Characterization of a TetR/AcrR family transcriptional repressor gene NpR3597 in

Nostoc punctiforme

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in Biology

By

Arash Seyed Yazdani

May 2012
The thesis of Arash Seyed Yazdani is approved by:

__________________________________________
Virginia Vandergon, PhD

__________________
Date

__________________________________________
Sean Murray, PhD

__________________
Date

__________________________________________
Michael L. Summers, PhD, Chair

__________________
Date

California State University, Northridge
ACKNOWLEDGEMENTS

I thank Dr. Michael L. Summers for this immensely enriching experience and for his guidance, patience, support and the vast knowledge I have acquired. Jenevieve Polin’s training, help, previous work and thorough notes played an integral part in keeping my research on track. I thank previous RA Peter Holmquist for contributing incredible insight and direction to my research; he taught me to think way outside of the box, which led me to great discoveries.

I very much appreciate all committee members for there help. I thank Dr. Sean Murray’s advisement in troubleshooting my experiments and Dr. Virginia Vandergon for all my priceless knowledge of bioinformatics that played a huge part in my research.

I thank my parents Jila and Behrouz Yazdani, and my sister Ava Yazdani for all their support and encouragement through this journey.

This work was partially supported by NIH SCORE grant GM48680 to MLS and an award to Arash S. Yazdani from the Graduate Thesis, Project or Performance Support Program 2011-2012.
TABLE OF CONTENTS

Signature page ii
Acknowledgements iii
List of Tables vi
List of Figures vii
Abstract x
Introduction 1

*Nostoc punctiforme* 1
The AcrR/TetR family of transcriptional regulators 1
Cyanobacterial phycobilisomes 3

Materials and Methods 5
Growth Conditions and Strain Maintenance 5
Chlorophyll a Readings 5
Cloning and PCR 5
Transformation 7
Plasmid purification and screening 7
Electroporation into *Nostoc punctiforme* 7
Epifluorescence microscopy 8
Bioinformatic analysis 8
PCR amplification 9
Microarray 9
Protein purification 10
Electric Mobility Shift Assay, EMSA 11
CCA experiment 12
Desiccation experiment screening for viable akinetes 13
LIST OF TABLES

Table 1. Primers used in this study 14
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Diagram of AcrBA-TolC multi-drug efflux pump complex</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>Diagram of phycobilisome under green/red light conditions</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>NCBI’s conserved domains output indicates putative HTH domain and homology to AcrR type proteins</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>Two homo-dimers of TetR binding to an inverted repeat DNA sequence</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>Analysis of the N-terminal conserved region within NpR3597</td>
<td>17</td>
</tr>
<tr>
<td>6</td>
<td>Loci of putative AcrR genes in <em>N. punctiforme</em> NpR3597 and NpR1932 with known <em>tetR</em> and <em>acrR</em> family transcriptional repressors in <em>E. coli</em> and AcrR family gene <em>pigZ</em> in <em>Serratia</em> 39006</td>
<td>18</td>
</tr>
<tr>
<td>7</td>
<td>Predicted crystal structure of both NpR3597 and NpR1931 compared to AcrR crystal structure</td>
<td>19</td>
</tr>
<tr>
<td>8</td>
<td>DNA microarray data for four <em>Nostoc punctiforme</em> genes up- and down-regulated in zwf mutant over 6 days in the dark</td>
<td>20</td>
</tr>
<tr>
<td>9</td>
<td>Coding sequence of NpR3597 inserted into MCS of pET28a vector</td>
<td>22</td>
</tr>
<tr>
<td>10</td>
<td>Translation of sequence after insertion of 3597 into pET28a vector</td>
<td>23</td>
</tr>
<tr>
<td>11</td>
<td>Colony PCR of transformed pET28a::3597-His into <em>E. coli</em></td>
<td>24</td>
</tr>
<tr>
<td>12</td>
<td>SDS-PAGE showing Induction of pET28a::NpR3597-His protein in <em>E. coli</em></td>
<td>25</td>
</tr>
<tr>
<td>13</td>
<td>SDS-PAGE gel of nickel column purification of NpR3597-His protein.</td>
<td>26</td>
</tr>
<tr>
<td>14</td>
<td>Truncation of upstream regulatory regions to show binding of AcrR protein NpR3597 to inverted repeat sites.</td>
<td>27</td>
</tr>
<tr>
<td>15</td>
<td>Truncation of upstream regulatory region of F1932 to show binding of AcrR protein R3597 to inverted repeat site.</td>
<td>28</td>
</tr>
<tr>
<td>16</td>
<td>PCR of truncated DNA fragments for use in EMSA.</td>
<td>29</td>
</tr>
</tbody>
</table>
Figure 17. Primer design for EMSA; inverted repeat binding site exclusion by truncation of NpF1932 upstream region

Figure 18. Primer design for EMSA; inverted repeat binding site exclusion by 2 truncations of NpF3598 upstream region

Figure 19. NpF1932 upstream regulatory region EMSA using truncated fragments and NpR3597 protein

Figure 20. NpF3598 upstream regulatory region EMSA using truncated fragments and NpR3597 protein

Figure 21. Graph of whole cell absorption spectrum from 400 nm – 730 nm of WT/pSCR119 vs. WT/pSCR119::3597 under grey light filters.

Figure 22. CCA results showing difference in phycocyanin/chlorophyll-a ratio between wild-type and overexpressor of NpR3597 under grey light

Figure 23. Graph of whole cell absorption spectrum from 400 nm – 730 nm of WT/pSCR119 vs. WT/pSCR119::3597 under red light filters.

Figure 24. CCA results showing difference in phycocyanin/chlorophyll-a ratio between wild-type and overexpressor of NpR3597 under red light

Figure 25. Graph of whole cell absorption spectrum from 400 nm – 730 nm of WT/pSCR119 vs. WT/pSCR119::3597 under green light filters.

Figure 26. CCA results showing difference in phycocyanin/chlorophyll-a ratio between wild-type and overexpressor of NpR3597 under green light

Figure 27. Expression of GFP in WT/pSUN119::p3597 in stationary phase vs. log phase under 40x magnification epifluorescent microscopy.

Figure 28. GFP expression of WT/pSUN119::p3597 in stationary phase vs. log phase under 100x magnification epifluorescent microscopy

Figure 29. Akinete viability test in wild-type, -3597 mutant, and pSCR119::3597 overexperssor

Figure 30. Regulation flowchart of NpR3597 on NpF3598 and NpF1932
ABSTRACT

Characterization of a TetR/AcrR family transcriptional repressor gene NpR3597 in Nostoc punctiforme

By

Arash Seyed Yazdani

Master of Science in Biology

*Nostoc punctiforme* is a filamentous cyanobacterium capable of differentiating its vegetative cells into spore-like akinetes that can withstand desiccation and cold. The NpR3597 gene was previously identified in a time-course DNA microarray experiment to be up-regulated during akinete formation, and up-regulated in dividing cells of the filament as indicated by GFP transcriptional reporter strains. Sequence similarity to characterized proteins indicates that this is a putative tetracycline repressor family protein similar to TetR/AcrR found in a vast array of bacteria species that represses divergently transcribed genes encoding for ABC transport proteins. The mechanism of AcrR repression typically involves binding to an inverted repeat in the promoter of the divergently transcribed gene, inhibiting attachment of RNA polymerase. Although no phenotype could be determined for a NpR3597 deletion mutant strain, an over-expression strain bearing multiple copies of this gene on a plasmid under control of its own promoter caused pigmentation changes. Desiccation-revival tests showed that over-expression of NpR3597 results in nonfunctional akinetes. Spectrophotometric examination of the over-expression strain indicated that the levels of phycocyanin, a light harvesting protein subunit of the photosynthetic phycobilisome apparatus, was significantly higher than normal. Through a second round of microarray analysis, it was found that over-expression of NpR3597 correlated to increased transcription of genes involved in phycocyanin synthesis as well as other photosynthetic genes, and down-regulation of upstream divergently transcribed multi-drug resistance efflux pump gene NpF3598. Strong down-regulation of a number of additional genes with inverted repeat sequences in their upstream intergenic region, particularly another multi-drug resistance efflux pump gene NpF1932, was also observed. To determine the binding site for NpR3597, a conserved inverted repeat sequence found upstream of both NpF3598 and NpF1932 was used in Electric Mobility Gel Shift Assays (EMSA). A plasmid encoding for a 6xHistidine tagged NpR3597 was constructed, and the His-tagged protein purified. DNA fragments containing or lacking the putative inverted repeat binding sites were generated by PCR, end-labeled with biotin and used for EMSA with the purified protein. Only fragments containing the following motif ANNNNACNN – N$_2$ - CNGTNTAGT in their inverted repeat sequence exhibited a gel mobility shift, indicating NpR3597 likely acts as a dimer and represses transcription by binding to these inverted repeat (IR) sequences.
INTRODUCTION

*Nostoc punctiforme*

Bacteria survive harsh environmental conditions by adaptively regulating their cell morphology and physiology in response to environmental changes. Some cyanobacteria can differentiate into four different states: vegetative cells, nitrogen-fixing heterocysts, temperature- and starvation-resistant akinetes, and motile hormogonia. Using the cyanobacterium model system, *Nostoc punctiforme*, we can study the currently unknown regulatory pathways that lead to these transformations. This knowledge can be used to develop manipulative controls over a broad range of not only photosynthetic, but also pathogenic bacteria.

Akinetes are the resting cells capable of surviving long periods of desiccation and cold that differentiate from normal vegetative cells following low light and potassium or phosphate starvation. A mutant strain of *N. punctiforme* lacking the glucose-6-phosphate dehydrogenase activity differentiates akinetes within 4 d of dark incubation in the presence, but not in the absence of fructose (Summers et al., 1995). After induction of akinete formation, as described for a single time point in Wong and Meeks (2002), microarray analysis was performed on all the genes of the *N. punctiforme* genome. Genetic expression was assessed each day for 6 d following akinete induction to identify genes that are upregulated during akinete formation (M.L. Summers, unpublished results). Using bioinformatics, gene-deletion mutants, over-expression strains, and deletion mutant complementation with specialized plasmids, we are characterizing these genes involved in the pathway of akinete development. One specific gene of interest was NpR3597 that was chosen because of its up-regulation during the 6 d time-course microarray of akinete induction and it being a possible regulator of akinete differentiation as a putative AcrR transcriptional repressor. GFP expression studies conducted by J. Polin (unpublished results) showed that NpR3597 exhibited cell-specific expression only two days after induction of akinetes and continued to be expressed throughout the time course of the study. The expression of NpR3597 was observed to have a periodic pattern throughout the entirety of the *Nostoc* filament, specific to the cells between heterocysts and completely absent in the heterocysts themselves.

The AcrR/TetR family of transcriptional regulators

NpR3597 is an AcrR type gene that belongs to the TetR family of transcriptional repressors. TetR is an abbreviation for Tetracycline Repressor and so names the TetR family of genes because it is the most extensively studied gene in the TetR family. First discovered in *Escherichia coli*, it is a regulator of a number of genes encoding for proteins that are involved in Tetracycline resistance, particularly the direct repression of the divergently transcribed *tetA* gene. Repression of the divergently transcribed gene is
achieved by the binding of the TetR dimer complex to its operator, tetO, in the -35 region of the divergently transcribed tetA gene. In the absence of tetracycline, TetR is maintained at very low levels. But, when very small levels of tetracycline are detected, transcription of tetR increases by a system of positive feedback on itself. After binding the tetracycline, the TetR protein is capable of releasing the upstream region of the divergently transcribed gene tetA, causing its expression by de-repression. In summary, TetR is a repressor of other genes and can be an inducer of itself by some kind of undefined interaction with RNA polymerase. (Ramos et al., 2008)

The characteristics that define a TetR family gene, first and foremost, are that it regulates gene transcription of a multi-drug efflux pump and second that it contains a Helix-Turn-Helix (HTH) DNA-binding motif at its N-terminal end that is involved in sequestering an inverted repeat (IR) DNA sequence that is contained in the operator of the repressible gene. The DNA-binding domain is comprised of 4 alpha helices, \( \alpha_1 \) through \( \alpha_4 \). The HTH DNA binding motif that is most conserved between species is found in \( \alpha_2 \) to \( \alpha_3 \), the very location that direct contact with the DNA nucleotides take place. The \( \alpha_4 \) helix has the role of being the link between the N-terminal DNA-binding domain and the C-terminal core of the protein (Muhl et al., 2009). Binding of the operator is achieved when two homo-dimers of TetR recognize and make contact with a specific site of DNA with an inverted repeat sequence. Each HTH motif of the four sub-units binds to one major groove of the palindrome, with one dimer sitting on one side of the DNA double helix and the other covering the opposite side. By this mechanism, RNA polymerase is blocked from binding to the promoter region of the multi-drug efflux pump gene and transcription is down-regulated (Ramos et al., 2008).

More specifically, the sequence of NpR3597 as a whole will be shown to have more sequence similarity to AcrR, one of the most studied TetR family repressors, rather than the actual tetR gene. AcrR was originally described in Escherichia coli as a transcriptional repressor of the divergently transcribed AcrAB-TolC multidrug efflux pump operon, which is a resistance-nodulation-cell-division (RND) family efflux pump. It is the source of E. coli’s intrinsic antibiotic resistance due to the particular quality that it can bind to and pump out a variety of toxic compounds, antibiotics and chemotherapeutic agents. This unusually high level of promiscuity in substrate recognition is characteristic of RND efflux pumps and creates great obstacles in the study of anticancer therapy (Li et al., 2007). The multiprotein complex is comprised of three main structural components that make up the functional efflux pump: AcrB, AcrA and TolC (Fig. 1). In Gram-negative bacteria, the RND efflux pump multiprotein complex penetrates through the inner and outer membranes allowing for transportation of toxic compounds across the whole cellular envelope and expulsion out of the cell. AcrB is responsible for substrate recognition and binding, AcrA is responsible for fusing the
AcrB and TolC complexes together and TolC is responsible for expulsion of the toxic compound out of the cell through the outer membrane (Gristwood et al., 2008).

![Diagram of AcrBA-TolC multi-drug efflux pump complex of E. coli in cross-section taken from Lomovskaya et al. (2007). The diagram shows AcrA subunits acting as the fusion protein that keeps the AcrB trimer and TolC “exhaust pipe” aligned.]

**Cyanobacterial phycobilisomes**

The photosynthetic system of *N. punctiforme*, like many cyanobacteria and red algae, has the ability to adapt to the color and intensity of light in different environments. Light-harvesting chromoproteins (phycobiliproteins) are assembled into “antenna-like” complex structures called phycobilisomes (Fig. 2) that are arranged along the photosynthetic thylakoid membrane. The phycobilisomes are separated into units of specific proteins or chromoproteins that are held together by linker peptides and are capable of capturing a specific range of the light spectrum because they are covalently bound to molecules called chromophores. There are many types of chromophores that are capable of absorbing certain wavelengths of visible light. Specifically, the ones that are incorporated into chromoproteins are referred to as phycobilins (Wolf and Schubler, 2005). Because the transcriptional regulator forming the focus of this work was found to have an effect on phycobilisome composition, a more thorough description of this light-harvesting protein complex is warranted.
In *N. punctiforme*, the phycobilisome is comprised of three different types of phycobiliproteins. Allophycocyanin (APC) makes up the center of the phycobilisome which has the role of transferring harvested light energy to chlorophyll *a* in photosystems II and I. The light-harvesting rods or “antennas” attached to APC are comprised of two different types of phycobiliproteins that look like stacked disks: Phycocyanin (PC), which is the layer of stacked disks proximal to the APC core, and Phycoerythrin (PE), which is distal to APC (Williams *et al.*, 1980). The PC and PE composition of the phycobilisome rods is directly dependent on the quality and intensity of light in the environment and this ability to adjust the number of PE and PC protein units is called Complimentary Chromatic Adaptation (CCA). Phycocyanin specifically absorbs light in the red range of the spectrum with an absorption peak at ~620nm and phycoerythrin absorbs the green with its highest absorption peak at ~566nm. CCA occurs when the ambient light conditions favors either the PC or PE absorption spectrum and excludes the other causing the cells to react by changing the number of PC and PE to better compliment the ambient light conditions (Fig. 2). This allows the cells to harvest the available light more efficiently. Under red light, phycocyanin’s absorption spectrum is favored and thus PC accumulates, making the cyanobacteria filaments a more green color. Under green light, PE will accumulate in the phycobilisomes, making the filaments a more reddish color. (Kehoe and Gutu, 2006)
MATERIALS AND METHODS

Growth Conditions and Strain Maintenance

Cultures of *Nostoc punctiforme* (ATCC 29133) (Rippka and Herdman, 1992) were grown in 125 ml Erlenmeyer flasks containing 50 ml of Allen and Aaron (AA/4) liquid media (Allen, 1955) or media was solidified using Nobel agar (Difco) for plates. 5 mM 3-(N-morpholino) propanesulfonic acid (MOPS), 2.5 mM ammonia and 5 mM nitrate (MAN) were added as a buffer and nitrogen source for the cells. All strains containing the pSCR119 and pSUN119 plasmids were grown in 10 µg/ml neomycin. Cultures were kept in incubators which shook the flasks at a speed of ~150 rpms. Light conditions were controlled to expose the cultures to a light intensity of 10-12 µmol of photons per square meter per second. A fresh inoculation into a new flask was administered if the culture was reaching early-stationary phase, which is ~12 µg Chla/ml, to keep the cultures in log phase.

*E. coli* DH5α MCR (Hanahan, 1985) cultures were grown on solid or liquid Luria Broth (LB) containing 30 µg/ml kanamycin for pET28a vector-containing strains. In the case where blue/white screening was done for pSCR119 vector cloning, 0.1 mM IPTG and 20µg/mL X-gal was incorporated into the solid LB plates. Liquid cultures, were grown at 37°C for a maximum of 24 hours in a rolling incubator after transformation of cells with plasmid vectors.

