THE UNIVERSITY OF CHICAGO

# GENE NETWORK ANALYSIS IDENTIFIES A CENTRAL POST-TRANSCRIPTIONAL REGULATOR OF CELLULAR STRESS SURVIVAL

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN BIOCHEMISTRY AND MOLECULAR BIOPHYSICS

BY

MATTHEW TIEN

CHICAGO, ILLINOIS

JUNE 2018

## Table of Contents

**List of Figures**

## List of Tables

# Chapter 1: Introduction

## 1.1. Gene regulation dictates the physiology of the cell

Cells employ gene expression programs to maintain homeostasis with their environment. François Jacob and Francis Monod first developed this seminal model in their studies of diauxic growth in *Escherichia coli* cultures. They formulated that bacteria sense the levels of specific carbon sources and adapt their transcriptional programming to persist under this new growth regime (Jacob and Monod 1961). Today, this concept largely remains unchanged: gene regulation in response to extrinsic or intrinsic stimuli underlies almost all biological processes.

Defining sets of genes that are shown necessary for a response is now routine with global measurements of steady state levels of RNA and protein; however, interpretation of this information at best provides correlations of hundreds of gene output levels. For example, the general stress response (GSR) pathway in bacteria is a ubiquitous transcriptional response that is necessary for bacteria to persist under numerous non-optimal conditions (Battesti, Majdalani, and Gottesman 2011, Fiebig et al. 2015, Francez-Charlot et al. 2015, Hecker, Pane-Farre, and Volker 2007). The GSR affects the expression of hundreds of genes; however, how do these transcriptional changes allow cells to persist under non-optimal growth?

To better understand the broad transcriptional response of the GSR, I focused on a unique observation that the GSR in *C. crescentus* is necessary for cell survival during hydrogen peroxide stress (Alvarez-Martinez et al. 2007). The GSR-regulated genes in *C. crescentus* at the

transcriptional level show no obvious roles in hydrogen peroxide stress, unlike the analogous system in *E. coli* (Schellhorn and Hassan 1988). Thus, how does the general stress response in *C. crescentus* (an obligate aerobe) increase resistance to a common endogenous and exogenous metabolite of aerobic respiration?

Using a bioinformatic analysis, we uncovered the function of a small non-coding RNA (named GsrN) that is under the control of the GSR and elucidates GSR's role in peroxide stress survival. Surprisingly, we found that GsrN is important for other stress responses in *C. crescentus*, an extremely rare observation in known sRNA systems in other bacteria.

### 1.2. Stress response activates alternative sigma factor(s)

The bacterial stress response is a historic and well-studied gene regulation system. Early proteomic studies on carbon, nitrogen, and phosphorous starvation as well as heat shock in *E. coli*, found that protein abundances and species shifted quite significantly upon stress induction (Lemaux et al. 1978, Groat et al. 1986). Moreover, some of the proteomic shifts seemed consistent in all conditions. Decades later, scientists recognized that the several distinct stress-response mutations converged on the general stress response (GSR) alternative sigma factor, RpoS, in *E. coli*.

The GSR in Alphaproteobacteria echoes many of the principles from the RpoS system in Gammaproteobacteria (Fiebig et al. 2015, Francez-Charlot et al. 2015). Thus it is of importance to highlight the historical precedence of RpoS to address how the thesis will approach studying

the GSR in *C. crescentus*. Given that GSR transcriptional regulation relies heavily on sigma factors, it is appropriate to begin with the function of these proteins.

### 1.2.I. Introduction to bacterial transcription and sigma factors

RNA production from a DNA template occurs in two distinct steps: initiation and elongation (Feklistov et al. 2014). This two-step process is conserved across all organisms; however, the general components and mechanics of initiation and elongation by RNA polymerase vary between prokaryotes and non-prokaryotes (Struhl 1999). Regardless, the initiation step of transcription is the predominant mechanism in which genes are transcriptionally regulated.

Initiation of transcription in prokaryotes requires a modular protein called a sigma factor (Burgess et al. 1969). Although they are not necessary for the elongation of a transcript, they are required for the RNA polymerase to bind and initiate the elongation reaction (Burgess and Travers 1970). When a sigma factors binds to RNA polymerase, specific amino acid residues in the sigma factor are exposed and determine the sequence specific recognition of regions along DNA (Campbell et al. 2002, Murakami et al. 2002, Murakami, Masuda, and Darst 2002). Thus, the sigma factor is the predominant form in which bacteria determine the transcriptional output of a cell.

With the exception of one known bacterial species, all bacteria have multiple varieties of sigma factor genes (Gruber and Gross 2003). Thus, each sigma factor might drive a distinct regulon in the cell. This gives rise to the idea that modular components, such as sigma factors, can activate hundreds of genes and dictate the transcriptional capacity of the cell at a sequence specific level.

However, how does a cell determine the abundance and activity of each sigma factor? How does sigma factor balance determine the optimal programming of a cell?

## 1.2.II. RpoS discovery, regulation, and importance

During platonic exponential growth conditions in bacteria, the "house-keeping" sigma factor (RpoD) drives the majority of transcriptional programming within the cell (Jishage and Ishihama 1995). Upon broad stress-induction, the alternative sigma factor RpoS becomes activated and produces physiological effects for the cell to persist. This model of gene regulation under stress took decades to develop.

RpoS in *E. coli* has moonlighted under several identities: *katF*, *appR*, and *csi2* to name a few. Historically, in mutational screens of distinct stress conditions in *E. coli*, mutations within RpoS gave rise to multiple stress sensitivities. Thus, independent studies on oxidative stress induced by hydrogen peroxide, acid tolerance, or carbon starvation (respective to the aforementioned RpoS aliases) re-discovered the importance of RpoS in various stress responses (Loewen and Triggs 1984, Touati, Dassa, and Boquet 1986). It was not until a seminal paper of characterization of the *csi2* (carbons starvation induced) *lacZ* mutant in stringent response that the culmination of seemingly separate stress response systems converged on a single transcriptional regulator, RpoS (Lange and Hengge-Aronis 1991). Since then, *rpoS* has been implicated in over a dozen stress responses (Battesti, Majdalani, and Gottesman 2011) and its activation and regulon is referred to as the general stress response (GSR).

### 1.2.II.A. RpoS regulation and regulon

Mis-regulation of RpoS-dependent transcription has been suggested to be deleterious for the cell (Zambrano et al. 1993). Thus, several systems ensure RpoS-dependent transcription only occurs during appropriate growth conditions. The dominant form of *rpoS* transcription is a classical feed forward loop: RpoS bound to RNA polymerase directly transcribes the *rpoS* gene (Lange and Hengge-Aronis 1994). Post-transcriptionally, several sRNAs activate the translation of *rpoS* mRNA by exposing the ribosome binding site hidden in structured RNA hairpins, which will be discussed in Chapter 3.1 (Majdalani et al. 1998, Majdalani, Hernandez, and Gottesman 2002, Papenfort et al. 2009). The most predominant levels of regulation is probably the sigma factor adaptor, RssB, a protein that sequesters RpoS from binding to RNA polymerase and leads RpoS to be degraded by the ClpX protease (Hengge-Aronis 2002). Lastly, anti-adaptors can disrupt the RssB-RpoS interactions and allow RpoS to bind RNA polymerase (Battesti et al. 2013).

The inputs and outputs of the RpoS regulon are not always consistent across all stress conditions. The RpoS regulon can be tuned where subsets of genes are only activated during specific stress exposure; however, the mechanisms that dictate the logic of the subset activation is still an area of exploration. On the other hand, many RpoS-dependent genes are always activated despite different chemically and physically distinct stress signals (Weber et al. 2005). The broad induction allows for stress resistance to multiple stress conditions, termed cross-resistance.

The cross-resistance feature of the RpoS regulon has made it difficult to assign stress-specific functions to genes activated in this process. Although the broad regulon has been identified, the

mechanisms by which direct and indirect regulation of these genes confer adaptation to stress is largely unknown.

### 1.2.II.B. Physiological importance of RpoS

The prevalence of the RpoS system is conserved throughout the in Beta-, Gamma-, and Deltaproteobacteria (Battesti, Majdalani, and Gottesman 2011). However, the role of RpoS in bacterial native environments remains an open area of exploration. One avenue to explore the ecological role of RpoS is to understand how RpoS affects virulence in pathogenic bacteria.

RpoS seems essential for virulence in some pathogenic bacteria, but does not seem to affect colonization or infection in others (Dong and Schellhorn 2010). For instance in *Borrelia burgdorferi* and *Salmonella enterica*, RpoS is essential for virulence (Fang et al. 1992, Caimano et al. 2004). However, in *Shigella flexneri* and *Yersinia enterocolitica*, RpoS removal has no effect on virulence or invasion (Mogull et al. 2001, Badger and Miller 1995). In all cases, the RpoS regulon under laboratory conditions still manifest phenotypes in terms of sensitivity to environmental perturbations.

RpoS is a great model system to understand the stress survival capacity of several bacterial species, however it is not the only stress response system known in bacteria. Based on the essentiality of RpoS in bacterial survival in host organisms and in harsh environmental conditions, it is not surprising that both Firmicutes and Alphaproteobacteria have evolved the same strategies to activate a sigma factor under multiple stress conditions (Hecker, Pane-Farre, and Volker 2007, Fiebig et al. 2015, Francez-Charlot et al. 2015).

## 1.2.III. General stress response in Alphaproteobacteria

In Alphaproteobacteria, the functionally analogous GSR system also relies on the activation of an alternative sigma factor. Unlike the RpoS sigma factor, however, the sigma factor(s) involved in the GSR in Alphaproteobacteria are in the class 3 sigma family, also known as the extracytoplasmic function (ECF) sigma factors. It is common for Alphaproteobacteria to have multiple paralogues of GSR ECF sigma factors, some of which are redundant and some of which are necessary (Fiebig et al. 2015, Francez-Charlot et al. 2015). Compared to the RpoD and RpoS class 1 sigma factors, ECF sigma factors are missing two domains, a regulatory N-terminal domain and an internal domain that provides increased specificity to the extended -10 element (Lane and Darst 2006).

Although structurally distinct from RpoS, ECF stress response sigma factor(s) (EcfG) in Alphaproteobacteria are governed by some of the same principles of the RpoS system. Strains without EcfG transcription in Alphaproteobacteria cause severe viability defects under stressful conditions, compromises colonization of natural niches, and affects host-bacterial interactions (Fiebig et al. 2015, Francez-Charlot et al. 2015). Moreover, hyper EcfG transcription has been shown to be lethal to the cell much like mis-regulation of RpoS (Kim et al. 2014, Kulkarni, Wu, and Newman 2013, Sauviac et al. 2007, Kaczmarczyk et al. 2011). Not surprisingly, the regulation of EcfG transcription is tightly controlled and is activated during specific response.

Regulation and function of EcfG(s) across the Alphaproteobacteria are remarkable conserved. The known conserved activation of EcfG occurs through a partner-switch mechanism, in which a response regulator (PhyR) that contains a sigma-like domain binds the anti-sigma factor (NepR).

Under optimal growth, NepR sequesters EcfG from binding RNA polymerase. Under specific conditions, including stress induction, NepR switches its protein partner to PhyR and allows for EcfG to bind RNA polymerase(Herrou et al. 2010, Francez-Charlot et al. 2009, Gourion et al. 2009, Herrou and Crosson 2011). EcfG bound to RNA polymerase transcribes a specific set of genes with a well-defined, highly conserved DNA binding-site located at the canonical -35 and -10 promoter elements (Alvarez-Martinez et al. 2007).

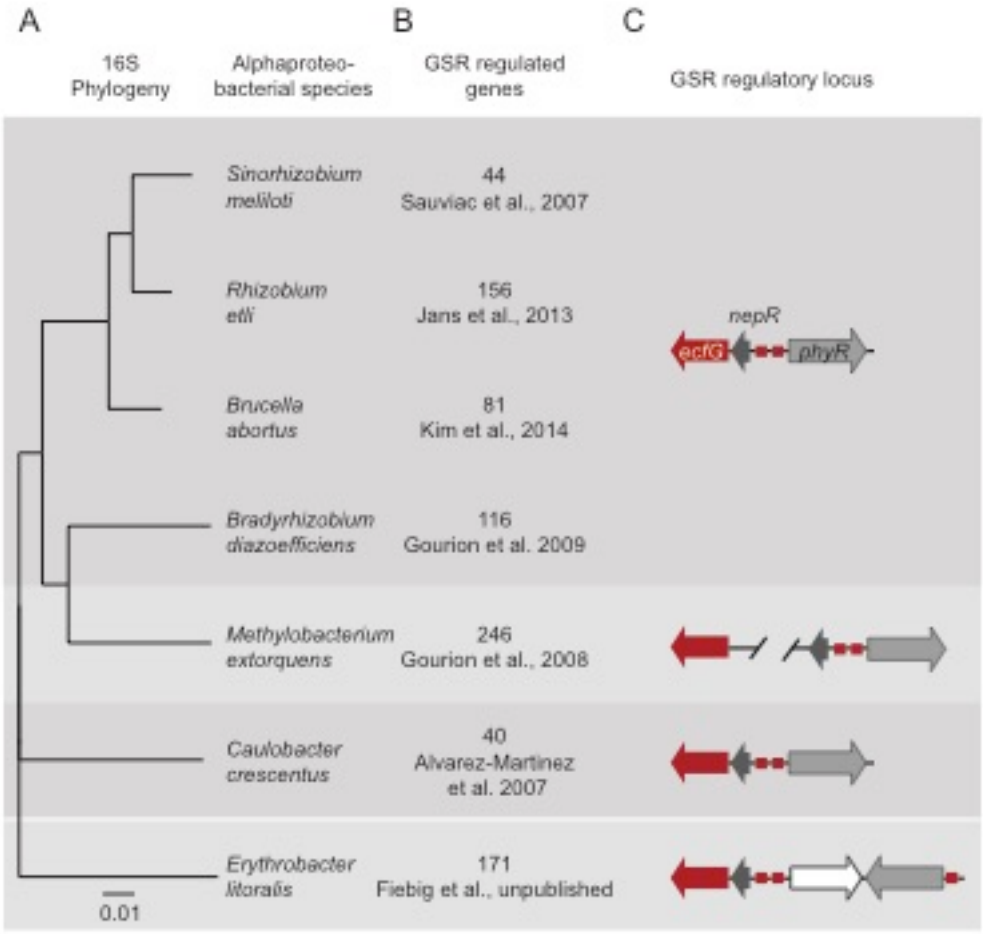


Figure 1. Features of the general stress response in Alphaproteobacterial species.
(A) Phylogenetic analysis of 16S rRNA of alphaproteobacteria with defined GSR regulons.
(B) Number of reported GSR associated genes for each species
(C) The gene architecture of each respective GSR-core components: ecfG (red), nepR (dark grey), and phyR (light grey). Red boxes indicate SigT-Binding sites.

Akin to the transcriptional feed-forward regulation of RpoS, EcfG bound to RNA polymerase initiates the transcription of *ecfG*. Most often *nepR* lies directly upstream of *ecfG* and both genes are co-transcribed from the same EcfG promoter. The partner-switch response regulator gene, *phyR*, also has strong synteny with *nepR* and *ecfG*. It is also not uncommon for the kinases that phosphorylate PhyR, which causes the sigma-like domain to become exposed, to also be proximal to the core regulatory gene locus (*nepR*, *ecfG*, and *phyR*) (Figure 1C).

In pathogenic alphaproteobacteria, such as *Brucella* and *Bartonella*, removal of GSR affects chronic infection and host colonization (Kim et al. 2013, Kim et al. 2014, Tu et al. 2016, Abromaitis and Koehler 2013). Similarly in *Rhizobia* and *Methylobacteria*, root nodule formation and leaf colonization are also compromised upon removal of the GSR (Gourion et al. 2009, Gourion, Rossignol, and Vorholt 2006, Kaczmarczyk et al. 2011). In the oligotrophic model organism *Caulobacter crescentus*, genetic deletion of GSR sensory system hinders survival in the face of several stresses (Foreman, Fiebig, and Crosson 2012, Alvarez-Martinez et al. 2007).

Although phenotypic characterization of the GSR in Alphaproteobacteria has been established in several systems, the signals that input and activate PhyR phosphorylation and the gene outputs that are regulated by EcfG directly or indirectly are current areas of research. Studies on signal inputs into the GSR and the resulting gene output across Alphaproteobacteria demonstrate a variety of mechanisms across ecologically distinct species. Moreover, the number of reported EcfG regulated genes vary by organism. Systematic characterization of these downstream genes

strongly regulated by the EcfG in several organisms are reported, however, a majority of these studies have yet to assign functions that can elucidate phenotypic characterization (Figure 1B).

### 1.2.IV. General stress response in *Caulobacter crescentus*

In the oligotrophic fresh-water bacterium, *Caulobacter crescentus*, the general stress response contains the known Alphaproteobacterial core-regulatory machinery: *phyR*, *nepR* and *ecfG,* known as *sigT* in *C. crescentus*. The partner kinase gene, *phyK*, in *C. crescentus* lies downstream of *sigT* and *nepR*. Moreover, there is strong evidence that another two-component system, *lovK* and *lovR*, that is under the direct regulation of SigT and seems to negatively regulate PhyR activation. Lastly, *C. crescentus* contains another EcfG sigma factor gene, *sigU*, which is a paralogue of *sigT* (Foreman, Fiebig, and Crosson 2012, Alvarez-Martinez et al. 2007, Lourenco, Kohler, and Gomes 2011).

The GSR regulon has been defined in a variety of global transcriptome experiments (Foreman, Fiebig, and Crosson 2012, Alvarez-Martinez et al. 2007, Lourenco, Kohler, and Gomes 2011). These studies revealed several classes of genes whose annotations seemed to be involved in stress response. For instance, a subset of genes upregulated under osmotic stress (induced with either sodium chloride or sucrose) has annotations associated with known membrane and envelope stress relief proteins (*blc* and *ompA*) (Wang 2002, Campanacci et al. 2004). Interestingly the nucleoid-inducing gene, *dps*, which is also regulated by RpoS in *E. coli*, is in the regulon of the GSR (Martinez and Kolter 1997).

