small inverted repeat sequence (GGGTCTAgCCCagCA) upstream of a few of these genes; however, HetR has been shown to bind promoters that do not contain this sequence, such as the promoter of *patS*, a peptide involved in pattern formation (27). Therefore, our understanding of the HetR target site and its DNA binding is incomplete.

The crystal structure for HetR has been solved, and shows that the protein consists of a DNA binding pocket, a flap which might enhance the DNA binding by extending the binding site, and a hood that may work to recruit other proteins to the binding site (63). HetR's structure is unlike that of any previously solved protein, suggesting that this novel transcription factor may function in new or intriguing ways. In addition, the prevalence of two protein binding domains in addition to HetR's DNA binding domain suggests that HetR's activity is modified and regulated by other proteins in the cell.

HetR is conserved in both heterocystous and nonheterocystous cyanobacteria, always in tandem with the peptide PatS (133). However, little work has been done to identify HetR's role outside of heterocyst development. It is known that PatS is a negative regulator of HetR and is important for even spacing of heterocysts. Since HetR and PatS are only conserved in filamentous cyanobacteria, it is possible that these two proteins govern cell to cell signaling or patterning during vegetative cell growth in addition to their role in governing heterocyst spacing.

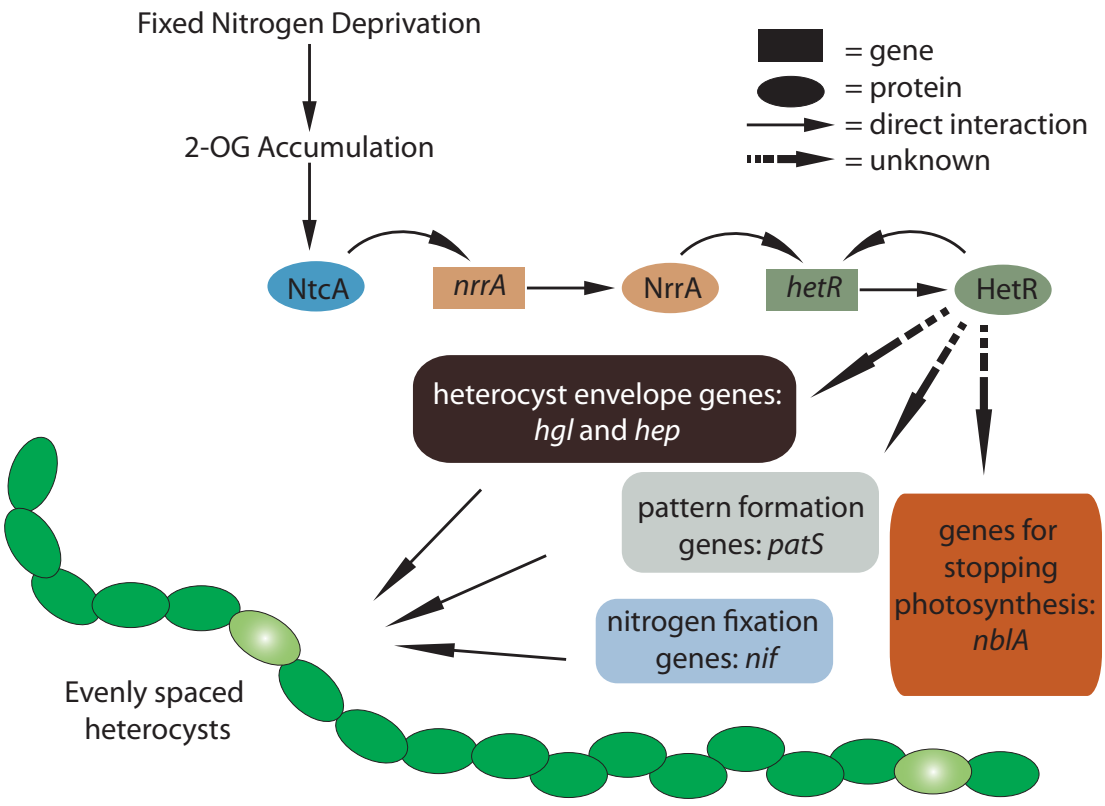



**Figure 1.3: The transcriptional cascade that regulates heterocyst development.**
The cascade starts with the accumulation of 2-OG, the backbone of nitrogen assimilation. 2-OG enhances the DNA-binding activity of NtcA, which activates expression of NrrA and, in turn, HetR. HetR activation results in a cascade of transcription of genes that regulate heterocyst morphological and physiological changes, patterning along the filament, and nitrogen fixation.

### 1.2.4 Pattern Formation: PatS, HetN, PatA, and PatN

HetR's DNA-binding is inhibited by a small pentapeptide, RGSGR which is present in the protein PatS (56). Work from our lab supports a model in which PatS is a diffusible peptide originating in proheterocysts (cells that will become heterocysts after prolonged nitrogen deprivation) (129). In addition, a C-terminal 6 amino acid motif in PatS alone (ERGSGR) blocks HetR's DNA-binding site *in vitro* (38, 56). It remains unclear whether the PatS protein is truncated in cells or acts in its full-length form.

The protein HetN, which is required for maintenance of the heterocyst pattern, is turned on after *patS* transcription decreases and contains the RGSGR motif {Li, 2002 #178}. Taken together, this data suggests a model whereby HetR regulates heterocyst differentiation by binding the promoters of key genes involved in cellular morphogenesis and signaling. HetR's DNA binding activity is blocked by PatS and, eventually, HetN. However, HetR produced in heterocysts is somehow immune to inhibition by the RGSGR-containing protein, which diffuses away from proheterocysts to create and maintain a pattern of evenly spaced heterocysts with 10-15 vegetative cells in between.

Recent work suggests that pattern formation and spacing is also regulated by the protein PatN. When *patN* is deleted, heterocysts are spaced closer together, although evenly and in a pattern (102). A *patN-gfp* fusion showed that PatN localizes to one side of the cell prior to cell division and,

therefore, one daughter cell is transiently devoid of PatN. Eventually, both cells accumulate PatN; however, this transient lack of PatN may poise certain cells for heterocyst development, should nitrogen deprivation occur.

PatN appears to negatively regulate expression of *patA*, another gene involved in pattern formation. PatA is thought to couple heterocyst patterning with cell division and to be regulated by HetR (72, 131). *patA* deletion results in heterocyst formation only at the ends of filaments, suggesting that *patA* is a positive regulator of heterocyst formation. *patA* is turned off in a Δ*hetR* mutant and is upregulated in a Δ*patA* mutant. PatA-GFP localizes to the septum of *Anabaena* cells and overexpression of *patA* results in aberrant vegetative cell morphology, suggesting that PatA plays a role in governing cell division. It is possible that PatN and PatA regulate one another to govern prepatterning, programming certain cells to respond to nitrogen deprivation by forming heterocysts.

### 1.2.5 Degradation of the Phycobilisomes

In response to nitrogen deprivation, heterocysts degrade the light harvesting complex proteins, called phycobilisomes. Phycobilisomes, in conjunction with chlorophyll α, are responsible for the cyanobacteria's blue-green color. The phycobilisome complex transfers electrons to photosystems I and II and makes up as much as 50% of the cyanobacterial cellular protein mass (18, 43). Therefore, phycobilisomes are often degraded as a nutrient

stress response, slowing metabolism and light absorption while providing amino acids for biosynthesis.

Within 6 hours after the removal of nitrogen from the medium, *Anabaena* expresses a small peptide, NblA, responsible for degradation of the phycobilisomes. NblA is thought to trigger degradation of phycobilisomes by recruiting the Clp protease directly to the phycobiliproteins (59). In a ∆*nblA* background, phycobilisomes are not degraded after the removal of nitrogen, although heterocysts still fix nitrogen (7). This mechanism is conserved in many species of cyanobacteria, including those that do not fix nitrogen, as NblA is expressed in response to nutrient starvation in general, not specifically during heterocyst development. Interestingly, NblA expression is abundant in all cells after fixed nitrogen deprivation, but its expression is transient in vegetative cells, which go on to re-build their phycobilisomes (21). Heterocysts, on the other hand, maintain high levels of NblA and no longer form phycobilisomes, as seen by a lack of autofluorescence.

### 1.2.6 The Role of DevH in *hgl* gene expression

Immediately preceding expression of the nitrogen fixation (*nif*) genes, the heterocyst specific glycolipid (*hgl*) and heterocyst exopolysaccharide (*hep*) genes are expressed, resulting in a thick cell envelope around heterocysts that limits the entry of oxygen (5). Expression of the *hglE1* and *hglE2* genes requires the transcription factor DevH, which is also turned on

late during heterocyst development (49, 100). DevH is a helix-turn-helix DNA binding protein in the cyclic AMP receptor protein family. Its structure is similar to that of NtcA, the protein that triggers heterocyst development upon sensing 2-OG accumulation.

Expression of *devH* requires HetR and NtcA and increases dramatically by 12 and 21 hours after the removal of combined nitrogen from an *Anabaena* culture, supporting DevH's role late during heterocyst development. A Δ*devH* strain is only capable of diazotrophic growth and nitrogen fixation under anoxic conditions and does not form the heterocyst specific-glycolipid layer, confirming the importance of DevH in creating the heterocyst-specific cell envelope (100). Gene-specific analysis of DevH DNA binding (such as mobility shifts or ChIP) has not been published to date and it is still unclear whether DevH binds directly to the *hgl* gene promoters or affects *hgl* gene expression through a secondary mechanism.

## 1.3. BACTERIAL TRANSCRIPTOMES

### 1.3.1 Deep Sequencing

Deep sequencing technologies, such as the ABI Solid system, Roche 454 Sequencing, Life Technologies' Ion Torrent, and Illumina GAII sequencing, have revolutionized how genome and transcriptome data are collected (19, 82, 115). Unlike microarrays, they do not rely on our current understanding of gene boundaries and can accurately measure extremely low

or high levels of transcription. In this work, we will focus on the Illumina GAII sequencing technology as a window into bacterial transcription. Illumina sequencing takes total-RNA samples, sheared for sequencing reads of either 40bp or 100bp in size, and converts them to single stranded cDNAs with both sample-specific DNA adapters and short adapters for annealing to the sequencing cell. After annealing the cDNA library to the cell (with up to 200 million cDNA strands per cell on a 7 cell chip), the library is sequenced via the addition of fluorescently labeled nucleotides in a stepwise process. A camera reads the entire chip after each nucleotide addition, determining the location and abundance of fluorescence for each added nucleotide. This process provides less biased data for the entire transcriptome, including small RNAs, antisense RNAs, and novel genes.

After sequencing, RNA-seq reads are aligned to the sequenced genome using either free or commercially available algorithms such as Bowtie (67) or SOAP (71). Reads can be assembled into large single transcript contigs and compared to the sequenced genome, which allows automated discovery of new transcripts using the algorithms Cufflinks, Cuffdiff, and Cuffcompare (103). The abundance of various transcripts is measured as **r**eads **p**er **K**b gene length per **m**illion reads in the sample set, or RPKM. This normalized value controls for both the varying size of each defined gene and the size of a given sample set, allowing comparison of gene expression for each gene in the genome.

### 1.3.2 Transcriptional Studies in Cyanobacteria

Deep sequencing has been applied to study gene structure and expression with much success in many bacterial transcriptomes. Early RNA-seq work showed that, while this new method confirmed earlier microarray data on gene expression, it is also highly effective at identifying new genes, sRNAs, and more subtle changes in gene expression in response to environmental stress (97). In the world of clinical microbiology, where samples are often small, RNA-seq studies have been used when microarrays would require too much input RNA. One such study determined the difference between genetically identical clinical and environmental isolates of *Burkholderia cenocepacia*, a pathogen involved in cystic fibrosis (128). Recently, strand specific RNA-seq studies have discovered antisense transcription and have identified new transcripts in organisms such as *Salmonella enterica serovar typhi (98).*

One key difference between RNA-seq in bacteria and RNA-seq in eukaryotic cells is the lack of polyadenylated tails on mRNAs. Eukaryotic RNA-seq studies often enrich for messages by selecting for only polyadenylated RNA molecules, while prokaryotic RNA-seq studies must either sequence the entire RNA pool, which is between 60 and 95% ribosomal and transfer RNA, or enrich for mRNAs with chemical methods. Liu et al showed that selection for small RNAs gave them highly specific

transcriptome profile for sRNAs while excluding the tRNA and 5S RNA background (75). However, work in our lab has shown that known rRNA depletion methods for cyanobacteria remove less than half or the rRNA while preferentially depleting known coding messages; similar conclusions have been made by others (77) Therefore, in this work, we take advantage of the deep coverage offered by Illumina's sequencing platform and have sequenced the entire RNA population, including as much as 95% ribosomal and transfer RNA.

### 1.3.3 Bacterial transcriptomes encode abundant antisense transcription

Noncoding transcription in bacteria and its molecular significance have been studied for decades on a gene-by-gene basis. Antisense RNAs, both short and long, are thought to regulate transcription of gene targets (which they often overlap) by interfering with transcription or translation (44). Regulatory RNAs that are not antisense can bind proteins, anneal to other RNAs through base pairing and even act as a cellular pathogen response, as is the case with CRISPR RNAs (for a review see (124).

Recent developments in high-throughput sequencing technologies have shown that noncoding RNAs, particularly antisense RNAs, are more abundant in bacterial transcriptomes than previously thought. Dornenburg et al. saw as many as 1,000 antisense transcripts in *E. coli*, while as much as 65% of the *Synechocystis* PCC 6803 transcriptome is antisense (25, 83). It is

unclear how much of this antisense transcription results in regulation of target genes and how much is simply not selected against during evolution. Nonetheless, understanding the location and abundance of antisense transcripts is vital to genetic engineering and transcriptome studies in cyanobacteria.

While the molecular mechanisms of many of the noncoding transcripts have been studied in *E. coli* and other model bacteria, only a handful of well-studied examples exist in cyanobacteria. For example, in *Synechocystis*, the antisense RNA IsiR is thought to base pair with *isiA*, an iron response protein, to decrease translation of this transcript (29). In *Anabaena*, levels of the protein FurA, another iron stress protein, are downregulated by an antisense transcript running through the coding region of *furA* (50). In addition, recent work has shows that the **n**itrogen **s**tress-**i**nduced **R**NA **1** (NsiR1), is developmentally regulated in response to nitrogen deprivation – however the role of this RNA is not yet understood (57). Nonetheless, all published directional RNA-seq datasets in cyanobacteria discovered abundant antisense transcription, suggesting that antisense transcription is common to these clades and highlighting the need for more molecular and high throughput work to identify the role of these transcripts.

### 1.3.4 ChIP in Bacteria

Chromatin immunoprecipitation, or ChIP, is a method of analyzing the DNA targets of a transcription factor or DNA-binding protein. The protein of interest is chemically fixed to its target DNA and then immunoprecipitated, allowing analysis of DNA bound to the protein. (Like others, we use "ChIP" even if the protein-DNA complex is collected by methods other than immunoprecipitation.) With the advent of deep sequencing, the immunoprecipitated DNA can then be sequenced as a whole to allow analysis of the entire regulon of a particular transcription factor in a given sample.

While this technology has been used extensively in eukaryotic organisms (9, 16, 17), there are fewer examples in prokaryotes and almost none in cyanobacteria. In *E. coli*, the RNA polymerase, nucleoid associated proteins, as well as transcription factors such as RutR, MelB, IHF, HU, and LexA are some of the proteins that have been successfully “ChIPped” using both deep sequencing and microarrays for downstream analysis (52, 99, 107, 122). In general, these studies showed that, in addition to the expected DNA binding profile, chromatin immunoprecipitation reveals unconventional transcription factor binding sites that may not result in gene regulation, such as binding sites within coding regions, between convergent genes, or far away from transcriptional start sites. However, confirming ChIP data for a specific transcription factor with other studies of gene expression, such as RNA-seq, promoter *gfp* fusions, and genetic studies tended to give the most biologically reproducible results (135).

The analysis of ChIP data after deep sequencing is usually done with a "peak finding" algorithm, which search for areas with unusually high enrichment in coverage, suggesting the target transcription factor associated with that DNA region. One commonly used algorithm, MACS or Model-based analysis of ChIP-Seq, analyzes aligned ChIP-seq reads while correcting for sequencing biases (136). First, MACS searches for areas of high enrichment along the genome, excluding regions that are present in both the experiment sample and control. Then, because sequencing along the genome is extremely biased with respect to local chromatin structure and GC content, MACS calculates a Poisson distribution around each peak to control for local biases. Finally, MACS shifts any peaks that pass these controls to more accurately predict the binding site. Shifts are based on both the size of sheared DNA and the expected "twin peak" distribution of reads around a putative binding site. Because protein bound to DNA protects that DNA from shearing, real ChIP peaks look like two high coverage peaks surrounding a low or no coverage area centered over the transcription factor binding site.

The analysis of ChIP peaks must take into account the sequencing biases introduced by varying GC content, transcriptional activity of a region of DNA, and difficulty in sequencing highly structured DNA and DNA/protein complexes (74). Because of this, ChIP datasets work best when a negative control is used, such as ChIP DNA from a sample where the protein does not contain the epitope tag (or, if a tagged protein is not available, then a sample

where antibody is not added). This allows for a normalization control that is equally susceptible to sequencing bias, thus eliminating false peaks in areas that are simply more enriched for technical, not biological reasons. Finally, twin peak appearance and local bias should be confirmed by eye after algorithm analysis to reduce the false positive rate (15, 62, 93).

In recent work, two labs have successfully used ChIP to look at transcriptomics in cyanobacteria. ChIP of RNA polymerase was coupled with RNA-seq and microarrays to give a global view of the transcriptome of *Synechococcus elongatus* PCC 7942, and unpublished work in this same lab elucidated the regulon of a key transcription factor involved in circadian rhythms, RpaA (119). Also in *Synechococcus elongatus* PCC 7942, the regulon of RpaB in response to high light stress was analyzed with ChIP, giving a global view of high-light induced genes. However, chromatin immunoprecipitation has not been published to date in other cyanobacteria, possibly due to the expense involved in developing robust ChIP protocols and analysis tools (47).

## 1.4 SUMMARY OF THIS WORK

Thus far, genes involved in heterocyst development have been identified through genetic screens and microarray analysis. In this work, we studied the *Anabaena* transcriptome through deep sequencing, specifically RNA-seq and ChIP-seq. First, we mapped the *Anabaena* transcriptome in

response to nitrogen deprivation using directional RNA deep sequencing. This work identified new genes involved in the response to nitrogen deprivation, mapped 5' and 3' UTRs and operons across the genome, and identified abundant antisense transcription in *Anabaena*. We then followed up by studying the antisense transcription of a key transcript in the coding region of the phycobilisome degradation protein NblA. Finally, we coupled our RNA-seq studies with ChIP-seq to map the regulon of two key regulators of heterocyst development, HetR and DevH, in response to nitrogen deprivation. Taken together, these datasets not only contribute to our overall understanding of cyanobacterial transcriptomes, but also act as a starting point for future genetic engineering and molecular biology studies in *Anabaena*.

# Chapter 2: Gene Expression after Nitrogen Deprivation

## 2.1 INTRODUCTION

Cyanobacteria are photosynthetic prokaryotes that have evolved a wide array of metabolic capabilities (51). Because of their high photosynthetic efficiency, variety of metabolic pathways, and genetic manipulability, they are a potential source of "green" chemicals and fuels (24, 28). Some cyanobacteria reduce atmospheric nitrogen to ammonia to support growth in nitrogen-deficient conditions (12). Because nitrogen is often a limiting resource for growth, nitrogen-fixing strains have a competitive edge in some environments. Understanding the response to nitrogen deprivation, nitrogen fixation, and diazotrophic growth in cyanobacteria will shed light on basic mechanisms of bacterial genetic regulation and physiology. In addition, it may help to develop better strains of cyanobacteria for the production of renewable chemicals and biofuels.

The cyanobacterium *Anabaena* (*Nostoc*) sp. strain PCC 7120 grows as long filaments of photosynthetic vegetative cells in the presence of a source of combined nitrogen. In the absence of combined nitrogen, up to 10% of the cells terminally differentiate into nitrogen-fixing heterocysts. Heterocysts provide a microoxic environment for the expression of the oxygen-sensitive nitrogenase enzyme (41, 65). Single heterocysts are spaced about every 10-15 cells along filaments and they supply fixed nitrogen, probably in the form of

amino acids, to neighboring vegetative cells (41). Vegetative cells provide heterocysts with products of carbon fixation, probably as sucrose (76, 116), thus creating a multicellular organism with two mutually dependent cell types. Heterocyst development involves the response of vegetative cells to nitrogen deprivation, the formation and maintenance of the pattern of the two cell types, differentiation of heterocysts from vegetative cells, and the adaptations made by vegetative cells to adjust to diazotrophic growth.

The differentiation of a vegetative cell into a heterocyst involves substantial changes in cell morphology and physiology (41, 65). Heterocysts deposit glycolipid and polysaccharide layers outside of their cell wall to limit the entry of atmospheric oxygen (5, 55, 92). They lack photosystem II activity, which normally produces $O_2$, and increase respiration to consume $O_2$ that enters the cell. Heterocyst differentiation requires dramatic changes in transcription and some of the key components of this regulation are known. Nitrogen limitation is sensed by accumulation of 2-oxoglutarate (2-OG), the backbone for nitrogen assimilation. 2-OG enhances the DNA-binding activity of the transcription factor *ntcA* (137), which regulates expression of the response regulator *nrrA*, which is partially responsible for upregulation of *hetR* (30, 31). HetR, a key regulator of heterocyst development, regulates the expression of many genes, including the glycolipid genes (*hgl*), exopolysaccharide genes (*hep*), and the *patS* gene, which encodes a peptide involved in heterocyst pattern formation (132).

Factors other than those described above are known to be involved in heterocyst development and have been identified through DNA microarrays and genetic screens (34, 36, 54, 73, 81, 91). While these methods are powerful, microarrays often overlook unannotated regions of the genome and antisense or noncoding transcripts and genetic screens often miss subtle (or lethal) effects of the mutation of a gene. In addition, DNA microarrays lack sensitivity and do not provide information on UTR length or operon structure. Therefore, we have employed directional RNA-seq to analyze the transcriptome of *Anabaena* filaments during nitrogen deprivation to identify and map all transcripts during heterocyst development (39, 95, 97, 109, 112).