Chlorophyll *a* Readings

The density and growth phase of cell cultures were determined by calculating the amount of chlorophyll *a* present per volume of culture. 1 mL is extracted from liquid medium into a 1.5 mL Eppendorf tube and centrifuged at ~13,000 rpm for 5 minutes. Following centrifugation, 90% (900µl) of the supernatant is removed and 900µL of 100% methanol is replaced into the tube. The pellet is then re-suspended by vigorous vortexing for 30 seconds – 1 minute, then placed in the dark for 5 minutes for the purpose of extracting the chlorophyll *a* into solution. The tube is then centrifuged at ~13,000 rpm again for 5 minutes to pellet the dead cells and debris leaving chlorophyll *a* in the supernatant. The supernatant is then analyzed in a spectrophotometer by taking an absorption reading at 665 nanometers, which is the wavelength at which chlorophyll *a* absorbs light, against a 90% methanol blank. The absorption reading is then multiplied by a factor of 12.7 to obtain µg of chlorophyll *a* per mL. (Meeks, 1971)

Cloning and PCR

Cloning of NpR3597 into pET28a His-fusion plasmid and pSCR119 high-copy number plasmid. NpR3597 was first PCR amplified from genomic *N. punctiforme* DNA using P1 with a *NcoI* site primer and P2 with a *XhoI* site–reverse primer for cloning of NpR3597
PCR amplifications of DNA fragments were all carried out using the Robocycler® Gradient 96 Thermal Cycler (Stratagene). Each PCR reaction contained: 35.5 uL of dH$_2$O, 10 uL of 5X HF Buffer, 1 uL of dNTP (10mM), 1 uL of genomic template DNA (100 ng/ul), 2 uL of P1/P2 primers (10 uM), and .5 uL of Herculase® II fusion DNA polymerase. The PCR reaction conditions were: 1 cycle of 2 min. at 95°C; 30 cycles of 30 sec at 95°C, 30 sec at 56°C, and 40 sec at 72°C; and final extension for 3 min at 72°C. PCR fragments were confirmed by agarose gel electrophoresis and then cleaned with QIAquick® PCR purification kit (Qiagen). Plasmids were purified from E. coli DH5α MCR using Qiagen plasmid mini kit. The pET28a plasmid prep and NpR3597 PCR amplification product were then digested with NcoI and XhoI. Each of the two reactions contained: 22 uL dH$_2$O, 5 uL 10X Fast Digest Buffer, 20uL (or 2 ug) of plasmid or 20 uL (or 1 ug) of insert, 1 uL of NcoI, 1 uL of XhoI, and 1 uL of FastAP Thermosensitive Alkaline Phosphatase in the plasmid reaction to dephosphorylate plasmid DNA overhangs after digestion and inhibit recircularization during ligation. The pSCR119 plasmid prep and NpR3597 PCR amplification product were then digested with PstI and KpnI. Each of the two reactions contained: 22 uL dH$_2$O, 5 uL 10X Fast Digest Buffer, 20uL (or 2 ug) of pSCR119 plasmid or 20 uL (or 1 ug) of insert, 1 uL of PstI, 1 uL of KpnI, and 1 uL of FastAP Thermosensitive Alkaline Phosphatase. Reactions were then incubated for 37°C for 10 min to initiate digestion and stopped by heating at 65°C for 15 min. Ligation reaction contained: 100 ng of plasmid with 50 ng of NpR3597, 2 uL of 10X ligation buffer, 1 uL of FastLigase (3U/uL), and QS to 20 uL with dH$_2$O. The ligation was incubated for 5 minutes at room temperature and stored at -20°C until transformation into E. coli.

**Generation of DNA fragments for use in EMSA.** For NpR3598 upstream truncated fragments for use in the EMSA: fragment P1 was made with primer pairs 3598-Pal2.1-P1 and f3598 Qp2, P2 was made with primer pairs 3598-Pal1-P2 and f3598 Qp2, and P3 was made with 3598-pal0-P3 and f3598 Qp2. For NpR1932 upstream truncated fragments for use in the EMSA: fragment P1 was made with primer pairs 1932-Pal1-P1 and 1932-emsma P2, and fragment P2 was made with primer pairs 1932-Pal0-P2.1 and 1932-emsma P2. PCR amplifications of DNA fragments were all carried out using the Robocycler Gradient 96 Thermal Cycler (Stratagene). Each PCR reaction contained: 5 uL of 10X PCR buffer, 3 uL of 25 mM MgCl$_2$, 1 uL 100 ng/uL genomic DNA, 2 uL of F/R primers (10uM), 2 uL of Taq polymerase, 1 uL of dNTP (10mM), and QS to 50 uL with dH$_2$O. The PCR reaction conditions were: 1 cycle of 3 min. at 95°C; 30 cycles of 30 sec at 95°C, 30 sec at 52°C (48°C for both NpR1932 primer pairs due to low melting temperatures), and 40 sec at 72°C; and final extension for 7 min at 72°C. PCR fragments...
were confirmed by gel electrophoresis and then cleaned with QIAquick® PCR purification kit (Qiagen).

**Transformation**

CaCl$_2$-competent E. coli DH5α MCR cells were transformed using 4 μL of ligation reaction per 100μl of E. coli cells in 1.5 mL Eppendorf tubes. The mixture was then set on ice for 20 minutes and then heat shocked at 42-43°C for 90 seconds. They were then returned to ice for 1 minute and 1 mL of SOC (2% w/v bacto-tryptone, 0.5% w/v Yeast extract, 8.56 mM NaCl, 2.5mM KCl, 10mM MgCl$_2$, 20mM glucose) was added. The Eppendorf tube was then put in a test tube and placed in a rolling incubator so that the cells may recover for 1 hour at 37°C. Following incubation, dilutions were made of ratios 1:1, 1:5, 1:10, 1:25 and 100 μL of each were plated on Km$^{30}$ plates and incubated overnight in 37°C incubator to get isolated colonies. Colony PCR using forward and reverse primers outside of the plasmid multiple cloning site was performed on 8 colonies and an electrophoresis gel was run to identify colonies that were positive for plasmids that contained inserts. The positive colonies were then grown overnight under Km$^{30}$ selection in Luria Broth.

**Plasmid purification and screening**

Plasmids were purified from E. coli cultures grown from isolated colonies using the Promega Wizard® SV 96 Plasmid DNA Purification System. The concentrations of plasmid DNA were assayed using a Nanodrop 1000. Samples were sent for sequencing using forward and reverse primers outside of the multiple cloning site to confirm that there were no mutations in the sequence and that they were in the right orientation. Colonies of the samples that didn’t have any mutations were grown in 5 mL LB Km$^{30}$ broth, a sample made with 50% cryogenic solution in a 1.5 mL Eppendorf tube and stored at (-80°C).

**Electroporation into Nostoc punctiforme**

For electroporation of a plasmid into a particular strain, that strain is first grown to late log phase (~10μg Chla/ml). The culture was then poured into a 50 mL falcon tube and concentrated to 3-5 mL, or a density of ~100 μg Chla/mL, in growth medium. *Nostoc* cells were then sonicated for 15 seconds at 12% duty using a Labline Inc GEX600 sonicator with a 1/8$^{th}$ inch microtip. After the cells are broken up, they were placed back into 50 mL AA/4 with MOPS and ammonia for a little more than 4 hours under low light to recover. Prior to electroporation, the cells were washed 3 times with 20 mL room temperature double distilled water, vortexing between washes. After the third wash, cells were resuspended to a density of 50-100 μg Chla/mL and cooled on ice.
1μg of DNA was aliquoted to a sterile microfuge tube and set on ice along with the sterile 0.2 cm Electroporation cuvette. 400 μL of concentrated, washed culture was added to the microfuge tube containing DNA, pipetted up and down, and left on ice for 30-60 seconds to allow DNA to adsorb to cells. The cells were then transferred to the ice cold 0.2 cm electroporation cuvette and electroporated at a resistance of 600Ω, voltage of 1.6 KeV and a capacitance of 25μF. Maximum efficiency of electroporation was achieved with a time constant of between ~12.5-13.2 milliseconds. Immediately following electroporation, cells in the cuvette were diluted with 400 μL of AA/4 media containing MOPS, ammonia, 20 mM magnesium chloride and transferred to a 50 mL flask of the same composition. The culture was left to incubate overnight under dim light with gentle shaking.

The overnight culture was then concentrated by centrifugation in a 50 mL falcon tube and re-suspended to a volume of 1 mL. It was then spread onto selective media (A&A) plates containing MAN and 10μg/mL Neomycin. The first spread plate was 100μL of the undiluted culture, and following that was 1:5 and 1:10 dilutions of the concentrated culture. The plates were sealed using parafilm and placed in a CO₂ incubator under low light at 25°C for about 11 days or until colonies became visible. Once colonies became visible, they were patched onto new selective media plates and placed back into the incubator. Once the patches reached an adequate density to be harvested, a single uncontaminated patch was chosen and used to inoculate a liquid AA/4 culture containing MAN and neomycin.

**Epifluorescence microscopy**

Epifluorescence microscopy was performed using a Zeiss AxioLab microscope containing a 100x oil immersion objective lens. GFP fluorescence pictures were obtained using a long-pass blue excitation filter (395 nm) and a green band pass (509 nm) filter set (Omega Optical) and captured with a DVC 1312 digital camera. Pictures showing natural fluorescence from phycobilisomes were obtained using a texas red filter set (Omega Optical).

**Bioinformatic analysis**

Gene sequences of the entire *Nostoc* genome were analyzed using Artemis ([www.sanger.ac.uk/Software/Artemis](http://www.sanger.ac.uk/Software/Artemis)) (Rutherford et al., 2000) and the BLAST program was used through the National Center for Biotechnology Information (NCBI) ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) (Altschul et al., 1990 and 1997). *It is important to note that databases used “NpR3597” and “Npun_R3597” annotations interchangeably.

Upstream region analysis was done using CyanoBIKE ([http://biobike.csbc.vcu.edu/](http://biobike.csbc.vcu.edu/)). Multiple Sequence Alignments (MSAs) of NpR3597 homologous genes were done using
ClustalW v3.2 (Thompson et al., 1994) which was accessed through the San Diego Biology Workbench website (http://workbench.sdsc.edu).

Protein function analysis. Genes were selected for the MSA by choosing the closest homologs to NpR3597 from the results presented after running the nucleotide sequence through BLAST (www.ncbi.nlm.nih.gov). Putative conserved domains of the genes were designated using NCBI’s Conserved Domain Database (Marchler-Bauer, 2005).

Tertiary structure extrapolation. The amino acid sequence of all relevant genes were used to extrapolate their putative protein crystal structures by comparison to their closest structural homologues using the program FUGUE (Shi et al., 2001) which was accessed through the National Institute of Biomedical Innovation’s (NIBIO) website (http://tardis.nibio.go.jp/bioinfopro/). Fold library and substitution tables are based on the HOMSTRAD database. Crystal structure matches with a Z-score greater than 6 indicated a structural prediction with a 99% level of confidence.

Protein motifs. Sequence motifs were also analyzed using the GenomeNet Database “MOTIF search” (http://www.genome.jp/) which searched the BLOCKS (Henikoff, 1999), Pfam (Bateman, 2002), PRINTS (Attwood, 2002) and PROSITE (Falquet, 2002) databases that were used to establish NpR3597’s TetR family specific helix-turn-helix motif.

PCR amplification

PCR amplifications of DNA fragments were all carried out using the Robocycler® Gradient 96 Thermal Cycler (Stratagene).

Microarray

RNA Isolation. RNA was extracted by bead breakage followed by a Qiagen RNeasy® Mini protocol. ~100 mL of Log phase cell cultures (~3 ug/uL; ~300 ug total) were harvested in 50 mL conical vials for 5 minutes in clinical centrifuges at max speed ~13,000 rpm. Supernatant was removed and vials were centrifuged again for another 2 minutes. 1 mL of concentrated culture was put into 2 mL Mini-BeadBeater™ screw-cap tubes and microfuged for 2 minutes at max speed. The supernatant was removed and sample reduced to 700 uL to be snap frozen in dry-ice/ethanol bath then stored at -80°C until ready to isolate RNA. 0.58 g of 0.5 mm glass beads and 583 uL Tris-buffered phenol (pH 4.3) were added to the Mini-BeadBeater™ screw-cap tubes (Biospec) and shaken at high speed for 160 seconds in the Mini-BeadBeater™ (Biospec) to extract RNA from concentrated cells. Samples were then spun for 15 minutes at 4°C and ~650 uL of RNA containing supernatant (aqueous top layer) was transferred to microfuge tubes containing 600 uL of chloroform and vortexed for 15 seconds. Samples in chloroform were then spun for 10 minutes at 4°C and 500 uL of aqueous layer from cell extract was
added into new tubes containing 500 uL of precipitation solution (4 M lithium chloride, 20 mM Tris, pH 7.4, 10 mM EDTA, pH 8). Tubes were then mixed by inversion and precipitated on ice overnight. They were then spun for 15 minutes at 4°C and supernatant was discarded. 1 mL of cold 70% EtOH was then added to resuspend the RNA pellet by pipetting and then spun again for 10 minutes at 4°C. The supernatant was then removed, spun for an extra 1 minute and supernatant removed again. The pellet was resuspended in 100 uL TE buffer and then purified using the Qiagen RNeasy® Mini Kit with on-column DNase treatment following manufacturers instructions. RNA was quantified using a NanoDrop® 1000.

Microarray. Slides, labeling, amino blocking and hybridization, and analysis were done according to Campell et al. (2007). Three biological replicates of each strain were used for RNA extraction, each with a dye swap. After cDNA was made from RNA samples, each cDNA sample was coupled with both Alexa Fluor 647 (red) and Alexa Fluor 555 (green). 6 slides were used for hybridization. Three slides were hybridized with red WT/pSCR119 cDNA and green WT/pSCR119::3597 cDNA. The other three were dye swapped, which was hybridization with green WT/pSCR119 cDNA and red WT/pSCR119::3597 cDNA. Each slide had two technical replicates for each gene resulting in 12 data points for each gene.

Protein purification

Optimizing induction of His fusion protein. pET28a::3597-His was first transformed into E. coli Rosetta (DE3) pLysS competent cells. They were then plated on kanamycin selective LB media and left to grow for 24 hours at 37°C. 2-3 colonies were picked and inoculated into 5 mL of LB with kanamycin at 37°C overnight in a rolling incubator for aeration. The next day, a 1:10 dilution sample was made and the OD600 reading against pure LB blank was taken. We then diluted the culture to OD~0.15-0.2 in a total of 50 mL of fresh LB with kanamycin selection and grew it to an OD600 of ~0.6 with shaking at 37°C. 3 mL was removed before induction after target OD was reached. Induction of pET28a:3597-His was initiated by addition of IPTG to a concentration of 0.4 mM. 2 samples of the uninduced culture were cryogenically stored at -80°C to be used in large scale induction.

After induction, the culture was incubated for 4 hours at 35°C with vigorous shaking. 1.5 mL sample of culture was taken at time 0, 2, and 4 hours for SDS-PAGE analysis. The samples were centrifuged at maximum speed for 10 seconds down into a pellet, while the supernatant was discarded and the pellet was frozen on dry ice. The pellets from the three time points were then suspended in 300 µL of ice cold PBS and lysed using a sonicator with a microtip. Each sample was sonicated until the cloudy E. coli suspension became translucent which took about 10 seconds. 10-20 µL of this lysis solution was removed to analyze the soluble and insoluble (TOTAL) protein content via SDS-PAGE.
The sonicated solution was then centrifuged at 13,000 rpm for 5 minutes at 4°C to pellet down the insoluble material. 10 µL (20 for time 0) of this supernatant was taken and the soluble fraction of proteins was analyzed by SDS-PAGE. The supernatant was then removed and 300 µL of PBS was added to the insoluble pellet. It was then re-suspended by sonication and 10 µL (20 for time 0) was removed to analyze the insoluble fraction by SDS-PAGE to see if the induced protein was forming inclusion bodies. All samples were frozen in -20°C freezer until SDS-PAGE was to be performed. The three protein fractions from each time point were run on SDS-PAGE gel and visualized after staining with Coomassie blue.

**Induction.** For full scale induction of NpR3597 in preparation for purification, the above protocol was carried out using the *E. coli* Rosetta *(DE3)* p*lys*/*pET28a::3597-His strain in 500 mL of LB with kanamycin, instead of just 50 mL, grown to OD600 of ~0.6, and harvested 4 hours after induction by centrifugation at 4000 x g for 20 minutes. Pellets from 250 mL of culture was stored at -80°C until purification.

**Purification.** His-tagged 3597 protein was purified from pET28a::3597-His containing *E. coli* Rosetta(DE3)pLys strain using Ni-NTA Fast Start Kit and protocol (Qiagen). The mg/ml volume of protein was obtained using the A280 absorbance applied to the Beer-Lambert equation done by the NonoDrop 1000. 99.5 mg/ml NpR3597-His protein concentration was converted to 4.235 mM (4.235 moles/ul) by dividing 99.5 mg/ml by the molecular weight of the NpR3597-His protein (23.5 kg/mol) which resulted in 4.235 mM. The concentration of NpR3597-His in the EMSA DNA binding reactions was 2.4 mM.

**Electric Mobility Shift Assay, EMSA**

**EMSA.** The R3597-F3598 intergenic region DNA (P1) and truncated fragments (P2 & P3) were made by PCR using their corresponding primers 3598-Pal2.1-P1, 3598-Pal1-P2, and 3598-pal0-P3 with the common downstream primer f3598 Qp2 (Table 1). The R1931-F1932 intergenic region DNA (P1) and truncated fragment (P2) were made by PCR using their corresponding primers 1932-Pal1-P1 and 1932-Pal0-P2.1 along with the common downstream primer 1932-emsa P2 (Table 1). They were then Biotinylated after PCR amplification using the Biotin 3’ End DNA Labeling Kit (Thermo Scientific) run on 10% acrylamide gel using mini-protean tetracell apparatus (Bio-Rad) and physically excised from the gel on a transilluminator after staining in an ethidium bromide bath. The biotinylated DNA was then extracted from gel section by incubation in 50 uL of TE buffer pH 8 at room temperature overnight. Electric mobility-shift assay (EMSA) was carried out according to Chemiluminescent EMSA Kit protocol (Thermo Scientific) in native gels containing 9% acrylamide. Each reaction contained, 2 uL of binding buffer, 1µg/µL Poly (dI•dC), 250nM of designated DNA which was biotinylated, designated reactions positive for protein contained a 2.4mM concentration of R3597-His protein in
native form, and QS to 20 uL. These reactions were incubated for 20 minutes at room temperature before addition of 5µL of 5X loading buffer. Gels were run in mini-protean tetracell apparatus (Bio-Rad) with 5X TBE running buffer (450mM Tris, 450mM boric acid, 10mM EDTA, pH 8.3). The current was set to 100V for 8 × 8 × 0.1 cm gel and samples were electrophoresed until the bromophenol blue dye from the loading buffer had migrated approximately 2/3 to 3/4 down the length of the gel. Western blot was done using nylon membrane, Mini Trans-Blot Electrophoretic Transfer Cell and protocol (Bio-Rad). Visualization of Western Blot was carried out using LightShift® Chemiluminescent EMSA Kit (Thermo Scientific) and 20.3 x 25.4 cm Chemiluminescence BioMax® Light X-Ray Film (Kodak).