*C. crescentus* strains with a compromised GSR show hypersensitivity to oxidative stress (Alvarez-Martinez et al. 2007, Foreman, Fiebig, and Crosson 2012). This hypersensitivity could be contributed to *dps*, since strains of *E. coli* without *dps* are also hypersensitive to hydrogen peroxide. However, *dps* seems to have multiple stress-resistant phenotypes Nair and Finkel, 2004 (Nair and Finkel 2004). Recent studies in *dps* in *C. crescentus* show a modest susceptibility to hydrogen peroxide and do not seem to mimic the hypersensitivity of a GSR compromised strain (de Castro Ferreira et al. 2016).

Much like the stationary phase induction of RpoS, SigT is also been known to be highly upregulated under stationary phase and carbon starvation. Moreover there exists an interest cross talk between the GSR in *C. crescentus* and known global regulators of cell cycle progression. Specifically, SigT contributes to the degradation of CtrA (an important regulator of cell cycle in *C. crescentus*) during carbon starvation (Britos et al. 2011).

## Chapter 2: Gene regulatory network of cell-cycle regulation in *Caulobacter crescentus* and analysis of regulators of the general stress response

Cell cycle regulation is one the best studied features of *Caulobacter crescentus* (Chapter 7). Utilization of data generated from several decades of research focused on cell-cycle regulation can be quite powerful. The chapter demonstrates how leveraging publically available RNA-Seq data on *C. crescentus'* cell cycle lead to the creation of a hypothesis driven bioinformatics analysis that revealed an unknown component of the general stress response.

### 2.1. Components of the general stress response are cell cycle regulated.

In several studies of synchronized cultures of *C. crescentus*, the general stress response sigma factor gene, *sigT*, shows a consistent pattern of RNA expression (Figure 2A) (Laub et al. 2000, McGrath et al. 2007, Schrader et al. 2014, Fang et al. 2013). Transcript levels of *sigT* increase during the swarmer-to-stalk transition and decrease gradually as division progresses. Since *sigT* is necessary for its own transcription (Alvarez-Martinez et al. 2007, Lourenco, Kohler, and Gomes 2011), this suggests that SigT-dependent transcription is active during the swarmer-to-stalk transition.

Genes under the GSR regulon show variability in cell-cycle transcript levels compared to *sigT* (Figure 2B-D). Although this could be a consequence of genes directly dependent on SigT transcription versus genes under indirect regulation, the pattern of transcript levels do not correlate with the presence or absence of a SigT binding site (Figure 2C). Interestingly, the experimentally verified regulators of sigT-dependent transcription do follow this pattern (Figure 2B). These genes include SigT's cognate anti-sigma factor, *nepR*, the positive regulatory two-

component system *phyK-phyR*, the negative regulatory two-component system *lovK-lovR*, and the paralogue GSR sigma factor *sigU*.

It is interesting that the upstream regulators of the GSR in *C. crescentus'* demonstrate cell-cycle regulation, given that 20-30% of genes in the genome fall into this category. Thus, one could argue that there is a selective pressure for *sigT* to be expressed in a particular transition in *C. crescentus'* life style. If so, what is the function of the GSR during this transition? Moreover, can we leverage the strong cell-cycle transcriptional signatures of the GSR regulators to understand the role of the GSR during stress response?

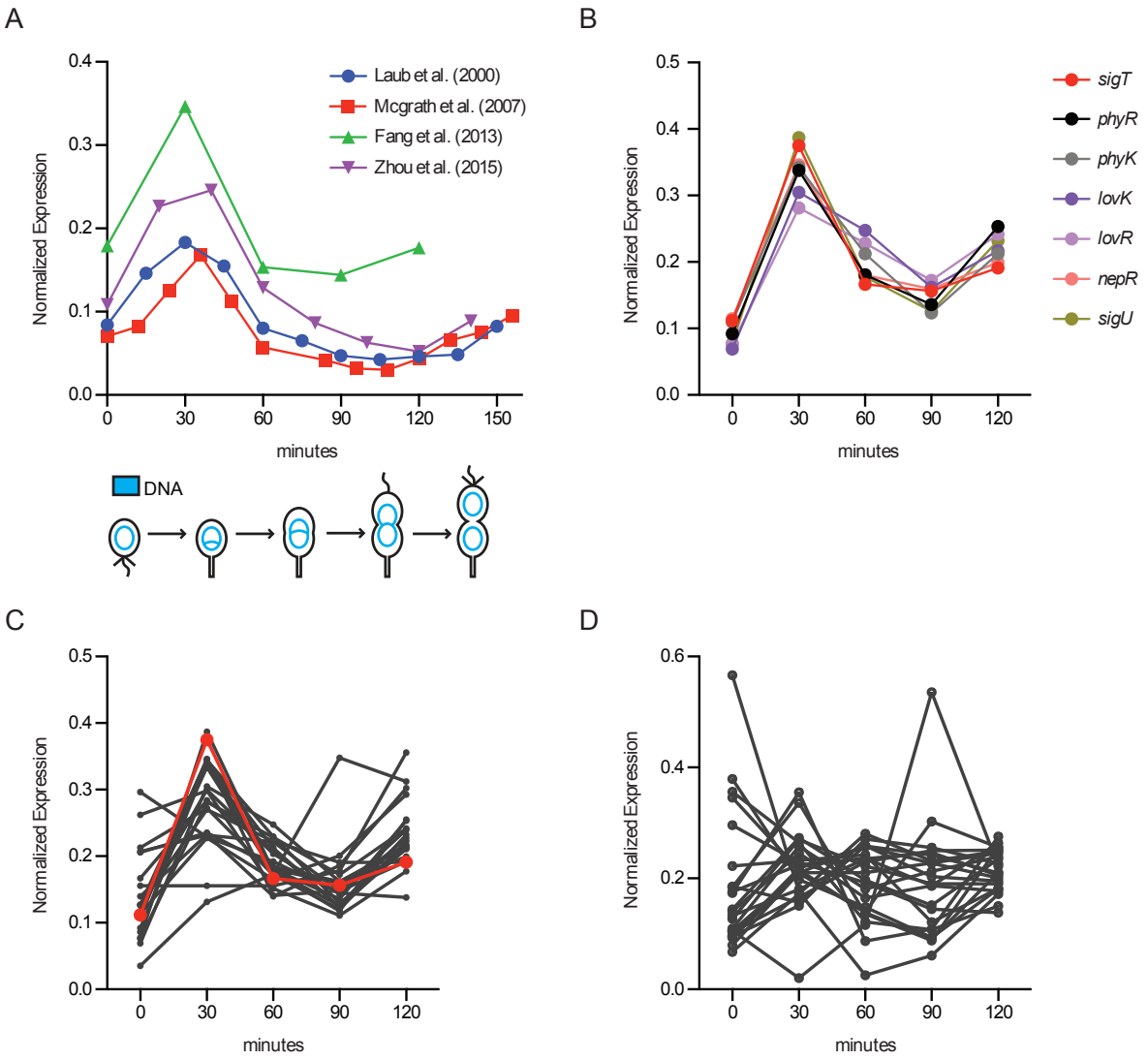


Figure 2. Components of the general stress response are cell cycle regulated.
(A) Normalized transcript levels of *sigT* from 4 different transcriptome studies of the cell cycle in *C. crescentus*. Levels are plotted as a function of time after synchronization. RNA-levels were measured by microarrays (blue and red (Laub et al. 2000, McGrath et al. 2007)) and measured by RNA-Seq (green and purple (Fang et al. 2013, Zhou et al. 2015b)).
(B) Normalized transcript levels from (Fang et al. 2013) of known GSR regulator genes are plotted as a function of cell cycle/time after synchronization. These genes are indicated in the legend on the right.
(C) Normalized transcript levels from (Fang et al. 2013) of known genes with a SigT-binding site are plotted as a function of cell cycle/time after synchronization. *sigT* normalized transcript levels are plotted in red as a reference. (continued on next page)
(D) Normalized transcript levels from (Fang et al. 2013) of known GSR-regulated genes that do not have a SigT-binding site are plotted as a function of cell cycle/time after synchronization. Genes were selected based on their association with the GSR from two studies (Alvarez-Martinez et al. 2007, Foreman, Fiebig, and Crosson 2012).

## 2.2. Simple network analysis reveals regulatory features of the general stress response

Gene regulatory networks (GRN) can organize multiple dimensions of information across hundreds of genes into simplified matrix-based relationships. For instance, two genes' transcript levels as function of the cell cycle can be represented as two nodes linked by an edge (Figure 3C). The edge can encompass a quantified relationship between the two genes' transcript levels during the cell cycle. In Figure 3A-C, the Pearson's correlation coefficient between two genes' transcript levels during the cell cycle encompasses an edge's weight.

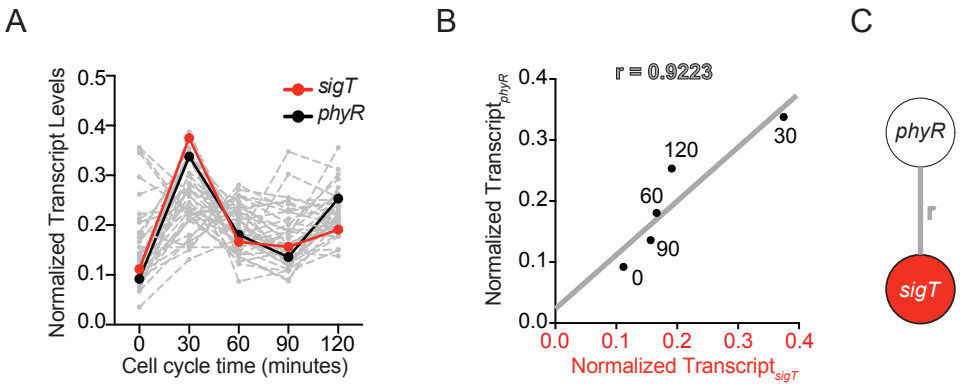


Figure 3 Construction of a gene regulatory network based on cell cycle transcription data.
(A) Normalized transcript levels from (Fang et al. 2013) of known GSR regulated genes are plotted as a function of cell cycle time. The core GSR regulators, *sigT* and *phyR*, are highlighted in red and black respectively.
(B) *sigT* and *phyR* transcript levels are correlated as a function of cell cycle progression, Pearson's correlation coefficient $r = 0.92$.
(C) Representation of a two-node gene regulatory network. Each gene is represented as a node and every node is connected by a weighted-edge from the Pearson's correlation coefficient calculated in (B)

As a proof of concept, differences in cell-cycle regulation within the GSR regulon can be captured in a GRN (Figure 4A). One simple network analysis that is a variant of the Highly Connected Subgraphs (HCS) clustering algorithm is to remove edges that have a low edge value

(Hartuv and Shamir 2000). Figure 4B demonstrates that removing edges that are 0.95 or less creates specific clusters of GSR genes.

The network structure visually simplifies the cell cycle relationship between GSR genes. For instance, coloring genes that have a strong SigT-binding site demonstrates that promoter sequence elements alone cannot predict co-expression within the GSR. Moreover, the regulatory components of the GSR form a unique clique along with some other genes in the GSR regulon. All of these other genes are annotated as genes of unknown function. However, given their association with this cluster, they may be play more critical role in the GSR, perhaps specifically in the cell cycle.

Given the strong cell cycle regulation of the core regulatory genes of the GSR, determining the genome-wide clique that these GSR genes are co-regulated with would more explicitly define the function of the GSR. Identification of this clique could also provide mechanisms that control GSR's cell cycle regulation and predict other genetic players that might be interacting with the GSR. Thus, we constructed a genome-wide GSN based on cell-cycle transcript data to analyze GSR through a different lens.

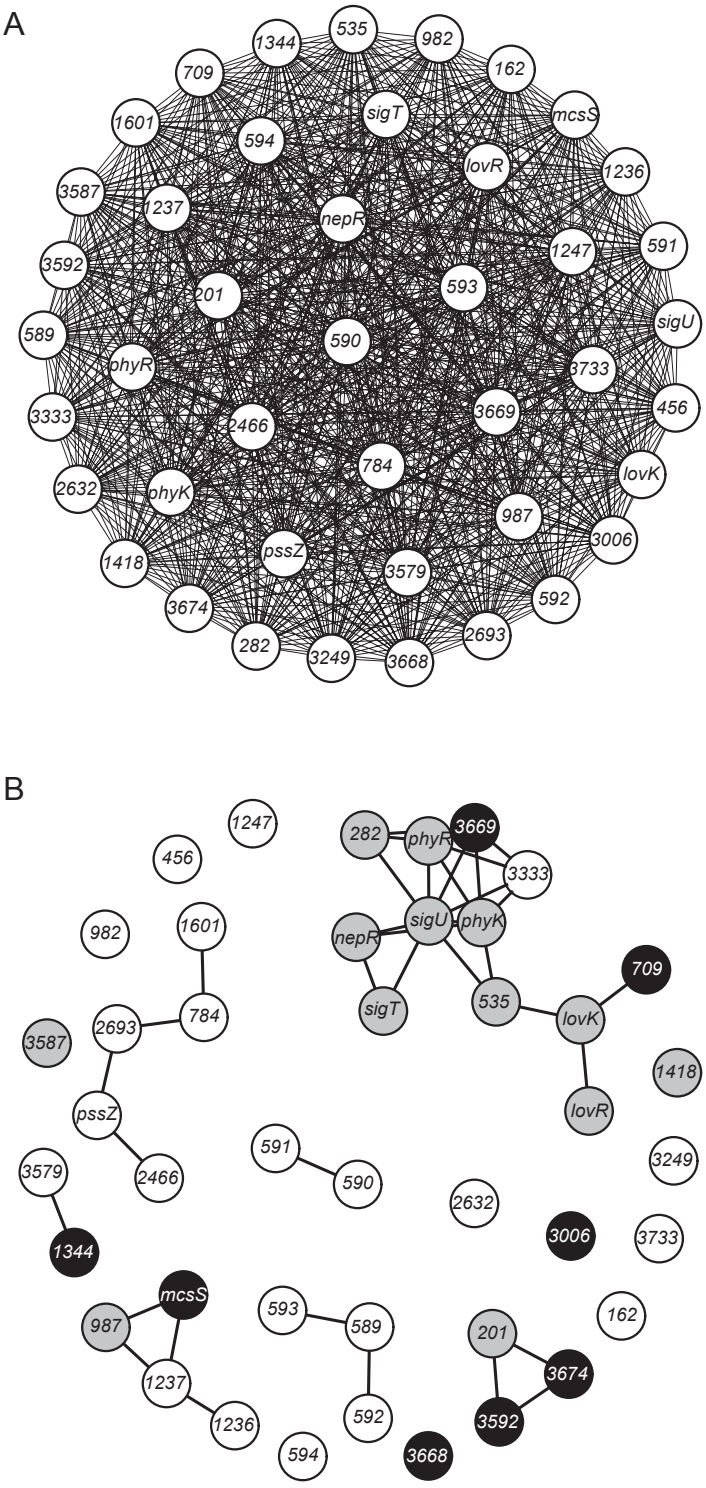


Figure 4 Simple network analysis of the general stress response shows an interesting cluster of the upstream regulators of the general stress response (GSR).
(A) Gene regulatory network (GRN) of the known GSR-regulated genes (identified in(Foreman, Fiebig, and Crosson 2012)) was constructed based on cell cycle RNA-Seq data from (Fang et al. 2013). Network structure is outlined in Figure 3C
(B) Edge reduction analysis of the GSR GRN layered with additional pieces of information. Black gene nodes contain a SigT-binding site in their promoter. White gene nodes have been shown to be regulated by the GSR (Foreman, Fiebig, and Crosson 2012). Grey gene nodes are identified components of the GSR-regulon and have SigT-binding sites. *sigT*, *sigU*, *phyR*, *phyK*, *nepR*, *lovR*, and *lovK* are known regulators of the GSR.

## 2.3. Iterative Rank analysis reveals a sRNA of unknown function to be the strongest co-expressed gene with the regulators of the general stress response

Identification of the cliques that matches the GSR fall into the NP-complete clique problem, a long standing issued in determining subgraphs in large networks (Arora et al. 1998). Given that the genome-wide complete network encompasses information from over 4000 nodes, I decided to approach the problem from the perspective of our original genes of interest. Instead of identifying cliques and finding the GSR regulatory genes, I used a hypothesis-driven algorithm to obtain analogous information to identify genes that are strongly co-expressed with the GSR regulatory genes.

Iterative rank, also known as PageRank (Brin and Page 1998), is a guilt-by-association network method where specified nodes of interest (known genes in the genetic network) are given initial weight. Through systematic iterations, weight is distributed to each node in the network as a function of weight distribution at the previous iteration and the strength of connection between nodes (i.e. the edges of the network). The iteration solution as time approaches infinity converges on a stable solution and provides a list of ranked nodes based on the final accumulated weight.

The GRN can be expressed as a matrix of correlation coefficients between nodes in the network where each row and column correspond to the same gene. This matrix, P, is symmetric with it's diagonal equal to 1. For the purposes of iterative rank, let *f* represent the single-dimensional weight-matrix, where the rows and columns of P match the entries in *f*. Let *t* denote the iteration step. For any given *t*, the weight-matrix *f* can be expressed in equation 1.

$$f^t = \propto f^0 + (1-\propto)\mathrm{P}f^{t-1} \quad (1)$$

In equation 1, let α represent a dampening factor to balance the initial weight, $f^0$, applied at $t$=0 and the weight to be gained as a function of the P and the previous t-1 iteration. As t approaches infinity, the solution converges to a stable solution.

$$f^\infty = \propto [I - (1-\propto)\mathrm{P}]^{-1} f^0 \quad (2)$$

Algorithm and solution information was adapted from (Wang and Marcotte 2010).

2.3.I. Iterative Rank parameter optimization through predictability of *phyR*

Iterative rank parameters were optimized through the self-prediction of the known associated co-expressed GSR regulatory genes. Variables tuned for exploration were the α parameter and the reduction of the number of edges based on correlation cut-offs (previously outlined in Figure 4). Parameters that best predicted the gene *phyR*, when initializing the weight-matrix with *sigT, sigU*, *nepR*, *phyK*, *lovR*, and *lovK* with a value of 1 were chosen. These two parameters showed that an edge reduction of a correlation coefficient of greater than 0.9 and an alpha factor greater than 0.5 yielded the highest rank for *phyR* (Figure 5A).