Our RNA-seq data provide information on the UTR lengths of each mRNA transcript and on the changes in expression of all transcripts whether or not they carry an annotation. The data show long 5' UTRs for many genes, likely with multiple transcriptional start or processing sites. In addition, our study identifies antisense transcription in the coding region or UTR of many genes known to be involved in heterocyst development, as will be outlined in Chapter 3. Finally, we detected new genes that are significantly upregulated in response to nitrogen deprivation.

## 2.2 MATERIALS AND METHODS

### 2.2.1 Preparation of RNA samples

For deep sequencing, total RNA was prepared from *Anabaena* (*Nostoc*) sp. strain PCC 7120 cultures grown in 100 ml of liquid medium in 250-ml flasks with cotton plugs as previously described with slight modifications (46). Briefly, 100-ml liquid cultures were grown to an $OD_{750}$ of 0.5 in BG-11($NH_4$) medium, which lacked sodium nitrate and contained 2.5 mM ammonium chloride and 5 mM MOPS (pH 8.0). For the 0 hour sample, cells were collected before deprivation of combined nitrogen. For nitrogen-deprived samples, cells were collected by centrifugation and washed 3 times in BG-$11_0$, which lacks sodium nitrate, resuspended in BG-$11_0$ to an $OD_{750}$ of approximately 0.05, and incubated for 6, 12, or 21 hours. These times were chosen because at 6 hours cells are beginning to establish a pattern of differentiating cells, at 12 hours proheterocysts are committed to becoming heterocysts (130), and at 21 hours nitrogen-fixing heterocysts are fully differentiated. For each sample, cells were rapidly cooled and collected by pouring 50 ml of culture over 100 g crushed ice and centrifugation at 4,000 x g for 10 minutes at 4°C. Cell pellets were transferred to a 2-ml tube and collected by centrifugation at 11,500 × *g* for 2 minutes at 4°C. Supernatant was removed and cell pellets were flash frozen in liquid nitrogen for storage at -80°C. RNA was isolated from filaments with the Ambion RiboPure RNA isolation kit according to the manufacturer's protocol and total RNA was used for deep sequencing.

**2.2.2 Preparation of a cDNA library**

Ribosomal RNA was not removed from total RNA to avoid any depletion of coding transcripts. The presence of ribosomal and transfer RNAs in our sequencing sample did result in a decreased yield of transcript information from coding RNAs (only ~10% of the sample at each time point was non-rRNA and non-tRNA transcripts); however, this protocol avoided the known biases introduced by current rRNA depletion methods (48). Furthermore, samples were not multiplexed, which allows the sequencing of multiple samples in the same lane, to avoid biases that result from using different adapters for different samples. A directional RNA-sequencing library was prepared from 1 µg of total RNA from each sample (0, 6, 12, and 21 h). The RNA samples were purified using a Qiagen RNeasy MinElute Cleanup Kit, fragmented for 180 seconds using a Covaris S2 sonicator set at 10% duty cycle, 5 Intensity, 200 cycles/burst in 120 µl of 1 mM Tris-EDTA pH 8.0 and purified again with the cleanup kit. The fragmented RNA (100 ng) was dephosphorylated using Antarctic phosphatase (2 units, 37 °C, 30 min), 5' phosphorylated using T4 polynucleotide kinase (20 units, 37 °C, 60 min), and purified with the cleanup kit. cDNA library preparation steps were performed as described in the Illumina Small RNA v1.5 Sample Preparation Guide except that during size selection on 4% agarose gels fragments of approximately 250 bp were obtained, which are suitable for up to 100 base sequencing reads. We saw no evidence of genomic DNA contamination of

the RNA samples prior to cDNA synthesis, as there were many regions with no read coverage across the genome.

### 2.2.3 Illumina sequencing and analysis

For sequencing, the library was denatured and diluted following standard Illumina-recommended protocols to a final concentration of 9 pM before being loaded onto an Illumina single read flow-cell for massively parallel sequencing on an Illumina GAIIx. Raw sequences were obtained from GA Pipeline software using CASAVA v1.7. The reads were further processed to remove any adapter sequence using the Illumina Flicker add-on. Flicker v2.7 trims the adaptor sequence from each read and does iterative alignment to the reference genome using ELAND.

One sample was sequenced per lane, yielding an average of 17 million high quality 36-bp reads per sample for the 0-, 12-, and 21-hour time points, 1.7 million of which were from non-rRNA and non-tRNA transcripts. The 6-hour sample was sequenced with longer 100-bp reads; however, these reads were trimmed to 40 bp for our analysis. RNA-seq data were aligned and analyzed with CLC Genomics Workbench 4 to create SAM files that were further analyzed with the Cufflinks software suite (112). CLC genomics workbench 4 was used to generate expression profiles and clustering.

RNA-seq data are available through the NCBI Gene Expression Omnibus (GEO) database, accession number #GSE26633. The raw sequence reads as well BAM files of reads at each time point aligned to

NCBI's current build of the *Anabaena* sp. strain PCC 7120 genome are included in the accession. Raw reads are .txt files and can be opened with a FASTA viewer. Aligned reads are in .BAM format and can be analyzed with free or commercial software suites.

## 2.3 RESULTS AND DISCUSSION

### 2.3.1 Analysis of the transcriptional response to nitrogen deprivation

We obtained RNA-seq data for total RNA isolated from *Anabaena* filaments grown with ammonium as a nitrogen source and at three times after nitrogen deprivation from ammonium to dinitrogen in air. The nitrogen deprivation produces relatively synchronous induction of heterocyst development. RNA-seq data were acquired from filaments at 0, 6, 12, and 21 hours after nitrogen deprivation, which provides detailed transcriptome data at important stages of heterocyst development (GEO accession #GSE26633). At 0 hours, all cells in the filaments are actively growing photosynthetic vegetative cells. At 6 hours, the cells have responded to the nitrogen deprivation and are expressing early heterocyst differentiation genes such as *hetR*, the master regulator of heterocyst development, and *patS*, which is involved in pattern formation. By 12 hours, proheterocysts are committed to complete differentiation and are expressing genes required for altering their morphology and physiology to become microoxic. By 21 hours, nearly all

heterocysts appear fully formed, contain polar cyanophycin granules, and are actively fixing nitrogen.

RNA-seq expression data are presented as RPKM, or **r**eads **p**er **k**ilobase (kb) of CDS (coding sequence) model per **m**illion mapped reads in the sample (86), with the CDS model defined as the CDS plus 100 bp of 5' UTR. RPKM values and changes in RPKM value for the chromosome and six plasmids are published as additional files 1-7 in the literature (40). These data can be examined and filtered in many ways. For example, for genes on the chromosome (additional file 1: Chromosome.xlsx) with a RPKM value of at least 2 (which includes only those genes with good read coverage) and a fold change of at least 5, there are 22 genes with increased expression by 6 hours after nitrogen deprivation, 434 genes upregulated by 12 hours (including many known heterocyst morphogenesis genes), and 396 genes upregulated by 21 hours (including the nitrogen fixation genes). For genes on the chromosome with decreased expression, there are 6 genes downregulated at 6 hours after nitrogen deprivation, 32 genes downregulated at 12 hours, and 35 genes downregulated at 21 hours.

Upregulation of nitrogen-fixation genes is the culminating event of heterocyst differentiation. The RNA-seq data provide detailed information on the expression of the known nitrogen-fixation genes as well as hypothetical and unknown genes that show the same pattern of regulation. For example, the data show very low levels of reads for *nifHDK* and other *nif* operons at 0,

6, and 12 hours after nitrogen deprivation (when heterocysts are not yet fully formed). The reads for all *nif* operons and especially for the *nifHDK* genes are dramatically increased in the 21 h sample, when most heterocysts are fully differentiated (Table 2.1).

Formation of the heterocyst envelope involves deposition of a polysaccharide outer layer followed by deposition of an inner glycolipid layer (5, 55). The RNA-seq data show that the genes responsible for heterocyst exopolysaccharide synthesis (*hep* genes) were upregulated by 12 hours after nitrogen deprivation (Table 2.1). However, strong upregulation of the genes required for heterocyst glycolipid synthesis (*hgl* genes) did not occur until the 21-hour sample. These data show that during heterocyst morphogenesis, the polysaccharide genes are expressed first, likely depositing the stabilizing exopolysaccharide later; subsequently, *hgl* genes are expressed to produce the underlying glycolipid envelope layer, which together are required to help create a microoxic environment within the heterocyst.

In addition to identifying single genes that respond to nitrogen deprivation, we mapped gene clusters that were upregulated in response to nitrogen deprivation. As expected, the region containing the major *nif* operons from *fdxH* (all1430)-*nifB* (all1517) (with the exception of the *nifD* and *fdxN* DNA elements present in the vegetative cell chromosome (65)) was strongly upregulated by 21 hours in response to nitrogen deprivation (Figure 2.1). Another cluster of genes in the *patB* (all2512)-alr2524 region, which contains

the cytochrome oxidase genes *coxBAC*, and many genes annotated as unknown or hypothetical, was strongly upregulated by 21 hours after nitrogen deprivation. Genes in the region alr2816-all2838, which contains *hetC*, *hetP*, and *hepA*, as well as many genes annotated as encoding hypothetical, glycosyltransferases, and metabolic proteins, were upregulated by 12 hours after nitrogen deprivation. Finally, genes in the region alr5340-alr5370, which contains several *hgl* (heterocyst glycolipid) genes along with a number of hypothetical genes, were upregulated by 21 hours after nitrogen deprivation. Each of these regions contains a number of genes and operons that are important for heterocyst development and nitrogen fixation.

**Table 2.1: Temporal response of the *nifHDK* operon and selected heterocyst glycolipid (*hgl*) and polysaccharide (*hep*) genes to nitrogen deprivation**

| Gene Name | Locus Name | RPKM | | | |
|---|---|---|---|---|---|
| | | 0 h | 6 h | 12 h | 21 h |
| *nifH* | all1455 | 0.89 | 0.77 | 6.51 | 1940.13 |
| *nifD* | all1454 | 0.00 | 0.04 | 0.42 | 108.34 |
| *nifK* | all1440 | 0.11 | 0.00 | 0.89 | 242.20 |
| *hglC* | alr5355 | 0.12 | 0.26 | 0.39 | 5.83 |
| *hglD* | alr5354 | 0.24 | 0.22 | 0.08 | 7.09 |
| *hglE* | alr5351 | 0.04 | 0.05 | 0.63 | 21.05 |
| *hepA* | alr2835 | 0.00 | 0.03 | 18.84 | 5.85 |
| *hepB* | alr3698 | 0.00 | 0.15 | 6.84 | 2.78 |
| *hepK* | all4496 | 5.27 | 5.68 | 12.80 | 10.79 |

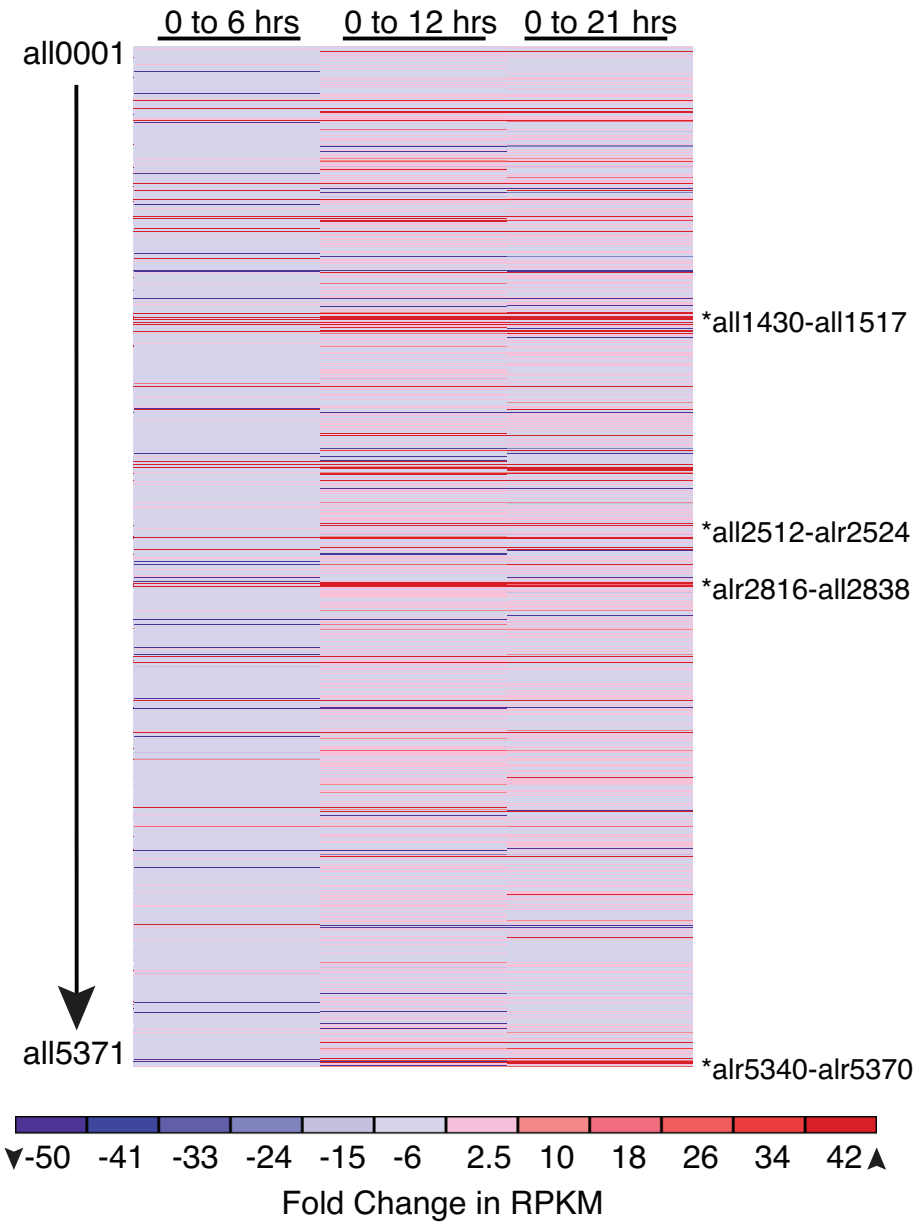



**Figure 2.1: Gene clusters upregulated during nitrogen deprivation.** The fold change in RPKM from 0 to 6, 12, and 21 hours after removal of combined nitrogen is represented in a heat map across the chromosome. Genes are shown in order from all0001 at the top to all5371 at the bottom. Clusters of genes that are upregulated and discussed in the text are marked with an asterisk and locus identifiers. The heat map was produced with The Broad Institute's GenePattern software and the dataset (additional file 8: RPKM GenePattern.gct) can be analyzed with the free GenePattern software (101).

### 2.3.2 Identification of unstudied genes regulated in response to nitrogen deprivation

The RNA-seq data showed regulation of numerous genes in response to nitrogen deprivation; including many that had not been previously identified as nitrogen-responsive in microarrays or genetic experiments. These genes are new candidates for the study of *Anabaena* heterocyst differentiation. We identified several new genes transcribed in response to nitrogen deprivation that have GO terms associated with regulation, including transcriptional regulators, two-component regulators, and kinases (Table 2.2). These genes may be involved in the regulatory pathways and transcriptional changes responsible for coordinating the expression of proteins required for heterocyst morphogenesis and nitrogen fixation.

We also identified many transposase genes that are upregulated by 6 hours after nitrogen deprivation (Table 2.3). Transposases are highly similar within a family and RNA-seq reads cannot always assign a sequence to a particular transposase locus for transposases within the same family. However, we can see that distinct families of transposons are upregulated in response to nitrogen deprivation. Table 2.3 shows the average fold change in RPKM during nitrogen deprivation for four families of transposons (126) that are upregulated at least 2-fold by 6 hours after nitrogen deprivation. It seems likely that transposase genes are turned on as a stress response to nitrogen deprivation.

**Table 2.2: Unstudied regulatory genes with 5-fold or greater increase in expression by 6, 12, or 21 hours after nitrogen deprivation[1]**

| Gene name | Gene description | RPKM fold change 6 h | 12 h | 21 h |
|---|---|---|---|---|
| *typA* (all4140)[2] | GTP-binding protein TypA/BipA | -1.2 | 6.8 | 5.9 |
| all0723[2] | probable GTP-binding protein | 1.1 | 6.1 | 6.2 |
| all2564[3] | pyruvate kinase | 2.3 | 50.5 | 25.2 |
| all4008[3] | pyruvate kinase | 1.0 | 4.3 | 5.3 |
| all0192[2] | serine/threonine kinase | 6.5 | 45.4 | 52.8 |
| alr1336[3] | serine/threonine kinase | 1.9 | 3.2 | 6.3 |
| all4838[3] | serine/threonine kinase | 1.3 | 6.6 | 3.8 |
| alr1044[2] | transcriptional regulator | 1.5 | 6.1 | 2.1 |
| all2237[2] | transcriptional regulator | 1.1 | 5.5 | 4.6 |
| alr2479[3] | transcriptional regulator | -1.2 | 7.9 | 2.8 |
| alr2769[2] | transcriptional regulator | -2.5 | 3.0 | 8.3 |
| alr3646[3] | transcriptional regulator | 1.1 | 9.0 | 2.1 |
| all3728[2] | transcriptional regulator | 2.4 | 7.2 | 6.7 |
| alr4564[2] | transcriptional regulator | 1.6 | 3.9 | 5.3 |
| alr0546[2] | two-component sensor histidine kinase | 1.0 | 5.3 | 2.6 |
| all3359[2] | two-component sensor histidine kinase | -1.3 | 5.7 | 8.4 |
| alr4878[2] | two-component hybrid sensor and regulator | 1.4 | 6.2 | 3.8 |
| alr5188[3] | two-component response regulator | 1.1 | 5.5 | 2.8 |

[1]Only unstudied genes with at least one-fold read coverage and annotated with the GO term "regulatory function" are presented.
[2]Genes that were not identified as being significantly upregulated in previous microarray analyses (31, 127).
[3]Genes that were previously identified as being regulated in microarray data but that show a greater degree of upregulation in our RNA-seq data (31, 127).

**Table 2.3: Transposase gene families with a 2-fold or greater increase in expression by 6 hours after nitrogen deprivation[1]**

| | Average Fold Change in RPKM | | |
|---|---|---|---|
| Transposon Family | 0 to 6 h | 0 to 12 h | 0 to 21 h |
| IS5/IS1031 | 6.28 | 1.67 | 1.01 |
| IS630 | 5.64 | 1.68 | 1.62 |
| IS982 | 3.50 | -0.28 | 0.51 |
| ISL3 | 2.57 | 0.26 | 2.30 |

[1]The average fold change in RPKM for all members of each transposon family is shown. Most RNA-seq reads align to multiple members of a highly conserved transposon family; therefore, we cannot report fold change in RPKM for single members of a family. Gene members of the IS5/IS1031 family are all2692, all2693, alr3610, alr3611, all4399, all4400, alr4438, alr4439, all4816, all4817, alr5157, and alr5158. Gene members of the IS630 family are alr0018, alr0019, all0362, all0363, alr0552, alr0553, alr1726, alr1727, alr1853, alr1854, alr1858, alr1859, all1971, all1972, asl1992, all2066, all2067, alr2773, alr2774, alr4628, all4867, all4868, alr5227, and alr5228. Gene members of the IS982 family are asl0588, alr0590, alr0999, all2664, alr2683, alr2694, alr3384, all3624, and alr4082. Gene members of the ISL3 family are alr1609 and alr2698.

### 2.3.3 Transcript mapping

Unlike previous whole-transcriptome analyses in *Anabaena*, deep sequencing provides information on all transcripts and can help identify 5' and 3' ends and characterize operon structure; however, transcriptional start sites versus processing sites cannot be differentiated with these methods. We used RNA-seq to identify distinct 5' ends for many transcripts, characterized by a set of reads with a common 5' end and the absence of upstream reads (GEO accession #GSE26633). These 5' ends often corresponded with published transcriptional start and/or processing sites; for example *psbB*, *petF*, *nrrA*, *psbAI*, and *nifB*, all of which were previously analyzed via primer extension (1, 66, 87, 89, 114, 121) (Figure 2.2 and 2.3).

However, for other genes such as *atp1* (ATP synthase) and *rbcL* (carbon fixation), the pattern of reads showed evidence for transcripts extending upstream from published 5' ends (Figure 2.2). For both *atp1* and *rbcL*, there is a considerable drop in read coverage at the previously identified 5' end, but there is still significant coverage upstream of these sites. This suggests that the previously identified 5' end may be a processing site or a site of RNA secondary structure that affects primer extension results on full-length transcripts, and that some of these transcripts originate at upstream start sites. Our analysis of transcript 5' ends suggests that transcription initiation and transcript processing in *Anabaena* is often complex, and that many transcripts have long 5' UTRs with unclear transcriptional start sites or

processing sites. These "trailing" 5' ends make identification of transcriptional start sites via primer extension or RACE difficult for some genes, and RNA-seq will be the method of choice for mapping transcripts.