CCA experiment

Growth conditions. Nine WT/pSCR119::3597 and 9 WT/pSCR119 cultures of Nostoc were grown in 125 mL Erlenmeyer flasks containing 50 mL of AA/4 liquid media (Allen, 1955) supplemented with MAN and 10 µg/ml neomycin. They were grown to a density of ~6 µg Chla/ml under 10-12 umol of photons per square meter per second, shaken at a speed of ~150 rpms and kept at a temperature of ~25°C. After 6 µg Chla/ml was reached, the WT/pSCR119::3597 flasks and WT/pSCR119 flasks were wrapped with grey, red and green light filter gels. 3 of each of the 2 different strains was wrapped with grey filters, 3 of each with red and 3 of each with green. They were then incubated again under the same conditions for 10 days to late-log phase ~12 µg Chla/ml. All flasks were taken out of incubation at the same time and a 1 ml sample of each was extracted. The spectrophotometer was blanked against AA/4 with the same MAN concentration as starting flasks, and full cell absorption reading was taken for each sample from 400 nm to 730 nm. Before each reading, the samples were diluted to normalize density at an OD\textsubscript{730} of ~0.10.

Data analysis using student t-test. After the 400 – 730 nm full cell absorption spectrum was acquired for the 18 samples, the results of the red, green and grey light trials were put into separate excel spread sheets. The beta-carotene (466 nm), phycoerythrin (564 nm), phycocyanin (620 nm) and chlorophyll a (678 nm) absorptions were extracted from the data. Ratios of significant pairs of the photopigment peaks were calculated. 3 replicates of a particular ratio in WT/pSCR119 and 3 of WT/pSCR119::3597 were averaged and subject to a two-tailed student T-test using the Excel formula to find out if the ratio of photopigments in the cell were significantly different between the two strains. A P-value of 0.01 or lower for any comparison between the two strains was considered to be significantly different.
Desiccation experiment screening for viable akinetes

Wild-type, WT/pSCR119::3597, and Δ3597 mutant strains were grown through 2 bottle changes in log phase with inoculation to ~1 Chl\textsubscript{a} μg/ml in AA/4 with MAN (and 10 μg/ml of neomycin for WT/pSCR119::3597). They were then inoculated into AA/4 flasks with MAN and no antibiotic selection. After growing to 6-8 μg/ml Chl\textsubscript{a}, under 8-12 umol photons per meters squared per second and shaken at 150 rpms at 25°C, cells were visualized under microscope to confirm filaments ~100 cells long. The 50 mL culture was then pelleted by centrifugation in a 50 mL Falcon tube, and supernatant was decanted. The pellet was resuspended and washed with 40 mL of AA/4 supplemented with MOPS by gentle shaking 3 times. After the final wash, the pellet was resuspended in 50 mL of AA/4 with MOPS and set back in shaking incubator for 7 days for heterocyst differentiation. After heterocysts were confirmed using microscopy, cultures were put into 50 mL Falcon tubes and set out for 15 minutes to let heterocyst containing filaments settle to the bottom. The top layer containing hormogonia was decanted and the remaining culture was pelleted by centrifugation. The supernatant was discarded and the pellet was resuspended and washed three times as before but with AA/4 –Pi and MOPS. –Pi media deprives the cells of phosphates and induces akinete formation. After the last wash, the pellet was resuspended and added to a flask containing 50 mL of AA/4 –Pi and MOPS with the addition of 80 μL of 0.01X filter-sterilized +Pi stock used to make AA medium to aid in the acclimation of the culture to phosphate deprived conditions and reducing cell death. Cultures were maintained under low heat and low light conditions, approximately 23°C, with 50 rpm shaking and ~5 μmol photons per meter squared per second. Akinetes were visualized approximately 14 days later located between the heterocysts that were spaced 10-15 cells apart.
<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’-3’)</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>3597compl P1-Pst</td>
<td>GTCACTGCAAGGCATTATG</td>
<td>Forward primer with PstI site for construction of pSCR119::3597 over-expression plasmid</td>
</tr>
<tr>
<td>3597complP2-Kpn</td>
<td>GTTAGGTACCGATCGCTGGA</td>
<td>Reverse primer with KpnI site for construction of pSCR119::3597 over-expression plasmid</td>
</tr>
<tr>
<td>3597His Neo P1</td>
<td>TAATTCATGGTACGCATCAA</td>
<td>Forward primer with NcoI site for construction of pET28a::3597-6xHis</td>
</tr>
<tr>
<td>3597His Xho P2</td>
<td>TAGCTTCGAGATCGGGCGCATT</td>
<td>Reverse primer with XhoI site for construction of pET28a::3597-6xHis</td>
</tr>
<tr>
<td>3598-Pal2.2-P1</td>
<td>CAAGTTTAGTTATATCATCTTGACTA</td>
<td>3597 – 3598 forward primer for truncation EMSA</td>
</tr>
<tr>
<td>3598-Pal2.1-P1</td>
<td>ATCATCTTGACTAAACTAGTCCGTTT</td>
<td>3597 – 3598 forward primer for truncation EMSA</td>
</tr>
<tr>
<td>3598-Pal1-P2</td>
<td>GTTTAGCATAAATTTGAACTGAACAGT</td>
<td>3597 – 3598 forward primer for truncation EMSA</td>
</tr>
<tr>
<td>3598-pal0-P3</td>
<td>CAAGTTGCTACTGGTGTAAGG</td>
<td>3597 – 3598 forward primer for truncation EMSA</td>
</tr>
<tr>
<td>f3598 Qp2</td>
<td>ATAAACGCTGATTCGGCCTA</td>
<td>3597 – 3598 reverse primer for truncation EMSA</td>
</tr>
<tr>
<td>1932-Pal1-P1</td>
<td>CTTGTTTATTGGAGAATCATAACCA</td>
<td>1931 – 1932 forward primer for truncation EMSA</td>
</tr>
<tr>
<td>Primer</td>
<td>Sequence (5’-3’)</td>
<td>Function</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>1932-Pal0-P2.1</td>
<td>TATTTAATGATAAAAAATATA GGGTGGTATC</td>
<td>1931 – 1932 forward primer for truncation EMSA</td>
</tr>
<tr>
<td>1932-Pal0-P2.2</td>
<td>TGATAGTTCCAGTAAATATT TAATGATA</td>
<td>1931 – 1932 forward primer for truncation EMSA</td>
</tr>
<tr>
<td>1932-ems P2</td>
<td>CCACCTGCTATAACCAGCAA</td>
<td>1931 – 1932 reverse primer for truncation EMSA</td>
</tr>
<tr>
<td>T7 promoter primer</td>
<td>TAATACGACTCACTATAGGG</td>
<td>Forward primer confirming NpR3597 insert into pET28a</td>
</tr>
<tr>
<td>T7 terminator primer</td>
<td>GCTAGTTATTGCTCAGCGG</td>
<td>Reverse primer confirming NpR3597 insert into pET28a</td>
</tr>
<tr>
<td>M13 forward primer</td>
<td>GTTTTCCCAGTCAGAC</td>
<td>Forward primer confirming NpR3597 insert into pSCR119</td>
</tr>
<tr>
<td>M13 reverse primer</td>
<td>CAGGAAACAGCTATGAC</td>
<td>Reverse primer confirming NpR3597 insert into pSCR119</td>
</tr>
</tbody>
</table>
RESULTS

Bioinformatic Analysis

Protein motifs and structures of NpR3597. According to NCBI’s Conserved Domains Database, amino acids 21-67 of *N. punctiforme* gene NpR3597 are characteristic of a helix-turn-helix domain specific to the TetR family proteins. The sequence as a whole is homologous to AcrR type proteins (Fig. 3). Since the HTH domain makes up the region that is responsible for DNA sequence recognition and binding it (Fig. 4) is not a surprise that this functional domain is the amino acid region that is the most conserved.

Figure 3. NCBI’s conserved domains output indicates putative HTH domain and homology to AcrR type proteins.

In order to look more closely at the HTH domain, the GenomeNet Database’s “MOTIF search” (http://motif.genome.jp) was used to analyze the protein motifs of NpR3597. It showed more evidence of conservation in the 20aa-60aa region and defined the conserved domain found by the multiple sequence alignment (Fig. 5A and Appendix C). This website was used to search the Blocks, Pfam, PRINTS, and PROSITE databases, and resulted in identification of a TetR family specific helix-turn-helix domain from aa 21-63 or 64 (Fig. 5B and C). This is the region that comprises the second and third alpha helices from the N-terminal end of the protein that actually comes in contact and binds to the DNA-binding sequence (Fig. 4). Since the AcrR/TetR protein makes up a homodimer that binds an inverted repeat DNA recognition site that another homodimer of itself binds to on the the opposite side of the double-helix, this 20-60aa region is involved in binding to the distal major groove of the dsDNA of the outer one-half of the inverted repeat, while the other subunit will be turned 180 degrees to recognize the inner part of the other half of the inverted repeat at a major groove (Fig. 4). Since everything after the fourth alpha helix from the N-terminus is speculated, using crystal structure analysis, to be involved in ligand binding and since the ligands this family of protein bind to are...
different in varying species, there were no other significant protein motifs found in the downstream aa sequence.

A. Blocks database

<table>
<thead>
<tr>
<th>Position</th>
<th>Sequence</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>21..63</td>
<td>ILQGAMQEFLQNGYAGTSMDRVAVAAGVSKATVYSHFQDKEGL</td>
<td>1251</td>
</tr>
</tbody>
</table>

B. Pfam database identification

<table>
<thead>
<tr>
<th>Position</th>
<th>Sequence (Upper = NpR3597 / Lower = Database)</th>
<th>Conditional E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>21..65</td>
<td>ILQGAMQEFLQNGYAGTSMDRVAVAAGVSKATVYSHFQDKEGLFK</td>
<td>4.3e-20</td>
</tr>
</tbody>
</table>

Amino acid sequence of NpR3597:
MVRIKTGEVDRDNSVDKVEQILQGAMQEFLQNGYAGTSMDRVAVAAGVSKATVYSHFQDKEGLFKVLLEQLTSSKNSSIFGTEPIEPEAAILRQIVTKALEQMINDQEHSAFMRVLIGESGRFPELAIQCVRVMIKPVTENLTIQYLEAPELKGIDTPARILLGTLVHFHITQNYM岗HGPVDMPESDRLLDUALTHLITKCAD

Figure 5. Analysis of the N-terminal conserved region within NpR3597. A) Amino acid sequence multiple alignment and conserved regions of NpR3597 with 19 of its closest homologs. B) Motifs search of NpR3597 amino acid sequence in BLOCKS identifying a region from amino acid number 21-63 as IPB001647; Bacterial regulatory protein TetR, HTH motif. C) Motifs search of NpR3597 amino acid sequence in the Pfam database identifying it as TetR_N; Bacterial regulatory proteins, tetR family showing identification of the same conserved HTH domain (highlighted in red text within the complete amino acid sequence).

Gene locus organization of NpR3597 compared to TetR-family genes in other organisms, and its paralog, NpR1931. The putative AcrR gene NpR3597 is composed of a 615 bp protein coding sequence and is divergently transcribed from NpF3598, a gene with homology to a multi-drug efflux pump protein. The intergenic region between them is 300 bp. The orientation of NpR3597 transcription relative to its upstream divergently
transcribed efflux pump gene’s ORF is similar to that of its homolog acrR in *E. coli* (Fig. 6). Gene NpR1931 was found to be a paralog of NpR3597 in the *N. punctiforme* genome, also coding for a putative AcrR transcriptional repressor. NpR1931 is 639 bps long and codes for a 212 aa long AcrR protein. When aligned with NpR3597, NpR1931 had 51% identical aa sequence matches and 72% similar matches. Interestingly, the gene organization of NpR1931 was in a similar orientation to divergently transcribed MDR ABC transporter genes NpF1932, NpF1933, NpF1934 and NpF1935 (Fig. 6). NpR3597 and NpR1931 chromosomal architecture is also similar to that of tetR, which is also divergently transcribed from the tetracycline associated efflux pump genes tetA and tetC that it regulates in *E. coli*. The acrR homolog, pigZ, found in *Serratia* sp. ATCC 39006 is also similarly arranged, and has about the same number of base pairs separating it from its divergently transcribed 4 component efflux pump system with 247 bps. Also, the coding sequence of pigZ (612 bp), is similar to that of NpR3597 (615 bp). Although suspiciously similar in size, identity matches in there 204 aa sequences is only 36 out of a covered 130 aa’s or 28%, and positives for amino acid similarities is only 67/130 or 52%, mainly covering the N-terminal helix-turn-helix domain.

![Figure 6. Loci of putative AcrR genes in *N. punctiforme* NpR3597 and NpR1932 with known tetR and acrR family transcriptional repressors in *E. coli* and AcrR family gene pigZ in *Serratia* 39006 (Gristwood *et al.*, 2008) and their orientations relative to divergently transcribed MDR genes.](image-url)
In many instances it is the tertiary or quaternary structure, not the primary sequence, of a protein that is conserved. To test this, the amino acid sequence of AcrR homolog NpR3597 and its paralog, NpR1931, were submitted individually to the National Institute of Biomedical Innovation (NIBIO) in Japan (http://tardis.nibio.go.jp/bioinfopro/). At this site, the predicted crystal structure of the two homologs was obtained using the FUGUE program that utilizes the HOMSTRAD databases fold library and substitution tables to output a predicted tertiary structure of the protein in question based on primary aa sequence and secondary structure fold alignment with a previously crystallized protein. FUGUE output was a RasMol file that was the predicted 3D tertiary structure of NpR3597 (Fig. 7, left) and NpR1931 protein (Fig. 7, left) as compared to Cmer in Campylobacter jejuni, AcrR family repressor of multidrug efflux pump CmeABC (Fig. 7, right). NpR3597 amino acid sequence alignment with Cmer (shown in Appendix D) had a Z-score of 30.44 for its crystal structure prediction and NpR1931 amino acid sequence alignment (shown in Appendix E) had a similar Z-score of 31.88 for its prediction value. A Z-score equal to or greater than 6 indicates a 99% confidence in the prediction. By enabling labels on the FUGUE output of the predicted crystal structure of the NpR3597 and NpR1931 (which have a AA sequence similarity = 71%, and AA sequence Identity = 51%), I found the N-terminus end to contain the blue groups of the alpha helices. Turning the protein to match the right sub-unit of the AcrR dimer of E. coli, showed an amazing amount of similarity to the predicted crystal structure of NpR3597 and NpR1931.
Time-course microarray showing changes in gene expression during akinete induction

Figure 8. DNA microarray data for four *Nostoc punctiforme* genes up- and down-regulated in zwf mutant over 6 days in the dark relative to expression at time 0. Relative expression data was taken DNA microarray experiments comparing 6 time points (Day 1,2,3,4,5,6) to expression prior to akinete induction (M.L. Summers, unpublished results). The data points at day 0, although not actually measured, are all plotted as an expression level of 1 ($2^0$ on the log scale) to illustrate changes occurring during the first day of induction.

Time-course microarray data of NpR3597 encoding a putative AcrR, and MDR genes NpF3598, NpF3599, NpF3600 that represent the repressor with its divergently transcribed efflux pump system, was plotted to show expression changes following akinete induction. An inverse relationship was seen between NpR3597 and divergently transcribed NpF3598 over the 6 days after akinete induction by light deprivation of the zwf mutant. At day 4 of akinete induction, NpR3597 had its peak up-regulation with a 1.5-fold increase in expression. NpF3598 had a 2-fold decline in expression just after 24
hours. There was no significant change in expression of NpF3599 or NpF3600 in the time-course array.

Expression of the NpR3597 paralog NpR1931 was not significantly altered in the time course array. Of the 4 divergently transcribed genes, on the second, NpF1933 encoding for an ABC type multi-drug transporter showed a change in expression, exhibiting a 2- to 4-fold increase during the first 3 days, peaking at day 2 following akinete induction.

**Microarray of WT/pSCR119::3597 vs. WT/pSCR119**

A DNA microarray was conducted to compare gene expression of a WT/Pscr119::3597 over-expression strain to a WT/Pscr119 control strain in late-log-phase Six slides were used for hybridization. Three slides were hybridized with the three biological replicates red WT/pSCR119 cDNA and three replicates of green WT/pSCR119::3597 cDNA. The other three slides were dye swapped. Each slide had two technical replicates for each gene resulting in 12 data points for each gene. The array showed that transcription of NpR3597 was increased 135-fold over the control. Although not measured, it was assumed that protein expression for this gene product is also up-regulated in the over-expression strain.

The over-expression strain showed 1.5-fold higher transcriptional expression of 136 genes (Appendix B). Sixty-six of these were increased > 2-fold and included the known akinete marker gene avaK (NpF5452; 4.5-fold) (Zhou and Wolk, 2002), and NpF4153 (1.6-fold), an ECF sigma factor, SigE, others in our lab have linked to cell development and survival following envelope damage (M.L. Summers, pers. com). Several other genes also encoding for adaptive functions were up-regulated, such as the two-component regulator NpF1453 and another putative transcriptional regulator, NpR6520, both of which contain a helix-turn-helix DNA binding domain. Neither of these has a known function, however NpF1453 is currently under study in our lab and has been shown to have increased expression in developing akinetes and during stationary phase of growth (M.L. Summers, pers. com).