Edge reduction reduced the number of edges for each node (Figure 5B). Moreover, the number of edges was reduced from 10225998 edges to 946558 (Figure 5C). Only 19 nodes (.46%) were completely disconnected from the network (zero number of edges).

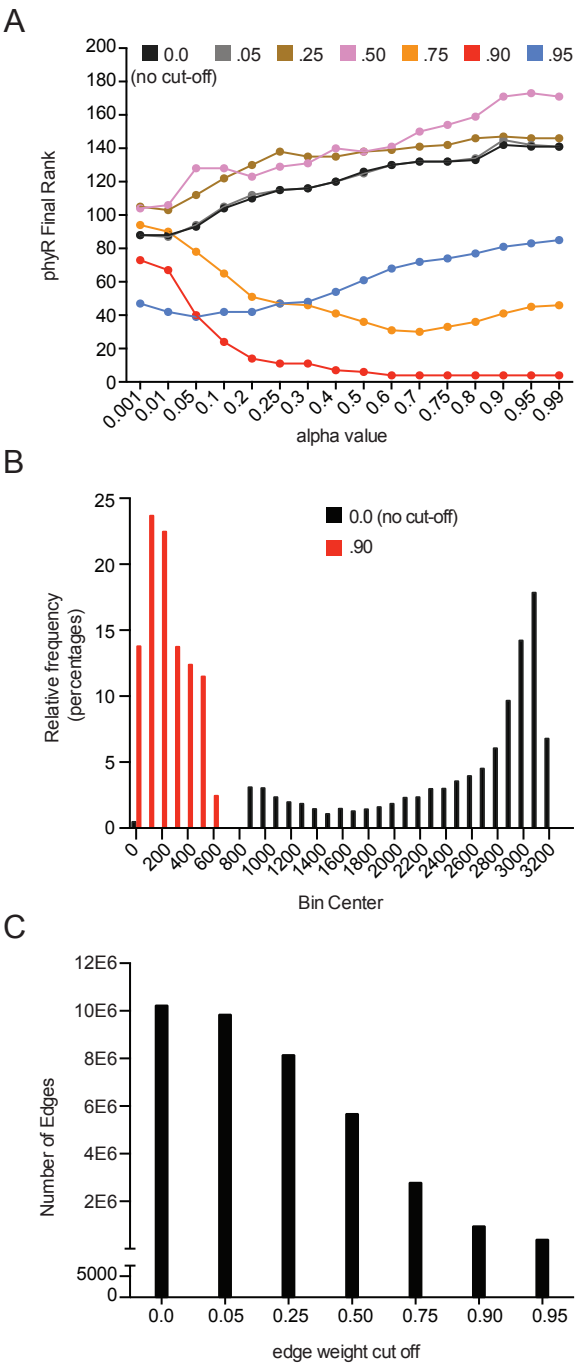


Figure 5. Parameter optimization of iterative rank through predicting *phyR* demonstrates edge-reduction as an important parameter.
(A) Systematic parameter exploration of the alpha value and edge-reduction, where *sigT*, *sigU*, *phyK*, *nepR*, *lovK*, and *lovR* are initialized with weight. Alpha value is plotted on the x-axis while the final rank of *phyR* on the y-axis. Colors indicate network where edges less than the indicated value were removed from the network.
(B) The number of edges drawn for a given node shows that an edge reduction of 0.9 dramatically shifts the average number of edges per node.
(C) The number of total edges in a network shows that an edge reduction of 0.9 is reduced by 10-fold.

## 2.3.II. *gsrN* is a SigT-dependent gene strongly co-expressed with the regulators of the general stress response

The highest ranked gene from an iterative rank search with *sigT, sigU*, *nepR*, *phyK*, phyR, *lovR*, and *lovK* and the two optimized parameters was a gene encoding an sRNA (*CCNA_R0081*) (Landt et al. 2008) with a consensus $\sigma^T$ binding site (Figure 6A), which we will refer to as *gsrN*.

We first wanted to validate the consensus SigT binding site and examine if GsrN is dependent on the presence of the general stress response transcription. To test whether *gsrN* transcription requires the GSR sigma factor, SigT, we generated a transcriptional reporter by fusing *gsrN* promoter to *lacZ* and measured β-galactosidase activity. The activity of the *gsrN* transcriptional reporter is dependent on the presence *sigT* (Figure 6A&B), which validates *gsrN* as a bona fide member of the GSR-regulon. To directly measure the levels of GsrN in a Δ*sigT*, we performed a Northern blot assay. With radiolabeled oligo-probes that are complementary and specific to GsrN, we were unable to detect GsrN probe signal in a Δ*sigT* strain. Surprisingly, we detected two GsrN isoforms in wild-type RNA samples.

GsrN comes in three types of isoforms: a full-length ~104nt RNA, a more abundant 5'isoform ~54nt RNA, and two low abundant 3 'isoforms. We identified the 5'isoform and 3'isoforms of GsrN with probes specific to the 5' and 3' portions of GsrN (Figure 7A). Size estimation of GsrN was done through identifying the 5S rRNA (118nt) and Tyr-tRNA (84nt) in our Northern blot assays. RNA-seq read densities mapped onto the *gsrN* locus also show evidence of a more abundant 5'isoform (Figure 7B).

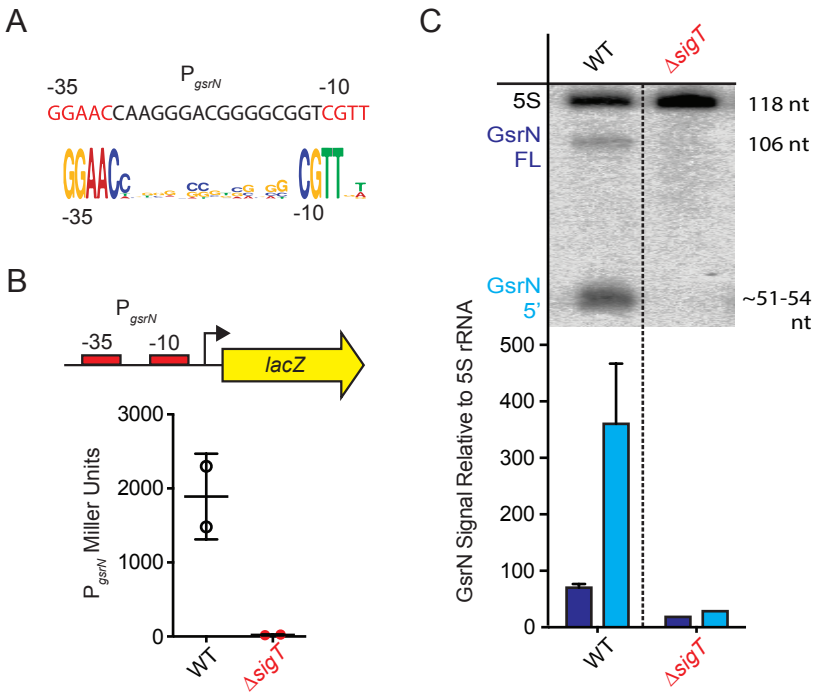


Figure 6. *gsrN* has a *sigT* binding site and its transcription depends on *sigT*.
(A) Promoter of *gsrN* contains a consensus SigT-binding site (nucleotides in red) (McGrath et al. 2007, Staron et al. 2009). SigT-binding motif (bottom) generated from twenty-one SigT-dependent promoters using WebLogo (Crooks et al. 2004).
(B) Schematic of *lacZ* transcriptional fusions to the promoters of *gsrN*. β-galactosidase activity from the $P_{gsrN}$*lacZ* transcriptional fusion in *Caulobacter* wild-type and Δ*sigT* backgrounds measured in Miller Units. Bars represent mean ± SD from 2 independent cultures.
(C) Northern blot of RNA isolated from wild type (WT) and Δ*sigT* cultures probed with oligos complementary to GsrN. 5S rRNA was blotted as a loading control. Cells were harvested in exponential phase. Blots were quantified by densitometry. GsrN signal from the full-length (FL; dark blue) and 5' isoform (5'; cyan) are normalized to 5S rRNA in each lane and multiplied by 100. Bars represent mean ± SD of triplicate extractions, each representing biologically independent samples.

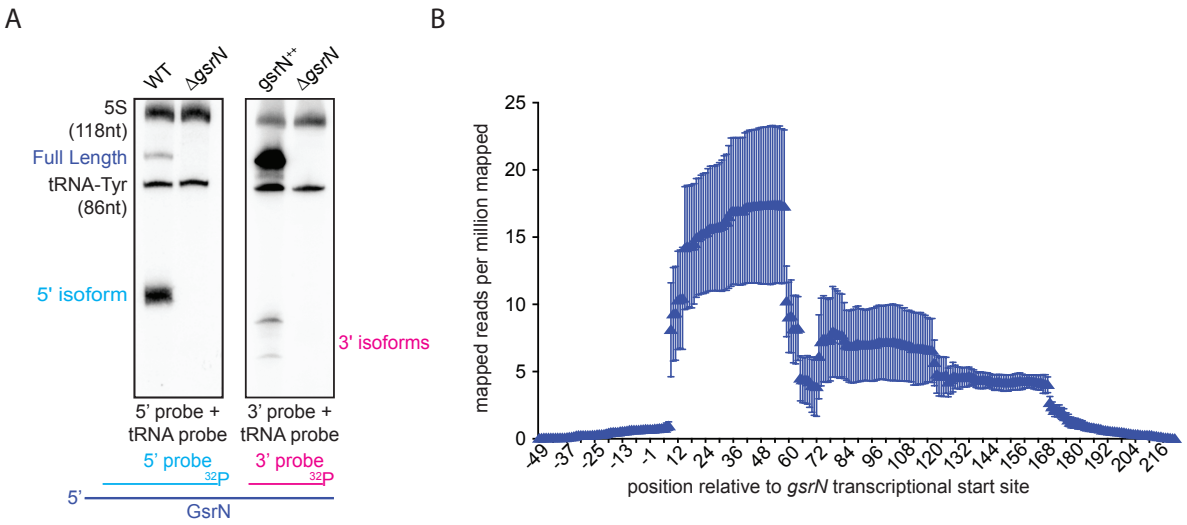


Figure 7. *gsrN* has two isoforms.
(A) Northern blots of total RNA from cultures ($OD_{660} \approx 1.0$) of wild type, Δ*gsrN*, and *gsrN*$^{++}$. Blots were probed with $^{32}$P-labeled oligonucleotides complementary to either the 5' or 3' end of GsrN. Probes to 5S rRNA and tRNA-Tyr were used to estimate the size of full-length GsrN and its 5' and 3' isoforms.
(B) RNA-seq read density from total wild-type RNA mapped to the *gsrN* locus. Chromosome position (x-axis) is marked in reference to the annotated transcriptional start site (TSS) of *gsrN* (position 3,830,130 in GenBank accession CP001340). Reads per million reads mapped is plotted as a function of nucleotide position. Mean ± SD from three independent biological replicates samples is plotted (GEO: GSR106168, read files: GSM2830946, GSM2830947, and GSM2830948).

To understand the implication of a GSR regulated sRNA, we wanted to test if GsrN had any phenotypes similar to those associated with *sigT* and *phyR* in *C. crescentus*. The strong co-expression could indicate a regulatory role for GsrN much like the sRNA regulators that feed into *rpoS* mRNA translational activation. However, GsrN might also act as a secondary activator and help elucidate some of the phenotypes associated with the SigT-regulon, much like the RpoD-dependent sRNAs discussed in Chapter 3.1.

## Chapter 3: Genetic analysis of *gsrN* in *Caulobacter crescentus*

### 3.1. Introduction to sRNA regulators in bacteria

As previously mentioned in Section 1.2.I.A, trans-encoded, small non-coding RNA (sRNA) regulators have been historically associated with general stress responses. Three non-coding RNA elements (RprA, DsrA, and ArcZ) help activate the translation of *rpoS* mRNA (Majdalani, Hernandez, and Gottesman 2002, Sledjeski, Gupta, and Gottesman 1996, Mandin and Gottesman 2010). Each sRNA is induced under specific conditions: DsrA during cold shock, RprA during envelope stress, ArcZ during aerobic growth. Moreover, there are two characterized sRNAs that are transcribed directly by RpoS, SdiA and SdrA. SdiA and SdrA repress inorganic nitrogen utilization genes and a porin gene *ompD*, respectively (Hao et al. 2016, Frohlich et al. 2012).

Mechanisms of sRNA involve base pairing to their mRNA targets to degrade or stabilize their target transcript or to modulate the translational activity of their targets (Wagner and Romby 2015). The sRNAs that activate *rpoS* translation bind the 5' untranslated region of *rpoS* mRNA and expose its ribosome-binding site for translation initiation (McCullen et al. 2010). Conversely, another well-studied sRNA, RybB, regulates the 5'UTR of its targets in the opposite manner, repressing translation initiation (Bouvier et al. 2008).

Mutations of sRNAs usually manifest physiological effects under distinct laboratory conditions. Moreover, only a handful of sRNAs show phenotypic consequences upon deletion (Papenfort et al. 2013). For instance, deletion of all three RpoS activating sRNAs in *E. coli* causes susceptibility to acid stress (Bak et al. 2014). In *C. crescentus*, the sole characterized sRNA,

CfrA, is induced under carbon starvation and inhibits growth when overexpressed in rich media (Landt et al. 2010). Thus, it was quite surprising that deletion and overexpression of GsrN showed complementary viability phenotypes in the face of oxidative stress.

### 3.2. GsrN is a necessary during peroxide stress in *Caulobacter crescentus*

To test whether *gsrN* plays a role in stress survival, we subjected strains lacking *gsrN* or the core GSR regulators, *sigT*, *phyR*, or *phyK*, to hydrogen peroxide, a known stress under which GSR regulatory mutants have a survival defect. Δ*sigT*, Δ*phyR*, and Δ*phyK* strains had a ≈4-log decrease in cell survival relative to wild type after exposure to hydrogen peroxide, as previously reported (Alvarez-Martinez et al. 2007, Foreman, Fiebig, and Crosson 2012). Cells lacking *gsrN* (Δ*gsrN*) had a ≈3-log viability defect relative to wild type (Figure 8A&B). Insertion of *gsrN* with its native promoter at the ectopic *vanA* locus fully complemented the peroxide survival defect of Δ*gsrN* (Figure 8C). To assess the effects of *gsrN* overexpression, we inserted constructs containing either one or three copies of *gsrN* under its native promoter into the *vanA* locus of wild-type and Δ*gsrN* strains (Figure 8C).

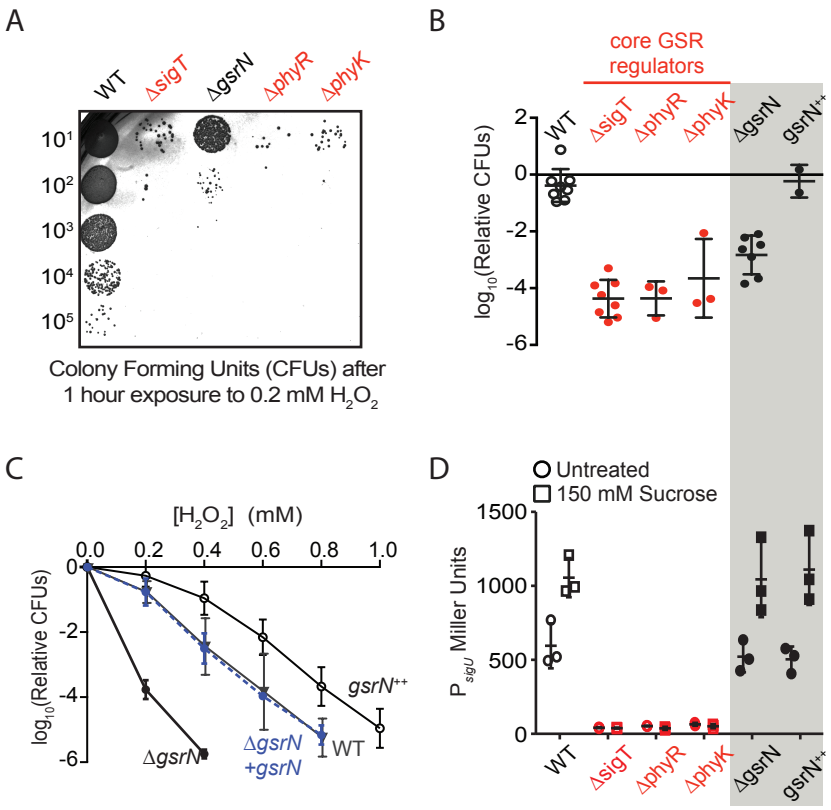


Figure 8. *gsrN* is necessary for peroxide stress survival, but does not affect SigT-dependent transcription.

(A) Colony forming units (CFU) in dilution series ($10^{-1}$ to $10^{-5}$) of wild-type, Δ*sigT,* Δ*phyR,* Δ*phyK* and Δ*gsrN C. crescentus* strains after 0.2 mM hydrogen peroxide treatment for 1 hour. Red labels indicate known regulators of the GSR.

(B) Relative survival of strains in (A) treated with 0.2 mM hydrogen peroxide for 1 hour. $Log_{10}$ relative CFU (peroxide treated/untreated) is plotted for each strain where independent experiments are plotted as points and bars represent the mean ± SD.

(C) *Caulobacter* wild type (WT), *gsrN* deletion (Δ*gsrN*), complementation strain contains *gsrN* under its native promoter integrated at an ectopic locus (Δ*gsrN*+*gsrN*), and *gsrN* overexpression (4*gsrN*, $gsrN^{++}$) strains were subjected to increasing concentrations of hydrogen peroxide for one hour and tittered on nutrient agar. $Log_{10}$ relative CFU (peroxide treated/untreated) is plotted as a function of peroxide concentration. Δ*gsrN* and WT strains carried the empty integrating plasmid (pMT552) as a control. Mean ± SD, n=3 independent replicates.

(D) β-galactosidase activity from the $P_{sigU}$*lacZ* transcriptional fusion in a set of the genetic backgrounds in (B). GSR transcription was induced by exposure to 150 mM sucrose (final concentration) for three hours before measuring β-galactosidase activity. Time and concentration of our reporter were chosen based on past experiments with this reporter (Foreman, Fiebig, and Crosson 2012). Bars represent mean ± SD from 3 independent cultures.