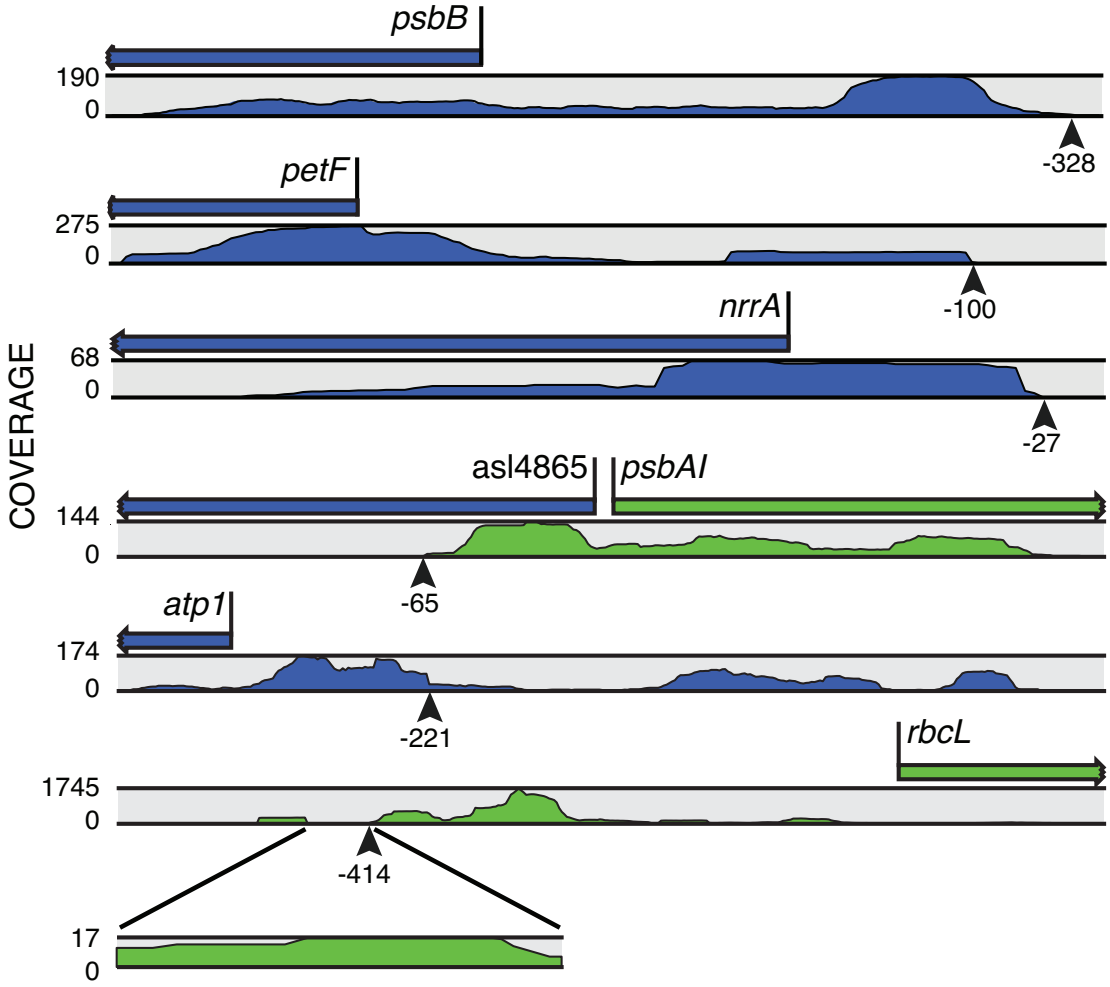



**Figure 2.2: Transcript 5' ends of *psbB*, *petF*, *nrrA*, *psbAI*, *atp1*, and *rbcL*.** RNA-seq coverage at each base along a gene's ORF and UTR are mapped below the gene. RNA-seq reads for *psbB, petF, nrrA,* and *psbAI* show clear 5' ends (characterized by an abrupt drop to 0 coverage) at -328, -100, -27, and -65, respectively, which correspond with published results obtained with other methods. However, *atp1* and *rbcL* show trailing reads without a complete break upstream of the mapped 5' ends at -221 and -414, respectively, indicating that transcription originates further upstream. RNA abundance along a transcript varies with message stability, often resulting in higher coverage at the 5' end of messages. Scales on left are reads per base position. All data shown are from the 0 h time point with the exception of *nrrA*, which is from the 21 h time point. Green and blue, ORF orientation and RNA-seq reads to the right and left, respectively.

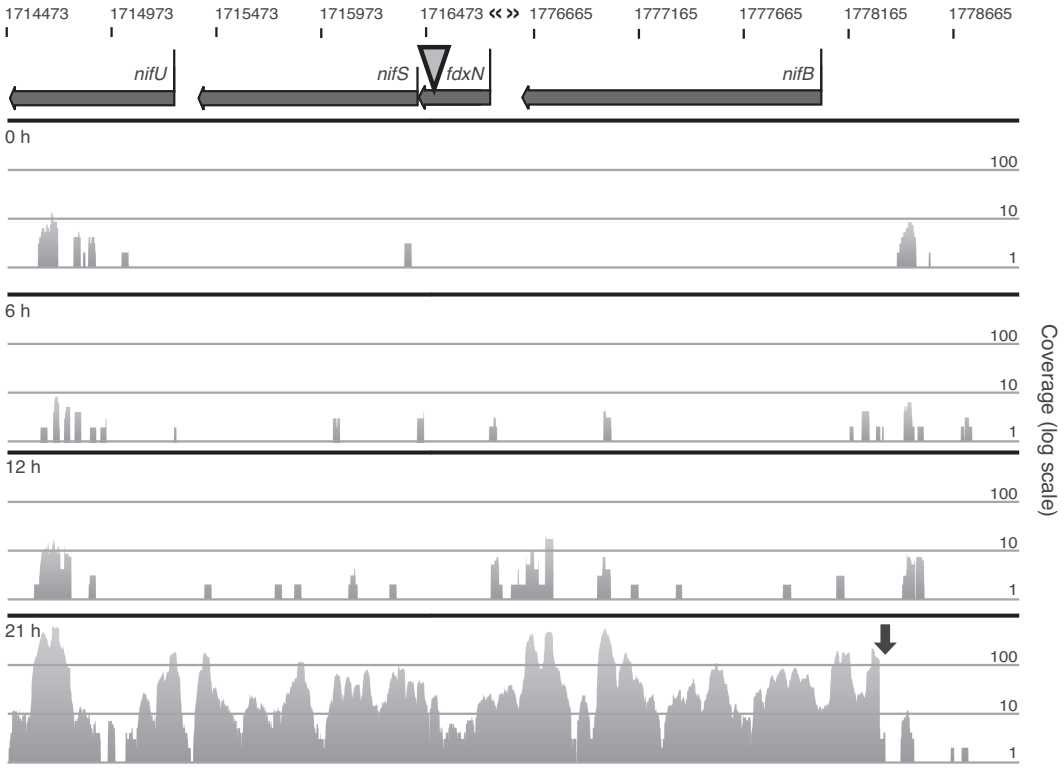



**Figure 2.3: Analysis of the *nifB-fdxN-nifS-nifU* operon structure, 5' end, and expression levels in response to nitrogen deprivation.**
RNA-seq read coverage is shown on a log scale across the *nifB-fdxN-nifS-nifU* operon in heterocyst chromosomes for 0, 6, 12, and 21 hours after the removal of combined nitrogen. The log scale is required to allow depiction of both the low numbers of reads at earlier time points and the large increase in reads at 21 h. The 59,428-bp *fdxN* element (denoted by the gray triangle) is excised from the chromosome in heterocysts (65), and the break in nucleotide numbering for the vegetative-cell chromosome is marked with «». The position of the previously mapped mRNA 5' end for the operon is at 282 bases upstream of the *nifB* start codon (marked by the black downward arrow in the 21 h graph) (87).

The RNA-seq data can be used to map transcripts and examine the coordinated expression of genes that form operons. For example, the genes in the *nifB-fdxN-nifS-nifU* operon showed very low numbers of reads at 0, 6, and 12 hours during nitrogen deprivation, when proheterocysts have yet to create a microoxic environment for nitrogen fixation (Figure 2.3). By 21 hours, there was a large increase in the number of reads for all four genes, indicating that a single transcript for all four genes originates from a promoter upstream of *nifB*. The RNA-seq data showed a clear 5' end at position -282 upstream of the *nifB* start codon, which validates previously published results (87). Further research will be required to fully describe the promoters and operon structure for the entire set of nitrogen fixation genes, which may be unexpectedly complex (113). Additional RNA-seq data or other types of approaches will be required to clarify transcription in certain regions because of fluctuations in read coverage present in our data set. These fluctuations in reads across ORFs and a general increase in coverage around the 5' ends of ORFs are consistent with other RNA-seq datasets (58, 85, 86, 96). It is likely that RNA stability and secondary structure contribute to these fluctuations in coverage.

## 2.4 CONCLUSIONS

Overall, our data confirm directional RNA deep sequencing as a more thorough method for analyzing transcriptional regulation in cyanobacteria and indicate that further studies using different environmental conditions and mutant strains will yield novel information about *Anabaena* gene regulation. Our RNA-seq data can be used to improve gene annotation and map RNA ends and operon structure. In addition, directional RNA-seq data provide superior information compared to 5' RACE and primer extension experiments for mapping RNA transcripts and identifying potential promoter regions. Because we used a strand-specific sequencing protocol, the dataset can be used to identify antisense and other noncoding RNAs potentially involved in gene regulation, which will be discussed in detail in Chapter 3. Finally, our work has provided a systems-level view of the *Anabaena* transcriptome during the response to nitrogen deprivation. Together, these features of the directional RNA-seq data can be used to define future directions for studying heterocyst development.

## ACKNOWLEDGMENTS

Chapter 2 is, in part, a reprint of the material as it appears in BMC Genomics, *2011* B. Flaherty, F. Van Nieuwerburgh, S.R. Head, and J.W. Golden. The dissertation author was the primary author of this publication. We thank Rodrigo Mella-Herrera for help with RNA purification, Arnaud Taton for

bioinformatics help, and Eric Allen for help with experimental design, software, and use of his facilities. This work was supported by National Science Foundation Grant 0925126 (JWG). BLF was supported by National Institutes of Health Training Grant T32GM007240.

# Chapter 3: Antisense Transcription in *Anabaena*

## 3.1 INTRODUCTION

Recent studies have shown that antisense transcription is abundant in prokaryotic organisms as well as eukaryotes. However, its mechanisms of action and biological importance are not yet fully elucidated (25, 44, 68). In many cases, individual antisense RNAs have important roles in gene regulation in bacteria. For instance, antisense RNAs have been shown to promote degradation of the sense mRNA transcript (94), interfere with translation (42), inhibit transcriptional activation (4), and inhibit processing and termination (125). While algorithms to predict the location of antisense transcription have made advances, sequencing and individual antisense RNA analysis still give the most accurate view of where and when antisense transcription occurs (3, 20, 23, 106).

Recent high-throughput transcriptomics have shown that cyanobacteria are no exception when it comes to the abundance of antisense transcription (40, 83, 84, 119). In fact, one study showed that as much as 65% of all transcripts in *Synechocystis* sp. strain PCC 6803 are noncoding (83). One well-studied example is the *isiA*/IsiR sense/antisense duo in *Synechocystis*, sp. strain PCC6803 (29). IsiA is an iron stress protein that promotes a large rearrangement of the photosynthesis apparatus. Therefore, aberrant expression of *isiA* would be detrimental to the cells under normal growth conditions. IsiR is the antisense RNA that is transcribed in opposition

to and in inverse proportions to *isiA.* IsiR regulates *isiA* through titration of the *isiA* mRNA and it is thought that the antisense regulation not only promotes a delay in *isiA* expression but also yields faster recovery from stress after transcription of *isiA* decreases.

In *Anabaena* sp. strain PCC 7120, two examples of antisense transcription have been well elucidated. The **F**erric **u**ptake **r**egulation protein FurA is involved in the integration of iron metabolism to environmental stress (6). *furA* transcription increases in response to iron stress, as does transcription of an antisense RNA in the coding region of *furA* (50). Removal of the *antisense furA* results in excess FurA protein, suggesting *antisense furA* is responsible for repressing FurA under normal growth conditions. In a second example, the **n**itrogen **s**tress-**i**nduced **R**NA **1** is a 60 nucleotide RNA with an unusual double hairpin secondary structure that is regulated in response to nitrogen deprivation in *Anabaena* (50). NsiR1 contains multiple tandem repeats upstream of the heterocyst development gene *hetF* and is highly conserved in *Nostocales* as well as other heterocyst-forming cyanobacteria. NsiR1 expression is specific to nitrogen-fixing heterocysts and was the first example of a heterocyst-specific antisense RNA, although its role and mechanism of action are not yet fully understood.

In this work, we used directional RNA-seq to get a global view of antisense transcription in response to nitrogen deprivation in *Anabaena*. Our high throughput sequencing discovered abundant antisense transcription at

varying levels throughout the genome. One striking example, an antisense RNA in the coding region of the protein NblA, is highly transcribed in N+ conditions and may be involved in regulation of sense *nblA* expression. We focus on this RNA, as well as others transcribed in the coding regions and UTRs of genes involved in the response to nitrogen deprivation.

## 3.2 MATERIALS AND METHODS

### 3.2.1 Antisense RNA mapping

We used directional RNA deep sequencing on the Illumina platform as described in Chapter 2 to map all sense and antisense transcription in *Anabaena* at 0, 6, 12, and 21 hours after the removal of combined nitrogen from the media. This protocol is outlined in the previous chapter. Briefly, RNA was extracted from total filaments, prepared into a cDNA library, and sequenced as 36-bp reads on the Illumina GAIIx platform. Information on the direction of transcription was maintained by ligation of strand-specific adapter sequences prior to cDNA synthesis. Reads were mapped and counted with CLC genomics workbench and antisense transcription was identified by manual scanning of the genome. Software suites such as Cufflinks did not efficiently identify antisense (or sense) transcripts, likely due to low coverage through many areas in the genome.

### 3.2.2 *gfp*-fusion to antisense promoters

The predicted promoter of *antisense-nblA* was cloned in front of *gfp* in the vector pAM1956 (129) via traditional cloning to produce pAM4615. The antisense promoter was amplified with anti-NblAXmaIforward (5' – agtcgtcccgggATGATCATTATATTTATTG -3') and anti-NblASacIreverse (5' – agtcgtgagctcGACTGTCTGGATTATCTG – 3') from 7120 genomic DNA. PCR products were digested, gel purified, and cloned into the SacI/XmaI site of pAM1956. Constructs were verified by sequencing and restriction digest. Clones were conjugated into wild-type *Anabaena* and at least 3 exconjugants (to allow detection of second-site mutations) were maintained in BG-11 N+ liquid medium. Clones were then washed three times in BG-$11_0$ or BG-11 and grown in BG-$11_0$ or BG-11 for 24 hours prior to imaging. Fluorescence microscopy was performed in the DIC, TRITC (autofluorescence), and GFP channels with an Olympus IX-71 inverted microscope with a 60X objective and images were obtained with Applied Precision's softWoRx software. Images were false colored red and green in ImageJ 64.

## 3.3 RESULTS AND DISCUSSION

### 3.3.1 Identification of antisense RNAs

The small RNA prep protocol maintains information on the direction of transcription by adding different adaptors to the 3' and 5' ends of each RNA molecule in the sample prior to cDNA synthesis. Therefore, we were able to identify antisense RNAs in ORFs or 5' UTRs of annotated genes and also identify transcripts from unannotated regions of the genome (data available at

GEO accession #GSE26633). The antisense transcripts would not have been identified with standard microarray or RNA deep sequencing methods because these methods do not normally distinguish between sense and antisense transcripts.

Our directional RNA-seq data showed antisense RNAs throughout the *Anabaena* transcriptome (Table 3.1, Figure 3.1, and GEO accession #GSE26633). For example, we identified novel antisense transcripts in the 5' ends of key developmental genes such as *hetR* (the master regulator of heterocyst differentiation) and *hetC* (a gene involved in early heterocyst development). Furthermore, we confirmed the presence of the noncoding RNA NsiR1 in the upstream region of *hetF* (another heterocyst regulatory gene) (13, 57, 61, 90), and our directional RNA-seq data suggest that the NsiR1 transcript is antisense to the 5' UTR of *hetF*. Other potential noncoding RNAs identified by our directional RNA-seq data include, for example, antisense reads in the region from alr0091 to alr0094 and from alr0709 to alr0710; and abundant rightward reads between alr0249 and all0250, and leftward reads between alr1199 and alr1200 (GEO accession #GSE26633).

## Table 3.1: Antisense RNAs transcribed within the ORF or 5' UTR of genes involved in heterocyst differentiation

*NsiR1 has been previously characterized (57).

| Gene | Gene Function | Gene Expression After Nitrogen Deprivation | RNA Location | Antisense RNA Expression After Nitrogen Deprivation |
|---|---|---|---|---|
| *narB* (alr0612) | nitrate reductase | increased by 12 h | *narB* ORF and 5' UTR | no change |
| *fraH* (alr1603) | cell-cell connection | increased by 12 h | alr1603 ORF | slight decrease by 21 h |
| alr3649 | heterocyst specific ABC transporter | increased by 12 h | 5' end of alr3649 ORF into upstream gene | increased by 12 h |
| alr3479 | similar to nitrogen regulation protein NtrR | decreased slightly at 6 h only | alr3479 ORF | increased by 21 h |
| *glnB* (all2319) | nitrogen assimilation | no change | *glnB* 5' UTR | no change |
| *hetR* (alr2339) | heterocyst differentiation regulator | increased by 6 h | *hetR* ORF and 5' UTR | decreased by 21 h |
| *hetC* (alr2817) | heterocyst differentiation | increased by 12 h | two RNAs, one in *hetC* ORF and one in *hetC-hetP* intergenic region | increased by 12 h |
| *hetF* (alr3546) | heterocyst differentiation | no change | *hetF* ORF and 5' UTR; NsiR1* | increased by 6 h |
| all3558 | nitrogen assimilation regulation | no change | all3558 ORF | increased by 21 h |
| *hepK* (all4496) | exopolysaccharide synthesis | increased by 12 h | *hepK* ORF; contains repeat sequence at 5' end | no change |
| *nblA* | phycobilisome degradation | increased by 12 h; largest increase at 21 h | *nblA* ORF and 5' and 3' UTRs | antisense transcription until 12 h only |
| *hglE* (alr5351) | heterocyst-specific glycolipid | increased by 12 h; largest increase at 21 h | *hglE* 3' end of ORF | increased by 12 h |
| *hglD* (alr5354) | heterocyst-specific glycolipid | increased by 21 h | *hglD* ORF and 5' UTR | no change |
| *hglC* (alr5355) | heterocyst specific glycolipid gene | increased by 6 h; largest increase at 21 h | *hglC* ORF | increased at 6 h only |

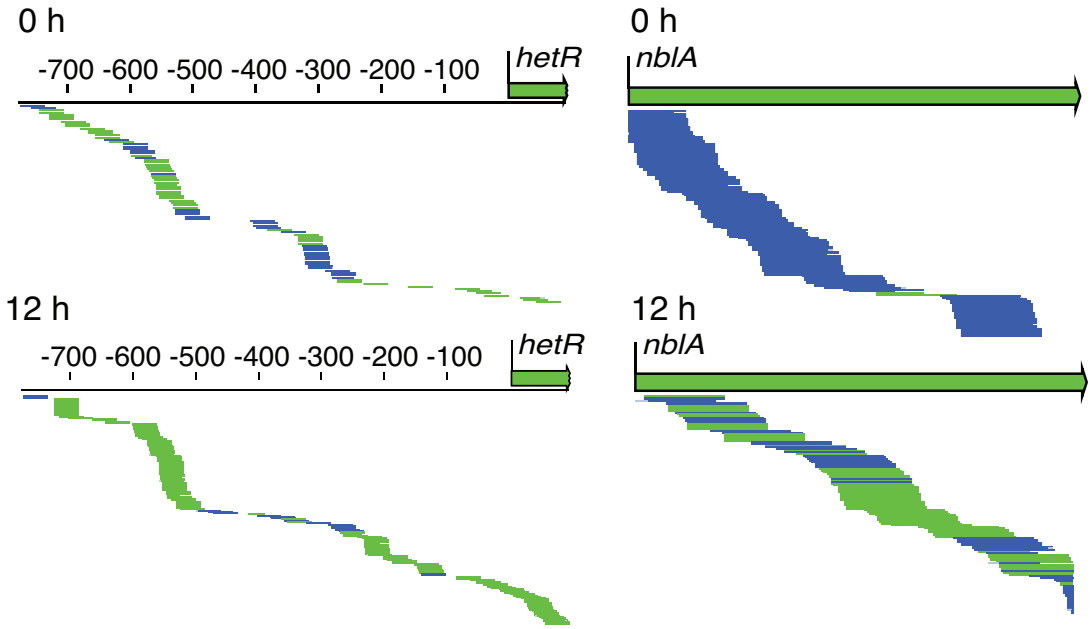



**Figure 3.1: Examples of antisense RNAs.**
RNA-seq reads for 0 and 12 hour time points are shown underneath the ORF or 5' UTR as single lines, with blue lines representing reads to the left and green lines representing reads to the right. For the 5' UTR of *hetR* (on left), the ratio of antisense (blue) to sense (green) reads was decreased at 12 hours after nitrogen deprivation. For *nblA* (on right), antisense reads in the coding region were dominant before nitrogen deprivation; however, at 12 hours, *nblA* sense reads were abundant and antisense reads were decreased.

### 3.3.2 Analysis of *anti-nblA*

Some genes showed striking changes in antisense RNAs. For example, *nblA*, which is involved in degradation of phycobilisome proteins in response to nitrogen deprivation (7), showed extensive antisense reads at 0 hours (Figure 4). NblA accumulates in all cells in the filament after nitrogen deprivation and results in degradation of phycobilisomes, the light harvesting proteins that funnel electrons to the photosystems (7). However, after phycobilisomes are re-synthesized in vegetative cells, which continue to fix carbon through photosynthesis, the NblA protein accumulates specifically in heterocysts which do not rebuild their phycobilisomes (21). NblA interacts directly with a chaperone of the Clp protease as well as with the phycobiliproteins themselves, suggesting that it recruits the protease to the phycobiliproteins for specific degradation (59). Functional NblA is required for phycobilisome degradation in *Anabaena*, but functional heterocysts can develop even in a ∆*nblA* background (7).

After nitrogen deprivation, the antisense reads decreased as *nblA* sense reads increased (Figure 3.1). A fusion of the antisense *nblA* promoter to *gfp* on the plasmid pAM1956 produced GFP fluorescence in vegetative cells and a decrease in fluorescence in heterocysts after nitrogen deprivation, confirming that the decreased antisense RNA expression in response to nitrogen deprivation is specifically localized to heterocysts (Figure 3.2). A fusion of the sense-*nblA* promoter to *gfp* showed heterocyst specific

upregulation of *nblA* (data not shown), confirming previous work that suggests that NblA accumulates specifically in heterocysts. We hypothesize that, similar to the *furA* gene, it is critical for vegetative cells to avoid expression of even small amounts of NblA protein during normal cell growth and that the antisense RNA ensures no expression from the *nblA* gene. The ratio of antisense to sense *nblA* RNA decreases after nitrogen deprivation, likely because heterocysts begin to express the coding transcript. Bioinformatics analysis of the antisense transcript did not identify any open reading frames, suggesting that it functions fully as an RNA and is not translated to protein. Future work on the characterization of an *anti-nblA* knockout may help further define the molecular role of this antisense RNA.

Overall, antisense transcription within genes involved in heterocyst development (Table 3.1) was abundant, although the levels of antisense RNAs were often quite low (see HetR in Figure 3.1). We do not know if the antisense transcription is simply spurious and not selected against during evolution of this AT-rich organism, or if these antisense RNAs may play a role in gene regulation during the response to combined nitrogen deprivation. Future analysis of these noncoding RNAs will shed light on the general mechanisms of antisense gene regulation. Therefore, our lab is focusing on *anti-nblA* as a model for antisense transcription that is highly regulated. It will be enlightening to see if knocking out *anti-nblA* as well as overexpressing this

transcript leads to misregulation of NblA and phycobilisome degradation in N+ conditions as well as slower recovery of phycobilisomes in vegetative cells.