Approximately 64 % (87/136) of the up-regulated genes are classified as unknown, hypothetical, or that grouped into no definitive physiological category. Of the remaining genes, grouping of multiple genes into specific physiological functions indicated potential up-regulation of peptidoglycan biosynthesis, sugar polymerization and glycogen synthesis associated with carbon storage or envelope biosynthesis, as well as several transport systems with potential roles in heavy metal export (Appendix B). Several genes associated with reaction centers of photosystem I and II are also up-regulated. If the stringency for 1.5X or higher gene expression is lowered to any gene showing increased expression with statistical significance of less than one chance in all the genes in the array (P<1.4 X 10^-5), the list increases to 214 genes, including the over-expressed NpR3597. Two phycobilisome associated proteins are up-regulated when R3597 is overexpressed, which are the phycocyanin alpha-subunit (NpF5290; 1.3-fold) and a phycobilisome associated gene (NpF0736; 1.7-fold) containing an allophtycocyanin linker conserved domain (Marchler-Bauer, 2005).
There are 162 down-regulated genes using the lower P-value<1.4 X 10^{-5} only stringency (Appendix A), that is reduced to 66 genes if the higher >1.5-fold regulation is imposed. This is less than half the number of up-regulated genes at the same higher stringency level. When using the higher cut-off, many genes involved in adaptation are reduced, including a possible anti-sigma factor (NpR1192; 3.7-fold) and SigC (Npf0996; 3.3-fold) implicated in gene regulation during stationary phase (Asayama et al., 2004). Several predicted S-layer proteins are also reduced (NpR5769, NpR2279), as are both subunits of Ribulose-bisphosphate carboxylase (rbcS/rbcL; NpF4197/NpF4195; each at ~1.6-fold).

Of interest in relation to the AcrR-like proteins described above, the NpF1932 and NpF3598 genes, representing the first gene in the divergently transcribed operons of NpR1931 and NpR3597, are down regulated (4.1 and 1.9-fold, respectively). These results indicate that over-expression of NpR3597 may act to repress ABC-type transporters other than the one in its own genetic loci. As is also the case with the up-regulated genes in the array results, direct regulation of genes by NpR3597 cannot be differentiated from indirect regulation by other transcriptional regulators showing altered expression in the array.

Construction of HIS-tagged R3597

```
actatttttccccctttactccaccacagtacactttgtctactaaactgttccagtcaatattatgtct
aactaaacgcgactggttagtctcaagatgataaatatactgcatcactaatatttggttttttttctccaccattttttttttttttttttttttttattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
```

Figure 9. NpR3597 gene sequence (capital letters) with up-stream/down-stream sequence (lower case) showing primer sites for PCR amplification (underlined). The bases replaced by the upstream NcoI (CCATGG) and downstream XhoI (CTCGAG) restriction enzyme sites are indicated by bold text within the primer region. The transcriptional start site was determined by RACE analysis and indicated by the capital “A” in the intergenic region. The PCR fragment produced with these primers was 628 bp in length.
PCR with 3597His Nco P1 primer allowed for inclusion of the natural in frame ATG start site (Fig. 9, shown in bold text). The native transcriptional start site was not include so that the pET28a T7 promoter could act to appropriately regulate transcription of NpR3597-His in the *E. coli* Rosetta strain. The goal with primer 3597His Xho P2 was to simultaneously remove the TGA termination codon, create a *XhoI* restriction enzyme site, and stay in frame with the down stream CACCACCACCACCACCAC on the plasmid sequence coding for 6 histidines.

The NpR3597 coding region was successfully inserted into the pET28a his-tag vector with its ATG start site down stream of the T7 promoter. The TGA stop signal was mutated to CTC in the *XhoI* site so that translation could read through the multi-His region and terminate at the following TGA encoded within the vector (Fig. 10).

---

**Figure 10. Translation of sequence after insertion of NpR3597 into pET28a vector.** The plasmid sequence is not highlighted and the NpR3597 DNA insert is highlighted. The T7 transcriptional promoter is underlined, and the ATG start site is shown in bold. Translation was done using the ExPASy translation tool at the ExPASy Bioinformatics Resource Portal (http://web.expasy.org/translate/) website.
The ATG start site starts the in-frame translation through the insert coding region and also translates the six-histidines at the C-terminus of the insert and stops at the TGA stop site immediately proceeding (Fig. 10).

Figure 11. Ethidium bromide stained agarose gel containing colony PCR products from clones transformed with the pET28a::3597-His ligation.

Figure 11 shows PCR performed on colonies after transformation of the pET28a::3597-His ligation into E. coli DH5 alpha MCR. T7 forward and T7 reverse primers flanking the multiple cloning site would make a 310 bp PCR product if there were no insert. By subtracting 158 bp, which is the distance between the NcoI site and XhoI site, and adding the 614 bp R3597 insert (size of insert excludes the nucleotides that were excised from the PCR product by the digestion with the restriction enzymes NcoI and XhoI), the colony PCR should result in a 766 bp product if the insert was successfully cloned. This is evident in all the screened colonies shown in the figure with the size of the PCR product being between 700 bp and 1,000 bp as compared to the labeled DNA ladder.
Induction and Purification of the R3597-His Native Protein

Figure 12. PAGE-SDS showing optimization of induction conditions for His-tagged NpR3597 expression from *E. coli* bearing pET28a::NpR3597. The first lane shows the ladder with 20kD and 30kD markers labeled to the left. The next three lanes (Time 0) contain samples of extracted total fraction (T), sample extracted from soluble fraction (S), and sample extracted from resuspended in-soluble fraction, or pellet fraction (P). The same fractions were sampled after 2 hours and 4 hours. The over-expressed 23.5 kD protein of interest is apparent in the T and S lanes of the 4 hour samples.

Transcription of R3597-His in *E. coli* Rosetta (DE3) pLysS strain was induced at time 0 with IPTG and samples were taken of the lysed sonicated total sample, the supernatant after centrifugation of cellular debris and the resuspended pellet of cellular debris at time 0, 2, and 4 hours after induction. The SDS-PAGE gel (Fig. 12) indicates the approximately 23.5kD sized 3597-6xHis protein, can be found increasing in quantity, from time 0 h to 4 h. It is not present at time 0 because it has not yet been induced by IPTG. In the total fraction and the soluble fraction at time 2 h after induction it starts to appear, and can be seen optimally induced after 4 hours within the soluble fraction.
Figure 13. SDS-PAGE gel of nickel column purification of NpR3597-His protein. The first lane shows a protein ladder with the 20 kD and 30 kD protein markers labeled in red. The next lane shows the protein content of a 10 µL sample of the 10 mL lysate before being run through the nickel column. The lane labeled “Flow Through” shows the protein content of the collected lysate fraction after being run through the column. “Wash #1” and “Wash #2” represent 10 µL samples of collected fractions after two washes with 4 mL native Wash Buffer containing 20 mM imidazole. “Elute #1” and “Elute #2” represent 10 µL samples of collected fractions after two elutions with 1 mL aliquots of native Elution Buffer containing 250 mM.

After induction of pET28a::3597-6xHis by IPTG and lysis of E. coli Rosetta (DE3) pLysS cells after 4 hours, the protein was purified under native-conditions using a nickel column. Protein purification from the soluble fraction after lysis was analyzed by running 10 µL samples of the soluble fraction of lysed cells, the nickel column flowthrough, two low concentration imidazole washes, and two high concentration imidazole elutions (Fig. 13). The first addition of the soluble fraction of the lysed cells to the column was done so that the histidine tagged NpR3597 protein could stick to the nickel resin. The low concentration 20 mM imidazole washes were performed to rid the nickel column of any non-specific contaminant proteins. The high-concentration 250mM imidazole washes allow the imidazole to out-compete NpR3597-His for the nickel due to imidazole’s higher affinity for it, and NpR3597 is eluted out of the column. The 23.5 kD
His-tagged 3597 protein is found in the first and second elution, and the second elution, which is the most pure, was used in the DNA-Protein binding Electric Mobility Shift Assay.

**Truncated DNA fragments for use in EMSA**

![Diagram showing truncated DNA fragments](image)

Figure 14. Truncation of upstream regulatory regions of NpR3598 used in EMSA to test binding of AcrR protein NpR3597 to inverted repeat sites. The location of the putative -35 and -10 sites are boxed and the blue and red underlines indicate the 2 inverted repeat regions suspected to be the NpR3597 binding sites. The “L” shaped arrow indicates the F3598 translational start codon.

After biotinylation, the R3597-F3598 intergenic region DNA (P1) and truncated fragments (P2 & P3) were used for the DNA-Protein interaction gel shift assay. NpR3597 binding to the blue palindrome and not the red (Fig. 14) would show a mobility shift in the P1 fragment but not in P2 or P3. R3597 binding to the red palindrome but not the blue would show a mobility shift with P1 fragment and a similar sized shift with P2 fragment and no shift with P3. Binding to both red and blue inverted repeat regions would show a supershift in P1 with a smaller sized shift using P2 and no shift in P3. No shift at all in any fragment indicates no binding of the NpR3597 protein to any of the palindromes. If the protein is binding to P3 as well, then the interaction is non-specific.
Figure 15. Truncation of upstream regulatory region of F1932 used in EMSA to test binding of AcrR protein NpR3597 to inverted repeat site. The location of the putative -35 and -10 sites are boxed and the red underline indicate the inverted repeat region thought to be the NpR3597 binding site. The “L” shaped arrow indicates the F1932 start codon.

The NpR1931-F1932 intergenic region DNA (P1) and truncated fragment (P2) were used for the DNA-protein interaction with NpR3597 gel shift assay. Fragment P1 contains the inverted repeat region suspected of being the site of NpR3597 protein binding at the putative -35 site, while fragment P2 excludes the possible binding site. In the protein-DNA interaction mobility shift assay, the P1 fragment should show a mobility shift, if NpR3597 is binding to the palindrome, and there is no shift in P2. If there is no shift in P1 or P2 then the protein doesn’t bind. If there is a shift in both, then the protein binding is likely non-specific.
Figure 16. PCR of truncated DNA fragments for use in EMSA

Figure 16 shows the agarose gel electrophoresis of the PCR amplification of 1931-1932 intergenic region truncated fragments: 201 bp fragment P1 (done using forward primer 1932-Pal1-P1 and reverse primer 1932-emsa P2) and 149 bp fragment P2 (done using forward primer 1932-Pal0-P2.1 and reverse primer 1932-emsa P2) on the left of the DNA ladder. To the right of the DNA ladder, you can see PCR products of the NpR3597-F3598 intergenic region truncated fragments: 202 bp fragment P1 (done using forward primer 3598-Pal2.1-P1 and reverse primer f3598 Qp2), 175 bp fragment P2 (done using forward primer 3598-Pal1-P2 and reverse primer f3598 Qp2), and 143 bp fragment P3 (done using 3598-pal0-P3 and reverse primer f3598 Qp2). Refer to figure 14 and figure 15 for map of the contents in each fragment. There are some minor contaminant PCR products present above the dark correct-sized fragments, but after biotinylation of the fragments, the desired bands were excised from a 10% acrylamide gel (not shown), and contaminants were excluded.
Figure 17. Detailed location of primers used for EMSA; inverted repeat binding site exclusion by truncation of F1932 upstream region. Regions hybridizing to upstream and downstream primers are boxed and shaded, respectively. The ORF of NpF1932 is shown in green letters and the divergently transcribed AcrR homolog gene NpR1931 is shown in brown letters. Primer 1932-Pal1-P1 (P1) and primer 1932-Pal0-P2.1 (P2) both forward primers are shown with down-stream primer 1932-emsa P2 (PReverse,Yellow).

The 1931-1932 intergenic region contains a 25 bp long inverted repeat site at the putative -35 of Multi-drug resistance efflux pump gene NpF1932 as shown in the figure with converging red arrows on top of the inverted repeat (Fig. 17). Primer P1 and PReverse were used to PCR amplify the DNA fragment P1 (Fig. 15), which includes the inverted repeat site for the mobility band-shift assay. Primer P2 in the second box and PReverse were used to PCR amplify the DNA fragment P2, which excludes the inverted repeat site for the mobility band-shift assay. It is predicted that the fragment P1 including the inverted repeat site should show a band-shift when combined with NpR3597 native protein and show no band shift with P2, localizing DNA binding of the protein to the inverted repeat region.
Figure 18. Detailed location of primers used for EMSA; inverted repeat binding site exclusion by 2 truncations of F3598 upstream region. Regions hybridizing to upstream and downstream primers are boxed and shaded, respectively. The ORF for NpF3598 is shown in capital letters with the ATG translation start site underlined.

The 3597-3598 intergenic region contains two adjacent inverted repeat regions. The first one shown in blue is a 20 bp long inverted repeat site (Fig. 18). The second inverted repeat region that is shown in red is 23 bp long and located more proximal to the NpF3598 gene. Primer P1 and PReverse were used to PCR amplify the P1 DNA fragment, which includes both of the inverted repeat regions, to be used in the mobility band-shift assay. Primer P2 in the second box and PReverse were used to PCR amplify the P2 DNA fragment that includes the second inverted repeat site red but excludes the first blue inverted repeat, all except two downstream nucleotides, to define exactly which palindrome the NpR3597 protein can bind to. Primer P3 in the third box and PReverse were used to PCR amplify DNA fragment P3 for the mobility band-shift assay that excluded both inverted repeat sites, all except 3 downstream nucleotides of the red inverted repeat, and should not show any band-shift if protein binding is localized at the first and/or second palindromes.
Electric Mobility Shift Assay, EMS

<table>
<thead>
<tr>
<th>Protein</th>
<th>DNA Fragment</th>
<th>Palindrome Present</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P1</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>P1</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>P2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>P2</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 19. NpF1932 upstream regulatory region EMSA using truncated fragments and NpR3597 protein

NpR3597-His directly binds to the inverted repeat in the upstream region of NpF1932 (Fig. 19). DNA mobility band-shift assays were performed on fragments P1 and P2 using the NpR3597 nickel column purified protein in its native form. Once again, each lane contained 250nM of designated DNA which was biotinylated. Designated lanes positive for protein contained a 2.4mM concentration of NpR3597-His protein in native form. The first lane shows P1 DNA fragment without protein added to the sample. P1 with native NpR3597-His protein shows a band-shift indicating that NpR3597-His is bound to the P1 fragment. The reaction of NpR3597 with the truncation fragment P2 excluding the first palindrome shows no shift indicating that NpR3597 is definitely bound to the first inverted repeat region present in P1. Exclusion of the first palindrome shows a loss of band-shift in the higher molecular weight band, indicating that NpR3597 binding is localized at the first inverted repeat site and possible non-specific binding sites downstream of it, leaving a less distinct band-shift.
Figure 20. NpF3598 upstream regulatory region EMSA using truncated fragments and purified NpR3597-His protein.

NpR3597-His directly binds to both inverted repeats in the upstream region of NpF3598. DNA mobility band-shift assays (Fig. 19) were performed on fragments P1, P2, P3 (from Fig. 14) using the NpR3597 nickel column purified protein in its native form. Each lane contained 250 nM of designated DNA which was biotinylated. Designated lanes positive for protein contained a 2.4 mM concentration of NpR3597-His protein in native form. The first lane shows P1 DNA fragment without protein added to the sample. P1 with native NpR3597-His protein shows a band-shift indicating that NpR3597-His is bound to the P1 fragment. The reaction of NpR3597 with the truncation fragment P2 excluding the first inverted repeat shows a shift that is not as drastic as seen for fragment P1, indicating that 3597 is definitely bound to the second site and that the “super-shift” apparent for fragment P1 was caused by binding to both sites. Exclusion of the first and second palindrome shows no band-shift at all, indicating that NpR3597 binding is localized at both the first and second inverted repeat sites and there are no more binding sites down-stream of them. The third fragment also serves as a control indicating the shifts observed for the first two fragments is not due to non-specific interaction with purified NpR3597-His protein.
**Complimentary Chromatic Adaptation (CCA) assay,**

**Figure 21.** Graph of whole cell absorption spectrum from 400 nm – 730 nm of WT/pSCR119 vs. WT/pSCR119:3597 under grey light filters. Each line represents an average of 3 samples from 3 separate cultures. The whole cell absorption spectrum from 400 nm to 730 nm of WT/pSCR119 control strain (black line) and WT/pSCR119::3597 over-expression strain (red line) after growth under grey light filters with Phycoerythrin peak (PE), phycocyanin peak (PC) and Chlorophyll a peak (Chlα).

The phycoerythrin absorption signal peaks are greatest at 564 nm with the phycocyanin peak greatest at 620 nm and the chlorophyll a peaks greatest at 678 nm. Measurements were therefore taken at these specific wavelengths. In cells growing under a neutral density filter to control for the plastic gel used in these experiments, it is evident that the ratio of phycocyanin to chlorophyll a content in the representative cells is greater in the over-expression strain (red) than the wild-type control (black line in Fig. 21).
Figure 22. CCA absorption results showing difference in phycocyanin/chlorophyll-a ratio between wild-type and overexpressor of NpR3597.

Supporting the casual observations of the whole-cell scans, the calculated ratio of PC/Chlα was significantly greater in WT/pSCR119::3597 over-expression strain compared to WT/pSCR119 control strain under grey light according to statistical t-test (Fig. 22). The ratios of PC/Chlα (620nm/678nm) from 3 over-expression strain samples averaged 1.043 and 3 control strain samples averaged 0.948. The averages of those 2 sets were found to be significantly different with a P-value of 0.01 and 99% confidence. It is also evident that according to the PE/PC ratios, the PC content significantly increases in the over-expression strain relative to the PE content with a P-value of 0.0499.