To determine whether *gsrN* is a feedback regulator of GSR transcription, much like DsrA, RprA, and ArcZ, we utilized a well-characterized $P_{sigU}$*lacZ* reporter (Foreman, Fiebig, and Crosson 2012). Transcription from $P_{sigU}$ required *sigT* and other GSR regulators (*phyR*, *phyK*), but was unaffected by deletion or overexpression of *gsrN* (Figure 8D).

To more closely examine the relationship between oxidative stress survival and the levels of GsrN, we measured GsrN expression directly in these strains by Northern blot (Figure 9A&B) and tested their susceptibility to hydrogen peroxide (Figure 9C). Treatment with increasing concentrations of hydrogen peroxide revealed that strains overexpressing *gsrN* have a survival advantage compared to wild type. Measured levels of GsrN in the cell directly correlated (*r=0.92*) with cell survival, which provides evidence that the protective effect of *gsrN* under peroxide stress is dose dependent over the measured range (Figure 9C).

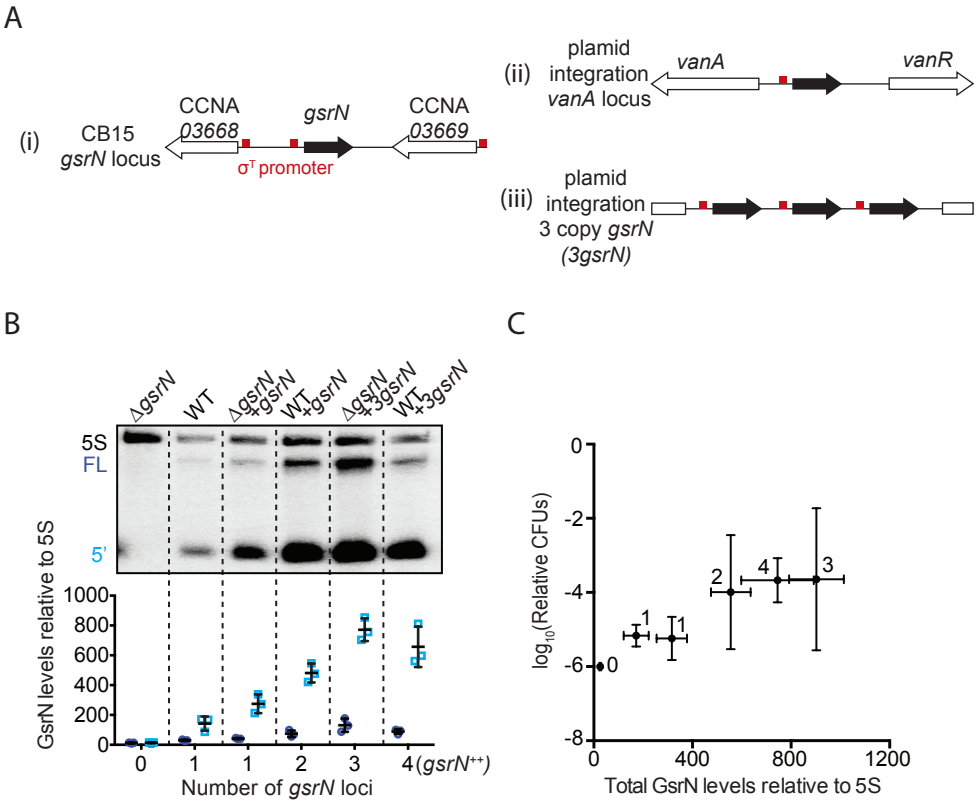


Figure 9. GsrN-dependent cell protection under oxidative stress is dose dependent.
(A) In this study, *gsrN* was expressed several different ways. (i) At the native locus, the *gsrN* promoter contains a consensus SigT-binding site (red box); *gsrN* is flanked by two genes also with predicted $\sigma^T$ promoters. Ectopic complementation and overexpression strains were created using pMT552-derived plasmids containing either (ii) one or (iii) three tandem copies of *gsrN* that were integrated into the chromosomal *vanA* locus.
(B) Northern blots of RNA isolated from strains expressing increasing copies of *gsrN* probed with oligos complementary to GsrN. 5S rRNA was blotted as a loading control. Cells were harvested in exponential phase. Blots were quantified by densitometry. GsrN signal from the full-length (FL; dark blue) and 5' isoform (5'; cyan) are normalized to 5S rRNA in each lane and multiplied by 100. Bars represent mean ± SD of triplicate extractions, each representing biologically independent samples.
(C) Relationship between GsrN levels and peroxide stress survival. Total GsrN levels quantified by Northern blot (B) plotted against relative cell survival after 0.8 mM hydrogen peroxide treatment (CFU determined as outlined in Figure 1- figure supplement 2B). 0.8 mM hydrogen peroxide provided the best dynamic range of quantifying survivability between *gsrN* mutants. The Δ*gsrN* strain has zero CFUs after one hour treatment with 0.8 mM hydrogen peroxide, and no detectable GsrN by Northern blots, thus the y-axis point for this strain was plotted at $10^{-6}$, the detection limit of our assay.

### 3.3. GsrN is sufficient to rescue strains with a compromised general stress response during peroxide stress

Given the necessity of *gsrN* in hydrogen peroxide stress survival, we wanted to test if *gsrN* expression, independent of GSR transcription, can allow cell survival under peroxide stress. To decouple *gsrN* transcription from SigT, we created two promoters (P1 and P2) controlled by the primary sigma factor, RpoD, to express *gsrN* in Δ*sigT*. P1 is derived from the identified RpoD binding site from *vanA* (Thanbichler, Iniesta, and Shapiro 2007). P2 is derived from the identified RpoD binding site from *xylX* (Meisenzahl, Shapiro, and Jenal 1997) (Figure 10A).

*gsrN* expression from P1 was 15% higher, and expression from P2 50% lower than *gsrN* expressed from its native SigT-dependent promoter (Figure 10B). Expression of *gsrN* from P1, but not P2, rescued the Δ*sigT* peroxide survival defect (Figure 10C). We conclude that *gsrN* is the major genetic determinant of hydrogen peroxide survival regulated by the GSR under these conditions. The differences in RpoD expression and susceptibility to hydrogen peroxide is consistent with the dose dependent protection by GsrN as seen in Figure 9C. These data demonstrate that a threshold level of *gsrN* expression is required to protect the cell from hydrogen peroxide.

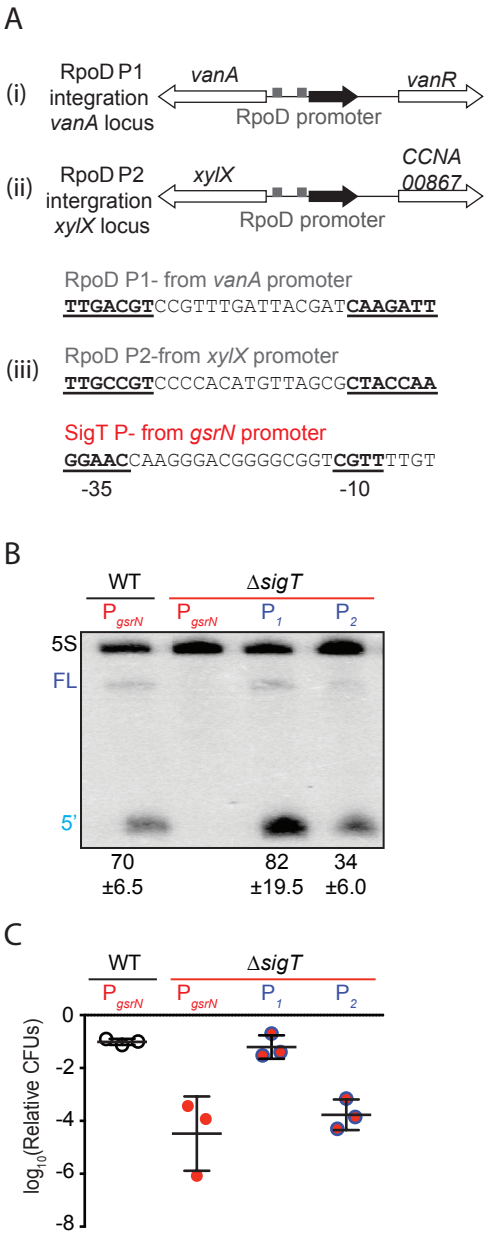


Figure 10. *gsrN* expression is sufficient to rescue the peroxide survival deficient of Δ*sigT*.
(A) Strains in which *gsrN* expression was driven from one of two distinct RpoD-dependent promoters integrated in the chromosome. (i) The RpoD P1 promoter was taken from the predicted RpoD-binding site directly upstream of the *vanA* transcriptional start site (Thanbichler, Iniesta, and Shapiro 2007) and directly replaced the consensus SigT-binding site of *gsrN*; (ii) RpoD P2 promoter was taken from the predicted RpoD-binding site upstream of *xylX (Meisenzahl, Shapiro, and Jenal 1997)* and directly replaced the consensus SigT-binding site of *gsrN* (iii) Sequence of the RpoD and SigT binding sites used to drive expression of *gsrN*.
(B) Northern blot of total RNA isolated from WT and Δ*sigT* strains expressing *gsrN* from its native promoter ($P_{sigT}$) or from two constitutive $\sigma^{RpoD}$ promoters ($P_1$ or $P_2$); probed with $^{32}$P-labeled oligonucleotides specific for GsrN and 5S rRNA as a loading control. 5S label refers to 5S rRNA, FL refers to full-length GsrN, and 5' refers to the 5'isoform of GsrN. Quantified values are mean ± SD of normalized signal, n=3 independent replicates.
(C) Relative survival of strains in (B) treated with 0.2 mM hydrogen peroxide for 1 hour normalized as in (A). Mean ± SD from 3 independent experiments (points) is presented as bars.

### 3.4. 5' portion of GsrN is necessary and sufficient for peroxide stress survival

Given the presence of two major isoforms of GsrN, the full-length and 5' portion as discussed in Chapter 2.3.II, we decided to test the function of the 5' half of GsrN, we integrated a *gsrN* allele that contains only the first 58 nucleotides (Δ59-106), and lacks the transcriptional terminator (*gsrN*Δ3') into the *vanA* locus (Figure 11A). This short *gsrN* allele complemented the Δ*gsrN* peroxide survival defect (Figure 11B). The *gsrN*Δ3' allele produced a 5' isoform that was comparable in size and concentration to the wild-type 5' *gsrN* isoform. Since the transcriptional terminator of *gsrN* was removed, we also observed a run-on ~200nt transcript from *gsrN*Δ3' (Figure 11C).

To test the necessity of the 5' portion of GsrN in peroxide stress survival, we deleted nucleotides 10-50 from *gsrN* at its native locus (Figure 11D). The *gsrN*Δ5' strain had a peroxide viability defect that was equivalent to Δ*gsrN*. Ectopic expression of either full-length *gsrN* or *gsrN*Δ3' in the *gsrN*Δ5' strain complemented its peroxide survival defect (Figure 11E).

Given the importance of the 5' half of GsrN and the presence of two isoforms, it is of interest to investigate isoform production. The 5' isoform of GsrN could be produced from two distinct processes: intrinsic termination or endonucleolytic processing. The precedence of sRNAs coming in multiple isoforms has been observed in RprA and ArcZ; however, these sRNA produce more stable 3' isoforms that arise from an endonucleolytic-processing event (Papenfort et al. 2009). Moreover, the 3' isoform of RprA does not seem to influence *rpoS* mRNA translation initiation and potentially regulates a subset of mRNAs in RprA's second identified base-pairing region (Papenfort et al. 2015).

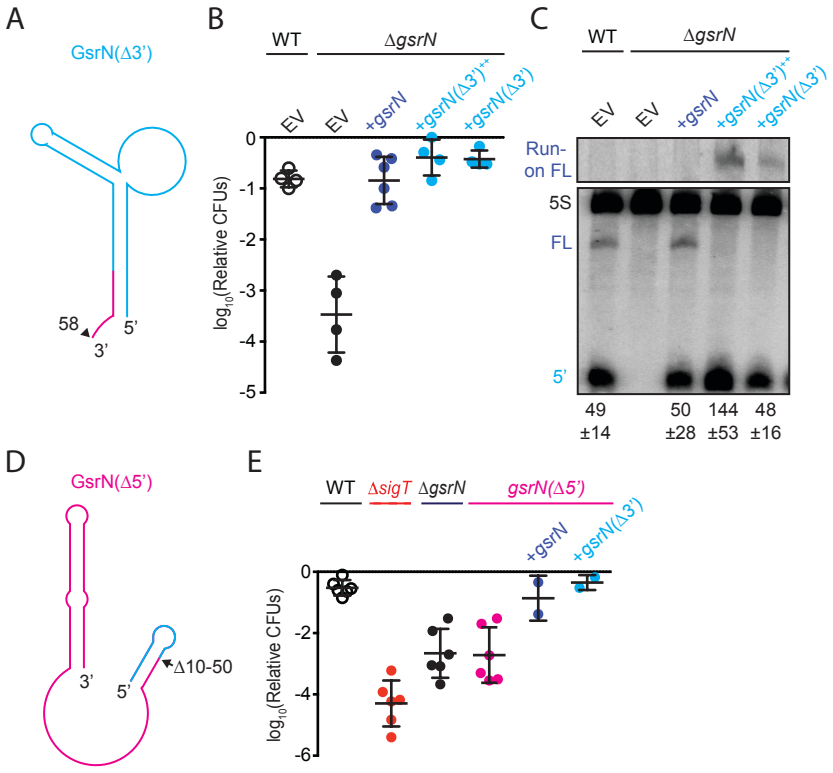


Figure 11. 5' portion of GsrN is necessary and sufficient for peroxide survival.
(A) Schematic diagram of GsrN(Δ3'), which lacks nucleotides 59-106, which includes GsrN's intrinsic terminator.
(B) Relative survival of strains treated with 0.2 mM hydrogen peroxide for 1 hour. WT and Δ*gsrN* strains carry empty integrated plasmids (EV), integrated plasmids harboring full-length *gsrN*, *gsrN*(Δ3'), or multiple copies of *gsrN*(Δ3') (labeled *gsrN*(Δ3')$^{++}$). Bars represent mean ± SD from 4 independent experiments (points).
(C) Northern blot of total RNA from strains in panel (B) harvested during exponential growth phase. Blots were hybridized with probes complementary to the 5' end of GsrN and 5S rRNA. Mean ± SD of total GsrN signal from 3 independent samples.
(D) Schematic diagram of GsrN(Δ5'), which lacks nucleotides 10-50, but contains GsrN's intrinsic terminator.
(E) Relative survival of strains treated with 0.2 mM hydrogen peroxide for 1 hour. Genetic backgrounds are indicated above the line; the GsrN(Δ5') strain was complemented with either *gsrN* (dark blue) or GsrN(Δ5') (cyan). Bars represent mean ± SD from at least 2 independent experiments (points).

## Chapter 4: Biochemical characterization of GsrN in *Caulobacter crescentus*

Direct mRNA targets of sRNAs are challenging to identify (Sharma and Vogel 2009). To complement traditional RNA-seq approaches for sRNA identification, we developed a forward biochemical approach to identify molecular partners of GsrN. Our method inserted aptamer sequences of *Pseudomonas* phage7 hairpin (PP7hp) within *gsrN* in order affinity-purify protein and mRNA partners.

Our approach is similar sRNA studies that studies implemented the MS2-aptamer, taken from the bacteriophage MS2 (Corcoran et al. 2012, Said et al. 2009). However, efficient purification often requires multiple MS2 aptamer insertions for efficient study (Lalaouna and Masse 2015). Moreover, multiple MS2 aptamers would often disrupt the native function of the sRNAs. Lastly, the MS2-aptamer system is more sensitive to temperature, buffer ionic strength, and pH than the analogous bacteriophage aptamer system, PP7 (Carey, Lowary, and Uhlenbeck 1983, Lim, Downey, and Peabody 2001, Lim and Peabody 2002, Hogg and Collins 2007).

Based on the shortcomings of the MS2 approach, we wanted to design the PP7-affinity purification of GsrN with some specific properties. First, we wanted to minimize gratuitous insertion and try to achieve purification with one insertion. Second, we wanted to screen for insertions that did not disrupt the function of GsrN. Based on the importance of the 5' portion of GsrN in hydrogen peroxide survival, we sought to understand the properties of the 5'isoform to better inform the insertion of PP7hp sequences. Specifically, what leads to the formation of the 5'isoform and what is the exact sequence of the 5' isoform?

## 4.1. GsrN is endonucleolytically processed

The short isoform of GsrN could arise through two biological processes: alternative transcriptional termination of *gsrN* or endonucleolytic processing of full-length GsrN. To test these two possibilities, we inhibited transcription with rifampicin, and monitored levels of both GsrN isoforms over time (Wehrli 1983). If alternative transcriptional termination produces the 5' isoform, we would expect to see a constant decay of both isoforms. If endonucleolytic processing gave rise to the 5' isoform, we would expect to see accumulation of 5'isoform as the full-length GsrN decays.

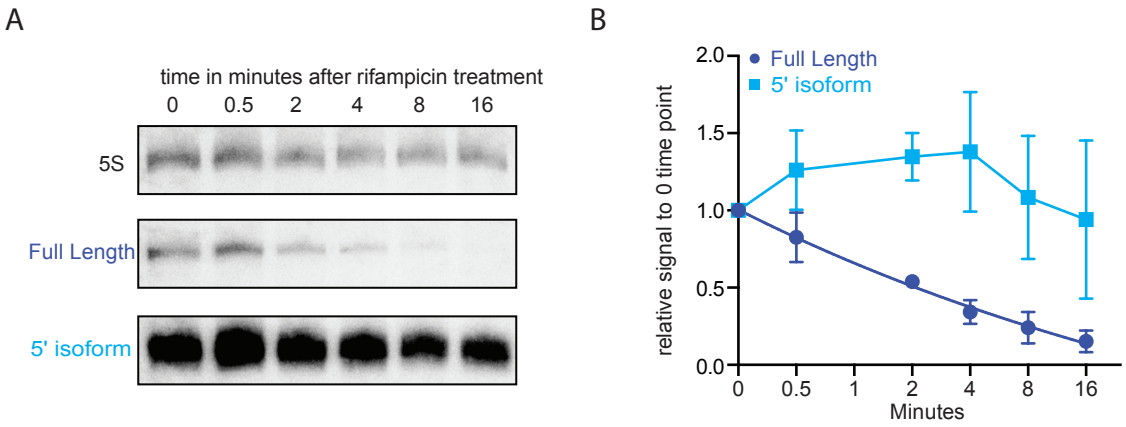


Figure 12. GsrN decay kinetics is consistent with the endonucleolytic cleavage model for isoform formation.
(A) Northern blot of total RNA extracted from wild type *Caulobacter* cells in exponential phase ($OD_{660} \approx 0.2$-$0.25$) 0 to 16 minutes after treatment with 10 µg/mL rifampicin (final concentration). Bands for full-length GsrN, 5' GsrN isoform, and 5S RNA loading control are shown.
(B) Quantification of blots from (A) of full-length GsrN and 5' GsrN isoform normalized to 5S rRNA levels in each lane. Signal at each time point is normalized relative to the zero minute time point. Data represent mean ± SD from three independent biological replicates.