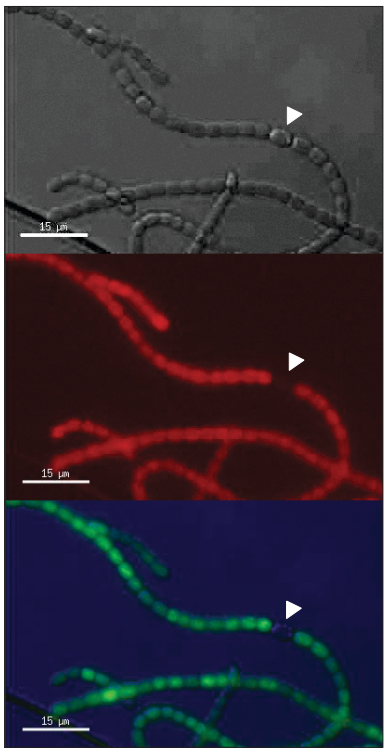



**Figure 3.2: *Anti-nblA* is expressed in vegetative cells**
A fusion of *gfp* to the promoter of antisense-NblA resulted in no GFP expression in N- conditions in heterocysts, but consistent GFP expression throughout vegetative cells. All images are of *Anabaena* cells grown in BG-$11_0$ liquid medium for 24 hours, then imaged with DIC (top), TRITC/autofluorescence (middle) or the GFP filter (middle). Heterocysts (marked with a white triangle) are slightly larger than vegetative cells and do not display autofluorescence as they have degraded their photosynthetic machinery.

## 3.4 CONCLUSIONS

Directional RNA-seq uncovered abundant antisense transcription in *Anabaena* sp. strain PCC 7120. In particular, strong antisense transcription was seen in the coding region of *nblA*, a gene involved in phycobilisome degradation in response to nitrogen deprivation. Antisense transcripts were also seen in a number of genes involved in heterocyst development, some of which seemed to be up or down regulated in response to nitrogen deprivation. However, the level of antisense transcription was relatively low compared to that of sense transcripts. We do not yet know if antisense transcription through any individual gene plays a role in the regulation of that gene. Further work elucidating the molecular mechanisms of antisense transcription for specific transcripts will shed light on both antisense transcription in cyanobacteria as a whole and on the transcriptional response during heterocyst development in *Anabaena*.

## ACKNOWLEDGMENTS

Chapter 3 is, in part, a reprint of the material as it appears in BMC Genomics, *2011* B. Flaherty, F. Van Nieuwerburgh, S.R. Head, J.W. Golden. The dissertation author was the primary author of this publication. We thank Rodrigo Mella-Herrera for help with RNA purification, Arnaud Taton for bioinformatics help, and Eric Allen for help with experimental design, software, and use of his facilities. This work was supported by National

Science Foundation Grant 0925126 (JWG). BLF was supported by National Institutes of Health Training Grant T32GM007240.

# Chapter 4: ChIP of HetR

## 4.1 INTRODUCTION

*Anabaena* fixes atmospheric $N_2$ into ammonia by forming specialized nitrogen-fixing cells called heterocysts. In response to nitrogen deprivation, a transcriptional cascade controls the differentiation and even spacing of heterocysts along the filament. HetR is a key transcription factor required for heterocyst development and diazotrophic growth. Overexpression of HetR results in multiple contiguous heterocysts, suggesting that HetR alone can trigger heterocyst development and override external queues. (10, 13, 56). HetR has been shown to bind DNA through electrophoretic mobility shift assays, but its binding site is still not fully understood (27, 54, 56). One study suggested that HetR's binding site is GGGTCTAgCCCagCA, but this site is not upstream of all known HetR targets, including the heterocyst pattern regulator PatS (27).

HetR is conserved in conjunction with PatS in both heterocystous and nonheterocystous cyanobacteria (133). PatS and HetR are thought to act together to regulate the spacing of heterocysts (129, 130). PatS can bind directly to the DNA-binding pocket of HetR, and a small pentapeptide RGSGR motif in PatS is sufficient to inhibit heterocyst formation altogether (38, 129). Taken together, this data suggests a model whereby HetR and PatS both are produced in heterocysts, but the heterocyst-specific HetR is immune to the inhibitory effects of PatS. PatS is thought to diffuse from heterocysts along

the filament to vegetative cells, setting up a gradient that regulates heterocyst spacing.

Because these proteins are conserved in cyanobacteria that do not form heterocysts, and because HetR is expressed at low levels in vegetative cells, it is likely that HetR and PatS serve a function in addition to their known roles in heterocyst development. HetR has two protein interaction domains, suggesting that other proteins modulate HetR's activity and DNA binding. However, very little is known about HetR's role outside heterocyst development or about potential changes in its regulon in response to environmental queues. Therefore, we used chromatin immunoprecipitation to give the first global view of HetR's DNA-binding sites in both vegetative cells and proheterocysts. We isolated his-tagged HetR bound to DNA at 6 hours after the removal of combined nitrogen and used deep sequencing to identify all regions of the genome enriched for HetR binding.

Chromatin Immunoprecipitation, or ChIP, is a powerful way to identify new DNA targets of transcription factors and DNA-binding proteins. ChIP involves first crosslinking protein to DNA in live cells, followed by shearing of the genomic DNA to produce short 100-200 bp regions bound by the protein under study. Immunoprecipitation or affinity precipitation of an epitope tag is used to purify the protein of interest along with the bound DNA fragment away from the cell lysate. Then, the crosslinks are reversed and the enriched DNA fraction is purified. For decades, this fraction was analyzed by northern blots,

microarrays, and quantitative PCR. However, recent developments in high-throughput sequencing technologies have allowed whole regulon sequencing for the enriched DNA sample. Sequencing after ChIP helps to identify novel binding targets and gives a more precise view of the protein's binding site.

ChIP has been used in eukaryotic systems to map chromatin architecture, identify regulons for key transcription factors, and define consensus binding sites (9, 16, 17). However, its use in prokaryotic organisms is more recent. In the past 8 years, ChIP was employed to map the regulon of both bacterial transcription factors and RNA polymerase, but most of these studies were carried out in *E. coli* (52, 99, 107, 122). In the past two years, ChIP assays were used to map RNA polymerase binding sites as well as those of two transcription factors in *Synechococcus elongatus* PCC 7942, but to the best of our knowledge there are no publications of ChIP studies in other cyanobacterial species to date (47, 119).

It is possible that the lack of ChIP studies in cyanobacteria has to do with the difficulty in applying this protocol to a new organism. ChIP relies heavily on protein chemistry for crosslinking and protein extraction as well as purification with high quality antibodies, all of which must be heavily optimized for the target organism and protein. Multiple crosslinkers, crosslinking times, buffers, and sonication procedures can be used, and finding the correct combination of factors requires time and resources. Furthermore, downstream analysis of ChIP data must be optimized in an organism-specific

manner. For unclear reasons, prokaryotic ChIP analysis require much deeper sequencing of the negative control sample than those in eukaryotes, and prokaryotic read coverage is less consistent across the genome than in eukaryotes. This makes it difficult to apply currently available tools and algorithms, developed in eukaryotic assays, to analyze ChIP-seq data from cyanobacteria.

In this work, we used ChIP to map the regulon of HetR, a key regulator of heterocyst development in *Anabaena* sp. strain PCC 7120. HetR is a problematic protein to work with outside of the cell, and mobility shifts have been historically difficult. Therefore, we believe that ChIP, which identifies the protein's *in vivo* binding sites*,* may help bypass some of the difficulty of working with HetR *in vitro*. We developed a refined ChIP protocol for *Anabaena* and optimized the downstream analysis, opening the door for future studies of transcription factor regulons in this cyanobacterium. Furthermore, our work refined the consensus-binding site for HetR and identified novel HetR gene targets that suggest that HetR plays a role in regulating both heterocyst and vegetative cell-specific gene expression.

## 4.2 MATERIALS AND METHODS

### 4.2.1 Cell Growth Conditions and Nitrogen Deprivation

*Anabaena* (*Nostoc*) sp. strain PCC 7120 cultures were grown in 100 ml or 2 ml of liquid medium in 250-ml flasks with cotton plugs as previously

described with slight modifications (46). Briefly, 100-ml or 2-ml liquid cultures were grown to an $OD_{750}$ of 0.02 in BG-11($NH_4$) medium, which lacked sodium nitrate and contained 2.5 mM ammonium chloride and 5 mM MOPS (pH 8.0). For nitrogen deprivation, cultures were spun down at 4,000 x g for 5 minutes and washed three times in BG-$11_0$ media by centrifugation and decanting of the supernatant. Cells were then resuspended in 100 ml or 2 ml BG-$11_0$ at a final $OD_{750}$ of between 0.02 and 0.05. Cells were grown for 6 hours in BG-$11_0$ shaking at 100 µmol photons $m^{-2}$ $s^{-1}$.

**4.2.2 Chromatin Immunoprecipitation**

6 hours after nitrogen deprivation, WT and HetR-his cells were spun down at 4,000 x g for 5 minutes, then resuspended in 5 mL BG-$11_0$. Cells were crosslinked by the addition of 4.1 mg disuccinimidyl glutarate (DSG) and 0.56 mg ethylene glycol bis (succinimidylsuccinate) (EGS) in 500 µL DMSO. Crosslinking occurred at room temperature, rocking, for 20 minutes. After 20 minutes, 135 µL of 37% formaldehyde was added for additional protein to DNA crosslinking and left rocking at room temperature for 15 minutes. To quench the reaction, 125 mM glycine was added for 5 minutes at room temperature.

Cells were then spun down at 4,000 x g for 5 minutes at 4°C and washed twice in 30 mL ice cold PBS (137 mM NaCl, 2 mM KCl, 10 mM $Na_2HPO_4$, 1.8 M $KH_2PO_4$, pH 7.4). Washed and fixed pellets were resuspended in 500 µL ice-cold binding/wash buffer (100 mM $NaHPO_4$, 600

mM NaCl, 0.02% Tween 20, 1 EDTA Proteinase Inhibitor Tab from Roche Biosciences in 10 mL total volume) on ice. Protein was extracted by bead beating 2 x 5 minutes with 2 minutes on ice in between. Complete lysis was confirmed by microscopy. Lysed cells were separated from beads via centrifugation and DNA was sheared via sonication on ice, 12 cycles of 20 seconds on, 15 second off at 14% power. Cell debris was pelleted via two cycles of centrifugation at 14,000 x g for 15 minutes at 4°C. Protein concentration was determined by the absorbance at 280 nm and normalized to 20 mg/mL for each sample by dilution in cold binding/wash buffer.

His-tagged HetR was bound to Dynabeads (Dynabeands His-tag Isolation and Pulldown beads, Invitrogen) following manufacturer's protocol at 4°C and eluted in 100 uL elution buffer (100 mM imidazole, 50 mM $NaPO_4$, 300 mM NaCl, 0.01% Tween 20). Crosslinks were reversed at 65°C for 18 hours in a PCR block. Input DNA was used to assess size distribution after shearing on a 1% agarose gel. IP efficiency was measured via western blotting of HetR-6xHis with the Qiagen Penta-His antibody, BSA Free. After crosslinks were reversed, proteins were digested by the addition of 250 µL TE, 4 µL of 20 µg/µL glycogen, and 10 µL of 10 µg/µL proteinase K for 2 hours at 37°C. DNA was column purified with the Promega SV DNA purification kit and resuspended in 30 µL nuclease free water.

### 4.2.3 DNA Library Preparation and Sequencing

DNA was prepared for sequencing with the Illumina ChIP-seq Sample prep kit by the Next Generation Sequencing Core at The Scripps Research Institute (La Jolla, CA) following the manufacturer's protocol. Sequencing was performed on the Illumina HiSeq platform with 4 samples multiplexed on one cell, yielding approximately 40 million 40-bp reads per sample. Sequence reads were filtered based on adapter sequences using the Illumina Flicker add-on and saved as fasta files for analysis in CLC Genomics Workbench 5.

### 4.2.4 Sequence Alignment and Peak Finding

Sequencing reads from the experimental HetR-6xhis sample were randomly assigned to three files and the full ChIP-seq analysis was performed on each sample as a technical replicate. ChIP-seq reads from the three HetR-his samples and the WT sample were aligned to NCBI's current build of the *Anabaena* genome with CLC Genomics Workbench 5. ChIP peaks were called using a 100 bp window and a false discovery rate of 1% with the WT sample as the control using CLC's ChIP Analysis pipeline. ChIP peaks were excluded from the final data set if they were not present in at least two of the three technical replicates. All peaks were verified by eye prior to further analysis.

### 4.2.5 *gfp*-reporter Fusion Construction

Promoters of six putative HetR target genes were defined based on RNA-seq datasets (40, 84) and amplified from the *Anabaena* genome with SacI and SmaI restriction sites using the oligos in Table 4.1. Truncated

versions of four of these promoters, missing the putative HetR binding site, were also amplified with a “truncated” reverse primer, as shown in Table 4.1. PCR fragments were cloned into the SacI/SmaI site of pAM1956 to yield pAM4653, pAM4654, pAM4658, pAM4659, pAM4660, pAM4661, pAM4662, pAM4695, and pAM4696 (Table 4.2). Plasmids were then transformed into *E. coli* strain AM1359 for conjugation into *Anabaena* wild type strain AMC1078 and the ∆*hetR* strain AMC1537/UHM103. Exconjugants were maintained in liquid BG-11 N+ in 2 ml cultures as described above. 2 ml cultures were grown in 24 well plates for nitrogen deprivation.

**Table 4.1: Oligos used in this study**

| | |
|---|---|
| PatAFSmaI | cagtaGcccgggACTAGAAAACATTCTTTCAA |
| PatARSacI | cagtaGgagctcCATTTTGCAAGTGATGGTCA |
| asr1469FSacI | cagtaGgagctcTGAGGAGCGCGAAAGAAAAAAGAG |
| asr1469RSmaI | cagtaGgggcccAAGGCACAAGAGTTTCCATC |
| asl2028FSmaI | cagtaGgggcccCATCTTGCTACAGAATAAAAAAGTTC |
| asl2028RSacI | cagtaGgagctcTGCCGAATTGTGTCGCGG |
| alr3758FSacI | cagtaGgagctcTTATTCAAGCTGTTTGGAGTG |
| alr3758RSmaI | cagtaGgggcccAGCTGTTTTTTTCTTTAAATAAATTG |
| patATruncSacIR | cagtaGgagctcTAACTATACGTGTAGTAC |
| asr1469TruncSacIF | cagtaGgagctcTTAAAAAGGTTTTAATTC |
| asl2028TruncSacIR | cagtaGgagctcTTACTGCCACCATCGCGCCCTAG |
| alr3758TruncSacIF | cagtaGgagctcTCAATAGTTCCTGGGCTG |

**Table 4.2: Strains and plasmids used in this study**

| | |
|---|---|
| **Plasmids** | |
| pAM4653 | pAM1956 carrying the promoter of *asr1469* driving *gfp*, $Km^rNm^r$ |
| pAM4654 | pAM1956 carrying the truncated promoter of *asr1469* driving *gfp*, $Km^rNm^r$ |
| pAM4658 | pAM1956 carrying the promoter of *asl2028* driving *gfp*, $Km^rNm^r$ |
| pAM4659 | pAM1956 carrying the truncated promoter of *asl2028* driving *gfp*, $Km^rNm^r$ |
| pAM4660 | pAM1956 carrying the promoter of *asr1469* driving *gfp*, $Km^rNm^r$ |
| pAM4695 | pAM1956 carrying the promoter of *alr3758* driving *gfp*, $Km^rNm^r$ |
| pAM4696 | pAM1956 carrying the truncated promoter of *alr3758* driving *gfp*, $Km^rNm^r$ |
| pAM4661 | pAM1956 carrying the promoter of *all5131* driving *gfp*, $Km^rNm^r$ |
| pAM4662 | pAM1956 carrying the promoter of *alr2242* driving *gfp*, $Km^rNm^r$ |
| pAM1956 | promoterless mut2 *gfp* in pAM505, $Km^rNm^r$ (129) |
| pAM4375 | pAM505 carrying C-terminal 6xHis tagged HetR, $Km^rNm^r$ |
| pAM505 | pDUI plasmid for expression in *Anabaena,* $Km^rNm^r$ (129) |
| pRL623 | conjugal helper plasmid carrying AvaI, AvaII, and AvaIII methylases, $Cm^r$ (35) |
| pRL443 | conjugal RP-4 derivative, $Ap^rTc^r$ (108) |
| | |
| **Strains** | |
| AMC1078 | Wild Type Anabaena sp. strain PCC 7120 |
| AMC1537 | AMC1078 carrying pAM4375, $Nm^r$ |
| UHM103 | clean deletion of *hetR* from the chromosome, AMC1537 {Borthakur, 2005 #146} |
| AM1359 | DH10B carrying pRL623 and pRL443, $Tc^rCm^rEm^rAp^r$ |

### 4.2.6 Imaging GPF promoter fusions

Promoter-*gfp* fusions of putative HetR targets in WT and Δ*hetR* backgrounds were grown in 2 ml BG-11($NH_4$) liquid media in a 24 well plate for 24 hours prior to nitrogen deprivation at an $OD_{750}$ of 0.01-0.03. After nitrogen deprivation was performed as previously described, cells were imaged in the DIC, TRITC (autofluorescence), and GFP channels on an Olympus IX-71 inverted microscope with a 60X objective using Applied Precision's softWoRx software. Exposure for each channel, GFP, TRITC, and DIC, is the same for each time point of a given strain, and the same for all strains except those carrying the promoter *alr2028*, where the GFP exposure was 1/5 that of the other strains. Images were false colored red and green in ImageJ 64 (105).

## 4.3 RESULTS AND DISCUSSION

### 4.3.1 ChIP Peaks

His-tagged HetR was crosslinked to its DNA binding targets and affinity precipitated to identify all HetR targets at 6 hours after nitrogen deprivation with ChIP-seq. After Illumina sequencing, ChIP DNA from WT and HetR-6xHis samples were aligned to the *Anabaena* PCC 7120 genome. The CLC Genomics Workbench 5 ChIP-seq algorithm was used to search for ChIP peaks with the WT samples as a negative control. The algorithm found 39 putative ChIP peaks with a very stringent 1% false discovery rate and a 100

bp window for the Poisson distribution. With more lenient parameters of an 80-bp window and a 5% false discovery rate, some 1,800 ChIP peaks were discovered. While these peaks may contain many real targets of HetR, for our targeted purposes we continued analysis of the smaller and more stringent dataset.

### 4.3.2 Identification of a Refined Putative HetR Binding Site

From the 39 identified ChIP peaks, we analyzed each ChIP peak by eye to confirm a 3-fold enrichment in reads of the ChIP peak over wild type and a twin peak morphology. In addition, we excluded peaks that were not upstream of a gene that is misregulated in a Δ*hetR* background (as determined by our lab's unpublished RNA-seq analysis of the Δ*hetR* mutant at 6 hours after nitrogen deprivation {Johnson, unpublished data #177}) or that were not near a putative transcriptional start site (as determined by our RNA seq data and a recently published map of transcriptional start site in *Anabaena* {Mitschke, 2011 #89}). We queried the resulting 27 ChIP peaks with FiMO (**F**ind **I**ndividual **M**otif **O**ccurrences) for the previously defined putative HetR binding site (8, 27) in our ChIP regions with a p value of $10^{-4}$. FiMO found putative HetR binding sites within 100 bp of all but 6 of the putative ChIP peaks (Table 4.3).

We then used the MEME suite to align the discovered putative HetR binding sites identified with FiMO to create a refined putative HetR binding

site (Figure 4.1). Key differences between this newly refined putative HetR binding site and the GGGTCTAgCCCagCA site identified earlier include a highly conserved “A” two bases upstream of the triple G repeat and an extra base in the spacer between the GGG and CCC inverted repeat. Furthermore, we see little to no conservation of the nucleotides downstream of the CCC. However, our newly refined binding site is present in all HetR targets known to bind HetR in a mobility shift, including *patS.*

## Table 4.3: High-Confidence ChIP Peaks

| Chromosomal coordinates | 5' gene (distance) | Fold Change expression in Δ*hetR* | 3' gene (distance) | Fold Change expression in Δ*hetR* | Putative HetR Binding Site found? |
|---|---|---|---|---|---|
| +(136127..136558) | all0131 (223) | no coverage | alr0132 (64) | -1.07 | yes |
| +(1729122..1729569) | all1467 | 1.3 | alr1468 | no coverage | yes |
| +(1732738..1733693) | asr1469 | -4 | alr1470 | -3.14 | yes |
| +(1734763..1735224) | all1471 | -2.8 | all1472 | -1.6 | yes |
| +(1788499..1788932) | alr1527 | -5.8 | none close | | yes |
| +(2427714..2428147) | asl2028 | -2.5 | none | | yes |
| +(2691579..2692024) | all2239 (199) | n/a | alr2240 | -7.1 | yes |
| +(2820376.2822006) | alr2338 | n/a | **hetR (583)** | n/a | yes |
| +(3431297..3432038) | hetC (99) | n/a | **hetP (325)** | -4.3 | yes |
| +(3627499..3627977) | alr2986 | -3 | none close | -1.14 | yes |
| +(3695599..3696450) | asr3053 | 2.06 | all3054 | N/A | yes |
| +(3821175..3821606) | alr3156 | n/a | aphA | -7.67 | yes |
| +(4383426..4383866) | avaIR (630) | -1.2 | avaIM | -1.9 | yes |
| +(4539927..4540364) | alr3757 (11) | n/a | alr3758 | infinite | yes |
| +(4740704..4741151) | alr3926 (1150) | n/a | all3927 | -3 | yes |
| +(5064797..5065242) | asr4228 (209) | n/a | alr4229 | -7.5 | yes |
| +(5303024..5303480) | all4424 (998) | n/a | all4425 | 1.7 | yes |
| +(5309952..5310383) | all4430 | 4.9 | all4431 | 10 | yes |
| +(5355615..5356047) | alr4469 | -1.3 | asr4470 | -10 | yes |
| +(613436..613919) | patA | -7.9 | alr0522 | -5.6 | yes |
| +(2274728..2275590) | alr1903 | -4.9 | none close | | no |
| +(2693489..2693944) | alr2241 (108) | n/a | alr2242 | No change | no |
| +(2706756..2707187) | gvpA (89) | n/a | asr2255 | n/a | no |
| +(374646..375090) | all0326 (145) | no coverage | all0327 | N/A | no |
| +(4857231..4857672) | alr4031 | n/a | alr4032 | no coverage | no |
| +(6129044..6129475) | all5131 (34) | -4.2 | all5132 (22) | n/a | no |




**Figure 4.1: Refined HetR Consensus binding site as identified by ChIP peaks.**

ChIP peaks were queried for the previously identified HetR binding site GGGTCTAgCCCagCA. Sites were then re-aligned to give a putative HetR binding site present in 22 of the 27 ChIP peaks identified in our analysis.