<table>
<thead>
<tr>
<th>nm</th>
<th>overex-1</th>
<th>overex-2</th>
<th>overex-3</th>
<th>pscr119-1</th>
<th>pscr119-2</th>
<th>pscr119-3</th>
<th>PE/PC</th>
<th>PC/Chlα</th>
<th>PE/Chlα</th>
<th>B-car/ChlαA</th>
<th>B-car/PC</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-car 466</td>
<td>0.156725</td>
<td>0.159924</td>
<td>0.166417</td>
<td>0.169719</td>
<td>0.16593</td>
<td>0.166136</td>
<td>1.061972</td>
<td>1.029032</td>
<td>1.046622</td>
<td>0.955236</td>
<td>0.942409</td>
</tr>
<tr>
<td>PE 564</td>
<td>0.14099</td>
<td>0.148478</td>
<td>0.144357</td>
<td>0.151079</td>
<td>0.144222</td>
<td>0.150073</td>
<td>0.963588</td>
<td>0.967382</td>
<td>1.016109</td>
<td>1.054559</td>
<td>1.063162</td>
</tr>
<tr>
<td>PC 620</td>
<td>0.182647</td>
<td>0.165317</td>
<td>0.165747</td>
<td>0.160938</td>
<td>0.156072</td>
<td>0.163467</td>
<td>1.087924</td>
<td>1.063467</td>
<td>1.063162</td>
<td>1.016331</td>
<td>0.027508 PE/Chlα</td>
</tr>
<tr>
<td>Chlα 678</td>
<td>0.153168</td>
<td>0.161027</td>
<td>0.158364</td>
<td>0.166902</td>
<td>0.163385</td>
<td>0.173458</td>
<td>0.920566</td>
<td>0.918943</td>
<td>0.911551</td>
<td>0.896314</td>
<td>0.882709</td>
</tr>
<tr>
<td>PE/Chlα 564/678</td>
<td>1.081972</td>
<td>1.029032</td>
<td>1.046622</td>
<td>0.955236</td>
<td>0.942409</td>
<td>0.010304 PC/Chlα</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE/Chlβ 564/678</td>
<td>0.920566</td>
<td>0.918943</td>
<td>0.911551</td>
<td>0.896314</td>
<td>0.882709</td>
<td>0.865195</td>
<td>0.027508 PE/Chlα</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-car/Chlα 466/678</td>
<td>1.023304</td>
<td>0.987631</td>
<td>1.063482</td>
<td>1.000159</td>
<td>1.015571</td>
<td>0.9578</td>
<td>0.029786 B-car/ChlαA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-car/PC 466/620</td>
<td>0.963588</td>
<td>0.967382</td>
<td>1.016109</td>
<td>1.054559</td>
<td>1.063162</td>
<td>1.016331</td>
<td>0.049867 B-car/PC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Red Filter](image)
Figure 23. Graph of whole cell absorption spectrum from 400 nm – 730 nm of WT/pSCR119 vs. WT/pSCR119:3597 under grey light filters. Each line represents an average of 3 samples from 3 separate cultures.

Figure 23 shows the whole cell absorption spectrum from 400 nm to 730 nm of WT/pSCR119 control strain (black line) and WT/pSCR119::3597 over-expression strain (red line) after growth under red light filters that cut-off the green spectrum and promote phycocyanin accumulation. The phycoerythrin absorption signal appears to be at the same level. It is evident that the ratio of phycocyanin to chlorophyll a content is greater in the over-expression strain (red) than the wild-type control (black).

<table>
<thead>
<tr>
<th></th>
<th>nm</th>
<th>overex-1</th>
<th>overex-2</th>
<th>overex-3</th>
<th>pscr119-1</th>
<th>pscr119-2</th>
<th>pscr119-3</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>B-car</td>
<td>466</td>
<td>0.151944</td>
<td>0.1575351</td>
<td>0.156219</td>
<td>0.149648</td>
<td>0.165258</td>
<td>0.170032</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE</td>
<td>564</td>
<td>0.137804</td>
<td>0.14020796</td>
<td>0.140186</td>
<td>0.135656</td>
<td>0.139294</td>
<td>0.144986</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC</td>
<td>620</td>
<td>0.161885</td>
<td>0.16634764</td>
<td>0.165466</td>
<td>0.157304</td>
<td>0.153969</td>
<td>0.167456</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chla</td>
<td>678</td>
<td>0.151224</td>
<td>0.15741828</td>
<td>0.15824</td>
<td>0.162343</td>
<td>0.156512</td>
<td>0.170064</td>
<td>P-Val</td>
<td></td>
</tr>
<tr>
<td>PE/PC</td>
<td>564/620</td>
<td>0.938596</td>
<td>0.9470234</td>
<td>0.944114</td>
<td>0.95079</td>
<td>1.073322</td>
<td>1.015379</td>
<td>0.186287PE/PC</td>
<td></td>
</tr>
<tr>
<td>PC/Chla</td>
<td>620/678</td>
<td>1.070499</td>
<td>1.05673717</td>
<td>1.045664</td>
<td>0.969511</td>
<td>0.93754</td>
<td>0.979482</td>
<td>0.001742PC/Chla</td>
<td></td>
</tr>
<tr>
<td>PE/Chla</td>
<td>564/678</td>
<td>0.91126</td>
<td>0.8906827</td>
<td>0.885903</td>
<td>0.835613</td>
<td>0.890991</td>
<td>0.848983</td>
<td>0.132256PE/Chla</td>
<td></td>
</tr>
<tr>
<td>B-car/ChlaA</td>
<td>466/678</td>
<td>1.004766</td>
<td>1.00075483</td>
<td>0.987226</td>
<td>0.921801</td>
<td>1.055885</td>
<td>0.994546</td>
<td>0.876798B-car/ChlaA</td>
<td></td>
</tr>
<tr>
<td>B-car/PC</td>
<td>466/620</td>
<td>0.938596</td>
<td>0.9470234</td>
<td>0.944114</td>
<td>0.95079</td>
<td>1.073322</td>
<td>1.015379</td>
<td>0.186287B-car/PC</td>
<td></td>
</tr>
</tbody>
</table>

Figure 24. CCA results showing difference in phycocyanin/chlorophyll-a ratio between wild-type and overexpressor of NpR3597.

Once again, when quantified, the ratio of PC/Chl a is significantly greater in WT/pSCR119::3597 over-expression strain compared to WT/pSCR119 control strain under red light according to statistical t-test. The ratios of PC/Chl a from 3 over-expression strain samples averaged to 1.058 and the 3 control strain samples averaged to 0.978. The averages of those 2 sets were found to be significantly different with a P-value of 0.0017 and 99.9% confidence. This is almost 10% more confidence in the difference than under grey light.
Figure 25. Graph of whole cell absorption spectrum from 400 nm – 730 nm of WT/pSCR119 vs. WT/pSCR119::3597 under green light filters

The graph above shows the whole cell absorption spectrum from 400 nm to 730 nm of WT/pSCR119 control strain (black line) and WT/pSCR119::3597 over-expression strain (red line) after growth under green light filters that cut-off the red spectrum and promote phycoerythrin accumulation. The phycoerythrin absorption signal peaks are greatest at 564 nm, and between the overexpressor and control strain the absorption appears to be increased compared to the other filter conditions. Not only has the PE peak increased in both, but they have done so to the same level. It is evident that the ratio of phycocyanin to chlorophyll a content is greater in the over-expression strain (red) than the wild-type control (Black). Ultimately, phycocyanin remains high and chlorophyll a remains reduced.
Figure 26. CCA results showing difference in phycocyanin/chlorophyll-a ratio between wild-type and overexpressor of NpR3597 grown under green light.

Once again, the ratio of PC/Chla is significantly greater in WT/pSCR119::3597 overexpression strain compared to WT/pSCR119 control strain under green light according to statistical t-test, and is so to an even greater degree than red and grey light conditions. The ratios of PC/Chla from 3 over-expression strain samples averaged to 1.019 and 3 control strain samples were averaged to 0.925, the averages of those 2 sets were found to be significantly different with a P-value of 0.0001 (Fig. 26) and 99.99% confidence. This is almost 10% more confidence in the difference than under red light, and almost 100% more confident in the difference than under grey light. It is apparent that causing the cells to perform chromatic adaptation by altering the quality of light exaggerated the phenotype of increased PC/Chla. This indicates initial observations were real, and that the over-expression of this transcriptional regulator has a pleiotropic phenotype that extends beyond regulation of transporters.
Figure 27. Expression of GFP in WT/pSUN119::P_{3597} in stationary phase vs. log phase under 40x magnification epifluorescent microscopy. A, shows periodic GFP expression in *Nostoc* filaments that are in log phase under the GFP filter. B, natural fluorescence of phycobilisomes in the filament shown in A under Texas Red filter. C, *Nostoc* filaments in stationary phase with minimal GFP expression under GFP filter. D, Natural fluorescence of phycobilisomes in filaments shown in C under Texas Red filter.

GFP expression tests showed that pSUN119::P_{3597} is ON during log phase when cells are dividing in a periodic pattern of expression along the filament (Fig. 27A). The NpR3597 promoter seems to be inactive during stationary phase, which is the time when cells have a reduced rate of growth and cell division.
Figure 28. GFP expression of WT/pSUN119::P₃₅₉₇ in stationary phase vs. log phase under 100x magnification epifluorescent microscopy. A, shows periodic GFP expression in pre-dividing cells in *Nostoc* filaments that are in log phase. B, natural fluorescence of filaments under Texas Red filter. C, *Nostoc* filaments in stationary phase with minimal GFP expression under GFP filter. D, Natural fluorescence of filaments under Texas Red filter.

GFP expression of pSUN119::p3597 seems to be on during log phase and with a closer look under 100x magnification (Fig. 28), expression is observed to be specific to the bigger cells on the vegetative filament that are just about to divide and off during stationary phase during the time cells are not dividing.
Akinete viability assay

Figure 29. Akinete viability test in wild-type, -3597 mutant, and pSCR119::3597 overexpression strain

A preliminary experiment was conducted to test if over-expression of NpR3597 caused a phenotype related to survival. After akinete induction, confirmation of akinete formation in filaments between heterocysts by microscopy and dessication of wild-type, WT/pSCR119:3597 over-expression strain and Δ3597 mutant strain; the three strains were grown on AA media plates containing MAN in 1:1, 1:2, 1:4, 1:8 dilutions. After incubation for 20 days, the mutant and wild-type grew similarly to the same degree indicating that the akinetes induced before desiccation were functional. The over-expression strain however, showed very limited growth and ultimately a defect in the functionality of its akinetes to survive such stress.
DISCUSSION

According to NCBI, the gene NpR3597 in *Nostoc punctiforme* PCC 73102 is classified as an AcrR type transcriptional regulator belonging to COG1309 because of its sequence similarity to other AcrR’s in its amino acids 19-201. Amino acid region 21-64 defines a helix-turn-helix (HTH) domain that is specific to TetR family of transcriptional repressors.

The AcrR dimer of *E. coli*, which has already been obtained through crystallography in the Li *et al.* (2007) paper, showed an amazing amount of similarity to the predicted crystal structure of NpR3597 and NpR1931 (Fig. 7). The first three alpha helices of the *Nostoc* homolog protein, Cmer, match very well with the first three AcrR alpha helices. The rest of the alpha groups are in a very similar conformation, capable of attaching to another identical subunit to form a homodimer. Also, the interior of the single subunit is observed to be similar in ligand binding capability to that of AcrR in *E. coli*.

The predicted structure can be related back to the protein’s TetR motif. The motif is the helix-turn-helix signature specific to TetR family proteins spanning amino acid 21- to about 65, which is the localization of the end of alpha helix 1 and includes alpha helix 2 and 3. These groups are seen in blue on the 3D FUGUE output (Fig. 7), and annotated in the AcrR figure to the right of it. The AcrR paper applied to this study (Li *et al.*, 2007) shows that these conserved groups are indeed conserved for the purpose of binding to a specific helix-turn-helix binding sequence of the compatible DNA site.

The function of AcrR proteins has been defined to transcriptionally regulate genes coding for multi-protein multi-drug resistance (MDR) efflux pumps that are sometimes set up as operons. They regulate these MDR genes by forming homodimers and binding to inverted repeat DNA with their HTH domains which comprises alpha-helices 1-4 of its crystal structure (Gu *et al.*, 2007; Gu *et al.*, 2008) In *E. coli*, divergently transcribed gene acrA is transcriptionally repressed by the binding of two AcrR homodimers to an inverted repeat site in the -35 promoter region of acrA. The efflux pump genes are usually, but not always, located upstream of the acrR and transcribed in the opposite direction (Li *et al.*, 2007; Su *et al.*, 2007). NpR3597 has a similar orientation to its divergently transcribed multidrug resistance efflux pump genes NpF3598, NpF3599 and NpF3600 (Fig. 6).

In the preceding microarray showing regulation of genes during akinete induction (Fig. 8), the timing of the up-regulation of 3597 does not correspond to the down-regulation of its presumptive target NpF3598. NpF3598 is quickly down-regulated 2-fold just after 1 day of induction, while NpR3597 transcriptional regulator is barely up-regulated ~1.15 fold. However, NpR3597 does continue to increase to 1.5 fold over the duration of the time-course. Since AcrR/TetR family proteins are known to exhibit change in functionality following ligand binding (Li *et al.*, 2007), this may indicate that an inducer of the NpR3597 protein that is already present in the cell is quickly produced to repress the upstream NpF3598 gene, and later more NpR3597 is transcribed to keep it off. Crude
extract from zwf mutant strain at day 1 following akinete induction might be the best source for the discovery of such an inducer molecule.

**Direct regulation**

By performing an Electrical Mobility Shift Assay (EMSA), we tested the hypothesis that NpR3597 has the same binding patterns characteristic to that of AcrR. The results showed that the NpR3597 protein product in its native form does indeed bind to 2 specific sites in the putative -35 region of divergently transcribed MDR gene NpF3598. Interestingly, the two sites have a 7 bp long consensus sequence GTTTAGT on the right side of each of the inverted repeat. We speculate that this is the sequence that the HTH domain of the first NpR3597 dimer recognizes, and that it is directing the binding of the second homodimer of NpR3597 on the upstream side of the inverted repeat consecutively.

TetR and AcrR family genes do not always regulate divergently transcribed MDR efflux pump genes located directly upstream. They are sometimes capable of regulating MDR genes elsewhere in the genome (Gristwood *et al.*, 2008; Lin *et al.*, 2005). The microarray of the WT/pSCR119::3597 vs. WT/pSCR119 strain showed a peculiar result. Although over-expression of transcriptional regulator NpR3597 led to a down-regulation of the divergently transcribed MDR gene NpF3598, there was another MDR gene, NpF1932, being down-regulated to a much higher degree (more than 4-fold). The NpR1931-NpF1932 system showed strong similarities in genetic architecture to the NpR3597-NpF3598. According to NCBI, NpR1931 also codes for an AcrR gene with a TetR N-terminal domain and an AcrR domain. The predicted crystal structure of NpR1931 was the same as NpR3597. NpF1932 was found to be divergently transcribed from the AcrR homolog and encoded an MDR gene. Consequently, the upstream promoter region of NpF1932 was subject to analysis and a 25 bp long palindrome was found ATCATACCATAACGATCTATGTATGAT. The inverted-repeat interestingly had the same 7 bp long consensus sequence on the right side of the inverted-repeat with the first inverted-repeat (Fig. 18, BLUE) of NpF3598 that was off by only 1 nucleotide (underlined above). But when the sequence was compared to just the second inverted-repeat (Fig. 18, red) of the NpF3598 upstream region, an 11 bp long sequence motif was found ACAGTNTAGTA (italicized above).

Naturally it was hypothesized that since the upstream palindrome of NpF1932 had strong sequence motifs similar to that of the two NpF3598 upstream palindromes, it would make a great candidate for an EMSA to test binding of NpR3597 to the inverted reapeat. The *in vitro* results proved the binding of NpR3597 native protein, to not only both NpF3598 upstream inverted repeats, but also to the NpF1932 upstream inverted repeat. The micro-array results showing down-regulation of both genes by NpR3597 combined with the binding assays resulted in the regulation flowchart shown below demonstrating transcriptional repression of NpF3598 and NpF1932 (Fig. 30). Since it is the characteristic of AcrR to bind to inverted repeats located in the -35 promoter regions, future studies should include RACE mapping to find the exact location of these inverted repeats relative to their -35 promoter regions.

**Phycobilisome regulation**

*Nostoc punctiforme* and other Cyanobacteria have a regulation pathway that responds to ambient light intensity and quality. This regulon helps them adjust the levels of phycoerythrin and phycocyanin units in their phycobilisomes to maximize efficiency in harvesting ambient light for use in Photosystem I and II. Under red light, phycocyanin accumulation is triggered and under green light phycoerythrin is accumulated (Grossman *et al.*, 2001). We used this information to better understand the physiological changes that were observed in the WT/pSCR119::3597 over-expression strain. Preliminary data (not shown) showed an odd whole cell absorption spectrum from 400 nm – 730 nm. Either the phycocyanin absorption peak was too high or the phycoerythrin absorption peak was too low. Using what we know about complementary chromatic adaptation, we were able to pinpoint what changes were occurring in the composition of the phycobilisome complex in WT/pSCR119::3597 over-expression strain vs. the WT/pSCR119 control strain.

Comparison of the two strains under the grey filter showed that when neither phycocyanin nor phycoerythrin is promoted to accumulate, phycocyanin levels are increased relative to Chlα in the over-expression strain. The ratios of PC/Chlα (620nm/678nm) from three over-expression strain samples averaged to 1.043±0.021 and
three control strain samples averaged to 0.948±0.006. Neither phycocyanin nor chl a change significantly relative to any of the other peaks because they are only significantly changing relatively to each other. This means that as phycocyanin levels go up, Chl a levels are going down. This could possibly be due to the fact that irregularly increased levels of phycocyanin could be over-harvesting light energy, causing the accumulation of electrons to reduce oxygen into superoxide anion radicals and other reactive oxygen species RAS (Latifi et al., 2008). Photosynthetic cells respond to this oxidative stress by reducing the amount of chlorophyll a to slow down transfer of electrons to photosystem I and II (Schagerl and Muller, 2005). This reduction in chlorophyll a could also be attributed to its ability to act as a photosensitizer, creating reactive oxygen species on its own, and thus damaging itself because of this accumulation of free radicals (Rastogi et al., 2010).