Full-length GsrN decayed exponentially with a half-life of ~105 seconds (Figure 12). The 5' isoform increased in abundance for several minutes after treatment, concomitant with the decay of the full-length product. This observation is consistent with a model in which the 5' isoform

arises from the endonucleolytic cleavage of the full-length product. This is also seen in ArcZ when *E. coli* is treated with rifampicin (Vogel et al. 2003).

To better inform our PP7hp insertions, we wanted to identify the endonucleolytic cleavage sites of GsrN. Thus, we conducted primer extension assays from the underlined sites in Figure 13E. Extension from an oligo complementary to the 5' portion of GsrN confirmed the annotated transcriptional start site (Figure 13C). Extension from the 3' portion identified two internal 5' ends (Figure 13D). The positions of these internal 5' ends are consistent with two small bands observed on Northern blots of high concentrations of total RNA hybridized with the 3' probe (Figure 7A). The terminus around C53 corresponds to a potential endonucleolytic cleavage site that would generate the abundant stable 5' isoform (Figure 13B).

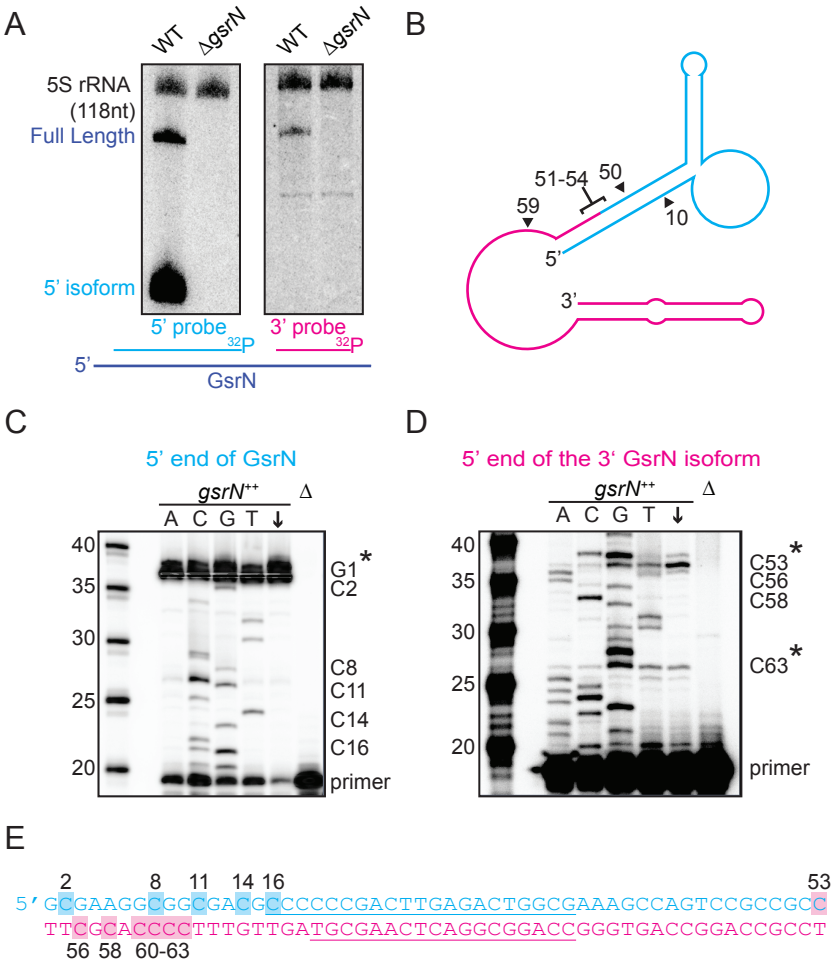


Figure 13. GsrN site of processing occurs at the 51st to 54th nucleotide.
(A) Northern blots of total RNA from wild-type and Δ*gsrN* cells hybridized with probes complementary to the 5'end (left) or 3' end (right) of GsrN, and to 5S rRNA as a loading control.
(B) Predicted secondary structure of full-length GsrN using RNA-specific folding parameters (Andronescu et al. 2007). Nucleotide positions labeled with arrows. Cyan indicates the 5' end of GsrN determined by primer extension. Pink represents the 3' end. The nucleotide residue reference numbers labeled in this diagram provides context to the mutants in Figure 4.
(C) Primer extension from total RNA extracted from *gsrN*$^{++}$ and Δ*gsrN* (negative control) cultures ($OD_{660} \approx 1.0$, a condition in which GsrN levels were observed to be the highest). Sequence was generated from a radiolabeled oligo anti-sense to the underlined cyan sequence in (E). Sanger sequencing control lanes A, C, G, and T mark the respective ddNTP added to that reaction to generate nucleotide specific stops. "C" labels on the right of the gel indicate mapped positions from the "G" lane. Arrow indicates lane without ddNTPs. Asterisk indicates positions of 5' termini.
(D) Primer extension from RNA samples as in (C). Sequence was extended from a radiolabeled oligo anti-sense to the underlined pink sequence in (E).
(E) GsrN coding sequence. Cyan and pink indicate the predicted 5' and 3' isoforms, respectively. Primers binding sites used for primer extension in (C) and (D) are underlined. Highlighted C positions correspond to ddGTP stops in the "G" extensions.

## 4.2. GsrN associates with several RNAs in *Caulobacter crescentus*

We designed the PP7 system with the goal of purifying tagged GsrN-PP7hp with its interacting partners from *C. crescentus* lysates by affinity chromatography Figure 15A. We first designed sequence insertions at the extreme 5' and 3' ends of *gsrN* akin to N-terminal and C-terminal insertions of affinity-tags in proteins. However, attempts to append PP7 to the 3' end resulted in premature termination of the *gsrN* transcription. Interestingly GsrN-PP7hp alleles tagged at the 5' end did not complement the Δ*gsrN* peroxide survival defect and showed compromised levels of the 5'isoform compared to its full-length levels (Figure 14B-C). This led us to examine several internal nucleotide positions: 37, 54, 59, 67, and 93 (Figure 14A).

We designed each insertion with a specific function in mind. The 37$^{th}$ insertion was placed in a predicted hairpin loop in the 5' end of GsrN. The 54$^{th}$ insertion was placed at the end of the endonucleolytic cleavage site in hopes of producing an extreme 3' insertion in the 5'isoform. The 59$^{th}$ insertion was placed right after the endonucleolytic cleavage site in hopes of producing an extreme 5' insertion in the 3'isoform. The 67$^{th}$ insertion was placed to be right before the intrinsic terminator hairpin of GsrN. The 93$^{rd}$ insertion was placed in the loop of the intrinsic terminator hairpin.

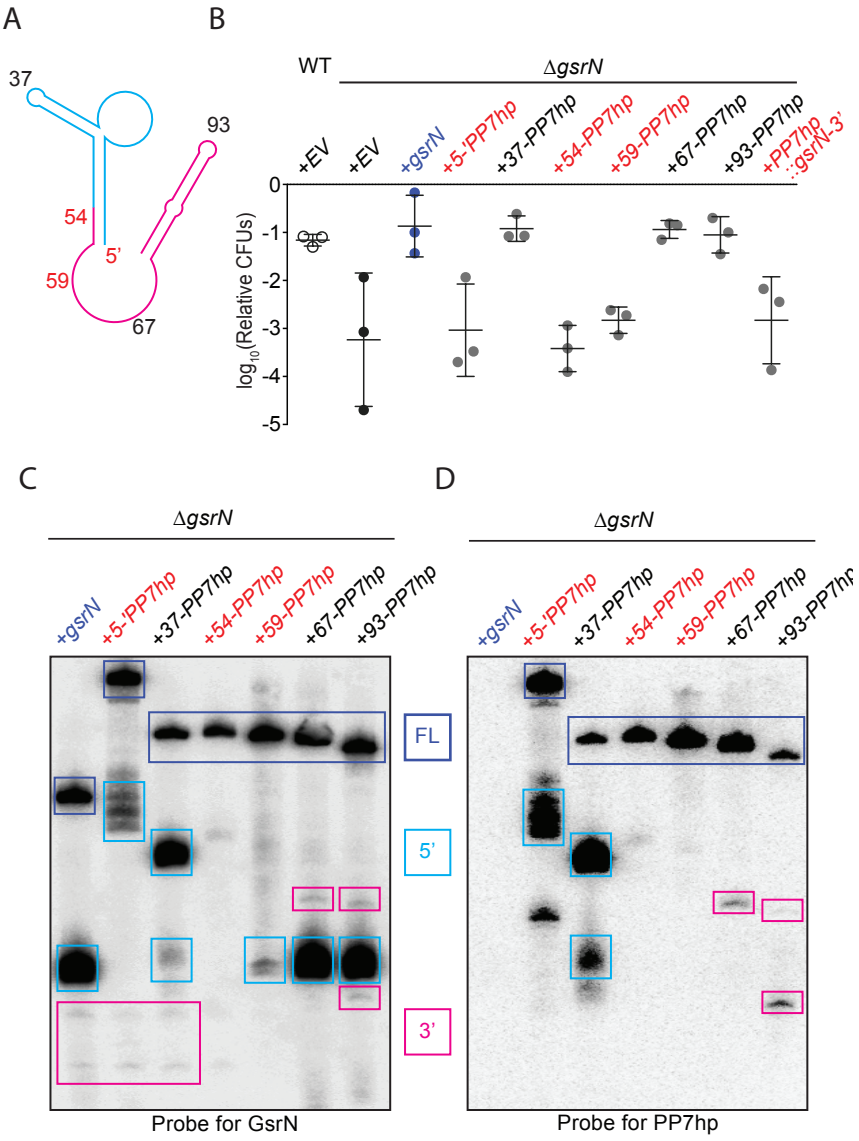


Figure 14 Mutational analysis of PP7hp insertions into GsrN shows a correlation between isoform formation and peroxide survival.
(A) Predicted GsrN secondary structure diagram from mFold (Zuker 2003). Cyan and pink represent the 5' and 3' products, respectively, determined by primer extension and northern blot analyses (Figure 13). Numbered positions along the secondary structure indicate where PP7 RNA hairpin sequences (PP7hp) were inserted into *gsrN*.
(B) Wild type, Δ*gsrN-EV*, and Δ*gsrN*+*gsrN*-PP7hp strains were subjected to hydrogen peroxide, diluted, and tittered as in previously described. Empty vector (EV) strains carry the integrating plasmid pMT552. The nucleotide position of each PP7hp insertion in *gsrN* is marked above each bar. Data represent mean ± SD of three independent trials.
(C) Northern blots of total RNA from stationary phase cultures ($OD_{660} \approx 1.0$, a condition in which GsrN levels were observed to be the highest) of Δ*gsrN* strains carrying *gsrN*-PP7hp fusions. Blots were probed with oligonucleotides complementary to both the 5' and 3' ends of GsrN. Blot is overexposed to reveal minor products. Purple boxes mark full length GsrN, cyan boxes mark 5' isoforms, and pink boxes mark 3' isoforms.
(D) Northern blots of same samples as in (C) ran in parallel but probed with oligonucleotides complementary to the PP7 hairpin sequence.

Given the results of our 5'PP7hp insertion, we functionally assessed the internal insertions abilities to complement Δ*gsrN* peroxide survival defect and the formation of the 5'isoform. GsrN-PP7hp alleles with insertion at nucleotide positions 54 or 59 did not complement the Δ*gsrN* peroxide survival defect and yielded lower steady-state levels of 5' isoform compared to wild type (Figure 14B-C). GsrN-PP7hp alleles with insertions at nucleotides 37, 67, and 93 restored peroxide resistance to Δ*gsrN* and produced more 5' isoform than non-complementing GsrN-PP7 constructs (Figure 14B-C). Lastly, we tested the ability for each isoform to be affinity purified using the experimental scheme depicted in Figure 15A (Figure 16A).

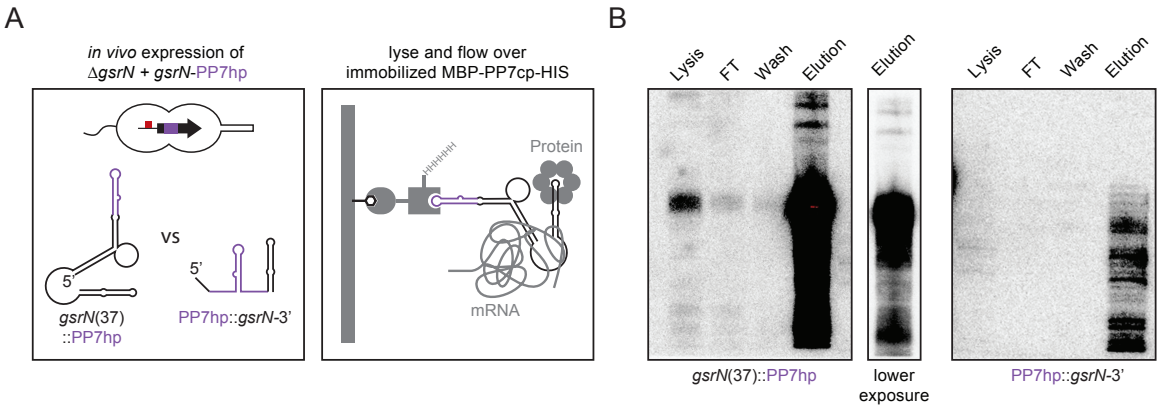


Figure 15. GsrN-PP7hp affinity co-purification scheme and proof-of-concept.
(A) GsrN-target co-purification strategy. GsrN(black)-PP7hp(purple) fusions were expressed in a Δ*gsrN* background. PP7 RNA hairpin (PP7hp) inserted at nucleotide 37 (*gsrN*(37)::PP7hp) was used as the bait. PP7hp fused to the 3' hairpin of *gsrN* (PP7hp::*gsrN*-3') served as a negative control. Stationary phase cultures expressing these constructs were lysed and immediately flowed over an amylose resin column containing immobilized PP7hp binding protein (MBP-PP7cp-His).
(B) GsrN-PP7hp purification from strains bearing *gsrN*(37)::PP7hp (left) and PP7hp::*gsrN*-3' (right) was monitored by Northern Blot with probes complementary to 5' end of GsrN and PP7hp, respectively. Lysate, flow through (FT), buffer wash, and elution fractions are blotted. Approximately 1µg RNA was loaded per lane, except for buffer wash (insufficient amount of total RNA).

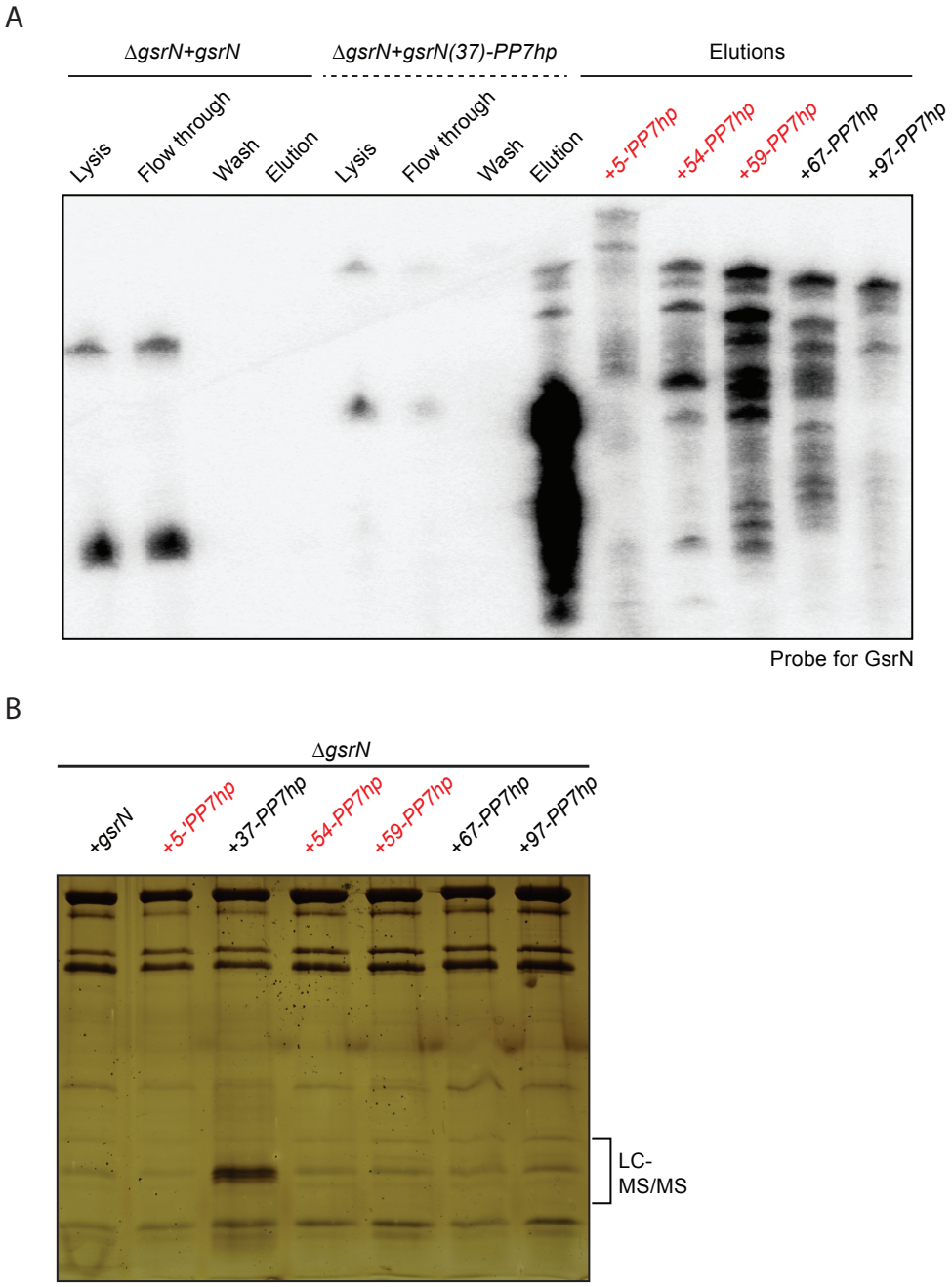


Figure 16. No significant protein profile differences between GsrN-PP7hp affinity-purifications. (A) Northern blot of RNA samples through purifications of several GsrN::PP7hp constructs. First four lanes demonstrate that untagged GsrN cannot bind and amylose resin with MBP-PP7cp-HIS, while the subsequent four lanes shows the purification of GsrN(37)::PP7hp. The elutions of the other GsrN::PP7hp constructs were able to bind and elute with MBP-PP7cp-HIS. Blots were probed with GsrN 5' and 3' specific probes.
(B) Silver-stained SDS acrylamide gel of the elutions in (A). Bands were cut, digested, extracted, and identified by LC-MS/MS

The PP7hp aptamer inserted at GsrN nucleotide 37 (GsrN(37)::PP7hp) was selected as the bait to identify molecular partners that co-purify with GsrN. We selected the GsrN(37)::PP7hp since our previous experiments showed the 5' end of GsrN is necessary for function and we did not want to rule out potential binding partners that could potentially only elute with just the 5' isoform. The pull-down fraction was compared to a negative control pull-down from cells expressing PP7hp fused to the last 50 nucleotides of GsrN including its intrinsic terminator (PP7hp::GsrN-3') (Figure 15A). Northern blots demonstrated GsrN-PP7hp fusion transcripts were enriched in our purification (Figure 15B). Electrophoretic separation of the eluate followed by silver staining revealed a potential significant protein differences between GsrN(37)::PP7hp and the other pull-downs (Figure 16B). However, upon LC-MS/MS analysis there was no significant difference in peptide signal. This band could very likely correspond to the amount of GsrN(37)::PP7hp that is purified versus the other constructs. However, we could also adjust our conditions in which we performed the pull-down. We identified and quantified co-eluting RNAs by RNA-seq.