HetR is thought to potentially interact with other factors, which could be different at different times and in different cell types. Since our ChIP sample was collected at 6 h after nitrogen removal, it seems likely that the consensus binding site we have defined for HetR is dependent on the binding partners present at that time. Furthermore, our ChIP data is from HetR expressed in both proheterocysts and in vegetative cells. Therefore, our binding site may be skewed towards HetR targets in vegetative cells compared to previous analyses, which focused on HetR's role in regulating heterocyst-specific genes. Further chromatin immunoprecipitation experiments to identify changes in the HetR consensus site before and during nitrogen deprivation may shed light on any changes in the HetR regulon in response to environmental queues. Furthermore, mass spectrometry of proteins which immunoprecipitated with HetR may help discover HetR's putative binding partners in vegetative and heterocyst cells.

### 4.3.3 GO Term Enrichment in HetR Targets

Genes in two **G**ene **O**ntology categories were enriched with respect to their abundance in the genome in our ChIP peak sample set: "heterocyst differentiation" and "transposon related functions" (Figure 4.2). It is not surprising that genes involved in heterocyst differentiation, which includes genes such as *hetR, patS, hetP*, and *patA*, would be enriched as targets for the key regulator of heterocyst development, HetR. However, we do not expect transposases to be involved directly in heterocyst development. It is

interesting to note that the transposase genes enriched in ChIP of HetR also appear to be upregulated in response to nitrogen deprivation (as outlined in Chapter 2), confirming that these genes are, in fact, highly transcribed during heterocyst development. The ChIP data suggests that HetR binds the promoters of transposases; however, it is unclear whether this DNA binding is part of the regulatory network that regulates the response to nitrogen deprivation or whether transposases have simply hijacked the HetR binding site as a signal of cellular stress. Further work to identify the relationship, if any, of transposases to the response to nitrogen deprivation may shed light on this issue.

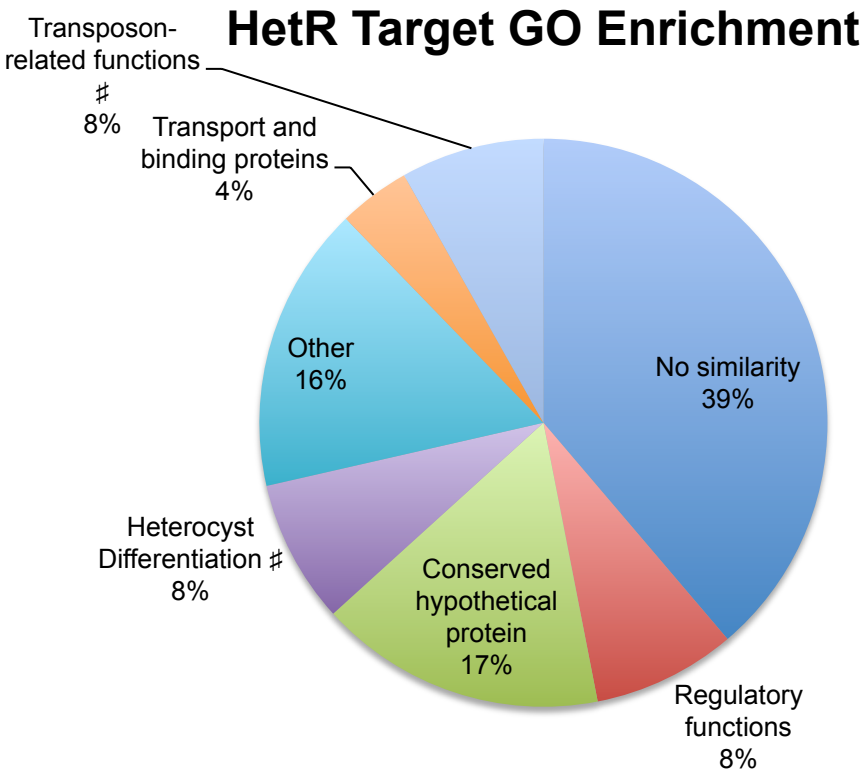



**Figure 4.2: GO Term enrichment in HetR target genes identified by ChIP.** Significant enrichment in the GO Term categories "Heterocyst Differentiation" and "Transposon Related Functions" was seen in putative targets of HetR as identified by ChIP. These two categories are at least 3-fold enriched over their relative proportion in the genome as a whole. Each GO term is labeled with the percent of ChIP hits that fall within that category.

### 4.3.4 *gfp*-Promoter Fusions for HetR Targets

We chose six putative HetR targets for further analysis. Four of these targets, *patA, alr3758, asl2028,* and *alr1469* all contain a putative HetR binding site while two of the targets, *all2242,* and *all5131* do not. Each of these six genes is downregulated in a Δ*hetR* strain and many of them are regulated in response to nitrogen deprivation (Table 4.4). To analyze the temporal and spatial expression of these genes in response to nitrogen deprivation, we fused the promoter of each gene (with the exception of *patA,* which has already been studied in the literature) to *gfp* in both wild-type *Anabaena* and UHM103, a markerless knockout mutant of *hetR* (11). For the four genes that have putative HetR binding sites, we also fused a truncated version of the promoter, missing this putative HetR binding site, to *gfp*.

## Table 4.4: HetR Targets Chosen for Further Analysis

| Region | Peak location | Possible regulated gene (fold change in hetR delta) | Role of regulated gene | HetR Binding site |
|---|---|---|---|---|
| +(1732738..1733693) | asr1469 promoter | asr1469 (-4) | unknown small peptide | TTGGGTCTAACTTAT |
| +(613436..613919) | patA promoter and 5' end | patA (-8) | pattern formation | TTGGCTCAAACCCAT |
| +(2427714..2428147) | asl2028 promoter and coding region | asl2028 (low coverage) | unknown hypothetical | ACGGGGTAACAATCC |
| +(4539927..4540364) | 5' end of and promoter of alr3758 | alr3758 (11 reads to 0) | anti sigma factor antagonist | TTGGGTAAACCTGCT |
| +(6129044..6129475) | all5131 promoter | all5131 (-4.2) | transcriptional regulator | none found |
| +(2693489..2693944) | 5' end of and promoter of alr2242 | alr2242 (1.3) | unknown possible ATPase | none found |

PatA is a well-studied protein involved in heterocyst development, and two previous studies have shown that expression of *patA* increases in a heterocyst specific manner after nitrogen deprivation (72, 79, 131). *patA-gfp* expression is low in a wild type background, but in a ∆*patA* strain in N- conditions *patA-gfp* is clearly on in all cells and upregulated in heterocysts (131). This suggests that *patA* is a target of HetR in both vegetative and heterocyst cells and that its accumulation in heterocysts may be due to an increase in HetR.

A PatA-GFP translational fusion shows that PatA forms FtsZ-like rings around cells. A ∆*patA* strain has aberrant cell morphology and forms bloated vegetative cells which only differentiate heterocysts at the ends of filaments in N- media (72). Therefore, it has been suggested that PatA is involved in the coupling of cell division and heterocyst differentiation. PatA contains both a helix-turn-helix DNA binding domain and an N-terminal PATAN protein interaction domain, suggesting its DNA binding activity may be regulated by binding partners or phosphorylation.

Previous work has shown that PatA promotes transcription of *hetR* while simultaneously limiting HetR accumulation in cells. Deletion of *patA* in conjunction with overexpression of HetR results in the formation of nearly three times the number of heterocysts compared with overexpression of HetR alone (14). The ∆*patA* phenotype can be bypassed by inhibition of the HetR inhibitors PatA and HetN or by or overexpression of the heterocyst

differentiation protein HetF (14). Therefore, it is hypothesized that PatA and HetN have similar roles in the promotion of HetR activity, although the molecular mechanism for the interaction between PatA and HetR is not yet understood.

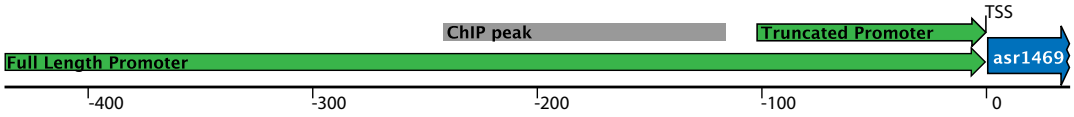
PatA is expressed in vegetative cells prior to nitrogen deprivation and its expression increases in heterocysts in N- media. Previous bioinformatic analysis suggested that *patA* has a strong HetR binding site and our ChIP dataset confirmed this interaction *in vivo*. RNA-seq analysis of a ∆*hetR* strain showed an 8-fold decrease in *patA* transcription in the absence of HetR {Johnson, unpublished data #177}. It is possible that *patA* and *hetR* regulate one another through a feedback loop both in vegetative cells and during heterocyst development, as both of these DNA-binding proteins appear to affect one another in overexpression or knockout experiments. Further work on the molecular role of PatA in cell division and on other possible PatA DNA binding targets may expand our understanding of the function of PatA during heterocyst development.

The *asr1469* gene encodes a small protein conserved in a few diverse cyanobacteria but with no known domain homologies. However, *asr1469* is upstream of a putative cyanophycin synthase gene and may be co-transcribed with this gene. Cyanophycin is a nitrogen storage molecule that is present in all cells, but accumulates in heterocysts and forms polar granules. A fusion of a truncated *asr1469* promoter containing 100 bases upstream of

the transcriptional start site but not the putative HetR binding site yielded no GFP fluorescence (Figure 4.2 and 4.3); however, it is possible that by truncating the promoter we also removed other binding sites important for regulation. Expression of the full-length promoter in WT cells yielded GFP fluorescence in all cells in N+ conditions. In contrast, expression of the full-length promoter of 450 bases yielded no GFP fluorescence in a Δ*hetR* background. In the WT strain after nitrogen deprivation, GFP fluorescence transiently increased in the majority of heterocysts at 21 hours. However, by 25 hours after the removal of combined nitrogen, most heterocysts showed no GFP fluorescence compared with vegetative cells. This suggests that regulation of the *asr1469* promoter in heterocysts is transient.

RNA-seq data showed a four-fold decrease in expression of *asr1469* in the Δ*hetR* strain, but no regulation of the gene in response to nitrogen deprivation. It is possible that the changes in transcription in *asr1469* seen in the GFP images are too subtle and transient to measure with RNA-seq. However, promoter-*gfp* fusions suggest that *asr1469* transcription is heavily dependent on HetR in vegetative cells as expression of GFP required both HetR protein and the full-length promoter containing the putative HetR binding site. Transient up-regulation of *asr1469* in heterocysts suggests that this gene and may play a role in development, possibly cyanophycin synthesis, and that it is not only on during the early stages of heterocyst differentiation, but that is on at a basal level in vegetative cells at all times.

A.




B.

| Peak location | Fold Change in Δ*hetR* | Fold Change after N step down | HetR Binding site |
|---|---|---|---|
| asr1469 promoter | 4 | low/no change | TTGGGTCTAACTTAT |

**Figure 4.2: Promoter-*gfp* fusions and RNA-seq expression of *asr1469*.**

Panel A shows the full length and truncated promoters used in this study. The transcriptional start site mapped here is from a previous TSS-specific RNA-seq analysis of *Anabaena* and is confirmed by our RNA-seq data (84). The ChIP peak associated with *asr1469* is marked with a grey bar. Panel B shows the expression of *asr1469* in the Δ*hetR* strain as well as its putative HetR binding site, located just upstream of the truncated promoter 5' end. There was no to low change in expression of this gene in response to nitrogen deprivation as measured with RNA-seq.

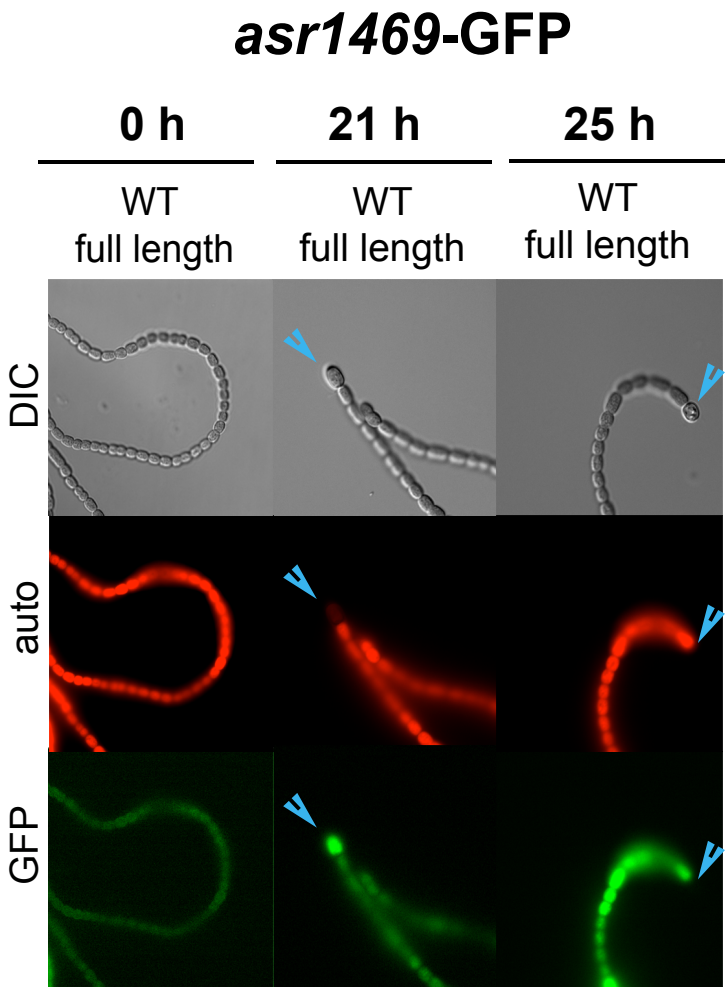
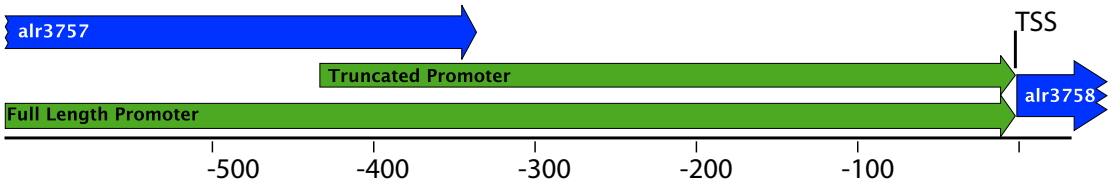



**Figure 4.3: Transient expression of *asr1469-gfp reporter fusion* in heterocysts.**
*Anabaena* WT cells were imaged at 0 (left), 21 (center), and 25 (right) hours after nitrogen deprivation in the DIC (top), TRITC/autofluorescence (middle), and GFP (bottom) channels. At 21 hours after nitrogen deprivation, the majority of heterocysts showed an increase in fluorescence when compared to vegetative cells (heterocysts are marked with a blue carrot). By 25 hours after nitrogen deprivation, the majority of heterocysts (the field of view was not subjected to a counting assay, but we can approximate more than 50% when we say “majority”) showed no GFP fluorescence, suggesting down-regulation of this promoter in a heterocyst-specific manner. The remaining heterocysts showed fluorescence similar to vegetative cells and it is possible that these heterocysts were somewhere in between the high and low states of expression of this gene.

A third putative HetR target, *alr3758*, is a putative anti-sigma factor antagonist. *alr3758* is off in a Δ*hetR* background and upregulated in response to heterocyst development in the RNA-seq data. *gfp* was fused to the full-length *alr3758* promoter as well as a truncated version of this promoter missing the putative HetR binding site (Figure 4.4). While we did not observe a change in GFP fluorescence of the full-length promoter in a WT background over the course of nitrogen deprivation in vegetative cells, we did see a dramatic increase in GFP fluorescence in heterocysts by 21 hours after the removal of combined nitrogen (Figure 4.5). This *gfp* fusion showed very low fluorescence in the Δ*hetR* background as well as when truncated in the WT background. This suggests that *alr3758* is upregulated in a heterocyst specific manner and may be involved in the inactivation of an anti-sigma factor to allow gene expression of heterocyst specific genes. Interestingly, there are a number of sigma factors that are regulated after nitrogen deprivation, and *alr3758* may play a role in allowing these new sigma factors to access their promoters by sequestering or inactivating an anti-sigma factor in heterocysts (2).

A.




B.

| Peak location | Read Change in Δ*hetR* | FOLD CHANGE IN RPKM 0 to 6 hours | 0 to 12 hours | 0 to 21 hours | HetR Binding site |
|---|---|---|---|---|---|
| 5' end of and promoter of alr3758 | 11 reads to 0 | 12 | 4 | 8 | TTGGGTAAACCTGCT |

**Figure 4.4: Promoter-*gfp* fusions and RNA-seq expression of *alr3758*.** Panel A shows the full length and truncated promoters used in this study. The transcriptional start site mapped here is from a previous TSS-specific RNA-seq analysis of *Anabaena* and is confirmed by our RNA-seq (84). The putative HetR binding site is located directly 5' of the site at which the truncated promoter begins, 431 base pairs upstream of the transcriptional start site of *alr3758.* Panel B shows the expression of *alr3758* in the Δ*hetR* strain as well as the fold change in RPKM from 0 to 6, 12, and 21 hours.

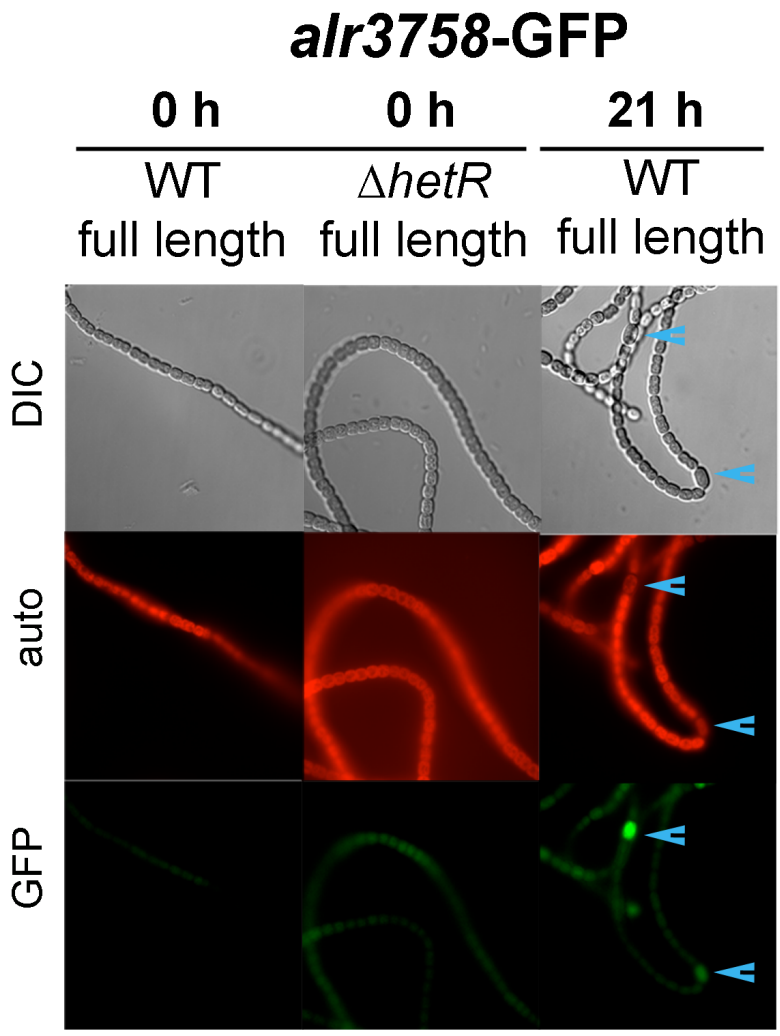


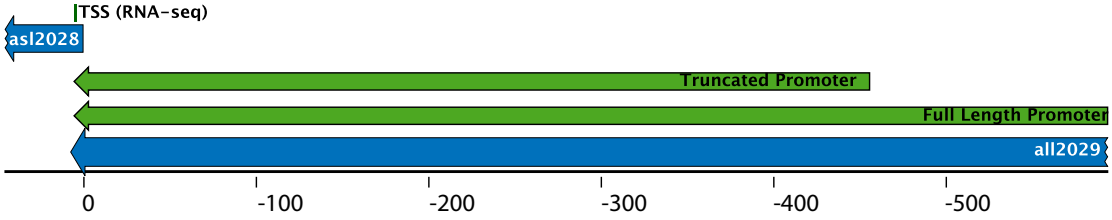

**Figure 4.5: Specific expression of *alr3758,* an anti-sigma factor antagonist, in heterocysts.**
*Anabaena* cells were imaged at 0 h (WT on left and Δ*hetR* in the center) and 21 h (right) after nitrogen deprivation in the DIC (top), TRITC/autofluorescence (middle), and GFP (bottom) channels. There is very low to no expression of this promoter in the WT and Δ*hetR* background at 0 h. At 21 hours after nitrogen deprivation, heterocysts showed an increase in fluorescence when compared to vegetative cells (heterocysts are marked with a blue carrot).

The *asl2028* gene encodes a hypothetical protein that is conserved in only a few species of cyanobacteria. However, *asl2028* is downstream of two nitrile hydratases, proteins involved in metabolism of nitriles as a fixed nitrogen source. At 6 and 12 hours after the removal of combined nitrogen in WT *Anabaena*, there is a slight increase in GFP fluorescence in the vegetative and proheterocyst cells in the entire filament, but the truncated promoter, as well as the full-length promoter in the Δ*hetR* strain, remain off (Figure 4.7). By 21 hours after the removal of combined nitrogen, *asl2028* promoter activity is still strong in WT vegetative cells, but is completely off in heterocysts. . Fusion of a truncated version of the *asl2028* promoter to *gfp*, missing the putative HetR binding site in a WT background, as well as the full length promoter in a Δ*hetR* background, yielded no fluorescence.