The two strains under red light both showed an increase in their phycocyanin peaks, but once again there was an even higher accumulation with the WT/pSCR119::3597 over-expression strain. The exaggerated phycocyanin once again came with a lowered chl a abundance. The ratios of PC/Chl a from three over-expression strain samples averaged to 1.058±0.012 and the three control strain samples averaged to 0.978±0.007. The significance in the difference of PC/Chl a ratios between the overexpressor of NpR3597 and the control rose to a P-value of 0.002. The most significant results came from the third test under green light. The green light filter blocks light in the red spectrum from hitting the cells, which should reduce the amount of phycocyanin and favor accumulation of phycoerythrin. The results show that the wild-type control strain successfully undergoes complimentary chromatic adaptation and reduces its phycocyanin content while increasing in phycoerythrin content. The over-expression strain, on the other hand, proves to be unable to undergo this reduction in phycocyanin content and an over accumulation of phycocyanin is still observed. The ratios of PC/Chl a from three over-expression strain samples grown in green light averaged to 1.019±0.003 and three control strain samples grown in green light were averaged to 0.925±0.006. The difference in ratios of PC/Chl a between the two strains rises to an even greater significance with a P-value of 0.0001. This is probably due to the effect over-expression of NpR3597 is having on the regulatory pathway that controls regulation of phycocyanin production. The presence of NpR3597, via some regulatory pathway, either disables the cells ability to undergo complimentary chromatic adaptation with respect to regulating and reducing phycocyanin content, or over-expression of NpR3597 promotes production of phycocyanin via some effector down the line of regulation. Implication of alternate regulatory pathways is supported by the DNA microarray results (Appendix A) that show down-regulation of NpF0996, homologous to a sigma factor associated with stationary phase growth in Synechocystis sp. Strain PCC 6803 (Asayama, 2004), and 8 genes with homology to members of two-component regulators (P<1.4 X 10^-5). Putative transcriptional regulators are also found in the up-regulated genes and include two sigma factors (NpF4153, NpR5797) and 5 additional genes involved in transcriptional control.

It is interesting to note in the array data the up-regulation of photosystem genes psaB and psbW (NpR0210, NpR3943) and simultaneous down-regulation of many ATPase subunits. This could indicate cell stress and photosystem damage. Transcription of both
subunits of RuBisCO (NpF4195, NpF4197) are reduced in the over-expression strain. From this data alone, one cannot tell if the inferred reduction of carbon fixation is a result of reduced energy and reductant generation caused by photosystem damage, or a cause of photosystem damage due to buildup of electrons in the photosynthetic electron transport chain.

**GFP studies**

The GFP expression studies concluded that the promoter of NpR3597 in WT/pSUN119:p3597 is highly active in a periodic pattern on the filament during log phase growth. The expression is specific to log phase because the GFP signal dissipates as cell cultures reach stationary phase. In stationary phase there is an absence of periodic activity of the NpR3597 promoter or any expression of GFP at all. Upon closer investigation of GFP fluorescing cells in log phase under 100X lens, it is evident that the larger cells that have yet to form a septum for division are the cells that show the most GFP expression. This would indicate that NpR3597 could be involved in cell division, or part of some supplemental physiological change that occurs in parallel with preparing for cell division. Cell division does involve the manipulation of the cell envelope, which MDRs are associated with, so it would be plausible that NpR3597 is expressed during division to regulate these MDRs and the cell envelope associated proteins, down-regulated by NpR3597 over-expression in the micro-array (Appendix A), to control overall composition of the cell envelope. Further studies following live division of WT/pSUN119:p3597 and -3597/pSUN119:p3597 could better fortify NpR3597 involvement in cell division.

It would also be interesting to know how the transcriptional reporters of NpF1932 and NpF3598 would act under these log and stationary conditions. GFP expression tests could also be done using their promoter regions that include and exclude the upstream inverted repeat NpR3597 binding sites to see how the expression of NpF3598 and NpF1932 would be affected. One would hypothesize higher transcription levels in promoters lacking the inverted repeats.

**Akinete viability**

Akinetes are the spore-like cells that form in *Nostoc punctiforme* that are capable of surviving harsh conditions and more importantly for this study, desiccation (Meeks *et al.* 2002). After side by side growth, heterocyst induction, akinete induction, confirmation of akinetes by microscopy and 14 days of desiccation in 37°C incubator, the wild-type and A3597 mutant strains re-grew side by side at the same rate on A&A plates with MAN. The NpR3597 overexpressor on the other hand had almost no growth. The desiccation test showed that the over-expression strain of NpR3597 had very limited growth, which indicates a defect in the cell’s ability to form akinete capable of withstanding such a stress. Since desiccation stress isolates akinetes by killing off the vegetative cells that are not equipped with the envelope protection, this could also mean the over-expression of NpR3597 hinders the strain from forming viable akinetes capable of re-growth into vegetative cells.
Microarray data shows over-expression of NpR3597 is responsible for the down-regulation of 7 cell-envelope specific proteins with a B-value greater than 1.5. 3 of these genes, NpR2279, NpR6096 and NpR5769 (down >2-fold) are putative proteins that are associated with the S-layer, which is the rigid glycoprotein layer of outer cell walls responsible for structure stabilization and environmental stress protection. Some cyanobacteria are known to have coarse S-layers made of p6 (or hexagonal) lattice structures (Smarda, 2002). This down-regulation of cell envelope proteins by over-expression of NpR3597 could be indicative of why the akinetes in the WT/pSCR119::3597 strain are showing defects in their functionality. Further testing using the control plasmid pSUN119 in wild-type strain and Δ3597 mutant strain will better support the claim that it is the over-expression of NpR3597 that is hindering the cells from forming viable akinetes during akinete induction and desiccation, not just the presence of pSCR119. Also, plating the akinete induced wild-type, overexpressor, and mutant strains before desiccation would be a proper test to see if the cultures are even capable of re-growth without being physically challenged by desiccation. And if so, this would indicate that sensitivity of the over-expression strain’s akinetes to desiccation is the factor causing the phenotype.
CONCLUSION

Bioinformatic analysis of the NpR3597, in parallel with the associated genes NpF3598, NpF1932, and NpR1931 genes, led to the hypothesis that NpR3597 is an AcrR family transcriptional repressor. Sequence alignments of NpR3597 with other AcrR genes showed great evidence that NpR3597 is indeed part of the AcrR family. Comparison of genomic architecture of the NpR3597-NpF3598 divergently transcribed system with other AcrR to MDR gene systems supported the mechanism in which NpR3597 works to repress NpR3598.

Through microarray analysis we found that NpR3597 over-expression down-regulates expression of NpF3598, which was in line with our hypothesis that NpR3597, being an AcrR, represses divergently transcribed NpF3598 multi-drug resistance (MDR) efflux pump gene. Interestingly, there was another MDR gene with sequence similarities to NpF3598 MDR, that was being down-regulated to an even greater degree. Microarray results also showed that over-expression of NpR3597 also causes an up-regulation of phycobilisome complex protein, phycocyanin.

Closer inspection of the upstream region of NpF3598 and NpF1932 MDR genes showed evidence of a hyphenated inverted repeat DNA sequence, characteristic of AcrR binding sites. An electric mobility shift assay indicated binding of NpR3597 to two inverted repeat sites upstream of NpF3598 in a putative -35 promoter region, and one inverted repeat site upstream of NpF1932 in the putative -35 region. Sequence alignment of the three inverted repeat sites showed a consensus sequence of CNGTNTAGT which is located on the right side, or downstream side, of each palindrome.

Through spectroscopy, we found that the composition of photosynthetic pigments had an irregular pattern in the NpR3597 over-expression strain compared to wild type. Complementary chromatic adaptation (CCA) assays concluded that composition of phycocyanin protein was in higher abundance in the NpR3597 over-expression strain, which was parallel with the microarray data indicating more transcription of the phycocyanin gene. The CCA assay also showed a decrease in chlorophyll a in the NpR3597 over-expression strain. This could possibly be attributed to oxidative stress caused by over-harvesting of electrons by the increased phycocyanin causing reduction of oxygen to form radical oxygen species that damage chlorophyll a.

Preliminary GFP expression experiments testing the activity of the NpR3597 promoter indicated that the gene is mostly activated during cell division and could possibly play a role in the process. Preliminary akinete viability tests showed that over-expression of NpR3597 renders Nostoc punctiforme incapable of forming functional akinetes with the ability to germinate following desiccation.
REFERENCES


Brune, I., Brinkrolf, K., Kalinowski, J., Pühler, A., Tauch, A.“ The individual and common repertoire of DNA-binding transcriptional regulators of *Corynebacterium glutamicum, Corynebacterium efficiens, Corynebacterium diphtheriae* and *Corynebacterium jeikeium* deduced from the complete genome sequences.” *BMC Genomics* (2005), **6**:86


Gristwood, T., Fineran, P.C., Everson, L. and Salmond, G.P.C. “PigZ, a TetR/AcrR family repressor, modulates secondary metabolism via the expression of a putative four-
component resistance-nodulation-cell-division efflux pump, ZrpADBC, in *Serratia* sp. ATCC 39006.” *Molecular Microbiology* (2008) 69(2), 418–435


Lin, J., Akiba, M., Sahin, O. and Zhang, Q. “CmeR functions as a transcriptional repressor for the multidrug efflux pump CmeABC in *Campylobacter jejuni*.” *Antimicrobial Agents Chemotherapy* (2005) 49:1067-1075


### APPENDIX A

Microarray results of down regulated genes in WT/pSCR119::3597 strain compared to WT/pSCR119 with a P-value less than 1.4E-5

<table>
<thead>
<tr>
<th>sub-function</th>
<th>Name-annotation</th>
<th>ID</th>
<th>logFC</th>
<th>P.Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-HK</td>
<td>ATP-binding region, ATPase-like</td>
<td>NpF3675</td>
<td>-1.374129027</td>
<td>2.07E-12</td>
</tr>
<tr>
<td>A-HK</td>
<td>BaeS, GAF:ATP-binding region, ATPase-like</td>
<td>NpF2781</td>
<td>-0.663436865</td>
<td>1.18E-05</td>
</tr>
<tr>
<td>A-HK RR in N term</td>
<td>Protein-glutamate methylsterase</td>
<td>pNPAR130</td>
<td>-0.942112018</td>
<td>1.57E-09</td>
</tr>
<tr>
<td>A-mot gvpN</td>
<td>gvpN, ATPase</td>
<td>NpF2153</td>
<td>-0.860296029</td>
<td>5.81E-10</td>
</tr>
<tr>
<td>A-mot pil</td>
<td>PilO, possible Tfp pilus assembly protein</td>
<td>NpF5007</td>
<td>-0.625251960</td>
<td>5.38E-08</td>
</tr>
<tr>
<td>A-secondary</td>
<td>possible Oligoketide cyclase/II</td>
<td>NpF2782</td>
<td>-0.850895232</td>
<td>1.98E-06</td>
</tr>
<tr>
<td>A-trans reg</td>
<td>spoIIAA, Sulfate transporter/antisigma-f</td>
<td>NpR1192</td>
<td>-1.914992209</td>
<td>5.88E-12</td>
</tr>
<tr>
<td>A-trans reg sig</td>
<td>sigC, RpoD, Sigma-70 region 3:Sigma-70 region</td>
<td>NpF0996</td>
<td>-1.772030731</td>
<td>3.20E-09</td>
</tr>
<tr>
<td>C-cell env</td>
<td>S-layer protein (SLH domain)</td>
<td>NpF5769</td>
<td>-1.001335627</td>
<td>1.95E-08</td>
</tr>
<tr>
<td>C-cell env</td>
<td>S-layer protein (SLH domain)</td>
<td>NpR2279</td>
<td>0.744146869</td>
<td>3.34E-08</td>
</tr>
<tr>
<td>C-cell env</td>
<td>WcaJ, Undecaprenyl-phosphate galactoseph</td>
<td>NpF4852</td>
<td>1.085645598</td>
<td>1.15E-05</td>
</tr>
<tr>
<td>C-cell env</td>
<td>S-layer protein (SLH domain)</td>
<td>NpR6096</td>
<td>0.688126051</td>
<td>4.30E-05</td>
</tr>
<tr>
<td>C-cell env adhesion</td>
<td>Integrins alpha chain</td>
<td>NpR6586</td>
<td>-0.98163737</td>
<td>6.45E-08</td>
</tr>
<tr>
<td>C-cell env LPS</td>
<td>galE, UDP-glucose 4-epimerase</td>
<td>NpF4853</td>
<td>1.069619523</td>
<td>6.31E-08</td>
</tr>
<tr>
<td>C-cell env LPS</td>
<td>GumC, Protein-tyrosine kinase.</td>
<td>NpR4851</td>
<td>-0.587484225</td>
<td>1.73E-05</td>
</tr>
<tr>
<td>C-cofactor</td>
<td>CoaD, Coenzyme A biosynthesis protein:Cy</td>
<td>NpR3950</td>
<td>-0.606194059</td>
<td>2.46E-05</td>
</tr>
<tr>
<td>C-energy atp</td>
<td>atpa, H+-transporting two-sector ATPase,</td>
<td>NpF4858</td>
<td>-1.248282211</td>
<td>8.59E-10</td>
</tr>
<tr>
<td>C-energy atp</td>
<td>atpC, ATP synthase F0, C subunit</td>
<td>NpF4859</td>
<td>-0.732276177</td>
<td>4.31E-08</td>
</tr>
<tr>
<td>C-energy op</td>
<td>Ndh, NADH dehydrogenase.</td>
<td>NpR3616</td>
<td>-0.640901644</td>
<td>6.23E-05</td>
</tr>
<tr>
<td>C-mono aa</td>
<td>pgdh, D-3-phosphoglycerate dehydrogenase</td>
<td>NpR5218</td>
<td>-0.640205272</td>
<td>4.31E-07</td>
</tr>
<tr>
<td>C-polymer (phospho)lipid</td>
<td>TesA, Lipolytic enzyme, G-D-S-L</td>
<td>NpF3822</td>
<td>-0.601782876</td>
<td>0.0001364</td>
</tr>
<tr>
<td>C-polymer DNA/RNA</td>
<td>McrA, HNH endonuclease</td>
<td>NpR2778</td>
<td>-0.947359243</td>
<td>7.04E-08</td>
</tr>
<tr>
<td>C-polymer prot prcs</td>
<td>RtcB, Protein splicing (intein) site:Pro</td>
<td>NpR6421</td>
<td>-1.060492671</td>
<td>1.14E-09</td>
</tr>
<tr>
<td>C-polymer prot syn</td>
<td>pth, Peptidyl-tRNA hydrolase</td>
<td>NpF4317</td>
<td>-0.709975704</td>
<td>2.13E-07</td>
</tr>
<tr>
<td>C-polymer prot turn</td>
<td>AprE, Thermitase.</td>
<td>pNPAPF039</td>
<td>-0.98100303</td>
<td>3.39E-10</td>
</tr>
<tr>
<td>C-polymer prot turn</td>
<td>LasB, Bacililysin.</td>
<td>NpF5602</td>
<td>-1.381697411</td>
<td>4.28E-10</td>
</tr>
<tr>
<td>C-polymer prot turn</td>
<td>Dcp, Oligopeptidase A.</td>
<td>NpF2866</td>
<td>-0.967584394</td>
<td>1.60E-05</td>
</tr>
<tr>
<td>C-polymer</td>
<td>TROVE</td>
<td>NpF0915</td>
<td>-1.70E-10</td>
<td></td>
</tr>
<tr>
<td>RNA</td>
<td>Description</td>
<td>Accession</td>
<td>Score</td>
<td>E-value</td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>C-porphyrin</td>
<td>cpeZ, PBS lyase HEAT-like repeat</td>
<td>NpR3805</td>
<td>1.613833702</td>
<td>1.08E-06</td>
</tr>
<tr>
<td>C-precursor</td>
<td>RbcS, Ribulose-bisphosphate carboxylase.</td>
<td>NpF4197</td>
<td>0.832498804</td>
<td>6.44E-07</td>
</tr>
<tr>
<td>C-precursor</td>
<td>RbcL, Ribulose-bisphosphate carboxylase.</td>
<td>NpF4195</td>
<td>0.717984206</td>
<td>3.31E-06</td>
</tr>
<tr>
<td>C-ROS</td>
<td>RimK, Glutathione synthetase, prokaryotic</td>
<td>NpF4803</td>
<td>0.580699958</td>
<td>8.30E-09</td>
</tr>
<tr>
<td>S-phage</td>
<td>NpR1095, possible Phage baseplate assembly</td>
<td>NpR1095</td>
<td>0.610136482</td>
<td>6.55E-06</td>
</tr>
<tr>
<td>S-phage tail</td>
<td>Phage tail sheath protein</td>
<td>NpF1112</td>
<td>0.895554759</td>
<td>2.55E-05</td>
</tr>
<tr>
<td>T-export SecI</td>
<td>EmrA, possible Multidrug resistance effl</td>
<td>NpF1932</td>
<td>2.044168341</td>
<td>2.74E-12</td>
</tr>
<tr>
<td>T-export SecI</td>
<td>EmrA, possible Multidrug resistance effl</td>
<td>NpF3598</td>
<td>0.943197274</td>
<td>3.54E-06</td>
</tr>
<tr>
<td>T-export SecII</td>
<td>PulE, Bacterial type II secretion system</td>
<td>NpR0118</td>
<td>1.255975817</td>
<td>1.03E-08</td>
</tr>
<tr>
<td>T-export SecII</td>
<td>HofQ, Bacterial type II and III secretion</td>
<td>NpF5008</td>
<td>0.583369569</td>
<td>1.29E-05</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>Smc, possible Chromosome segregation ATP</td>
<td>NpR1321</td>
<td>1.365333848</td>
<td>2.16E-12</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpR1423</td>
<td>1.480943558</td>
<td>9.75E-11</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>possible membrane protein</td>
<td>NpF6444</td>
<td>0.798312769</td>
<td>2.31E-09</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>possible Vesicle coat complex C</td>
<td>NpR5441</td>
<td>1.123425807</td>
<td>5.73E-08</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>putative similar to unknown protein</td>
<td>NpF1416</td>
<td>0.926656724</td>
<td>1.34E-07</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpF5573</td>
<td>-0.916859393</td>
<td>1.76E-07</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>cyano, conserved hypothetical protein</td>
<td>NpR2479</td>
<td>-0.635577389</td>
<td>1.89E-07</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>cyano hets, conserved hypothetical protein, Npun and 7120 (in chromosome)</td>
<td>pNPAF082</td>
<td>0.736409436</td>
<td>2.14E-07</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>cyano, conserved hypothetical protein, no interpro, cyano</td>
<td>NpR4122</td>
<td>-0.595489307</td>
<td>2.98E-07</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpF4077</td>
<td>-0.78115471</td>
<td>3.99E-07</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>putative putative secreted pro</td>
<td>NpF1405</td>
<td>-0.768370927</td>
<td>8.54E-07</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpR3283</td>
<td>1.052224772</td>
<td>1.18E-06</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>con hyp - no cog, all4835, P and T</td>
<td>NpF1418</td>
<td>0.580264047</td>
<td>1.52E-06</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical Npun and a few euks</td>
<td>NpF1407</td>
<td>0.809742741</td>
<td>1.98E-06</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpR1390</td>
<td>0.656535112</td>
<td>7.47E-05</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpR3926</td>
<td>0.773469397</td>
<td>9.62E-05</td>
</tr>
<tr>
<td>U-hyp</td>
<td>hypothetical</td>
<td>NpF1207</td>
<td>1.412196309</td>
<td>4.75E-10</td>
</tr>
<tr>
<td>U-hyp</td>
<td>hypothetical</td>
<td>NpF5045</td>
<td>1.339333315</td>
<td>4.69E-10</td>
</tr>
<tr>
<td>U-hyp</td>
<td>hypothetical</td>
<td>pNPAF077</td>
<td>2.27013721</td>
<td>6.32E-09</td>
</tr>
<tr>
<td>U-hyp</td>
<td>hypothetical</td>
<td>NpF1148</td>
<td>1.214911163</td>
<td>2.57E-08</td>
</tr>
<tr>
<td>U-hyp</td>
<td>hypothetical</td>
<td>pNPAF088</td>
<td>1.242218669</td>
<td>6.64E-06</td>
</tr>
<tr>
<td>U-hyp</td>
<td>hypothetical</td>
<td>NpF5190</td>
<td>0.677310671</td>
<td>1.47E-05</td>
</tr>
<tr>
<td>U-unassign</td>
<td>conserved hypothetical protein with ferredoxin signature</td>
<td>NpR3657</td>
<td>-1.331319771</td>
<td>4.29E-09</td>
</tr>
<tr>
<td>U-unassign</td>
<td>RsbW, ATP-binding region, ATPase-like</td>
<td>NpR0876</td>
<td>-0.591749368</td>
<td>3.98E-08</td>
</tr>
<tr>
<td>U-unassign</td>
<td>LcnDR2, possible Lantibiotic modifying enzyme, G-protein coupled receptors family 1 signature, hets only cyano</td>
<td>pNPAF076</td>
<td>-0.614833764</td>
<td>1.69E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>YSH1, Beta-lactamase-like</td>
<td>NpR1387</td>
<td>-0.679583003</td>
<td>7.96E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>von Willebrand factor, type A</td>
<td>NpF3694</td>
<td>-0.631734308</td>
<td>1.10E-06</td>
</tr>
<tr>
<td>U-unassign</td>
<td>putative glr4103 protein, phage tail?</td>
<td>NpF1114</td>
<td>-0.629419163</td>
<td>4.97E-05</td>
</tr>
</tbody>
</table>
### APPENDIX B