We applied IntaRNA 2.0 (Mann, Wright, and Backofen 2017b) to identify potential binding sites between GsrN and the enriched co-purifying RNAs. Of the 67 analyzed enriched genes and regions, 32 of the predicted RNA-RNA interactions involved the cytosine-rich 5' loop in the predicted secondary structure of GsrN (Figure 17C). Specifically 31 of these targets contained G-rich sequences. A sequence logo (Crooks et al. 2004) of the predicted target mRNA binding sites is enriched with guanosines (Figure 17D), consistent with a model in which 6 tandem cytosines in the 5' loop of GsrN determine target mRNA recognition. 27 of the predicted RNA-RNA interactions involved the 3' exposed region of GsrN. The remaining 8 enriched genes and regions did not have a significant binding site prediction with GsrN.

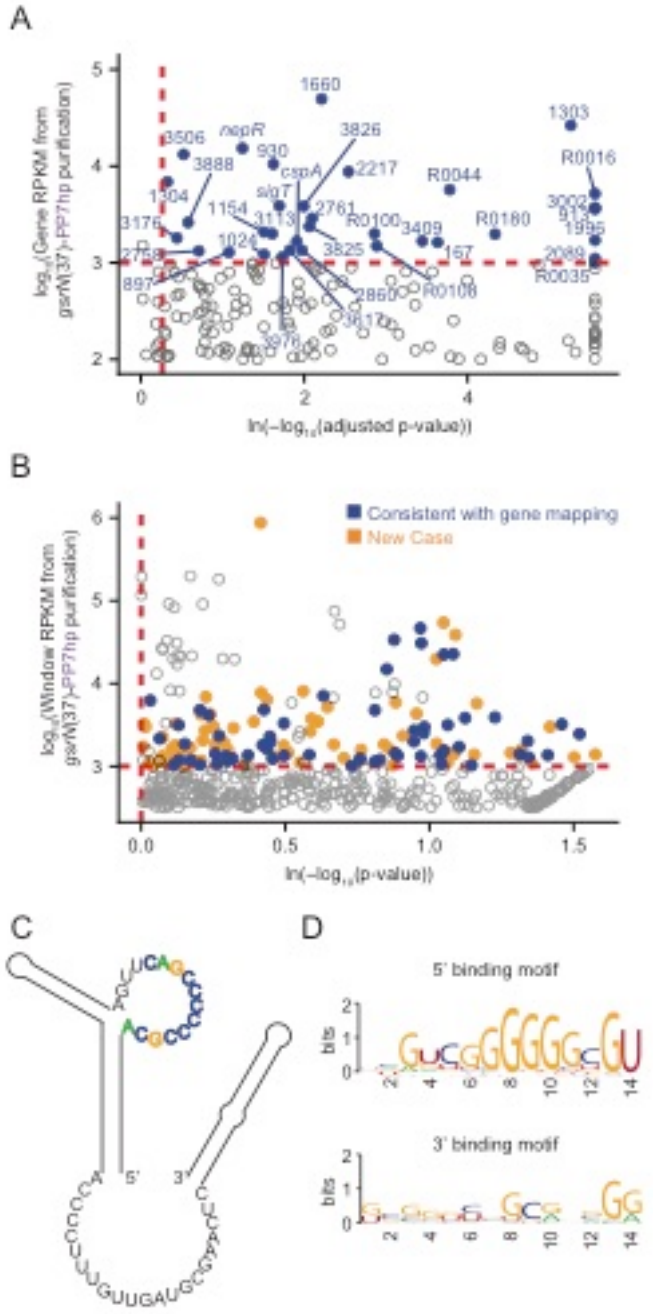


Figure 17. RNA-seq results show 67 co-eluting RNAs are enriched in the GsrN(37)::PP7hp pull-down.
(A) Annotation-based analysis of transcripts that co-purify with *gsrN*(37)::PP7hp. $Log_{10}$ reads per kilobase per million reads (RPKM) is plotted against the ln(-$log_{10}$(false discovery rate corrected p-value)). Dashed red lines mark the enrichment co-purification thresholds. Genes enriched in the *gsrN*(37)::PP7hp purification compared to PP7hp::*gsrN*-3' are blue; labels correspond to gene names or *C. crescentus* strain NA1000 CCNA GenBank locus ID. Data represent triplicate purifications of *gsrN*(37)::PP7hp and duplicate PP7hp::3'GsrN control purifications. Log adjusted p-values of zero are plotted as $10^{-260}$.
(B) Sliding-window analysis of transcripts that co-purify with *gsrN*(37)::PP7hp. Points represent 25-bp genome windows. RPKM values for each window were estimated by EDGE-pro; p-values were estimated by DESeq. Windows that map to genes identified in (A) are blue. Orange indicates windows with significant and highly abundant differences in mapped reads between *gsrN*(37)::PP7hp fractions and the PP7hp::*gsrN*-3' negative control fractions. Dashed red lines denote cut-off value for windows enriched in the *gsrN*(37)::PP7hp fractions. Grey points within the dashed red lines are signal that mapped to rRNA.
(C) Predicted loops in GsrN accessible for mRNA target base pairing are emphasized in colored text.
(D) A putative mRNA target site complementary to a cytosine-rich tract in the 5' GsrN loop is represented as a sequence logo. Similar logo was generated for the sequences that mapped to the 2nd exposed region of GsrN. Logo was generated from IntaRNA 2.0.2 predicted GsrN-binding sites in transcripts enriched in the *gsrN*(37)::PP7hp pull-down. 5' binding motif is present in 32 of the transcripts identified in (A) and (B) and 3' binding motif is present in 27 of the transcripts identified in (A) and (B).

We employed two approaches to identify RNAs enriched in GsrN(37)::PP7hp fractions relative to the negative control fractions. A conventional RNA-seq pipeline (Tjaden 2015) quantified mapped reads within annotated gene boundaries as a first pass (Figure 17A). To capture reads in non-coding and unannotated regions, and to analyze reads unevenly distributed across genes, we also developed a sliding window analysis approach. Specifically, we organized the Caulobacter genome into 25 base-pair windows and quantified mapped reads in each window using the EDGE-pro/DESeq pipeline (Anders and Huber 2010, Magoc, Wood, and Salzberg 2013). Together these two quantification strategies identified several mRNA, sRNAs, and untranslated regions enriched in the GsrN(37)::PP7hp pull-down fraction (Figure 17B).

Transcripts enriched in the GsrN(37)::PP7hp fraction encode proteins involved in proteolysis during envelope stress, enzymes required for envelope biogenesis, cofactor and nucleotide anabolic enzymes, and transport proteins (Table 1). We observed significant enrichment of rRNA in the GsrN(37)::PP7hp fractions; the functional significance of this signal is not known (grey points inside red cut-off Figure 17B). Interesting, *sigT* and its anti-σ factor, *nepR*, were also enriched in the GsrN(37)::PP7hp fraction, despite our lack of evidence for regulation of SigT/NepR by GsrN (Figure 8D). This discrepancy could arise from differences in conditions between the assays, the singularity of our SigT-dependent reporter, or subtle changes in SigT-levels that cannot be captured by our activity reporter assay. Lastly, *katG*, which encodes the sole catalase-peroxidase in the Caulobacter genome (Steinman, Fareed, and Weinstein 1997b), was the most compelling target identified from our purification in terms of peroxide survival.

| Table 1. GsrN(37)::PP7hp enriched RNAs from both annotation and annotation free analysis. Table 1 continued on page 45 and 46 | | | | | |
|---|---|---|---|---|---|
| Gene Locus ID | Gene Name | $log_2$ Fold | Identification Method | Regions(s) | Description |
| CCNA_00167 | - | 4.56, 6.95 | Rockhopper, Sliding window | 179311-180120 (+), 179500-179550 (+, I, S) | metallophosphatase family protein |
| CCNA_00416 | - | 7.2 | Sliding window | 429625-429725 (-, I, S) | conserved hypothetical membrane protein |
| CCNA_00587 | - | 4.87 | Sliding window | 616250-616300 (+, I, S) | alpha/beta hydrolase family protein |
| CCNA_00882 | - | 4.61 | Sliding window | 962875-962925 (-, U, S) | hypothetical protein |
| CCNA_00894 | - | 4.29 | Sliding window | 974800-974850 (+, I, S) | 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase |
| CCNA_00897 | - | 3.2 | Rockhopper | 976013-976177 (+) | hypothetical protein |
| CCNA_00913 | - | 7.64, 7.80 | Rockhopper, Sliding window | 993033-993209 (-), 993175-993225 (-, I, S) | hypothetical protein |
| CCNA_00930 | - | 3.72, 6.98, 3.81 | Rockhopper, Sliding window, Sliding window | 1006253-1006870 (+), 1006275-1006425 (+, I, S), 1006475-1006650 (+, I, S) | riboflavin synthase alpha chain |
| CCNA_01024 | - | 3.32 | Rockhopper | 1111617-1112111 (-) | hypothetical protein |
| CCNA_01058 | - | 5.81 | Sliding window | 1159075-1159125 (-, D, S) | helix-turn-helix transcriptional regulator |
| CCNA_01154 | - | 3.45, 6.81 | Rockhopper, Sliding window | 1257902-1258591 (+), 1257975-1258025 (+, I, S) | conserved hypothetical protein |
| CCNA_01303 | - | 5.87, 5.30, 8.22 | Rockhopper, Sliding window, Sliding window | 1430061-1430900 (+), 1430550-1430625 (+, I, S), 1430650-1430725 (+, I, S) | conserved hypothetical protein |
| CCNA_01304 | - | 2.9 | Rockhopper | 1431129-1431329 (+) | hypothetical protein |
| CCNA_01335 | - | 2.99 | Sliding window | 1448600-1448650 (-, I, S) | ABC-type multidrug transport system, ATPase component |
| CCNA_01344 | - | 4.62 | Sliding window | 1458550-1458725 (+, I, S) | conserved hypothetical protein |
| CCNA_01584 | - | 3.14 | Sliding window | 1699675-1699725 (+, I, A) | multimodular transpeptidase-transglycosylase PBP 1A |
| CCNA_01660 | - | 4.41, 6.21 | Rockhopper, Sliding window | 1781219-1781911 (-), 1781350-1781575 (-, I, S) | conserved hypothetical protein |
| CCNA_01966 | - | 11.3 | Sliding window | 2110225-2110275 (-, I, A) | vitamin B12-dependent ribonucleotide reductase |
| CCNA_01996 | - | 9.15, 8.86 | Rockhopper, Sliding window | 2142908-2143687 (-), 2143625-2143700 (-, I, S) | undecaprenyl pyrophosphate synthetase |
| CCNA_02034 | - | 7.24 | Sliding window | 2178500-2178550 (+, I, S) | luciferase-like monooxygenase |
| CCNA_02064 | lpxC | 3.6 | Sliding window | 2215450-2215550 (-, I, S) | UDP-3-O-(3-hydroxymyristoyl) N-acetylglucosamine deacetylase |
| CCNA_02089 | - | 8.52 | Rockhopper | 2237967-2238341 (-) | hypothetical protein |
| CCNA_02217 | - | 4.02 | Rockhopper | 2364081-2364383 (-) | hypothetical protein |

| Table 1 continued. | | | | | |
|---|---|---|---|---|---|
| CCNA_02286 | - | 3.26 | Sliding window | 2435450-2435500 (-, I, S) | hypothetical protein |
| CCNA_02595 | - | 6.85 | Sliding window | 2743525-2743625 (-, U, S) | Zn finger TFIIB-family transcription factor |
| CCNA_02758 | - | 2.93 | Rockhopper | 2921763-2922152 (+) | hypothetical protein |
| CCNA_02761 | - | 3.65 | Rockhopper | 2923673-2923918 (+) | hypothetical protein |
| CCNA_02846 | - | 5.44, 8.60 | Sliding window, Sliding window | 3000100-3000175 (-, I, S), 2999225-2999275 (-, I, S) | DegP/HtrA-family serine protease |
| CCNA_02860 | - | 3.70, 4.78 | Rockhopper, Sliding window | 3012116-3013060 (-), 3012500-3012550 (-, I, S) | DnaJ-class molecular chaperone |
| CCNA_02975 | - | 6.34 | Sliding window | 3130300-3130375 (-, I, A) | excinuclease ABC subunit C |
| CCNA_02987 | - | 7.26 | Sliding window | 3142700-3142800 (-, I, A) | hypothetical protein |
| CCNA_02997 | cspA | 3.61 | Rockhopper | 3152607-3152816 (-) | cold shock protein CspA |
| CCNA_03002 | - | 6.03, 4.48 | Rockhopper, Sliding window | 3155705-3156322 (-), 3155750-3155800 (-, I, S) | CDP-diacylglycerol--glycerol-3-phosphate 3 phosphatidyltransferase |
| CCNA_03105 | - | 9.15 | Sliding window | 3255775-3255850 (-, I, S) | DnaJ domain protein |
| CCNA_03113 | - | 3.50, 5.40 | Rockhopper, Sliding window | 3263780-3264499 (-), 3264400-3264450 (-, I, S) | membrane-associated phospholipid phosphatase |
| CCNA_03138 | katG | 3.35 | Sliding window | 3286000-3286050 (+, I, S) | peroxidase/catalase katG |
| CCNA_03176 | - | 2.83 | Rockhopper | 3335155-3335445 (-) | nucleotidyltransferase |
| CCNA_03338 | tolB | 5.27 | Sliding window | 3519425-3519475 (-, I, S) | TolB protein |
| CCNA_03409 | - | 4.46, 5.44 | Rockhopper, Sliding window | 3576740-3577696 (-), 3577550-3577600 (-, I, S) | alpha/beta hydrolase family protein |
| CCNA_03506 | - | 3.27, 4.10 | Rockhopper, Sliding window | 3664090-3664677 (+), 3664100-3664175 (+, I, S) | putative transcriptional regulator |
| CCNA_03589 | sigT | 3.58 | Rockhopper | 3743953-3744558 (-) | RNA polymerase EcfG family sigma factor sigT |
| CCNA_03590 | nepR | 3.43, 3.50 | Rockhopper, Sliding window | 3744561-3744746 (-), 3744675-3744725 (-, I, S) | anti-sigma factor NepR |
| CCNA_03590, CCNA_03589 | nepR, sigT | 4.42 | Sliding window | 3744500-3744575 (-, O, S) | anti-sigma factor NepR, RNA polymerase EcfG family sigma factor sigT |
| CCNA_03617 | - | 3.5 | Rockhopper | 3772262-3772717 (+) | Copper(I)-binding protein |
| CCNA_03618,CCNA_03617 | -,- | 6.99 | Sliding window | 3772700-3772750 (+, O, S) | SCO1/SenC family protein, Copper(I)-binding protein |
| CCNA_03681 | - | 5.11 | Sliding window | 3843700-3843750 (-, U, S) | ABC transporter ATP-binding protein |
| CCNA_03825 | - | 3.63 | Rockhopper | 3991412-3991774 (-) | hypothetical protein |

| Table 1 continued. | | | | | |
|---|---|---|---|---|---|
| CCNA_03825,CCNA_03826 | -,- | 8.01 | Sliding window | 3991750-3991825 (-, O, S) | hypothetical protein, conserved hypothetical protein |
| CCNA_03826 | - | 3.71 | Rockhopper | 3991771-3992325 (-) | conserved hypothetical protein |
| CCNA_03888 | - | 2.99 | Rockhopper | 761965-762324 (+) | conserved hypothetical protein |
| CCNA_03976 | - | 3.57 | Rockhopper | 2923462-2923683 (+) | hypothetical protein |
| CCNA_R0016 | - | 8.53 | Rockhopper | 844332-844401 (+) | small non-coding RNA |
| CCNA_R0035 | - | 6.64 | Rockhopper | 1549367-1549443 (+) | tRNA-Pro |
| CCNA_R0044 | - | 4.82 | Rockhopper | 2059848-2059942 (-) | complex medium expressed sRNA |
| CCNA_R0061 | - | 4.65 | Sliding window | 2800475-2800525 (-, I, S) | RNase P RNA |
| CCNA_R0089 | - | 3.3 | Sliding window | 3874375-3874425 (+, U, S) | tRNA-Ala |
| CCNA_R0100 | - | 4.4 | Rockhopper | 165492-165575 (+) | small non-coding RNA |
| CCNA_R0108 | - | 4.27 | Rockhopper | 472905-472973 (+) | small non-coding RNA |
| CCNA_R0180 | - | 5.14 | Rockhopper | 3266851-3266937 (-) | small non-coding RNA |

Gene Locus ID: GenBank locus ID
Gene Name: if available
$\log_2$Fold: calculated fold change of the given region
Identification Method: refers to what strategy identified the enriched gene in the PP7hp affinity purification RNA-Seq
Region(s): the region and strand used to calculate the $\log_2$Fold metric. Additionally for the sliding window analysis additional information is provided. First letter indicates the relative position of the region indicated to the annotated gene coordinates. Briefly: I-internal, U-upstream, D-downstram. Second letter indicates the direction in which th reads mapped. Briefly: S-sense, A-anti-sense.
Description: is the product description of the given gene(s)

## Chapter 5: GsrN regulates *katG* mRNA

To verify our pull-down approach, we wanted to follow up on our enrichment of the *katG* mRNA. This investigation would not only validate our approach to define the molecular partners of GsrN, but also begin to unravel how the GSR allows for survival under peroxide exposure.