The *gfp* fusion data suggests that *asl2028* is a vegetative cell specific target of HetR and that HetR is important for its expression. Because this gene is expressed at very low levels, the RNA-seq data could not reliably detect a change in expression across the time course (Figure 4.6). While HetR is known to be expressed in all cells in the filament, to the best of our knowledge vegetative cell targets of HetR have not been studied to date. We suggest that HetR may modify it's regulon based on interacting protein partners, and that *asl2028* is an example of a gene that is only turned on by HetR in vegetative cells. HetR is conserved in non heterocystous cyanobacteria and, in our hands, the Δ*hetR* strain grows slightly slower and is

more clumpy than wild type cells. Therefore, it is likely that HetR has a distinct during vegetative cell growth, and *asl2028* may be a HetR target that is part of that regulon.

A.




B.

| Peak location | Fold Change in Δ*hetR* | Fold Change 0 to 6 hours | Fold Change 0 to 12 hours | Fold Change 0 to 21 hours | HetR Binding site |
|---|---|---|---|---|---|
| asl2028 promoter and coding region | none/very low coverage | -3.5 | 1.2 | 1.2 | CGGGGTAACAATCCT |

**Figure 4.6: Promoter-*gfp* fusions and RNA-seq expression of *asl2028*.** Panel A shows the full length and truncated promoters used in this study. The transcriptional start site mapped here is from a previous TSS-specific RNA-seq analysis of *Anabaena* and is confirmed by our RNA-seq experiment (84). Since this TSS is downstream of the annotated ATG, it is possible that *asl2028* is misannotated, and that the functional translational start site is the ATG 91 bp after the annotated TSS. Panel B shows the extremely low expression of *asl2028* in the Δ*hetR* strain as well as its putative HetR binding site, located just upstream of the truncated promoter 5' end.

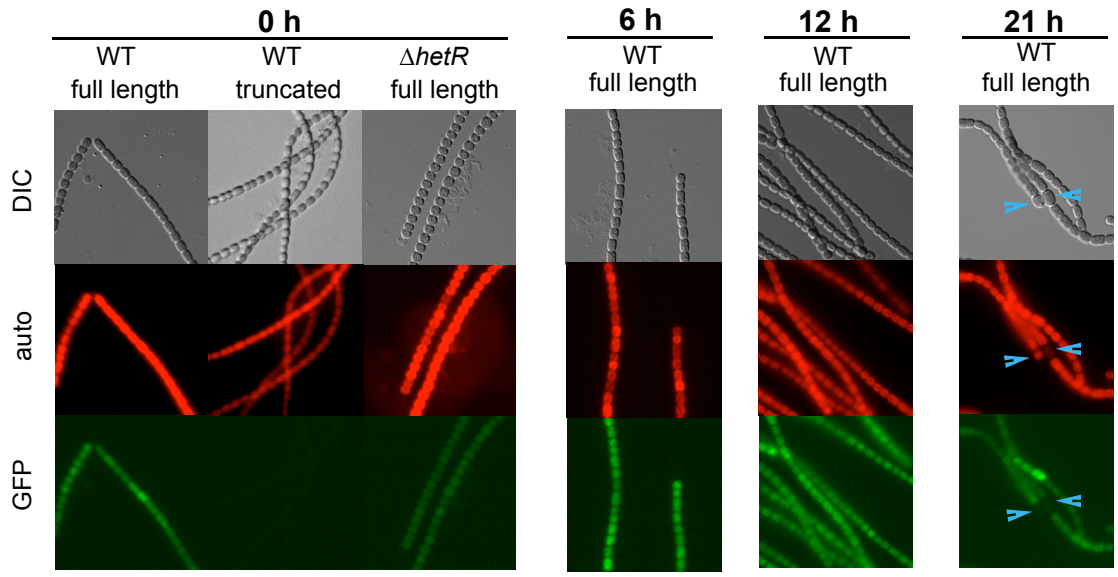


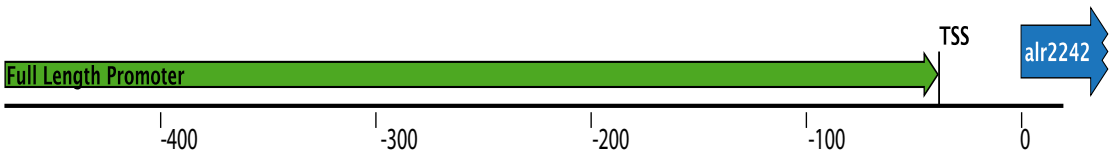

**Figure 4.7: *gfp* fusions show vegetative cell-specific expression of *asl2028*.**
The *asl2028* full length promoter fused to *gfp* in WT cells is expressed at a low level prior to nitrogen deprivation with a slight increase in expression at 6, 12, and 21 hours in vegetative cells. However, *asl2028* is off in heterocysts at 21 hours after nitrogen deprivation. A truncated *asl2028* promoter, missing the putative HetR binding site, is also off in WT cells and the full-length *asl2028* promoter is expressed at a slightly lower level in the Δ*hetR* mutant than in WT cells.

The *alr2242* gene encodes a putative HetR target in the highly-conserved NACHT family of NTPases. This domain includes proteins involved in signal transduction, DNA binding, and even some kinesin motor proteins. Although there was a strong enrichment of the *alr2242* promoter and 5' end in our ChIP data, we could not identify a putative HetR binding site upstream of *alr2242* through FiMO. It is possible that HetR binds a yet unidentified site in this promoter. We observed no change in *alr2242* in the Δ*hetR* strain in the RNA-seq data and a slight increase in expression of this gene late in heterocyst development (Figure 4.8).

Fusion of the full-length promoter of *alr2242* to *gfp* yielded no fluorescence in WT cells in N+ conditions, but strong GFP fluorescence in the Δ*hetR* mutant (Figure 4.9). By 21 hours after nitrogen deprivation, a small amount of GFP fluorescence was visible in WT vegetative cells but not in heterocysts. While this phenotype was conserved among three independent exconjugants, it is still possible that the GFP fusion data is an artifact. GFP fluorescence levels may be effected by a cell-type specific change in plasmid copy number or secondary changes in promoter regulation in the Δ*hetR* background that are not direct results of the deletion of HetR. However, it is also possible that *alr2242* is an example of a gene suppressed by HetR. In N- conditions, this suppression is slightly alleviated in vegetative cells. It is possible that HetR suppression of the *alr2242* promoter is inhibited by PatS or by other factors not present in vegetative cells in N+ conditions.

Expression of *alr2242* was not strong in the Δ*hetR* strain in the RNA-seq data, contrasting with our *gfp*-fusion experiments. It is possible that the *gfp* fusion data is an artifact, as stated above, and that deletion of HetR does not lead to an increase in alr2242 protein. To test this, we would need to assess copy number of the *alr2242-gfp* plasmid in both WT and Δ*hetR* strains via QPCR. In addition, to test whether the Δ*hetR* strain effects *alr2242* transcription through an indirect mechanism, we could pulse the Δ*hetR* strain carrying *alr2242-gfp* with HetR on an inducible copper promoter and look for suppression of the *alr2242-gfp* gene. However, the two pieces of data can also be explained if *alr2242* regulates its own expression, such that accumulation of alr2242 protein in the Δ*hetR* background in the RNA-seq experiments led to downregulation of the *alr2242* gene. In the *alr2242-gfp* fusion experiment, *alr2242* promoter activity is uncoupled from protein expression, and would not contribute to a negative feedback loop. Further work confirming the *gfp* promoter fusion data and assessing plasmid copy number will be necessary to confirm these results.

A.




B.

| Peak location | Fold Change in Δ*hetR* | FOLD CHANGE IN RPKM 0 to 6 hours | 0 to 12 hours | 0 to 21 hours | HetR Binding site |
|---|---|---|---|---|---|
| 5' end of and promoter of alr2242 | 1.3 | 1.4 | 1.2 | 2.2 | none |

**Figure 4.8: Promoter-*gfp* fusions and RNA-seq expression of *alr2242*.** Panel A shows the full length promoter used in this study. The transcriptional start site mapped here is from a previous TSS-specific RNA-seq analysis of *Anabaena* and is confirmed by our RNA-seq (84). Panel B shows the expression of *asl2028* in the Δ*hetR* strain compared to WT as well as in response to nitrogen deprivation.

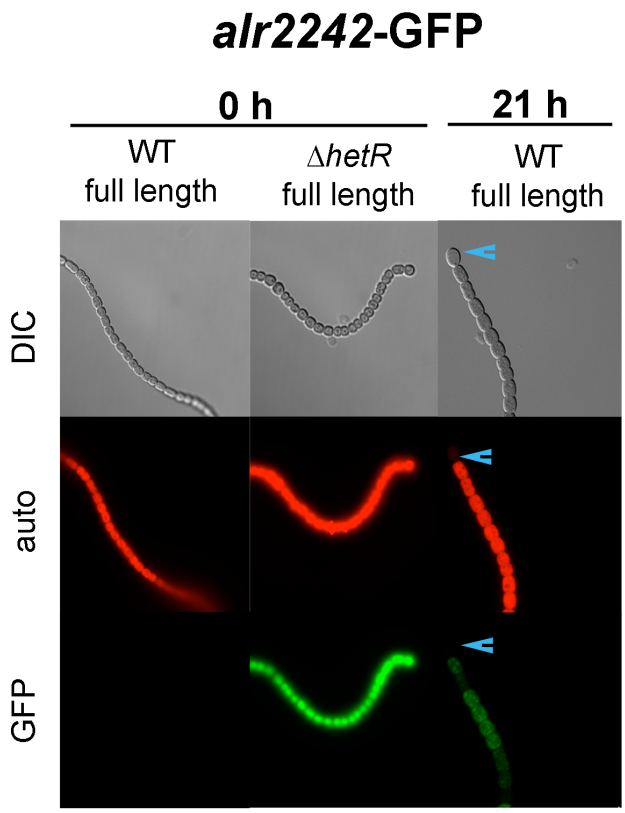
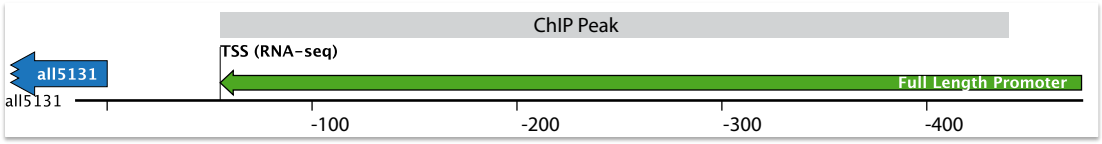



**Figure 4.9: *gfp* fusions show suppression of *alr2242* in Wild Type cells.** The *alr2242* full-length promoter fused to *gfp* in WT cells is not expressed prior to nitrogen deprivation. However, *alr2242* is highly expressed in the Δ*hetR* strain in N+ conditions, suggesting HetR may suppress *alr2242* expression. *alr2242* is off in heterocysts at 21 hours after nitrogen deprivation but slightly on in vegetative cells.

The *all5131* gene encodes a highly-conserved helix-turn-helix transcriptional regulator that is downregulated in the Δ*hetR* strain in our RNA-seq data but only mildly upregulated in response to nitrogen deprivation (Figure 4.10). A fusion of the promoter of *all5131* to *gfp* showed no change in regulation in response to nitrogen deprivation (data not shown). However, expression of this gene decreased in the Δ*hetR* strain, consistent with the RNA-seq data. Taken together, this data suggests that *all5131* encodes a transcriptional factor regulated by HetR in all cells, but expression of *all5131* is not entirely dependent on HetR. *all5131* may be an example of a gene regulated by HetR as part of HetR's basal activity in all cells and is a good candidate for the future study of the role of HetR outside of heterocyst development.

A.




B.

| Peak location | Fold Change in Δ*hetR* | FOLD CHANGE IN RPKM 0 to 6 hours | 0 to 12 hours | 0 to 21 hours | HetR Binding site |
|---|---|---|---|---|---|
| all5131 promoter | -4.2 | 1.7 | 1.4 | 1.7 | none |

**Figure 4.10: Promoter-*gfp* fusions and RNA-seq expression of *all5131*.** Panel A shows the full length promoter used in this study. The transcriptional start site mapped here is from a previous TSS-specific RNA-seq analysis of *Anabaena* and is confirmed by our RNA-seq (84). Panel B shows the expression of *all5131* in the Δ*hetR* strain compared to WT as well as in response to nitrogen deprivation.

## 4.4 CONCLUSIONS

We employed a combination of ChIP-seq, RNA-seq, and promoter-*gfp* fusions to elucidate the HetR regulon at one time point, 6 hours, after nitrogen deprivation. ChIP-seq expanded our understanding of the HetR regulon, allowing more precise definition of the (or a) putative HetR binding site as well as identification of promoters potentially regulated by HetR that do not contain the canonical HetR binding site. Furthermore, we identified specific genes that are regulated by HetR in a heterocyst specific or vegetative cell specific manner, or in both cell types.. This study sheds light on the HetR regulon in vegetative cells and may open the door for future studies to determine the role of this highly conserved protein outside of heterocyst development. Furthermore, our work adds to the growing use of chromatin immunoprecipitation couple with deep sequencing in cyanobacteria, specifically *Anabaena*, and we have developed an improved ChIP protocol to look at transcription factor regulons in this bacterium.

# Chapter 5: ChIP of DevH

## 5.1 INTRODUCTION

The nitrogenase enzyme, which breaks the $N_2$ triple bond, is inhibited in the presence of oxygen. In nitrogen fixing organisms such as *Anabaena* sp. strain PCC 7120, active nitrogenase is expressed in a microoxic environment. *Anabaena* forms specialized cells called heterocysts, which fix nitrogen. Heterocysts are evenly spaced along a filament and achieve low oxygen content by shutting down oxygenic photosynthesis, increasing respiration, and by creating an cell envelope that limits the diffusion of gases (92). This cell wall consists of an inner glycolipid layer, which decreases oxygen movement into cells, and an thick outer polysaccharide outer layer.

DevH is a helix-turn-helix transcription factor in the CRP family of transcription factors (49). DevH is required for nitrogen fixation in an oxic environment; however, in anoxic conditions, a Δ*devH* strain can fix atmospheric nitrogen. This ability to fix nitrogen only in the absence of oxygen is called a Fix+ Fox- phenotype, and is consistent with DevH's role in formation of the heterocyst-specific cell wall.

A DevH mutant does not form the heterocyst laminated glycolipid layer, suggesting that DevH is required for making the heterocyst cell envelope oxygen impermeable, thus resulting in a Fox- phenotype (100).

Furthermore, DevH is required for expression of the *hglE1* and *hglE2* genes, both of which are involved in synthesis of the heterocyst specific glycolipid layer. Expression of other *hgl* genes, such as *hglB, hglC,* and *hglD,* is dramatically reduced in a ∆*devH* strain (55, 100). Direct analysis of DevH DNA binding to the *hgl* genes has not been published to date.

It is possible that DevH binds directly to the promoters of the *hgl* genes or that DevH regulates heterocyst-specific glycolipid expression in a less direct manner, through other transcription factors, as has been suggested previously {Ramírez, 2005 #156}. Therefore, we employed ChIP-seq on 6xHis-tagged DevH to get a global view of its regulon in response to nitrogen deprivation. Our work identified new genes potentially regulated by DevH and suggests that DevH binds well to many targets that are not the *hgl* genes, including transcriptional regulators, identifying new candidates for the study of the transcriptional network that regulates glycolipid production.

## 5.2 MATERIALS AND METHODS

### 5.2.1 Cell Growth Conditions and Nitrogen Deprivation

*Anabaena* (*Nostoc*) sp. strain PCC 7120 cultures were grown in 100 ml or 2 ml of liquid medium in 250-ml flasks with cotton plugs as previously described with slight modifications (46). Briefly, 100-ml or 2-ml liquid cultures were grown to an $OD_{750}$ of 0.02 in BG-11($NH_4$) medium, which lacked sodium nitrate and contained 2.5 mM ammonium chloride and 5 mM MOPS (pH 8.0).

For nitrogen deprivation, cultures were spun down at 4,000 x g for 5 minutes and washed three times in BG-$11_0$ medium by centrifugation and decanting of the supernatant. Cells were then resuspended in 100 ml or 2 ml BG-$11_0$ at a final $OD_{750}$ of between 0.02 and 0.05. Cells were grown for 6 hours in BG-$11_0$ shaking at 100 µmol photons $m^{-2}$ $s^{-1}$.

### 5.2.2 Chromatin Immunoprecipitation

18 hours after nitrogen deprivation, WT and DevH-6xHis cells were spun down at 4,000 x g for 5 minutes, then resuspended in 5 mL BG-$11_0$. Crosslinking was performed by the addition of 4.1 mg DSG and 0.56 mg EGS in 500 µL DMSO. Crosslinking was performed at room temperature, rocking, for 20 minutes. After 20 minutes, 135 µL of 37% formaldehyde was added for additional protein to DNA crosslinking and left rocking at room temperature for 15 minutes. To quench the reaction, 125 mM glycine was added for 5 minutes at room temperature.

Cells were then spun down at 4,000 x g for 5 minutes at 4°C and washed twice in 30 mL ice cold PBS (137 mM NaCl, 2 mM KCl, 10 mM $Na_2HPO_4$, 1.8 M $KH_2PO_4$, pH 7.4). Washed and fixed pellets were resuspended in 500 µL ice-cold binding/wash buffer (100 mM $NaHPO_4$, 600 mM NaCl, 0.02% Tween 20, 1 EDTA Proteinase Inhibitor Tab from Roche Biosciences in 10 mL total volume) on ice. Protein was extracted by bead beating 2 x 5 minutes with 2 minutes on ice between pulses. Complete lysis was confirmed by microscopy. Lysed cells were separated from beads via

centrifugation and DNA was sheared via sonication on ice, 12 cycles of 20 seconds on, 15 second off at 14% power. Cell debris was pelleted via two cycles of centrifugation at 14,000 x g for 15 minutes at 4°C. Protein concentration was determined by absorbance at 280 nm and normalized to 20 mg/mL for each sample by dilution in cold binding/wash buffer.

His-tagged DevH was isolated with the Dynabeads His-Tag Isolation & Pulldown kit (Invitrogen) following manufacturer's protocol at 4°C and eluted in 100 µL elution buffer (100 mM imidazole, 50 mM $NaPO_4$, 300 mM NaCl, 0.01% Tween 20). Crosslinks were reversed at 65°C for 18 hours in a PCR block. Input DNA was used to assess size distribution after shearing on a 1% agarose gel. Pulldown efficiency was measured via western blotting of DevH-6xHis with Qiagen Penta-His antibody, BSA Free. After crosslinks were reversed, proteins were digested by the addition of 250 µL TE, 4 µL 20 µg/µL glycogen, and 10 µL 10 µg/µL proteinase K for 2 hours at 37°C. DNA was column purified with the Promega SV DNA purification kit and resuspended in 30 µL nuclease free water.

### 5.2.3 DNA Library Preparation and Sequencing

DNA was prepared for sequencing with the Illumina ChIP-seq Sample prep kit by the Next Generation Sequencing Core at The Scripps Research Institute (La Jolla, CA) following manufacturer's protocol. Sequencing was performed on the Illumina HiSeq platform with 4 samples multiplexed on one cell, yielding approximately 40 million 40 bp reads per sample. Sequence

reads were filtered based on adapter sequences using the Illumina Flicker add-on and saved as fasta files for analysis in CLC Genomics Workbench 5.

### 5.2.4 Sequence Alignment and Peak Finding

Sequencing reads from the experimental DevH-6xHis sample were randomly assigned to two files and the full ChIP-seq analysis was performed on each sample as a technical replicate. ChIP-seq reads from the two DevH-6xHis samples and the WT sample were aligned to NCBI's current build of the *Anabaena* genome with CLC Genomics Workbench 5. ChIP-seq peaks were called using a 100 bp window and a false discovery rate of 5% with the WT sample as the control using CLC's ChIP Analysis pipeline.

## 5.3 RESULTS AND DISCUSSION

ChIP-seq on 6xHis tagged DevH was performed 18 hours after the removal of combined nitrogen. Reads were randomly assigned to two sample sets and were normalized to a WT sample (as detailed in Chapter 4 for HetR-6xHis) to control for sequencing bias Initial peak finding analysis with CLC Genomics workbench 5.0 ChIP-seq pipeline of both DevH-6xHis randomized samples yielded 74 ChIP-seq peaks that were close to the transcriptional start site of a gene (Table 5.1). Many of these peaks are near a gene that is dramatically downregulated at 21 hours after nitrogen deprivation. It is possible that, in these cases, DevH acts as a negative regulator of transcription. Other genes, such as the 10 two-component system genes as

well as genes involved in exopolysaccharide synthesis and cell envelope remodeling, are upregulated at 21 hours after the removal of combined nitrogen. These genes are good candidates for future studies on the factors involved in creating the heterocyst-specific cell envelope.

Although a member of the *dev* operon, *devA*, (responsible for movement of glycolipids to the cell envelope) was identified in the 96 ChIP-seq peaks, no members of the *hgl* family were identified. However, DevH appears to regulate the expression of many other genes, including cell envelope modifying enzymes and transcriptional regulators in our ChIP dataset. It is possible that DevH's binding to *hgl* promoters is transient or weak and, therefore, not captured in this single 18 h time-point ChIP experiment. Our work suggests that DevH plays a role in the expression of many cell-envelope modifying genes that had not been previously analyzed in conjunction with heterocyst development.

A small subset of 10 of the 74 ChIP peaks were selected as high-confidence peaks (Table 5.2). High-confidence peaks are conserved in both randomized samples and are in the promoters of genes which are differentially regulated at 21 hours after the removal of combined nitrogen, as determined by previous RNA-seq studies. The 10 peaks are in the 5' UTR or coding region of a variety of genes, including a transcriptional regulator, a two-component sensor histidine kinase, and many metabolic genes and glycosyl transferases. Because these genes were identified in two random

samples, they are good candidates for future gene-specific work on the DevH binding site and regulon. Mobility shift experiments on portions of these promoters may help define an exact DevH binding site, and *gfp* fusions of these promoter may confirm when and where DevH is activating gene expression.