**Microarray results of up-regulated genes with a P-value less than 1.4E-5**

<table>
<thead>
<tr>
<th>sub-function</th>
<th>Name-annotation</th>
<th>ID</th>
<th>logFC</th>
<th>P.Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-alt C</td>
<td>XynA, Endo-1,4-beta-xylanase.</td>
<td>NpF1486</td>
<td>0.677674013</td>
<td>1.02E-08</td>
</tr>
<tr>
<td>A-dev aki</td>
<td>avaK PRC-barrel</td>
<td>NpF5452</td>
<td>2.159159893</td>
<td>5.49E-09</td>
</tr>
<tr>
<td>A-RR HTH</td>
<td>CitB, Bacterial regulatory protein, LuxR</td>
<td>NpF1453</td>
<td>1.165093702</td>
<td>4.04E-06</td>
</tr>
<tr>
<td>A-secondary</td>
<td>DNA repair</td>
<td></td>
<td>1.556784168</td>
<td>1.00E-05</td>
</tr>
<tr>
<td>A-stress polymer prot turn</td>
<td>LOC246307, Asparaginase, T</td>
<td>NpF0225</td>
<td>0.789555115</td>
<td>5.40E-05</td>
</tr>
<tr>
<td>A-trans reg</td>
<td>AR08, Aspartate transaminase in C-term and HTH in N-term.</td>
<td>NpR6520</td>
<td>0.802739771</td>
<td>1.44E-11</td>
</tr>
<tr>
<td>A-trans reg sig</td>
<td>RpOE, Sigma-70 region 2:Sigma-70 region</td>
<td>NpF4153</td>
<td>0.670943002</td>
<td>4.40E-05</td>
</tr>
<tr>
<td>C-cell env</td>
<td>NpR1877, Putative peptidoglycan binding</td>
<td>NpR1877</td>
<td>0.658908982</td>
<td>9.75E-06</td>
</tr>
<tr>
<td>C-cell env</td>
<td>murC, UDP-N-acetylmuramate--alanine ligas</td>
<td>NpF0466</td>
<td>0.888261963</td>
<td>5.34E-08</td>
</tr>
<tr>
<td>C-cell env</td>
<td>muR, UDP-N-acetylglucosamine 1-carboxyv</td>
<td>NpR5719</td>
<td>1.262405128</td>
<td>2.15E-11</td>
</tr>
<tr>
<td>C-cell env</td>
<td>manC GDP-mannose pyrophosphorylase, P</td>
<td>NpR5110</td>
<td>1.496173224</td>
<td>2.85E-12</td>
</tr>
<tr>
<td>C-cell env</td>
<td>Gif, UDP-galactopyranose mutase, T</td>
<td>NpF0220</td>
<td>1.756820755</td>
<td>3.03E-06</td>
</tr>
<tr>
<td>C-cell env</td>
<td>possible membrane protein</td>
<td>NpR2377</td>
<td>1.853062464</td>
<td>3.54E-08</td>
</tr>
<tr>
<td>C-cell env adhesion</td>
<td>conserved hypothetical protein</td>
<td>NpF5115</td>
<td>0.856532008</td>
<td>6.73E-05</td>
</tr>
<tr>
<td>C-cell env adhesion</td>
<td>Beta-Ig-H3/fasciclin:S-layer pr</td>
<td>NpR4210</td>
<td>0.862096983</td>
<td>2.87E-06</td>
</tr>
<tr>
<td>C-cofactor</td>
<td>NpR1005, Protozoan/cyanobacterial globin</td>
<td>NpR1005</td>
<td>0.680760039</td>
<td>4.30E-06</td>
</tr>
<tr>
<td>C-cofactor</td>
<td>SucC, Succinate--CoA ligase (ADP-forming)</td>
<td>NpF5183</td>
<td>0.768212439</td>
<td>3.28E-07</td>
</tr>
<tr>
<td>C-cofactor</td>
<td>possible Phytoene dehydrogenase</td>
<td>NpF3745</td>
<td>1.619374863</td>
<td>2.38E-08</td>
</tr>
<tr>
<td>C-energy ets</td>
<td>HcaE, Phthalate 4,5-dioxygenase, Rieske [2Fe-2S] region, 1 tm</td>
<td>NpR1926</td>
<td>0.617178667</td>
<td>7.35E-08</td>
</tr>
<tr>
<td>C-energy ps psbW</td>
<td>Photosystem II reaction centre W p</td>
<td>NpR3943</td>
<td>0.659212676</td>
<td>0.00011044</td>
</tr>
<tr>
<td>C-energy ps psaB</td>
<td>Protein of unknown function</td>
<td>NpR5455</td>
<td>0.688091508</td>
<td>2.86E-06</td>
</tr>
<tr>
<td>C-energy ps psaB</td>
<td>Photosystem I psaB, 11 tm</td>
<td>NpR0210</td>
<td>1.850928522</td>
<td>7.19E-12</td>
</tr>
<tr>
<td>C-polymer DNA dnaN</td>
<td>DNA polymerase III, beta chain</td>
<td>NpR0034</td>
<td>0.630055932</td>
<td>1.23E-05</td>
</tr>
<tr>
<td>C-polymer DNA ssb</td>
<td>Single-strand binding protein</td>
<td>NpF1842</td>
<td>1.0988768</td>
<td>1.34E-08</td>
</tr>
<tr>
<td>C-polymer DNA repair</td>
<td>pcrA, ATP-dependent DNA helicase PcrA</td>
<td>NpR1007</td>
<td>0.842340884</td>
<td>6.33E-10</td>
</tr>
<tr>
<td>C-polymer prot turn</td>
<td>HtpX, Peptidase M48, Ste24p</td>
<td>NpR5718</td>
<td>0.682013806</td>
<td>9.71E-05</td>
</tr>
<tr>
<td>C-polymer prot turn</td>
<td>DegQ, Peptidase S1, chymotrypsin:PDZ/DHR</td>
<td>NpR2556</td>
<td>0.735607434</td>
<td>2.00E-07</td>
</tr>
<tr>
<td>C-polymer prot turn</td>
<td>DegQ, PDZ/DHR/GLGF domain</td>
<td>NpR3310</td>
<td>1.679784777</td>
<td>3.86E-08</td>
</tr>
<tr>
<td>C-polymer prot turn</td>
<td>ThiJ, Peptidase C56, Pfp1</td>
<td>NpR6239</td>
<td>1.925185254</td>
<td>7.90E-07</td>
</tr>
<tr>
<td>C-porphyrin cpeE</td>
<td>Phycobilisome linker polypeptide:C</td>
<td>NpF0736</td>
<td>0.751506499</td>
<td>3.05E-05</td>
</tr>
<tr>
<td>C-precursor CemA</td>
<td></td>
<td>NpF4249</td>
<td>0.865741727</td>
<td>4.05E-08</td>
</tr>
<tr>
<td>C-ROS</td>
<td>Dyp-type peroxidase</td>
<td>NpR5915</td>
<td>1.036869584</td>
<td>5.66E-07</td>
</tr>
<tr>
<td>C-ROS</td>
<td>Dps, Ferritin and Dps</td>
<td>NpR3258</td>
<td>1.082460788</td>
<td>9.75E-07</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------------</td>
<td>---------</td>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>C-ROS</td>
<td>Gst, Glutathione transferase, unusual C-term.</td>
<td>NpF2351</td>
<td>1.096255964</td>
<td>1.01E-10</td>
</tr>
<tr>
<td>C-storage C</td>
<td>susA, Sucrose synthase.</td>
<td>NpF1876</td>
<td>0.714848338</td>
<td>2.77E-06</td>
</tr>
<tr>
<td>C-storage C</td>
<td>susB, Sucrose synthase.</td>
<td>NpF4951</td>
<td>0.791942715</td>
<td>6.11E-06</td>
</tr>
<tr>
<td>C-storage C</td>
<td>GlgB, 4-alpha-D-{(1-&gt;4)-alpha-D-glucano}</td>
<td>NpF5705</td>
<td>0.908642322</td>
<td>0.000134377</td>
</tr>
<tr>
<td>C-storage C</td>
<td>TreA, Alpha, alpha-trehalase.</td>
<td>NpF5707</td>
<td>0.943090093</td>
<td>7.10E-08</td>
</tr>
<tr>
<td>C-storage C</td>
<td>GlgB, 1,4-alpha-glucan branching enzyme.</td>
<td>NpK1027</td>
<td>1.892429463</td>
<td>1.27E-08</td>
</tr>
<tr>
<td>C-storage C</td>
<td>CphB, Peptidase S51, dipeptidase E</td>
<td>NpF1821</td>
<td>1.243718184</td>
<td>3.81E-06</td>
</tr>
<tr>
<td>T-ABC ATP bind</td>
<td>MglA, Molybdate-transporting ATPase.</td>
<td>NpR5325</td>
<td>0.690796872</td>
<td>7.48E-05</td>
</tr>
<tr>
<td>T-ABC pbp</td>
<td>substrate-binding protein ribose- hrm locus sp P and T</td>
<td>NpR5326</td>
<td>0.694157762</td>
<td>2.25E-05</td>
</tr>
<tr>
<td>T-cat ATPase</td>
<td>ZntA, Heavy metal-(Cd/Co/Hg/Pb/Zn)-transporter</td>
<td>NpR0548</td>
<td>1.318986191</td>
<td>3.56E-11</td>
</tr>
<tr>
<td>T-channel</td>
<td>TolQ, MotA/TolQ/ExbB proton channel</td>
<td>NpR0782</td>
<td>0.678523487</td>
<td>4.21E-06</td>
</tr>
<tr>
<td>T-export</td>
<td>ExbD, Bioplymer transport protein ExbD/TolR, 1 tm</td>
<td>NpR0781</td>
<td>0.613648713</td>
<td>8.49E-05</td>
</tr>
<tr>
<td>T-export</td>
<td>CadD, Cadmium resistance transporter, hets and gram positives</td>
<td>NpF5692</td>
<td>0.707782803</td>
<td>3.57E-05</td>
</tr>
<tr>
<td>T-export SecI</td>
<td>EmRA, possible Multidrug resistance efflux, interpro says Gram-negative bacterial RTX secretion protein D, 1 tm</td>
<td>pNPBF053</td>
<td>1.387912975</td>
<td>1.82E-06</td>
</tr>
<tr>
<td>T-porter</td>
<td>NhaP, possible NhaP-type Na+/H+ and K+/H+ transporter</td>
<td>NpF2956</td>
<td>0.755373242</td>
<td>5.08E-05</td>
</tr>
<tr>
<td>T-reg</td>
<td>ChaB, ChaB</td>
<td>NpR4442</td>
<td>0.745556198</td>
<td>2.19E-05</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpR1259</td>
<td>0.612761802</td>
<td>8.11E-07</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>putative patB and ORF2 genes all2511</td>
<td>NpF1872</td>
<td>0.613040566</td>
<td>3.21E-06</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>Protein of unknown function DUF</td>
<td>NpR4444</td>
<td>0.619242479</td>
<td>9.17E-05</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpF6242</td>
<td>0.62306268</td>
<td>2.20E-05</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpF5386</td>
<td>0.629728489</td>
<td>7.19E-05</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpR6236</td>
<td>0.65845321</td>
<td>1.17E-06</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>possible membrane protein, sp 4 tm</td>
<td>NpF3074</td>
<td>0.668852716</td>
<td>5.34E-08</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>pNPAR047</td>
<td>0.675025847</td>
<td>6.85E-06</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>Conserved hypothetical protein, related to HTH</td>
<td>NpF5021</td>
<td>0.679366414</td>
<td>4.84E-07</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpR4080</td>
<td>0.681974561</td>
<td>1.07E-06</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>putative gll1168 protein</td>
<td>NpF2218</td>
<td>0.688390576</td>
<td>5.60E-11</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>cyano hets, conserved hypothetical protein, Npun and 7120 plasmid</td>
<td>pNPDR020</td>
<td>0.695544223</td>
<td>7.80E-06</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>pNPAR116</td>
<td>0.695655459</td>
<td>1.43E-06</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>cyano, Uma2, Protein of unknown function DUF820, cyano</td>
<td>NpR3724</td>
<td>0.718166788</td>
<td>4.37E-05</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>pNPAR046</td>
<td>0.71959715</td>
<td>6.07E-08</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>cyano hets, conserved hypothetical protein, hets 7120 plasmid</td>
<td>pNPBR220</td>
<td>0.724831335</td>
<td>3.33E-07</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpF5066</td>
<td>0.748289174</td>
<td>1.41E-05</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>cyano hets, conserved hypothetical protein, 7120 plasmid</td>
<td>pNPDR023</td>
<td>0.761558853</td>
<td>3.32E-06</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>cyano hets, conserved hypothetical protein, hets</td>
<td>NpF5100</td>
<td>0.785614201</td>
<td>8.90E-10</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpR1956</td>
<td>0.788720636</td>
<td>5.04E-09</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>Uma2, Protein of unknown function DUF820, mostly cyano, not all</td>
<td>NpR3723</td>
<td>0.797991801</td>
<td>1.98E-10</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein, T</td>
<td>NpR2955</td>
<td>0.802655813</td>
<td>7.93E-08</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>pNPAR045</td>
<td>0.820656594</td>
<td>1.05E-07</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpF0470</td>
<td>0.857881275</td>
<td>6.81E-05</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>cyano hets, hypothetical, Npun and Avar</td>
<td>pNPBR057</td>
<td>0.904129296</td>
<td>2.11E-05</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpF6115</td>
<td>0.921928938</td>
<td>9.18E-07</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpF1617</td>
<td>1.02875991</td>
<td>2.85E-06</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>con hyp DUF427, alr4241, P</td>
<td>NpR3622</td>
<td>1.050674483</td>
<td>1.69E-07</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>Protein of unknown function DUF</td>
<td>NpR3528</td>
<td>1.058854559</td>
<td>7.65E-08</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpF1907</td>
<td>1.063063393</td>
<td>5.79E-05</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>con hyp - no cog, asr3390, P and T</td>
<td>NpR0652</td>
<td>1.13204318</td>
<td>1.34E-05</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpR4224</td>
<td>1.132972731</td>
<td>1.12E-06</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpF0169</td>
<td>1.146230102</td>
<td>1.50E-06</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpF5693</td>
<td>1.180602981</td>
<td>2.94E-09</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpR1889</td>
<td>1.199081099</td>
<td>2.37E-06</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpR6615</td>
<td>1.23749344</td>
<td>1.05E-05</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>hypothetical low temperature-in</td>
<td>NpR3259</td>
<td>1.252204766</td>
<td>3.61E-06</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>ChaB, possible Putative cation transport</td>
<td>NpR3257</td>
<td>1.260266032</td>
<td>3.21E-06</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>cyano hets, conserved hypothetical protein, hets, sp</td>
<td>NpF5465</td>
<td>1.334730295</td>
<td>7.73E-08</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpR6614</td>
<td>1.358270645</td>
<td>7.03E-08</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpF4819</td>
<td>1.418983409</td>
<td>3.58E-07</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpF5131</td>
<td>1.504909984</td>
<td>0.00012564</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>DedA, possible Uncharacterized membrane-</td>
<td>NpF3786</td>
<td>1.537063205</td>
<td>8.22E-08</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpF3788</td>
<td>1.541757439</td>
<td>2.88E-06</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>possible Uncharacterized conserved protein, T</td>
<td>NpR6240</td>
<td>1.565679424</td>
<td>2.48E-05</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpR6532</td>
<td>1.756745993</td>
<td>5.97E-08</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>conserved hypothetical protein</td>
<td>NpF3789</td>
<td>2.0843264</td>
<td>1.03E-07</td>
</tr>
<tr>
<td>U-con hyp</td>
<td>possible Uncharacterized conserved</td>
<td>NpR3624</td>
<td>2.870894325</td>
<td>8.57E-14</td>
</tr>
<tr>
<td>U-hyp</td>
<td>hypothetical</td>
<td>pNPAR050</td>
<td>0.664010373</td>
<td>0.00013789</td>
</tr>
<tr>
<td>U-hyp</td>
<td>hypothetical</td>
<td>NpF4653</td>
<td>0.72086956</td>
<td>4.61E-05</td>
</tr>
<tr>
<td>U-hyp</td>
<td>hypothetical</td>
<td>NpF3933</td>
<td>0.895491412</td>
<td>5.93E-06</td>
</tr>
<tr>
<td>U-hyp</td>
<td>hypothetical, 2 tm</td>
<td>pNPAR057</td>
<td>1.039177346</td>
<td>3.14E-09</td>
</tr>
<tr>
<td>U-hyp</td>
<td>hypothetical</td>
<td>NpF6451</td>
<td>1.071464595</td>
<td>6.46E-08</td>
</tr>
<tr>
<td>U-hyp</td>
<td>hypothetical</td>
<td>NpR0253</td>
<td>1.369915147</td>
<td>1.68E-08</td>
</tr>
<tr>
<td>U-hyp</td>
<td>hypothetical</td>
<td>NpF3932</td>
<td>1.381270218</td>
<td>5.19E-07</td>
</tr>
<tr>
<td>U-hyp</td>
<td>hypothetical</td>
<td>NpR2957</td>
<td>1.404528026</td>
<td>6.77E-06</td>
</tr>
<tr>
<td>U-hyp</td>
<td>hypothetical</td>
<td>pNFE7023</td>
<td>2.990251804</td>
<td>2.55E-10</td>
</tr>
<tr>
<td>U-unassign</td>
<td>HAD-superfamily hydrolase, subfamily IA, variant 1 &amp; 3</td>
<td>NpR2652</td>
<td>0.62935512</td>
<td>1.02E-06</td>
</tr>
<tr>
<td>U-unassign</td>
<td>IMP dehydrogenase.</td>
<td>NpR5794</td>
<td>0.637541491</td>
<td>1.58E-06</td>
</tr>
<tr>
<td>U-unassign</td>
<td>Tas, Aryl-alcohol dehydrogenase (NADP+), T.</td>
<td>NpF5918</td>
<td>0.666257536</td>
<td>6.76E-05</td>
</tr>
<tr>
<td>U-unassign</td>
<td>Pentapeptide repeat, sp 1 tm</td>
<td>NpR4054</td>
<td>0.677258446</td>
<td>0.000139943</td>
</tr>
<tr>
<td>U-unassign</td>
<td>Predicted transcriptional regulator, DNA binding HTH domain, only blast is npun P and T</td>
<td>NpF4701</td>
<td>0.779772787</td>
<td>0.000119916</td>
</tr>
<tr>
<td>U-unassign</td>
<td>Protein of unknown function, HH</td>
<td>NpR0971</td>
<td>0.86945529</td>
<td>2.93E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>CypX, Unspecific monoxygenase.</td>
<td>NpR1394</td>
<td>0.882634583</td>
<td>4.32E-06</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGR5S5, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>GlpG, Peptidase S54, rhomboid-like prote</td>
<td>NpR0461</td>
<td>0.917173973</td>
<td>1.39E-05</td>
</tr>
<tr>
<td>U-unassign</td>
<td>XseA, possible Exonuclease VII large sub, Dicto.1 tm</td>
<td>NpF3955</td>
<td>0.942045153</td>
<td>3.30E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
<tr>
<td>U-unassign</td>
<td>PE_PGRS55, Hemolysin-type calcium-bindin</td>
<td>NpR0879</td>
<td>0.908119613</td>
<td>8.98E-07</td>
</tr>
</tbody>
</table>
## APPENDIX C