### 5.1. Introduction to *katG* in *Caulobacter crescentus*

KatG is a catalase-peroxidase scavenging enzyme that degrades $H_2O_2$ into $H_2O$ through the oxidation of an intermediate haem (Smulevich et al. 2006). Although there are several scavenging enzymes in bacteria that can perform $H_2O_2$ degradation (Imlay 2013), *C. crescentus* seems to have only one *gene* within its genome to have this activity, *katG* (Steinman, Fareed, and Weinstein 1997b).

In most gram-negative bacteria, *katG* is under the transcriptional regulation of the OxyR transcription factor (Imlay 2013). OxyR is able to sense concentrations of $H_2O_2$ through key cysteine residues, subsequently binding upstream of *katG*, and activate *katG* transcription (Christman, Storz, and Ames 1989, Italiani, Neto, et al. 2011). Moreover, this activation also occurs as cultures reach stationary growth.

We sought to test whether GsrN affected the expression of *katG*. Since *katG* is the sole gene in *C. crescentus* that can breakdown $H_2O_2$, it is compelling that the linear relationship between GsrN levels and survival to peroxide exposure could be due to a potential activation mechanism

of GsrN on *katG* mRNA. We would expect the final protein levels of KatG to be under substantial influence by GsrN.

## 5.2. GsrN binds to the 5' leader of *katG* mRNA

Detailed analysis of our forward biochemical approach on the *katG* locus showed reads mapping to the first 60 nucleotides of *katG.* This region includes the 5' leader sequence and the first several codons of the open reading frame, based on the experimentally determined transcriptional start site (TSS) of *katG* (Zhou et al. 2015b). Reads in the GsrN(37)::PP7hp pull-down fractions relative to the negative control PP7hp-GsrN(3'term) were enriched over 3-fold (Figure 18).

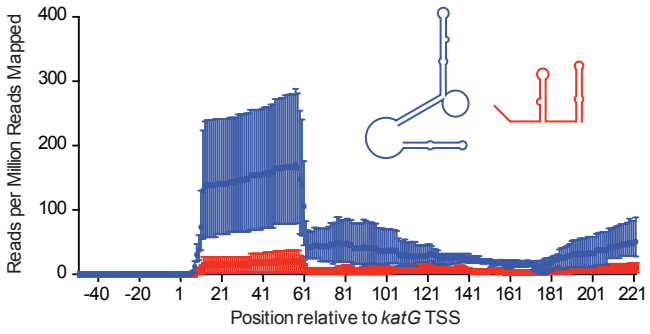


Figure 18. *katG* mRNA leader is enriched in RNA-seq reads of GsrN-PP7hp purifications. Density of reads that co-purified with *gsrN*(37)::PP7hp (blue) and PP7hp::*gsrN*-3' (red) and mapped to *katG*. Read density in each dataset represents read coverage at each nucleotide divided by the number of million reads mapped in that data set. Data represent mean ± SD of replicate purifications.

To verify this interaction, we created a variety of *katG* reporters that drove the production of the β-galactosidase enzyme (*lacZ*). Specifically, we sought to understand at what level of production, transcription and/or translation, that GsrN might influence KatG production. We first

constructed a transcriptional fusion of *katG* to rule out any indirect regulation of GsrN. GsrN did not effect *katG* transcription in exponential or stationary phases, or in the presence of peroxide as measured by a *katG-lacZ* transcriptional fusion (Figure 19A-C). We then constructed a strict translational *katG-lacZ* reporter, where the 5' UTR and the sequence of the first 50 codons of *katG* were N-terminally fused to *lacZ* under constitutively expressed from our P1 RpoD-dependent promoter. In both exponential and stationary phases, *katG-lacZ* activity is reduced in Δ*gsrN* and enhanced in $gsrN^{++}$ strains compared to wild type (Figure 19D&F). Hydrogen peroxide exposure did not affect *katG-lacZ* activity (Figure 19E). We conclude that GsrN enhances KatG protein expression, but not *katG* transcription.

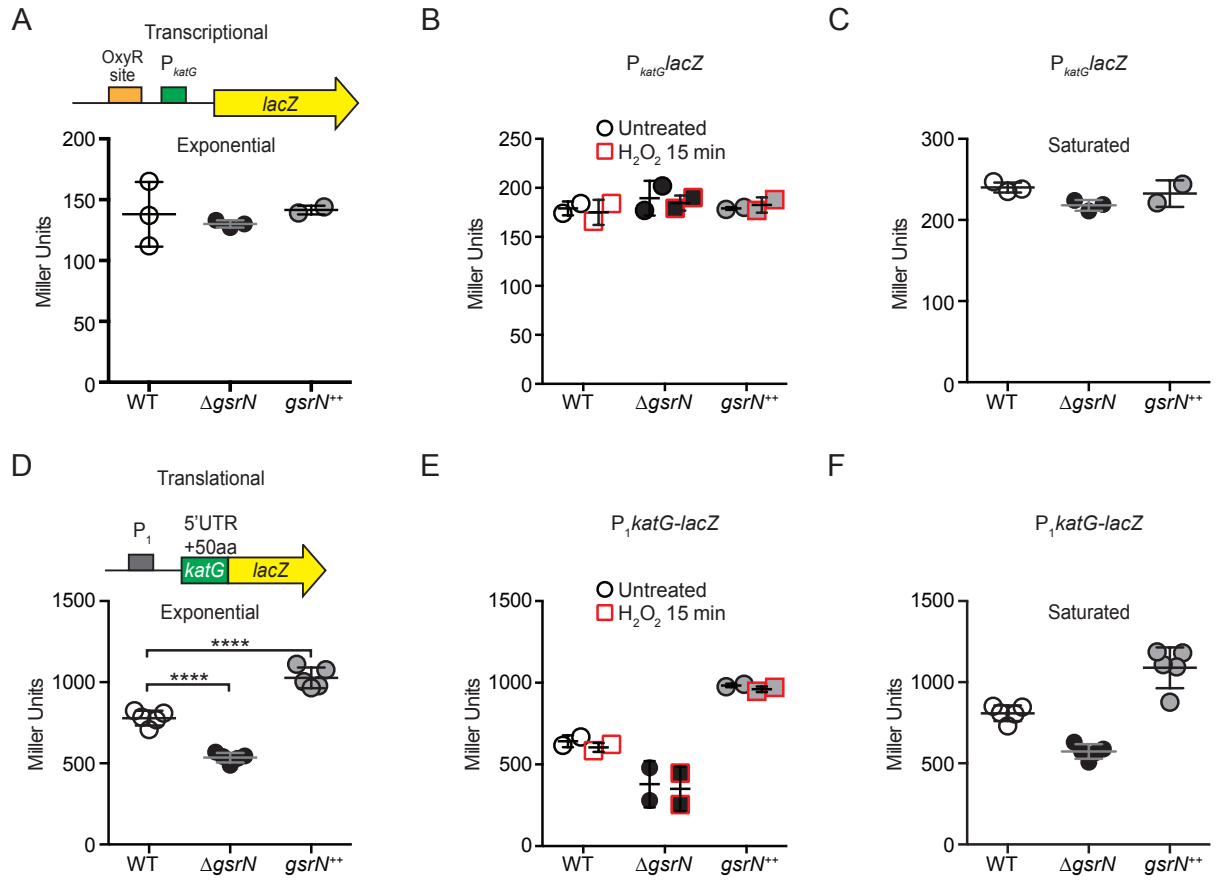


Figure 19. GsrN affects *katG* translation and does not affect *katG* transcription.
(A) *katG* transcriptional reporter construct contains the entire intergenic region upstream of *katG* fused to *lacZ* in pRKlac290. Transcription from this *katG* promoter ($P_{katG}$) reporter was assayed in wild type, Δ*gsrN,* and *gsrN*$^{++}$ backgrounds during exponential growth ($OD_{660} \approx 0.2$-$0.25$). Data represent mean ± SD of three independent trials.
(B) Activity from the *katG* transcriptional reporter with and without a 15-minute treatment with 0.2 mM hydrogen peroxide. This amount of time was chosen to sample peroxide effects and also ensure sufficient amount of signal could be obtained from Δ*gsrN* cultures. Cells were grown as in (A). Data represent mean ± SD of two independent trials.
(C) Activity from the *katG* transcriptional reporter in stationary phase cultures ($OD_{660} \approx 1.0$, a condition in which OxyR transcriptional activation w observed to be the highest (Italiani, da Silva Neto, et al. 2011)). Data represent mean ± SD of three independent trials.
(D) KatG translational reporter (top) assayed in exponentially growing cells (bottom). Reporter is constitutively expressed from the $P_{rpoD1}$ promoter. *katG* leader (1- 191 nt) region and the first 50 codons of *katG* are fused in-frame to *lacZ*. Mean ± SD β-galactosidase activity, measured in Miller Units, presented from five independent trials.

We then used this translational reporter to investigate a predicted binding interaction between the unpaired 5' loop of GsrN and a G-rich region at the 5' end of the *katG* transcript. Specifically, the first 7 nucleotides of *katG* mRNA are complementary to 7 nucleotides in the single-stranded 5' loop of GsrN, including 4 of the 6 cytosines (Figure 17C). We disrupted this predicted base pairing, mutating 5 of the 7 nucleotides in the putative *katG* target site and GsrN interaction loop. These mutations preserved GC-content, but reversed and swapped (RS) the interacting nucleotides (Figure 20A). We predicted that pairs of wild-type and RS mutant transcripts would not interact, while base pairing interactions would be restored between RS mutant pairs.

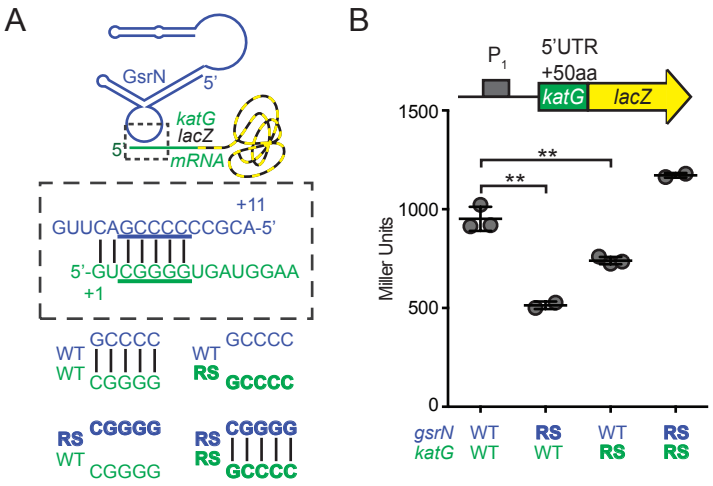


Figure 20. 5' C-rich loop of GsrN binds to the *katG-lacZ* translational reporter.
(A) Predicted interaction between GsrN (blue) and *katG* mRNA (green), with base-pairing shown in dashed box. Wild-type (WT) and reverse-swapped (RS) mutation combinations of the underlined bases are outlined below.
(B) Translation from *katG* and *katG-RS* reporters in Δ*gsrN* strains expressing 3*gsrN* (WT) or 3*gsrN(RS)* (RS). Measurements were taken from exponential phase cultures. Bars represent mean ± SD of at least two independent cultures (points) and p-value estimated by Student's t-test.

Mutating the predicted target site in the *katG* 5' leader ablated GsrN-dependent regulation of the *katG-lacZ* translational reporter; expression was reduced to a level similar to Δ*gsrN* with the katG-lacZ wild-type reporter (Figure 22). We further tested this interaction by assessing the

effect of the reverse-swapped *gsrN*(RS) allele on the expression of *katG-lacZ*. However, GsrN(RS) was unstable; total GsrN(RS) levels were ≈10-fold lower than wild-type GsrN (Figure 21B). To overcome GsrN(RS) instability, we inserted a plasmid with three tandem copies of *gsrN*(RS), 3*gsrN*(RS), into the *vanA* locus in a Δ*gsrN* background, which increased steady-state levels of GsrN(RS) approximately 4-fold (Figure 21B). *katG* target site or GsrN recognition loop mutations significantly reduced *katG-lacZ* expression (Student's t-test, $p=0.0026$ and $p=0.0046$, respectively). Compensatory RS mutations that restored base pairing between the *katG* target site and the GsrN loop rescued *katG-lacZ* expression (Figure 20B).

Based on the stability of the mutations to the GsrN C-rich loop compared to wild-type sequence, we decided to test the physiological role of the *gsrN(RS)* alleles in survival under peroxide stress. Expressing *gsrN*(RS) in one, three, or six tandem copies did not complement the peroxide survival defect of Δ*gsrN* (Figure 23B).

To assess the physiological consequence of mutating the G-tract in the *katG* mRNA leader, we replaced wild-type *katG* on the chromosome with the *katG*(*RS*) allele in both the Δ*gsrN*+3*gsrN* and Δ*gsrN*+3*gsrN*(RS) backgrounds, and measured survival after hydrogen peroxide exposure. Both *katG(RS)*and *gsrN(RS)* mutants had survival defects (Figure 24B&C). Strains harboring the *katG(RS)* allele phenocopy the survival phenotype of Δ*gsrN* under peroxide stress. While *katG(RS)* survival is compromised, the defect is not as large as a strain missing *katG* completely (Δ*katG*) (Figure 25). A strain expressing *katG(RS) and gsrN(RS),* which restores base pairing between the GsrN 5' loop and the *katG* 5' leader, rescued hydrogen peroxide stress survival (Figure 24B&C).

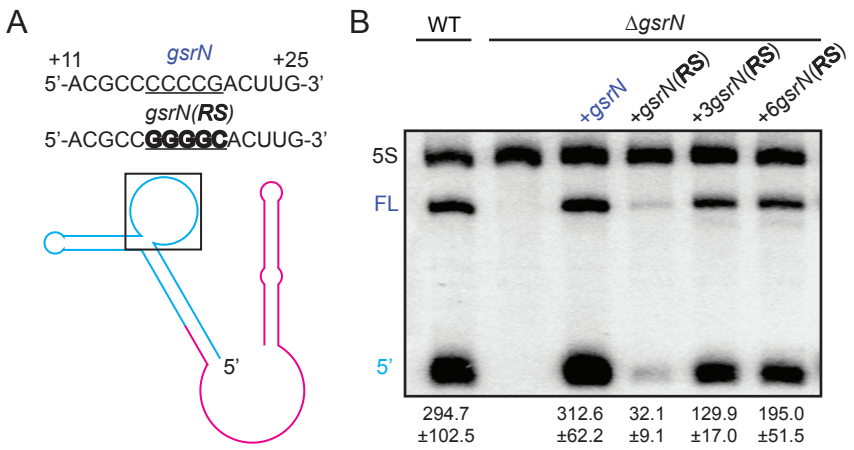


Figure 21. Mutations to the C-rich loop of GsrN cause instability.
(A) Predicted GsrN secondary structure diagram from Figure 3 with a black box is highlighting the nucleotides within the exposed 5' loop of GsrN. Blue labeled sequence represents the wild-type *gsrN* sequence, with the underlined nucleotides emphasizing the location of the RS mutation. The bolded labeled sequence represents the RS mutant *gsrN* sequence.
(B) Northern blot of RNA extracted from wild type, Δ*gsrN*, and Δ*gsrN* complementation strains during exponential growth phase. Complementation strains include wild-type *gsrN* and reverse-swapped (RS) *gsrN(RS)* mutants. Three different copy numbers of *gsrN(RS)* strains were tested. Blots were probed with oligonucleotides complementary to the 5' end of GsrN and to 5S rRNA. Quantified GsrN levels reported were normalized to the 5S rRNA signal in the same lane. Data represent mean ± SD from three independent biological replicates that were loaded, resolved, transferred, and hybridized on the same gel.

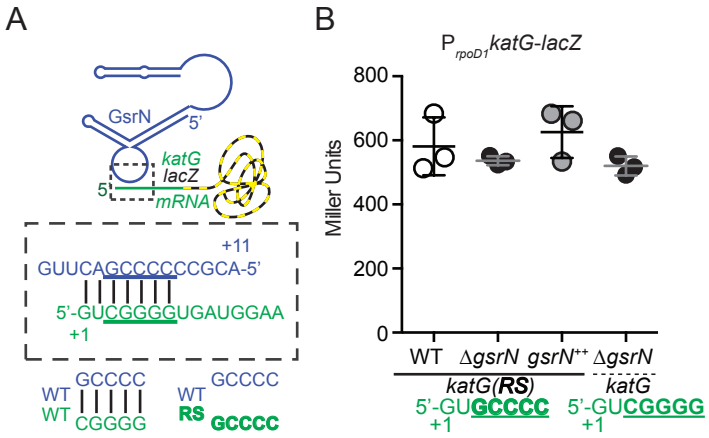


Figure 22. Mutations in the 5'leader of the *katG-lacZ* translational reporter disrupt GsrN regulation.
(A) Translational reporter activity from *katG* and *katG*(RS) leader fusions. The *katG(RS)-lacZ* construct is identical to that in (Figure 20B). Activity from *katG(RS)-lacZ* in wild type, Δ*gsrN,* and *gsrN*$^{++}$ compared to *katG-lacZ* in Δ*gsrN* during exponential growth phase. Bars represent mean ± SD from 3 independent cultures.