Analysis with MEME (as was done on HetR in Chapter 3) to identify a putative DevH binding site yielded no high-confidence sites. However, if the DevH binding site is similar to that of NtcA, then the site is ubiquitous throughout the genome and cannot be enriched in a statistically significant manner in a sample set of 74 large ChIP fragments. It is possible that a more directed search, possibly based on the NtcA binding site or on mobility shift work, may have more success in identifying a conserved DevH binding site. Furthermore, ChIP-seq at more time points or in a larger scale, with more protein, may increase the number of ChIP-seq peaks and increase the input sample size for MEME, increasing the likelihood of finding a DevH consensus binding sight with a significant p value. In addition, gene specific mobility shifts on portions of the “high confidence” ChIP hits, outlined above, may also narrow down the region of DNA used as an input for MEME analysis, again increasing the likelihood of discovering a DevH binding site in this dataset.

## 5.4 CONCLUSIONS

Overall, our work defined 74 genes that may be part of the DevH regulon and provides a small subset of 10 of these genes for future molecular studies. In addition, we have successfully “ChIPped” a second protein in *Anabaena*, demonstrating that our ChIP protocol is a relatively robust way to analyze transcription factor binding sites in this organism. Future detailed work on DevH DNA binding through promoter *gfp* fusions and mobility shifts may use this study as a platform to further analyze the DevH regulon and define the DevH binding site(s).

# Table 5.1: DevH ChIP Hits

| Region | Gene Regulated | Probable Function | Fold Change RPKM at 21 Hours |
|---|---|---|---|
| +(871925..872108) | alr0750 | hypothetical | -51.5 |
| +(736731..736917) | asr0636 | unknown protein | -33.3 |
| +(1298665..1298851) | alr1109 | hypothetical | -16 |
| +(1301485..1301692) | asl1111 or alr1112 | hypothetical, probable transglycosylase | -4.3, 2.3 |
| +(4868115..4868300) | all4041 | hypothetical protein | -3.5 |
| +(6347852..6348040) | alr5320 | hypothetical protein | -3.3 |
| +(1463352..1463547) | asr1234 | unknown | -2 |
| +(1029585..1029780) | all0889 | hypothetical | -1.8 |
| +(2964786..2964973) | ar2467 | hypothetical | -1.5 |
| +(2474904..2475089) | all2067 | transposase | -1.2 |
| +(4061841..4062031) | alr3356 | similar to phtochrome | -1 |
| +(1690431..1690621) | all1427 | hypothetical | -1 |
| +(2427688..2427869) | all2027 | nitrile hydratase | -1 |
| +(1456863..1457053) | alr1231 | two componenet hybrid sensor and regulator | 1 |
| +(1647109..1647296) | all1387 | unknown | 1 |
| +(3485335..3485516) | alr2861 | unknown | 1 |
| +(3757897..3758078) | alr3106 | unknown protein | 1 |
| +(3175147..3175343) | all2644 | peptide synthase | 1 |
| +(1456866..1457048) | alr1231 | two component hybrid sensor and regulator | 1 |
| +(656081..656269) | all0561 | similar to XisI protein | 1.1 |
| +(542148..542329) | all0456 | precorrin isomerase | 1.13 |
| +(3506708..3506889) | all2875 | two component hybrid sensor and regulator | 1.2 |
| +(4385430..4385629) | all3633 | glucanase | 1.2 |
| +(2599919..2600100) | all2164 | two component system regulatory protein | 1.2 |
| +(5355868..5356055) | asr4470 | hypothetical protein | 1.2 |
| +(4385430..4385634) | all3633 | glucanase | 1.2 |
| +(5317425..5317607) | all4433 | unknown protein | 1.29 |
| +(1142846..1143041) | all0978 | two component sensor histidine kinase | 1.4 |
| +(6142968..6143159) | alr5148 | unknown protein | 1.4 |
| +(1142848..1143029) | all0978 | two component hybrid sensor and regulator | 1.4 |
| +(1744601..1744793) | all1480 | unknown protein | 1.4 |
| +(4178820..4179005) | alr3466 | WD-40 repeat protein | 1.5 |
| +(5156129..5156313) | all4298 | unknown protein | 1.5 |
| +(343766..343949) | alr0303 | hypothetical | 1.6 |
| +(2036216..2036399) | all1696 | fibrinogen-binding protein | 1.7 |
| +(5837517..5837709) | all4896 | hypothetical protein | 1.7 |
| +(3882832..3883013) | alr3215 | unknown protein | 1.7 |
| +(2750742..2750924) | all2281 | two-component response regulator | 1.8 |
| +(476824..477040) | all0406 | unknown protein | 1.9 |
| +(1523692..1523878) | all1280 | two component sensor histidine kinase | 2.1 |
| +(1384729..1384913) | all1176 | hypothetical | 2.2 |
| +(4218214..4218402) | all3502 | unknown protein | 2.2 |
| +(2854010..2854191) | alr2372 | ABC transporter | 2.3 |
| +(5355869..5356052) | asr4471 | hypothetical | 2.6 |
| +(2479568..2479776) | alr2073 | N-acetyl glutamate transferase | 1.5, 2.6 |
| +(2479564..2479783) | alr2073 | glutamate N-acetyltransferase | 2.6 |
| +(3689234..3689418) | all3045 | unknown protein | 2.7 |
| +(928567..928758) | alr0803 | hypothetical protein | 2.9 |
| +(4768734..4769054) | alr3952 | transcriptional regulator | 3 |
| +(2144137..2144321) | all1784 | ribokinase | 3.1 |
| +(3942923..3943146) | all3279 | unknown protein | 3.2 |
| +(4457587..4457769) | alr3692 | hypothetical | 3.3 |
| +(3748947..3749165) | alr3097 | hypothetical protein | 3.4 |
| +(3715943..3716132) | alr3067 | glycosyl transferase | 3.5 |
| +(4180189..4180549) | all3465 | unknown protein | 3.6 |
| +(3690584..3690766) | all3047 | unknown protein | 3.8 |
| +(1611699..1611925) | all1357 | coproporphyrinogen III oxidase (*hemF*) | 4 |
| +(4820212..4820401) | isiA | photosystem II chlorphyll a-binding protein | 4 |
| +(1276756..1276963) | all1089 | hypothetical | 4.3 |
| +(3488270..3488451) | alr2864 | glycosyl transferase | 6.2 |
| +(4121557..4121739) | alr3410 | nicotinamide nucleotide transhydrogenase | 6.4 |
| +(4373178..4373364) | asr3621 | unknown protein | 7.8 |
| +(1774285..1774483) | asl1513 | hypothetical protein | 9.3 |
| +(3124694..3124885) | all2617 | hypothetical protein | 9.6 |
| +(5463760..5463988) | alr4570 | glutathione S-transferase | 12.5 |
| +(5305385..5305568) | all4426 | glucosyltransferase | 16.3 |
| +(65436..65631) | all0059 | exopolysaccharide biosynthesis | 19.2 |
| +(4481476..4481658) | devA | heterocyst specific ABC transporter | 70 |
| +(3104821..3105002) | alr2596 | ferrichrome-iron receptor | infinite (0 reads to 13) |
| +(5354074..5354255) | alr4468 | probable phosphinothricin N-acetyltransferase | infinite (0 reads to 8) |
| +(2320568..2320749) | asl1934 | unknown | no coverage |
| +(3922952..3923139) | alr3246 | hypothetical | no coverage |
| +(1252279..1252471) | all1071 | two component response regulator | no coverage |
| +(5555819..5556035) | alr4654 | transglycosylase-associated protein | no coverage |

## Table 5.2: High Confidence DevH ChIP Hits

| Region | Peak Location | Possible Gene Regulated | Fold change RPKM of regulated gene at 21 hours | Possible Role of Gene Regulated |
|---|---|---|---|---|
| +(928798..928990) | promoter of alr0803 | alr0803 | 2.9 | hypothetical |
| +(1276753..1276969) | coding region and 5' UTR of all1089 | all1089, all1088 | 4.3, 8.7 | hypothetical, two component sensor histidine kinase |
| +(245431..245626) | 5' end of alr0230 | alr0230 | 2.8 | unknown |
| +(1611699..1611925) | 5' end and promoter of all1357 | all1357 | 4 | coprophorphyrinogen III oxidase |
| +(3488270..3488451) | coding region of alr2863, upstream of alr2864 | alr2864 | 6.2 | glycosyl transferase |
| +(3715943..3716132) | coding region of alr3066 | alr3067 | 3.5 | probable glycosyl transferase |
| +(4768754..4769054) | between all3951 and alr3952 | all3951, alr3952 | 1.2, 3 | unknown, transcriptional regulator |
| +(1774284..1774487) | between alr1514 and asl1513 | alr1514, asl1513 | infinite (0 reads to 3 reads) | hypotheticals |
| +(871922..872108) | coding region or alr0750 | alr0751 | -38.6 | NADH dehydrogenase I chain E |
| +(2479568..2479776) | between divergent all2072 and alr2073 | all2072, alr2073 | 1.5, 2.6 | threonine synthase, glutamate N-acetyltransferase |

# Chapter 6: Conclusions

Cyanobacteria are a diverse group of photosynthetic microorganisms with the potential to be a viable source of biofuels or renewable chemicals. Genetic engineering of these organisms will rely heavily on our understanding of their gene expression and regulatory networks. Fortunately, cyanobacteria have been the subject of studies in bacterial cell biology, genetics, and microbiology for decades. Genetic tools and biochemical techniques developed in cyanobacteria from these earlier studies will be integral to the use of cyanobacteria in future applications in industry. However, there is still a great deal to learn about cyanobacterial molecular biology and the application of modern techniques to study cyanobacterial cells is vital to propelling the field of cyanobacterial research forward.

In this work, we applied deep sequencing technology to get a global view of the transcriptional networks that regulate heterocyst development. First, we used RNA-seq on the Illumina short read platform to sequence the *Anabaena* transcriptome at 0, 6, 12, and 21 hours after the removal of nitrogen from the media. With this “time course” analysis, we captured the transcriptional profile of *Anabaena* prior to nitrogen removal (0 h), as early heterocyst development transcriptional regulators like HetR are beginning to be expressed (6 h), just after proheterocysts have committed to becoming heterocysts (12 h), and when fully formed heterocysts are actively fixing nitrogen (21 h). In addition to identifying new genes that are up or down

regulated after combined nitrogen removal, our directional RNA-seq data can be used to map 5' UTRs and operons as well as identify antisense transcripts in genes of interest. These data can now be used as the foundation for future gene-specific studies in *Anabaena*, where knowledge of gene boundaries, operons, and gene expression during nitrogen deprivation may inform experimental design. In addition, our RNA-seq data is particularly useful as it was generated in a well-studied cyanobacterium with a fully sequenced and well-annotated genome. Therefore, these data can be used in combination with many other RNA-seq experiments as a "training set" as researchers look for robust computational ways of identifying transcription and transcriptional networks.

We focused on one particular antisense transcript identified in this study, *anti-nblA*, and suggest that this transcript is a negative regulator of the *nblA* gene. Since NblA triggers degradation of the phycobilisome light harvesting complexes, and this degradation would be harmful to a growing vegetative cell, it is possible that *anti-nblA* plays vital a role in keeping *nblA* expression off during vegetative cell growth. Future work outlining the mechanisms of this and other antisense RNAs in their potential regulation of target genes in cyanobacteria will expand upon our understanding of cyanobacterial gene networks. In addition, antisense RNAs have been used as genetic tools to decrease the expression of target genes, and understanding antisense RNA function in *Anabaena* may prove fruitful in the

development of antisense RNAs as a genetic tool in cyanobacterial genetic engineering {Collier, 1994 #107}.

We also applied chromatin immunoprecipitation coupled with deep sequencing, or ChIP-seq, to map the regulons of two transcriptional regulators involved in heterocyst development: HetR and DevH. HetR is a key regulator of heterocyst development and our work confirmed its role in regulating genes involved in heterocyst development, such as *hetP* and *patS*, while identifying new targets of HetR that are specifically regulated in heterocysts, such as *alr3758* and *as1469*. However, we also uncovered HetR binding upstream of genes that appear to play a role in vegetative cells. To the best of our knowledge, HetR's role outside heterocyst development has not been studied; however the fact that HetR is conserved in nonheterocystous cyanobacteria and is expressed in vegetative cells suggests that HetR may play an important role during vegetative cell growth. Therefore, our data provides the first concrete examples of HetR targets that are specifically expressed in vegetative cells, and may open the door for future work on the role of HetR in vegetative cell development, possibly in pattern formation. Finally, our ChIP-seq data helped to further refine the consensus HetR binding site, which was previously too stringent and could not be found upstream of genes known to be regulated by HetR. This newly defined binding site will aid future searches for HetR targets as we continue to expand our understanding of gene networks in *Anabaena.*

We also used ChIP-seq to elucidate the regulon of DevH, a transcription factor required for the expression of the heterocyst-specific glycolipid (*hgl*) genes, which are involved in morphogenesis of the cell envelope. Our ChIP-seq analysis found DevH binding sites upstream of many genes involved in cell wall synthesis as well as multiple transcriptional regulators, but not the *hgl* genes. It is possible that DevH's binding to the *hgl* genes is too transient to be captured at a single time point in ChIP; it is also possible that DevH regulates glycolipid gene expression via other transcriptional regulators. Further work on DevH's regulon, including mobility shift assays to define a consensus binding site and ChIP-seq on a larger sample set with multiple time points, will be required to complete our understanding of where and when DevH binds DNA to regulate deposition of the heterocyst-specific glycolipid layer. However, our data takes the first step in analyzing this transcription factor via ChIP-seq, and we believe the ChIP targets identified here will lay the groundwork for these future studies.

Cyanobacteria are a ubiquitous, ancient, and long-studied phylum, but these small and diverse organisms have enjoyed a new burst of interest as potential producers of biofuels and renewable chemicals. As industrial labs aim towards developing production strains, cyanobacteria like *Anabaena* may contribute vital properties, such as the ability to fix nitrogen and differentiate specialized cells. A thorough understanding of the regulatory networks and gene expression that govern how *Anabaena* senses its environment, alters

gene expression, and differentiates morphologically distinct heterocysts will contribute not only to studies which use *Anabaena* directly, but may be useful in the engineering of cyanobacterial production strains in general. For instance, a viable industrial strain may need to carry *Anabaena* nitrogen fixation genes to be competitive in the environment. Engineered strains may be able to hijack specific nitrogen transporters or cell envelope enzymes from *Anabaena* to create a strain that grows specifically in the industrial medium, thus outcompeting invasive species that do not carry these genes in the environment. In addition, *Anabaena* is currently used as a fertilizer for rice paddies, but with an increased knowledge of the regulation of nitrogen fixation in this organism, we may be able to increase its nutritional content and utilize fewer grams of fertilizer per acre. These and other strain engineering experiments will benefit from our increased understanding of the *Anabaena* transcriptome, and from future genetic experiments in *Anabaena.*

This study contributes to the growing body of deep sequencing data in cyanobacteria, helping to define gene networks and outline gene expression in response to environmental queues. Our work is an early step in the process of mapping transcriptomes and regulons on a global scale in one cyanobacterium, *Anabaena*, but this step and others like it are vital to propelling the field of cyanobacterial genetics forward and addressing the need for green energy and renewable chemicals.

# References


1. **Alam, J., R. Whitaker, D. Krogmann, and S. Curtis.** 1986. Isolation and sequence of the gene for ferredoxin I from the cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **168:**1265.

2. **Aldea, M. R., R. A. Mella-Herrera, and J. W. Golden.** 2007. Sigma factor genes *sigC, sigE*, and *sigG* are upregulated in heterocysts of the cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **189:**8392-8396.

3. **Alkan, C., E. Karakoc, J. H. Nadeau, S. C. Sahinalp, and K. Zhang.** 2006. RNA-RNA interaction prediction and antisense RNA target search. *Journal of Computational Biology* **13:**267-282.

4. **André, G., S. Even, H. Putzer, P. Burguière, C. Croux, A. Danchin, I. Martin-Verstraete, and O. Soutourina.** 2008. S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of *Clostridium acetobutylicum*. *Nucleic Acids Research* **36:**5955-5969.

5. **Awai, K., S. Lechno-Yossef, and C. P. Wolk.** 2009. Heterocyst Envelope Glycolipids, p. 179-202. *In* H. Wada and N. Murata (ed.), *Lipids in Photosynthesis: Essential and Regulatory Functions*. Springer, New York.

6. **Bagg, A., and J. Neilands.** 1987. Ferric uptake regulation protein acts as a repressor, employing iron (II) as a cofactor to bind the operator of an iron transport operon in *Escherichia coli*. *Biochemistry* **26:**5471-5477.

7. **Baier, K., H. Lehmann, D. P. Stephan, and W. Lockau.** 2004. NblA is essential for phycobilisome degradation in *Anabaena* sp. strain PCC 7120 but not for development of functional heterocysts. *Microbiology* **150:**2739-2749.

8. **Bailey, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li, and W. S. Noble.** 2009. MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research* **37:**W202-W208.

9. **Ballestar, E., and M. Esteller.** 2010. Examining DNA–Protein Interactions with Genome-Wide Chromatin Immunoprecipitation Analysis. *Modern Molecular Biology***:**33-45.

10. **Black, T. A., Y. Cai, and C. P. Wolk.** 2006. Spatial expression and autoregulation of *hetR*, a gene involved in the control of heterocyst development in *Anabaena*. *Mol. Microbiol.* **9:**77-84.

11. **Borthakur, P. B., C. C. Orozco, S. S. Young-Robbins, R. Haselkorn, and S. M. Callahan.** 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.

12. **Bothe, H., O. Schmitz, M. G. Yates, and W. E. Newton.** 2010. Nitrogen fixation and hydrogen metabolism in cyanobacteria. *Microbiol. Mol. Biol. Rev.* **74:**529-551.

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

14. **Buikema, W. J., and R. Haselkorn.** 2001. Expression of the *Anabaena hetR* gene from a copper-regulated promoter leads to heterocyst differentiation under repressing conditions. *Proc. Natl. Acad. Sci. USA* **98:**2729-2734.

15. **Chen, Y., N. Negre, Q. Li, J. O. Mieczkowska, M. Slattery, T. Liu, Y. Zhang, T.-K. Kim, H. H. He, J. Zieba, Y. Ruan, P. J. Bickel, R. M. Myers, B. J. Wold, K. P. White, J. D. Lieb, and X. S. Liu.** 2012. Systematic evaluation of factors influencing ChIP-seq fidelity. *Nat Meth* **9:**609-614.

16. **Collas, P.** 2010. The current state of chromatin immunoprecipitation. *Molecular biotechnology* **45:**87-100.

17. **Collas, P., and J. A. Dahl.** 2008. Chop it, ChIP it, check it: the current status of chromatin immunoprecipitation. *Frontiers in bioscience : a journal and virtual library* **13:**929-943.

18. **Collier, J. L., and A. Grossman.** 1994. A small polypeptide triggers complete degradation of light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria. *The EMBO journal* **13:**1039.

19. **Croucher, N. J., and N. R. Thomson.** 2010. Studying bacterial transcriptomes using RNA-seq. *Current Opinion in Microbiology* **13:**619-624.

20. **Danford, T., R. Dowell, S. Agarwala, P. Grisafi, G. Fink, and D. Gifford.** 2011. Discovering Regulatory Overlapping RNA Transcripts. *Journal of Computational Biology* **18:**295-303.

21. **de Alda, J. A., C. Lichtlé, J.-C. Thomas, and J. Houmard.** 2004. Immunolocalization of NblA, a protein involved in phycobilisome turnover, during heterocyst differentiation in cyanobacteria. *Microbiology* **150:**1377-1384.

22. **Deusch, O., G. Landan, M. Roettger, N. Gruenheit, K. V. Kowallik, J. F. Allen, W. Martin, and T. Dagan.** 2008. Genes of Cyanobacterial Origin in Plant Nuclear Genomes Point to a Heterocyst-Forming Plastid Ancestor. *Molecular Biology and Evolution* **25:**748-761.

23. **Ding, Y., and C. E. Lawrence.** 2001. Statistical prediction of single-stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond. *Nucleic Acids Research* **29:**1034-1046.

24. **Dismukes, G. C., D. Carrieri, N. Bennette, G. M. Ananyev, and M. C. Posewitz.** 2008. Aquatic phototrophs: efficient alternatives to land-based crops for biofuels. *Curr. Opin. Biotechnol.* **19:**235-240.

25. **Dornenburg, J. E., A. M. DeVita, M. J. Palumbo, and J. T. Wade.** 2010. Widespread antisense transcription in *Escherichia coli*. *MBio* **1**.

26. **Douglas, S. E.** 1998. Plastid evolution: origins, diversity, trends. *Current Opinion in Genetics & Development* **8:**655-661.

27. **Du, Y., Y. Cai, S. Hou, and X. Xu.** 2012. Identification of the HetR-recognition sequence upstream of *hetZ* in *Anabaena* sp. Strain PCC 7120. *J. Bacteriol.* **194:**2297-2306.

28. **Ducat, D. C., J. C. Way, and P. A. Silver.** 2011. Engineering cyanobacteria to generate high-value products. *Trends Biotechnol.* **29:**95-103.

29. **Dühring, U., I. M. Axmann, W. R. Hess, and A. Wilde.** 2006. An internal antisense RNA regulates expression of the photosynthesis gene *isiA*. *Proc. Natl. Acad. Sci USA* **103:**7054-7058.

30. **Ehira, S., and M. Ohmori.** 2006. NrrA directly regulates expression of *hetR* during heterocyst differentiation in the cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **188:**8520.

31. **Ehira, S., and M. Ohmori.** 2006. NrrA, a nitrogen responsive response regulator facilitates heterocyst development in the cyanobacterium *Anabaena* sp. strain PCC 7120. *Mol. Microbiol.* **59:**1692-1703.

32. **Ehira, S., and M. Ohmori.** 2006. NrrA, a nitrogen‐responsive response regulator facilitates heterocyst development in the cyanobacterium *Anabaen*a sp. strain PCC 7120. *Molecular Microbiology* **59:**1692-1703.

33. **Ehira, S., M. Ohmori, and N. Sato.** 2003. Genome-wide expression analysis of the responses to nitrogen deprivation in the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA research* **10:**97-113.

34. **Ehira, S., M. Ohmori, and N. Sato.** 2003. Genome-wide expression analysis of the responses to nitrogen deprivation in the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res.* **10:**97-113.

35. **Elhai, J., A. Vepritskiy, A. M. Muro-Pastor, E. Flores, and C. P. Wolk.** 1997. Reduction of conjugal transfer efficiency by three restriction activities of *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **179:**1998-2005.

36. **Ernst, A., T. A. Black, Y. Cai, J. M. Panoff, D. N. Tiwari, and C. P. Wolk.** 1992. Synthesis of nitrogenase in mutants of the cyanobacterium *Anabaena* sp. strain PCC 7120 affected in heterocyst development or metabolism. *J. Bacteriol.* **174:**6025.