### Amino Acid Multiple Sequence Alignment of NpR3597 in *N. punctiforme* with 19 of its closest homologs

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Anabaena variabilis</strong></td>
<td>~MAR</td>
<td>~MAR</td>
<td>~MAR</td>
<td>~MAR</td>
</tr>
<tr>
<td><strong>Nodularia spumigena</strong></td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
</tr>
<tr>
<td><strong>Nostoc punctiforme</strong></td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
</tr>
<tr>
<td><strong>Cyanothecae sp.</strong></td>
<td>~V</td>
<td>~V</td>
<td>~V</td>
<td>~V</td>
</tr>
<tr>
<td><strong>Lyngbya sp.</strong></td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
</tr>
<tr>
<td><strong>Microcoleus chthonoplastes</strong></td>
<td>~V</td>
<td>~V</td>
<td>~V</td>
<td>~V</td>
</tr>
<tr>
<td><strong>Arthrosira platensis_str._P</strong></td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
</tr>
<tr>
<td><strong>Stenotrophomonas sp._SKA14</strong></td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
</tr>
<tr>
<td><strong>Pseudomonas aeruginosa</strong></td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
</tr>
<tr>
<td><strong>Conexibacter woesei</strong></td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
</tr>
<tr>
<td><strong>Desulfovibrivulgaris_subs</strong></td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
</tr>
<tr>
<td><strong>Methylobacterium radiotoler</strong></td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
</tr>
<tr>
<td><strong>Methylococcus capsulatus_st</strong></td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
</tr>
<tr>
<td><strong>Citreicella sp.</strong></td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
</tr>
<tr>
<td><strong>Loktanella vestfoldensis</strong></td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
</tr>
<tr>
<td><strong>Granulibacter bethedensis</strong></td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
</tr>
<tr>
<td><strong>Agrobacterium vitis</strong></td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
</tr>
<tr>
<td><strong>Novosphingobium aromaticivo</strong></td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
<td>~M</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th></th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Anabaena variabilis</strong></td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
</tr>
<tr>
<td><strong>Nodularia spumigena</strong></td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
</tr>
<tr>
<td><strong>Nostoc punctiforme</strong></td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
</tr>
<tr>
<td><strong>Cyanothecae sp.</strong></td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
</tr>
<tr>
<td><strong>Lyngbya sp.</strong></td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
</tr>
<tr>
<td><strong>Microcoleus chthonoplastes</strong></td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
</tr>
<tr>
<td><strong>Arthrosira platensis_str._P</strong></td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
</tr>
<tr>
<td><strong>Stenotrophomonas sp._SKA14</strong></td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
</tr>
<tr>
<td><strong>Pseudomonas aeruginosa</strong></td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
</tr>
<tr>
<td><strong>Conexibacter woesei</strong></td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
</tr>
<tr>
<td><strong>Desulfovibrivulgaris_subs</strong></td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
</tr>
<tr>
<td><strong>Methylobacterium radiotoler</strong></td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
</tr>
<tr>
<td><strong>Methylococcus capsulatus_st</strong></td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
</tr>
<tr>
<td><strong>Citreicella sp.</strong></td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
</tr>
<tr>
<td><strong>Loktanella vestfoldensis</strong></td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
</tr>
<tr>
<td><strong>Granulibacter bethedensis</strong></td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
</tr>
<tr>
<td><strong>Agrobacterium vitis</strong></td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
</tr>
<tr>
<td><strong>Novosphingobium aromaticivo</strong></td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
<td>YAA</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th></th>
<th>90</th>
<th>100</th>
<th>110</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Anabaena variabilis</strong></td>
<td>KRF</td>
<td>KRF</td>
<td>KRF</td>
<td>KRF</td>
</tr>
<tr>
<td><strong>Nodularia spumigena</strong></td>
<td>KQF</td>
<td>KQF</td>
<td>KQF</td>
<td>KQF</td>
</tr>
<tr>
<td><strong>Nostoc punctiforme</strong></td>
<td>KEK</td>
<td>KEK</td>
<td>KEK</td>
<td>KEK</td>
</tr>
<tr>
<td><strong>Cyanothecae sp.</strong></td>
<td>KEK</td>
<td>KEK</td>
<td>KEK</td>
<td>KEK</td>
</tr>
<tr>
<td><strong>Lyngbya sp.</strong></td>
<td>KEK</td>
<td>KEK</td>
<td>KEK</td>
<td>KEK</td>
</tr>
<tr>
<td><strong>Microcoleus chthonoplastes</strong></td>
<td>NNS</td>
<td>NNS</td>
<td>NNS</td>
<td>NNS</td>
</tr>
<tr>
<td><strong>Synechococcus sp.</strong></td>
<td>EKR</td>
<td>EKR</td>
<td>EKR</td>
<td>EKR</td>
</tr>
</tbody>
</table>

---

Note: The alignment includes sequence information for each homolog, showing amino acid residues in the corresponding positions across the sequences.
Acaryochloris marina
Arthrospira platensis sub. P
Stenotrophomonas sp. SKA14
Pseudomonas aeruginosa
Conexibacter woesei
Desulfovibrio vulgaris subs
Methylobacterium radiotoler
Methylococcus capsulatus st
Citreicella sp.
Loktanella vestfoldensis
Granulibacter bethedens
Agrobacterium vitis
Novosphingobium aromaticivos

Anabaena variabilis
Nodularia spumigena
Nostoc punctiforme
Cyanotheca sp.
Lyngbya sp.
Microcoleus chthonoplastes
Synechococcus sp.
Acaryochloris marina
Arthrospira platensis sub. P
Stenotrophomonas sp. SKA14
Pseudomonas aeruginosa
Conexibacter woesei
Desulfovibrio vulgaris subs
Methylobacterium radiotoler
Methylococcus capsulatus st
Citreicella sp.
Loktanella vestfoldensis
Granulibacter bethedens
Agrobacterium vitis
Novosphingobium aromaticivos

Anabaena variabilis
Nodularia spumigena
Nostoc punctiforme
Cyanotheca sp.
Lyngbya sp.
Microcoleus chthonoplastes
Synechococcus sp.
Acaryochloris marina
Arthrospira platensis sub. P
Stenotrophomonas sp. SKA14
Pseudomonas aeruginosa
Conexibacter woesei
Desulfovibrio vulgaris subs
Methylobacterium radiotoler
Methylococcus capsulatus st
Citreicella sp.
Loktanella vestfoldensis
Granulibacter bethedens
Agrobacterium vitis
Novosphingobium aromaticivos
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anabaena_variabilis</td>
<td>PM~ESDRLI DALTHLIINC AE*-------</td>
</tr>
<tr>
<td>Nodularia_spumigena</td>
<td>PM~ESDRLI DALTHLIVNC AD*--------</td>
</tr>
<tr>
<td>Nostoc_punctiforme</td>
<td>PM~ESDRLI DALTHLITKC AD*--------</td>
</tr>
<tr>
<td>Cyanothecae_sp.</td>
<td>PM<del>SEEELV NSLIELLLK</del> -E*--------</td>
</tr>
<tr>
<td>Lyngbya_sp.</td>
<td>PM~DSDRLI EGMSCLTR* ~~~~~~~~~</td>
</tr>
<tr>
<td>Microcoleus_chthonoplastes</td>
<td>PL~ERDRMV NGLIQLLTAG EAEGAV*~~~</td>
</tr>
<tr>
<td>Synechococcus_sp.</td>
<td>PL<del>ESDRLI NSLIELLLK</del> -E*--------</td>
</tr>
<tr>
<td>Acaryochloris_marina</td>
<td>PF~ARERMI QGLVDLIVND SA*---------</td>
</tr>
<tr>
<td>Arthrosira_platensis_str._P</td>
<td>PM~ERDRVI KTLVDAIAPP AS*---------</td>
</tr>
<tr>
<td>Stenotrophomonas_sp._SKA14</td>
<td>SA~EERARVA ESAVAMFLGY YAI*-------</td>
</tr>
<tr>
<td>Pseudomonas_aeruginosa</td>
<td>DE~AEGERHV EEVVAFLRA FAVRD*------</td>
</tr>
<tr>
<td>Conexibacter_woesei</td>
<td>PDPEAPARLA AAAVEVFLAA YGTRAE*----</td>
</tr>
<tr>
<td>Desulfovibrio_vulgaris_subs</td>
<td>TE~EETERAV ESAVESVFLAA YAI*-------</td>
</tr>
<tr>
<td>Methylobacterium_radiotoler</td>
<td>GE~AEMRHQV DEAVRFLAG YGPAR*------</td>
</tr>
<tr>
<td>Methylococcus_capsulatus_st</td>
<td>SE~AEKISLV GAAVDFLAR HGYAESLCAR *</td>
</tr>
<tr>
<td>Citreicella_sp.</td>
<td>ED~DETGRVV DGAVENTLAR YGAA*------</td>
</tr>
<tr>
<td>Loktanella_vestfoldensis</td>
<td>TR~AEKDRVI NGAVEMFAR YGVP*-------</td>
</tr>
<tr>
<td>Granulibacter_bethesdensis</td>
<td>DP~AFRDYVV RETVRFDLR YAHPT*------</td>
</tr>
<tr>
<td>Agrobacterium_vitis</td>
<td>DQKAYMEAHI DMVQTLMRG KGE*--------</td>
</tr>
<tr>
<td>Novosphingobium_aromaticivo</td>
<td>PL~DVENMA ENHAGLILAG LRRGDVA*-----</td>
</tr>
</tbody>
</table>
APPENDIX D

NpR3597 AcrR family transcriptional repressor aligned with amino acid sequence of Cmer in *Campylobacter jejuni* (FUGUE output)

<table>
<thead>
<tr>
<th>hs2qcoa</th>
<th>( 6   )</th>
<th>QUERY 3597</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>tpstkvlargekIkavAlelFltkgYqetTs1sd1ik1Sggsy</td>
<td>MVRIKTGEVDRNSVDKVEQILQGAHGFPLQNYAGTSMDRVAVAAGVS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aaaaaaaaaaaaaaaaaaaaaaaaaa</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>hs2qcoa</th>
<th>( 48  )</th>
<th>QUERY 3597</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sniydgFkskegLFPeilddiickhflIyskTgeikngtikeILtsFGl</td>
<td>ATVYSHFQDEGLFKVLLEQQLTSSKKNSSIFGTE--PIEGEPAAILRQIVT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aaaaaaaaaaaaaaaaaaaaaaaaaa</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>hs2qcoa</th>
<th>( 98  )</th>
<th>QUERY 3597</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AFIeif--npeavafGkiysQvydkdrhanwiesnnisyniLmgeF</td>
<td>KALEQMINDQEHSAFMRLIGESGRFP--ELAQICVRMVIPKPVETLTQYL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aaaaaaaaaaaaaaaaaaaaaa</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>hs2qcoa</th>
<th>( 147 )</th>
<th>QUERY 3597</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>kqgnasyumknAeklAvlPCtmLkepyhhlNVliiaplnkkkeQkehVef</td>
<td>EA--PELKIDPEATARILGTLVHFHTQVNHQCD--IVPMESDLIDA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a aaaa aaaaaaaaa aaaaaa aaaaaaaa</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>hs2qcoa</th>
<th>( 197 )</th>
<th>QUERY 3597</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VVnflngins</td>
<td>LTHLITKCAD--</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aaaaaaa</td>
</tr>
</tbody>
</table>
## APPENDIX E

### NpR1931 AcrR family transcriptional repressor aligned with amino acid sequence of Cmer in *Campylobacter jejuni* (FUGUE output)

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>hs2qcoa</strong></td>
<td>( 6 )</td>
<td>tspkvlarqeIkvavAel1FltkgYqetTlsldIikI $</td>
<td>$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>QUERY 1931</strong></td>
<td></td>
<td>MPKTSGSQPETSRLRAMTQKQEIQILQGAMQVFLRDGYAGTSMDRVSAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>aaaaaaaaaaaaaaaaaaaaaa</td>
<td>aaaaaaaaaaaaaaaa</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>70</td>
<td>80</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td><strong>hs2qcoa</strong></td>
<td>( 44 )</td>
<td>ggsysnydgsFkskekLFFeilddickkhfh1IyskTqeikngt1keILt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>QUERY 1931</strong></td>
<td></td>
<td>GVSKQTYYSHFQOQGLFKALIERVTIASFQGIFCTE--DLHGEPAILLR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>aaaaaaaaaaaaaaaaaaaaaa</td>
<td>aaaaaaaaaaaa</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>110</td>
<td>120</td>
<td>130</td>
<td>140</td>
<td>150</td>
</tr>
<tr>
<td><strong>hs2qcoa</strong></td>
<td>( 94 )</td>
<td>sPClAPIeif-ngeavafkkiygsQvydkdrhlapwienngqnyfypI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>QUERY 1931</strong></td>
<td></td>
<td>EIAETLYKVAENSEYALFLRIITEQRP--ELAKLYQTIVQRGRLL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>aaaaaaaaaaaaaaaaaaaaaa</td>
<td>aaaaaaaaaaaa</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>160</td>
<td>170</td>
<td>180</td>
<td>190</td>
<td>200</td>
</tr>
<tr>
<td><strong>hs2qcoa</strong></td>
<td>( 143 )</td>
<td>mgfFkqmmnymkknAeklAv1FctmLkepyhl1NVlinap1kkkQke</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>QUERY 1931</strong></td>
<td></td>
<td>SQYLSH--PELGVSDPEATAQIFFGSLVSYVHQEMLYKE--MMPLSRDR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>aaaa aaaaa aaaaaaaa aaaaaaaa</td>
<td>aaaa</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>210</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>hs2qcoa</strong></td>
<td>( 193 )</td>
<td>hVefVvnvflngins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>QUERY 1931</strong></td>
<td></td>
<td>ILDSLMSLVLAHTSKDH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>aaaaaaaaaaaa</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>