37. **Falcon, L. I., S. Magallon, and A. Castillo.** 2010. Dating the cyanobacterial ancestor of the chloroplast. *ISME J* **4:**777-783.

38. **Feldman, E. A., S. Ni, D. D. Risser, J. L. Murakami, S. M. Callahan, and M. A. Kennedy.** 2010. Evidence that HetR from *Anabaena* binds PatS. *10th Cyanobacterial Molecular Biology Workshop* **ABSTRACT**.

39. **Filiatrault, M. J., P. V. Stodghill, P. A. Bronstein, S. Moll, M. Lindeberg, G. Grills, P. Schweitzer, W. Wang, G. P. Schroth, and S. Luo.** 2010. Transcriptome analysis of *Pseudomonas syringae* identifies new genes, noncoding RNAs, and antisense activity. *J. Bacteriol.* **192:**2359-2372.

40. **Flaherty, B., F. Van Nieuwerburgh, S. Head, and J. Golden.** 2011. Directional RNA deep sequencing sheds new light on the

transcriptional response of *Anabaena* sp. strain PCC 7120 to combined-nitrogen deprivation. *BMC Genomics* **12:**332.

41. **Flores, E., and A. Herrero.** 2010. Compartmentalized function through cell differentiation in filamentous cyanobacteria. *Nat. Rev. Micro.* **8:**39-50.

42. **Fozo, E. M., M. Kawano, F. Fontaine, Y. Kaya, K. S. Mendieta, K. L. Jones, A. Ocampo, K. E. Rudd, and G. Storz.** 2008. Repression of small toxic protein synthesis by the Sib and OhsC small RNAs. *Molecular Microbiology* **70:**1076-1093.

43. **Gantt, E.** 1981. Phycobilisomes. *Annual Review of Plant Physiology* **32:**327-347.

44. **Georg, J., and W. R. Hess.** 2011. cis-Antisense RNA, Another Level of Gene Regulation in Bacteria. *Microbiology and Molecular Biology Reviews* **75:**286-300.

45. **Georgianna, D. R., and S. P. Mayfield.** 2012. Exploiting diversity and synthetic biology for the production of algal biofuels. *Nature* **488:**329-335.

46. **Golden, J. W., L. L. Whorff, and D. R. Wiest.** 1991. Independent regulation of *nifHDK* operon transcription and DNA rearrangement during heterocyst differentiation in the cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **173:**7098-7105.

47. **Hanaoka, M., and K. Tanaka.** 2008. Dynamics of RpaB–promoter interaction during high light stress, revealed by chromatin immunoprecipitation (ChIP) analysis in *Synechococcus elongatus* PCC 7942. *The Plant Journal* **56:**327-335.

48. **He, S., O. Wurtzel, K. Singh, J. L. Froula, S. Yilmaz, S. G. Tringe, Z. Wang, F. Chen, E. A. Lindquist, R. Sorek, and P. Hugenholtz.** 2010. Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. *Nat. Meth.* **7:**807-812.

49. **Hebbar, P. B., and S. E. Curtis.** 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.

50. **Hernández, J. A., A. M. Muro-Pastor, E. Flores, M. T. Bes, M. L. Peleato, and M. F. Fillat.** 2006. Identification of a *furA* cis Antisense

RNA in the Cyanobacterium *Anabaena* sp. PCC 7120. *Journal of molecular biology* **355:**325-334.

51. **Herrero, A., and E. Flores (ed.).** 2008. Cyanobacteria: Molecular Biology, Genomics and Evolution. Caister Academic Press.

52. **Herring, C. D., M. Raffaelle, T. E. Allen, E. I. Kanin, R. Landick, A. Z. Ansari, and B. Ø. Palsson.** 2005. Immobilization of *Escherichia coli* RNA Polymerase and Location of Binding Sites by Use of Chromatin Immunoprecipitation and Microarrays. *J. Bacteriol.* **187:**6166-6174.

53. **Higa, K. C., and S. M. Callahan.** 2010. Ectopic expression of hetP can partially bypass the need for hetR in heterocyst differentiation by *Anabaena* sp. strain PCC 7120. *Molecular Microbiology* **77:**562-574.

54. **Higa, K. C., and S. M. Callahan.** 2010. Ectopic expression of *hetP* can partially bypass the need for *hetR* in heterocyst differentiation by *Anabaena* sp. strain PCC 7120. *Mol. Microbiol.* **77:**562-574.

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

56. **Huang, X., Y. Dong, and J. Zhao.** 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. USA* **101:**4848-4853.

57. **Ionescu, D., B. Voss, A. Oren, W. R. Hess, and A. M. Muro-Pastor.** 2010. Heterocyst-specific transcription of NsiR1, a non-coding RNA encoded in a tandem array of direct repeats in cyanobacteria. *Mol. Microbiol.* **398:**177-188.

58. **Jacob, N. M., A. Kantardjieff, F. N. K. Yusufi, E. F. Retzel, B. C. Mulukutla, S. H. Chuah, M. Yap, and W.-S. Hu.** 2009. Reaching the depth of the Chinese hamster ovary cell transcriptome. *Biotechnol. Bioeng.* **105:**1002-1009.

59. **Karradt, A., J. Sobanski, J. Mattow, W. Lockau, and K. Baier.** 2008. NblA, a Key Protein of Phycobilisome Degradation, Interacts with ClpC, a HSP100 Chaperone Partner of a Cyanobacterial Clp Protease. *Journal of Biological Chemistry* **283:**32394-32403.

60. **Katoh, H., R. Asthana, and M. Ohmori.** 2004. Gene expression in the cyanobacterium *Anabaena* sp. PCC7120 under desiccation. *Microbial ecology* **47:**164-174.

61. **Khudyakov, I. Y., and C. P. Wolk.** 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.

62. **Kidder, B. L., G. Hu, and K. Zhao.** 2011. ChIP-Seq: technical considerations for obtaining high-quality data. *Nat Immunol* **12:**918-922.

63. **Kim, Y., G. Joachimiak, Z. Ye, T. A. Binkowski, R. Zhang, P. Gornicki, S. M. Callahan, W. R. Hess, R. Haselkorn, and A. Joachimiak.** 2011. Structure of transcription factor HetR required for heterocyst differentiation in cyanobacteria. *Proc. Natl. Acad. Sci USA* **108:**10109-10114.

64. **Kneip, C., P. Lockhart, C. Voß, and U. G. Maier.** 2007. Nitrogen fixation in eukaryotes–New models for symbiosis. *BMC evolutionary biology* **7:**55.

65. **Kumar, K., R. A. Mella-Herrera, and J. W. Golden.** 2010. Cyanobacterial Heterocysts. *Cold Spring Harb. Perspect. Biol. 2009* **2:**a000315.

66. **Lang, J. D., and R. Haselkorn.** 1989. Isolation, sequence and transcription of the gene encoding the photosystem II chlorophyll-binding protein, CP-47, in the cyanobacterium *Anabaena* 7120. *Plant Mol. Bio.* **13:**441-456.

67. **Langmead, B., C. Trapnell, M. Pop, and S. Salzberg.** 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10:**R25.

68. **Lasa, I., A. Toledo-Arana, A. Dobin, M. Villanueva, I. R. de los Mozos, M. Vergara-Irigaray, V. Segura, D. Fagegaltier, J. R. Penadés, and J. Valle.** 2011. Genome-wide antisense transcription drives mRNA processing in bacteria. *Proc. Natl. Acad. Sci USA* **108:**20172-20177.

69. **Laurent, S., H. Chen, S. Bédu, F. Ziarelli, L. Peng, and C.-C. Zhang.** 2005. Nonmetabolizable analogue of 2-oxoglutarate elicits heterocyst differentiation under repressive conditions in *Anabaena* sp. PCC 7120. *Proc. Natl. Acad. Sci USA* **102:**9907-9912.

70. **Lesser, M. P., C. H. Mazel, M. Y. Gorbunov, and P. G. Falkowski.** 2004. Discovery of Symbiotic Nitrogen-Fixing Cyanobacteria in Corals. *Science* **305:**997-1000.

71. **Li, R., Y. Li, K. Kristiansen, and J. Wang.** 2008. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24:**713-714.

72. **Liang, J., L. Scappino, and R. Haselkorn.** 1992. The *patA* gene product, which contains a region similar to CheY of *Escherichia coli,* controls heterocyst pattern formation in the cyanobacterium *Anabaena* 7120. *Proc. Natl. Acad. Sci USA* **89:**5655.

73. **Liu, D., and J. W. Golden.** 2002. *hetL* overexpression stimulates heterocyst formation in *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **184:**6873.

74. **Liu, E., S. Pott, and M. Huss.** 2010. Q&A: ChIP-seq technologies and the study of gene regulation. *BMC Biology* **8:**56.

75. **Liu, J. M., J. Livny, M. S. Lawrence, M. D. Kimball, M. K. Waldor, and A. Camilli.** 2009. Experimental discovery of sRNAs in *Vibrio cholerae* by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res* **37:**e46.

76. **Lopez-Igual, R., E. Flores, and A. Herrero.** 2010. Inactivation of a heterocyst-specific invertase indicates a principal role of sucrose catabolism in heterocysts of *Anabaena* sp. *J. Bacteriol.* **192:**5526-5533.

77. **Ludwig, M., and D. A. Bryant.** 2011. Transcription profiling of the model cyanobacterium *Synechococcus* sp. strain PCC 7002 by Next-Gen (SOLiD™) sequencing of cDNA. *Frontiers in Microbiology* **2**.

78. **Machado, I. M. P., and S. Atsumi.** 2012. Cyanobacterial biofuel production. *J. Biotechnology*.

79. **Makarova, K. S., E. V. Koonin, R. Haselkorn, and M. Y. Galperin.** 2006. Cyanobacterial response regulator PatA contains a conserved N-terminal domain (PATAN) with an alpha-helical insertion. *Bioinformatics* **22:**1297-1301.

80. **McFadden, G. I., and G. G. van Dooren.** 2004. Evolution: Red Algal Genome Affirms a Common Origin of All Plastids. *Current Biology* **14:**R514-R516.

81. **Mella-Herrera, R. A., M. R. Neunuebel, and J. W. Golden.** 2010. *Anabaena* sp. strain PCC 7120 *conR* contains a LytR-CpsA-Psr

domain, is developmentally regulated, and is essential for diazotrophic growth and heterocyst morphogenesis. *Microbiology* **157:**617-626.

82. **Metzker, M. L.** 2010. Sequencing technologies - the next generation. *Nat Rev Genet* **11:**31-46.

83. **Mitschke, J., J. Georg, I. Scholz, C. M. Sharma, D. Dienst, J. Bantscheff, B. Voß, C. Steglich, A. Wilde, J. Vogel, and W. R. Hess.** 2011. An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci. USA* **108:**2124-2129.

84. **Mitschke, J., A. Vioque, F. Haas, W. R. Hess, and A. M. Muro-Pastor.** 2011. Dynamics of transcriptional start site selection during nitrogen stress-induced cell differentiation in *Anabaena* sp. PCC7120. *Proc. Natl. Acad. Sci USA* **108:**20130-20135.

85. **Mizrachi, E., C. Hefer, M. Ranik, F. Joubert, and A. Myburg.** 2010. De novo assembled expressed gene catalog of a fast-growing Eucalyptus tree produced by Illumina mRNA-Seq. *BMC Genomics* **11:**681-693.

86. **Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer, and B. Wold.** 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Meth.* **5:**621-628.

87. **Mulligan, M. E., and R. Haselkorn.** 1989. Nitrogen fixation (*nif*) genes of the cyanobacterium *Anabaena* species strain PCC 7120. The *nifB-fdxN-nifS-nifU* operon. *J. Biol. Chem.* **264:**19200-19207.

88. **Muro-Pastor, A. M., and W. R. Hess.** 2012. Heterocyst differentiation: from single mutants to global approaches. *Trends in microbiology*.

89. **Muro-Pastor, A. M., E. Olmedo-Verd, and E. Flores.** 2006. All4312, an NtcA-regulated two-component response regulator in *Anabaena* sp. strain PCC 7120. *FEMS Microbiology Letters* **256:**171-177.

90. **Muro-Pastor, A. M., A. Valladares, E. Flores, and A. Herrero.** 1999. The *hetC* gene is a direct target of the NtcA transcriptional regulator in cyanobacterial heterocyst development. *J. Bacteriol.* **181:**6664-6669.

91. **Nayar, A. S., H. Yamaura, R. Rajagopalan, D. D. Risser, and S. M. Callahan.** 2007. FraG is necessary for filament integrity and heterocyst maturation in the cyanobacterium *Anabaena* sp. strain PCC 7120. *Microbiology* **153:**601.

92. **Nicolaisen, K., A. Hahn, and E. Schleiff.** 2009. The cell wall in heterocyst formation by *Anabaena* sp. PCC 7120. *J. Basic Microbiol.* **49:**5-24.

93. **Nix, D., S. Courdy, and K. Boucher.** 2008. Empirical methods for controlling false positives and estimating confidence in ChIP-Seq peaks. *BMC Bioinformatics* **9:**523.

94. **Opdyke, J. A., E. M. Fozo, M. R. Hemm, and G. Storz.** 2011. RNase III Participates in GadY-Dependent Cleavage of the *gadX-gadW* mRNA. *Journal of Molecular Biology* **406:**29-43.

95. **Otto, T. D., D. Wilinski, S. Assefa, T. M. Keane, L. R. Sarry, U. Böhme, J. Lemieux, B. Barrell, A. Pain, and M. Berriman.** 2010. New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. *Mol. Microbiol.* **76:**12-24.

96. **Ozsolak, F., and P. M. Milos.** 2011. RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* **12:**87-98.

97. **Passalacqua, K. D., A. Varadarajan, B. D. Ondov, D. T. Okou, M. E. Zwick, and N. H. Bergman.** 2009. Structure and complexity of a bacterial transcriptome. *J. Bacteriol.* **191:**3203-3211.

98. **Perkins, T. T., R. A. Kingsley, M. C. Fookes, P. P. Gardner, K. D. James, L. Yu, S. A. Assefa, M. He, N. J. Croucher, D. J. Pickard, D. J. Maskell, J. Parkhill, J. Choudhary, N. R. Thomson, and G. Dougan.** 2009. A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. *PLoS Genet* **5:**e1000569.

99. **Prieto, A. I., C. Kahramanoglou, R. M. Ali, G. M. Fraser, A. S. N. Seshasayee, and N. M. Luscombe.** 2012. Genomic analysis of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and HU in *Escherichia coli* K12. *Nucleic Acids Research* **40:**3524-3537.

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

101. **Reich, M., T. Liefeld, J. Gould, J. Lerner, P. Tamayo, and J. P. Mesirov.** 2006. GenePattern 2.0. *Nat Genetics* **38:**500-501.

102. **Risser, D. D., F. C. Wong, and J. C. Meeks.** 2012. Biased inheritance of the protein PatN frees vegetative cells to initiate patterned heterocyst differentiation. *Proc. Natl. Acad. Sci USA*.

103. **Roberts, A., H. Pimentel, C. Trapnell, and L. Pachter.** 2011. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **27:**2325-2329.

104. **Sato, N., M. Ohmori, M. Ikeuchi, K. Tashiro, C. P. Wolk, T. Kaneko, K. Okada, M. Tsuzuki, S. Ehira, and H. Katoh.** 2004. Use of segment-based microarray in the analysis of global gene expression in response to various environmental stresses in the cyanobacterium *Anabaena* sp. PCC 7120. *The Journal of general and applied microbiology* **50:**1-8.

105. **Schneider, C. A., W. S. Rasband, and K. W. Eliceiri.** 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Meth* **9:**671-675.

106. **Sczakiel, G., M. Homann, and K. Rittner.** 1993. Computer-aided search for effective antisense RNA target sequences of the human immunodeficiency virus type 1. *Antisense research and development* **3:**45-52.

107. **Shimada, T., A. Ishihama, S. J. W. Busby, and D. C. Grainger.** 2008. The *Escherichia coli* RutR transcription factor binds at targets within genes as well as intergenic regions. *Nucleic Acids Research* **36:**3950-3955.

108. **Slutsky, A. M., P. M. Rabinovich, L. Z. Yakubov, I. V. Sineokaya, A. I. Stepanov, and V. K. Gordeyev.** 1980. Direct selection of DNA inserts in plasmid gene of kanamycin-resistance. *Molecular and General Genetics MGG* **180:**487-488.

109. **Sorek, R., and P. Cossart.** 2009. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat. Rev. Genet.* **11:**9-16.

110. **Sprent, J. I., and J. A. Raven.** 1985. Evolution of nitrogen-fixing symbioses. *Proceedings of the Royal Society of Edinburgh Section B: Biology* **85:**215-237.

111. **Thomas, J., J. C. Meeks, C. P. Wolk, P. W. Shaffer, and S. M. Austin.** 1977. Formation of glutamine from [13n]ammonia, [13n]dinitrogen, and [14C]glutamate by heterocysts isolated from *Anabaena cylindrica*. *J. Bacteriol.* **129:**1545-1555.

112. **Trapnell, C., B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold, and L. Pachter.** 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotech.* **28:**511-515.

113. **Ungerer, J. L., B. S. Pratte, and T. Thiel.** 2010. RNA processing of nitrogenase transcripts in the cyanobacterium *Anabaena variabilis*. *J. Bacteriol.* **192:**3311-3320.

114. **Valladares, A., A. M. Muro-Pastor, A. Herrero, and E. Flores.** 2004. The NtcA-dependent P1 promoter is utilized for *glnA* expression in $N_2$-fixing heterocysts of *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **186:**7337-7343.

115. **Van Vliet, A. H. M.** 2010. Next generation sequencing of microbial transcriptomes: challenges and opportunities. *FEMS Microbiology Letters* **302:**1-7.

116. **Vargas, W. A., C. N. Nishi, L. E. Giarrocco, and G. L. Salerno.** 2010. Differential roles of alkaline/neutral invertases in *Nostoc* sp. PCC 7120: Inv-B isoform is essential for diazotrophic growth. *Planta* **233:**153-162.

117. **Vaulot, D., D. Marie, R. J. Olson, and S. W. Chisholm.** 1995. Growth of *Prochlorococcus*, a Photosynthetic Prokaryote, in the Equatorial Pacific Ocean. *Science* **268:**1480-1482.

118. **Velazquez-Bermidez, M. F., A. Herrero, and E. Flores.** 2002. 2-Oxoglutarate increases the binding affinity of the NtcA (nitrogen control) transcription factor for the Synechococcus *glnA* promoter. *FEBS letters* **512:**71-74.

119. **Vijayan, V., I. Jain, and E. O'Shea.** 2011. A high resolution map of a cyanobacterial transcriptome. *Genome biology* **12:**R47.

120. **Villareal, T. A.** 1992. Marine nitrogen-fixing diatom-cyanobacteria symbioses. *Marine pelagic cyanobacteria: Trichodesmium and other diazotrophs. Kluwer***:**163-175.

121. **Vrba, J. M., and S. E. Curtis.** 1990. Characterization of a four-member *psbA* gene family from the cyanobacterium *Anabaena* PCC 7120. *Plant. Mol. Biol.* **14:**81-92.

122. **Wade, J. T., N. B. Reppas, G. M. Church, and K. Struhl.** 2005. Genomic analysis of LexA binding reveals the permissive nature of the

*Escherichia coli* genome and identifies unconventional target sites. *Genes & Development* **19:**2619-2630.

123. **Waterbury, J. B., S. W. Watson, R. R. L. Guillard, and L. E. Brand.** 1979. Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature* **277:**293-294.

124. **Waters, L. S., and G. Storz.** 2009. Regulatory RNAs in Bacteria. *Cell* **136:**615-628.

125. **Wen, Y., J. Feng, D. R. Scott, E. A. Marcus, and G. Sachs.** 2011. A cis-Encoded Antisense Small RNA Regulated by the HP0165-HP0166 Two-Component System Controls Expression of *ureB* in *Helicobacter pylori*. *J. Bacteriol.* **193:**40-51.

126. **Wolk, C. P., S. Lechno-Yossef, and K. M. Jager.** 2010. The insertion sequences of *Anabaena* sp. strain PCC 7120 and their effects on its open reading frames. *J. Bacteriol.* **192:**5289-5303.

127. **Xu, X., J. Elhai, and C. P. Wolk.** 2008. Transcriptional and developmental responses by *Anabaena* to deprivation of fixed nitrogen, p. 383-422. *In* A. Herrero and E. Flores (ed.), *The cyanobacteria: molecular biology, genomics and evolution*. Horizon Scientific Press, Norwich, United Kingdom.

128. **Yoder-Himes, D. R., P. S. G. Chain, Y. Zhu, O. Wurtzel, E. M. Rubin, J. M. Tiedje, and R. Sorek.** 2009. Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. *Proc. Natl. Acad. Sci USA* **106:**3976-3981.

129. **Yoon, H.-S., and J. W. Golden.** 1998. Heterocyst Pattern Formation Controlled by a Diffusible Peptide. *Science* **282:**935-938.

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

131. **Young-Robbins, S. S., D. D. Risser, J. R. Moran, R. Haselkorn, and S. M. Callahan.** 2010. Transcriptional regulation of the heterocyst patterning gene *patA* from *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **192:**4732-4740.

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

133. **Zhang, J.-Y., W.-L. Chen, and C.-C. Zhang.** 2009. *hetR* and *patS*, two genes necessary for heterocyst pattern formation, are widespread in filamentous nonheterocyst-forming cyanobacteria. *Microbiology* **155:**1418-1426.

134. **Zhang, W., Y. Du, I. Khudyakov, Q. Fan, H. Gao, D. Ning, C. P. Wolk, and X. Xu.** 2007. A gene cluster that regulates both heterocyst differentiation and pattern formation in *Anabaena* sp. strain PCC 7120. *Molecular Microbiology* **66:**1429-1443.

135. **Zhang, W., F. Li, and L. Nie.** 2010. Integrating multiple 'omics' analysis for microbial biology: application and methodologies. *Microbiology* **156:**287-301.

136. **Zhang, Y., T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, and X. S. Liu.** 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9:**R137.

137. **Zhao, M. X., Y. L. Jiang, Y. X. He, Y. F. Chen, Y. B. Teng, Y. Chen, C. C. Zhang, and C. Z. Zhou.** 2010. Structural basis for the allosteric control of the global transcription factor NtcA by the nitrogen starvation signal 2-oxoglutarate. *Proc. Natl. Acad. Sci. USA* **107:**12487-12492.