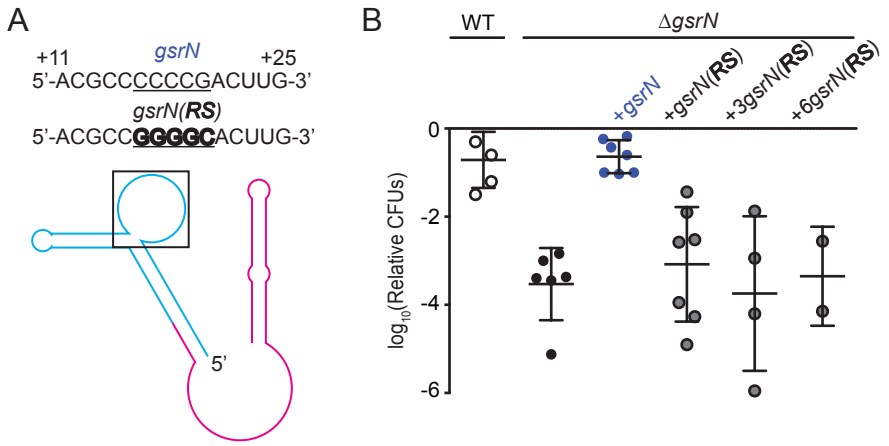


Figure 23. . Mutations to the C-rich loop of GsrN disrupt its role in peroxide survival.
(A) Black box indicates the nucleotides within the exposed 5' loop of GsrN. Blue labeled sequence represents the wild-type *gsrN* sequence, with the underlined nucleotides emphasizing the location of the RS mutation (bolded labels).
(B) Wild type, Δ*gsrN,* Δ*gsrN*+*gsrN* (complementation strain), and Δ*gsrN*+*3gsrN(RS)* strains were subjected to hydrogen peroxide, diluted, and tittered. Three different copy numbers of *gsrN(RS)* strains were tested. Bars represent mean ± SD of at least two independent cultures (points).

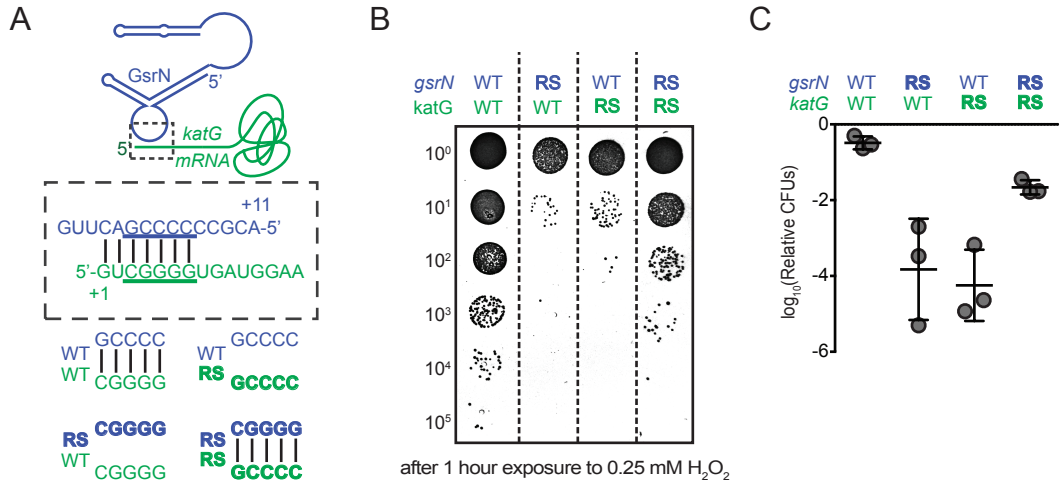


Figure 24. GsrN and *katG* mRNA interactions are important for peroxide stress survival.
(A) Predicted interaction between GsrN (blue) and *katG* mRNA (green), with base-pairing shown in dashed box. Wild-type (WT) and reverse-swapped (RS) mutation combinations of the underlined bases are outlined below.
(B) Colony forming units (CFU) in dilution series ($10^0$ to $10^{-5}$) of Δ*gsrN+3gsrN,* Δ*gsrN+3gsrN(RS),* Δ*gsrN+3gsrN katG(RS),* and Δ*gsrN+3gsrN(RS) katG(RS) C. crescentus* strains after 0.25 mM hydrogen peroxide treatment for 1 hour. Red labels indicate known regulators of the GSR.
(C) Relative hydrogen peroxide survival of RS strains from (B). Δ*gsrN* strains expressing 3*gsrN* or 3*gsrN(RS)* and encode *katG* or *katG(RS)* alleles. Bars represent mean ± SD from 3 independent experiments (points).

Genetic analysis of this proposed pathway would predict that katG is necessary for GsrN's function and overexpression of KatG should rescue the survival of a Δ*gsrN* under peroxide stress. Indeed, the protective effect of overexpressing *gsrN* is lost when *katG* is deleted, and overexpression of *katG* rescues the survival defect of Δ*gsrN* after peroxide treatment. We conclude that *katG* is necessary and sufficient to protect the cell from hydrogen peroxide (Figure 25).

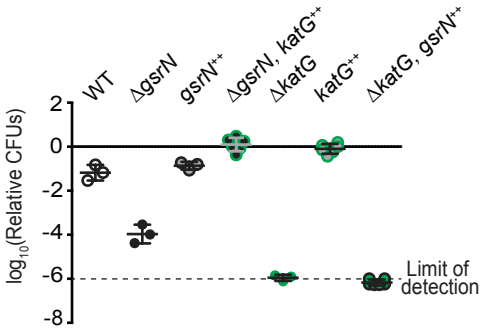


Figure 25. KatG is necessary for GsrN's role in survival under peroxide stress.
Wild type, Δ*gsrN, gsrN*$^{++}$, Δ*gsrN+katG*$^{++}$*(P*$_{xyl}$*katG),* Δ*katG, katG*$^{++}$*(P*$_{xyl}$*katG),* and Δ*katG+gsrN*$^{++}$*(4gsrN)* strains were subjected to hydrogen peroxide, diluted, and tittered. Bars represent mean ± SD of at least 3 independent cultures.

Given differences in steady state levels of GsrN and GsrN(RS) (Figure 21), we postulated that the capacity of GsrN to interact with its targets influences its stability *in vivo*. Indeed, mutation of the *katG* target site reduced GsrN by more than 2-fold (Student's t-test, $p<0.0001$). The compensatory *katG(RS)* allele partially restored stability to GsrN(RS) (Figure 26B). *katG(RS)* mutation or *katG* deletion did not influence *gsrN* transcription (Figure 26C). Thus, we attribute the differences in steady-state levels of the GsrN alleles to their ability to interact with mRNA targets via the 5' C-rich loop.

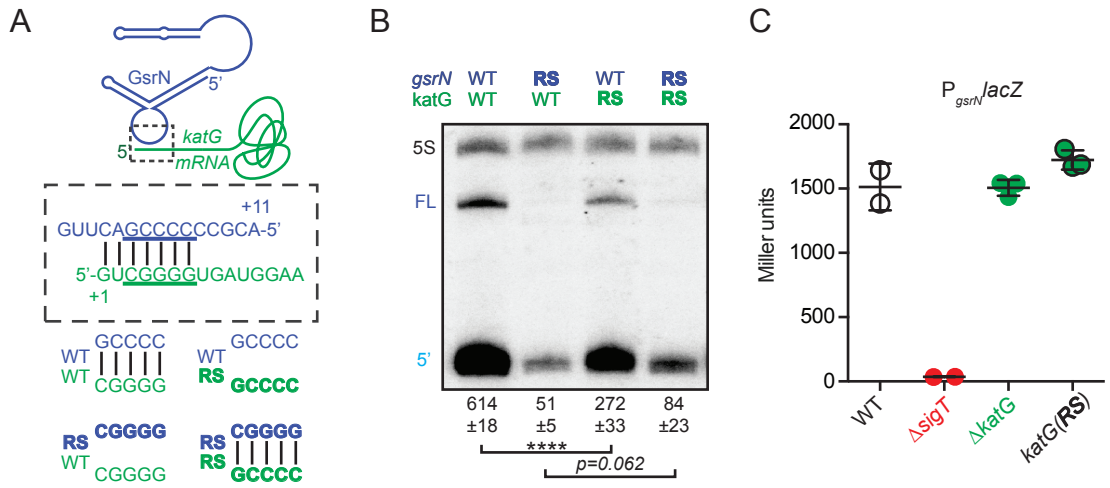


Figure 26. GsrN stability is dependent on binding to *katG* mRNA.
(A) Predicted interaction between GsrN (blue) and *katG* mRNA (green), with base-pairing shown in dashed box. Wild-type (WT) and reverse-swapped (RS) mutation combinations of the underlined bases are outlined below.
(B) Northern blot of total RNA from strains in (A) collected in exponential phase hybridized with probes complementary to 5' end of GsrN and 5S rRNA. Quantification is mean ± SD normalized signal from 3 independent experiments and p-value estimated by Student's t-test.
(C) Activity from the *gsrN* transcriptional reporter described in Figure 6 was assayed in Δ*katG* and *katG-RS* backgrounds during exponential growth. Mean ± SD of 3 independent cultures.

## 5.3. GsrN influences protein levels of KatG

To assess the relative effects of GsrN on *katG* transcript and protein levels *in vivo*, we directly measured both by dot blot and Western blot, respectively. In untreated and peroxide treated cultures, *katG* transcript levels trended lower in Δ*gsrN* and higher in $gsrN^{++}$ compared to wild type (Figure 27A). These differences are not statistically significant (Student's t-test, $p=0.39$) in untreated cultures; however, KatG protein tagged with the M2 epitope was reduced 2-fold in Δ*gsrN* lysates relative to wild-type in untreated cultures (Student's t-test, $p<0.0001$) (Figure 27B). Since GsrN does not influence *katG* transcription in untreated cultures (Figure 19A-C), GsrN may enhance KatG translation *in vivo*.

Steady-state *katG* transcript levels differ significantly between Δ*gsrN* and *gsrN*$^{++}$ in peroxide treated cultures (Student's t-test, $p<0.01$) (Figure 27A). KatG protein tagged with the M2 epitope was reduced 3-fold in peroxide treated cells in Δ*gsrN* lysates relative to wild-type; KatG-M2 levels in *gsrN*$^{++}$ were increased in both untreated and peroxide treated cells (Figure 27B). These data support a model whereby GsrN enhances KatG protein expression in the presence of peroxide by stabilizing *katG* mRNA and/or promoting *katG* translation.

It is interesting to note that in wild-type samples peroxide induction causes a 5-6 fold increase in *katG* mRNA levels, while peroxide induction only causes a more modest increase in KatG protein levels. Moreover, peroxide induction consistently causes *katG* mRNA levels to also increase in similar magnitude in Δ*gsrN* and in *gsrN*$^{++}$ samples, while KatG protein levels largely remain unchanged in Δ*gsrN* samples but increase 2-3 fold in *gsrN*$^{++}$ samples. These discrepancy highlights a possible mechanism for GsrN in the translation efficiency of *katG* mRNA. Given that KatG protein levels do not linearly scale with *katG* mRNA levels but are influenced heavily by the levels of GsrN demonstrates that GsrN most likely acts at promoting *katG* translation.

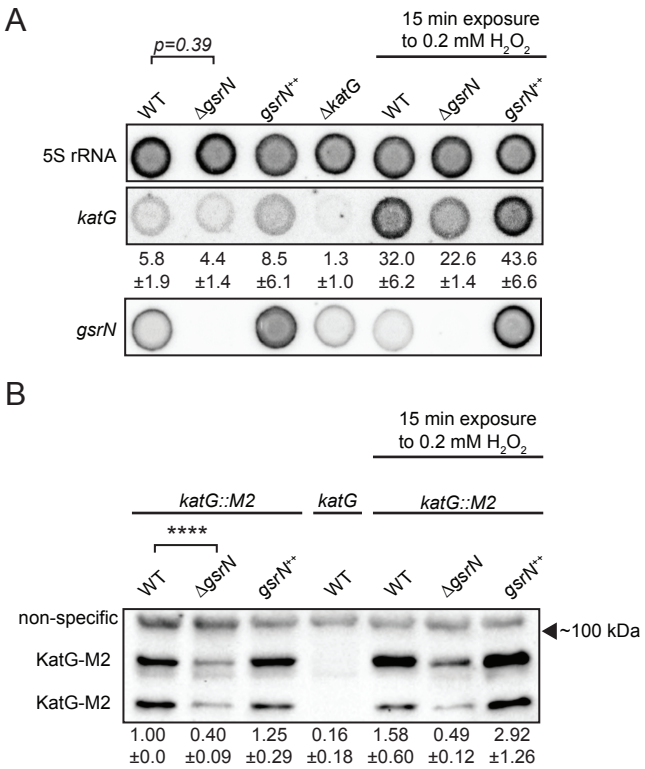


Figure 27. GsrN influences protein levels of KatG.
(A) Dot blot of total RNA of *gsrN* and *katG* mutants grown to early stationary phase ($OD_{660}$ 0.85-0.9, which reflects the density of culture used to start a viability assay). Samples on right were treated with 0.2 mM hydrogen peroxide before RNA extraction for 15 minutes. This amount of time was chosen to sample peroxide effects and also ensure sufficient amount of cells could be obtained from Δ*gsrN* cultures. Blots were hybridized with *katG* mRNA, GsrN or 5S rRNA probes. *katG* mRNA signal normalized to 5S rRNA signal is quantified (mean ± SD, n=3, p-value estimated with Student's t-test).
(B) Immunoblot of KatG-M2 fusion in wild type, Δ*gsrN*, and $gsrN^{++}$ strains in the presence and absence of peroxide stress probed with α-FLAG antibody. All indicated lanes contain a genetically encoded FLAG-tag at native *katG* locus. KatG migrates as two bands as previously reported (Italiani, da Silva Neto, et al. 2011). Normalized KatG-M2 signal (mean ± SD, n=4, **** $p<0.0001$ Student's t-test) is presented below each lane. Arrow indicates position of 100 kDa molecular weight marker

## Chapter 6: GsrN is a global regulator of stress response

Although we verified one of the targets identified in our forward biochemical approach, we wanted to potentially address the several targets enriched in the GsrN-PP7hp affinity-purification. Alternative to careful investigations into each particular target, we aimed to address the entire regulon of GsrN. Akin to the transcriptomic and proteomic studies performed on the RpoS and SigT regulons, we looked at how global transcript and protein levels changed as a function of GsrN. Moreover, the multiple GsrN targets suggested that GsrN might have regulatory roles beyond mitigation of peroxide stress. Thus, we examined another survival phenotype associated with the GSR, osmotic stress.

### 6.1. Transcriptomic and proteomic analysis of *gsrN* mutants

To globally define genes that are directly or indirectly regulated by GsrN, we performed RNA-seq and LC-MS/MS measurements on wild-type, Δ*gsrN* and $gsrN^{++}$ strains (Figure 28A). We identified 40 transcripts, including *gsrN*, with significant differences in mapped reads between the Δ*gsrN* and $gsrN^{++}$ samples (Figure 29A). 11 proteins had significant label free quantitation (LFQ) differences (FDR<0.05) between $gsrN^{++}$ and Δ*gsrN* (Figure 29B). Most genes identified as significantly regulated by transcriptomic and proteomic approaches did not overlap (Figure 6.1). Moreover there is low congruency between global measurements and enriched RNAs in the GsrN(37)-PP7hp sequencing; such discrepancies could arise from the lack of coverage in our proteomic studies. Since our proteomic measurements were done in near stationary cultures on the soluble fraction of total cell lysate, we were only able to survey 30% of *Caulobacter*

*crescentus*' apparent proteome. Nonetheless, these data provide evidence that GsrN can function as both a positive and negative regulator of gene expression, either directly or indirectly.

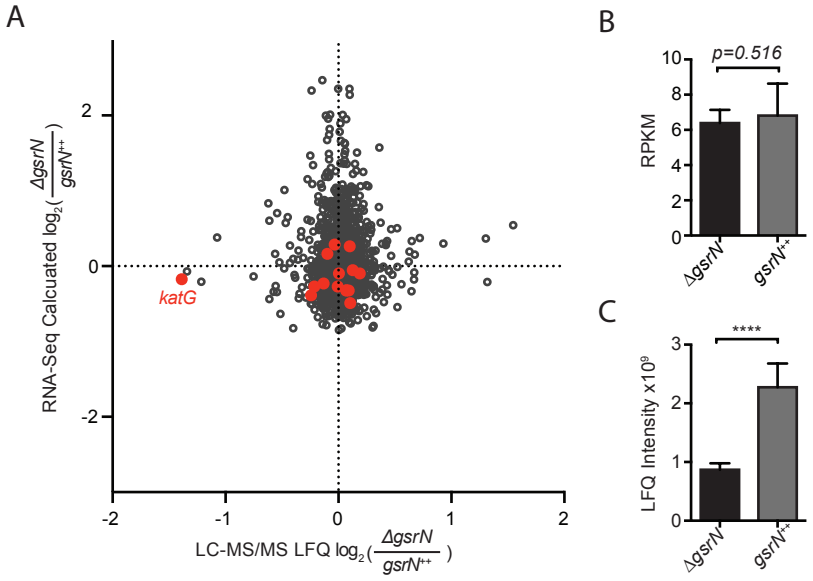


Figure 28. Combined analysis of RNA and protein levels of *gsrN* mutants verifies the GsrN's regulation of KatG.
(A) Transcriptomic and proteomic analysis of Δ*gsrN* (deletion) and *gsrN*$^{++}$ (overexpression) cultures in early stationary phase. Only genes detected in both analyses are plotted. Red indicates transcripts that co-purify with GsrN-PP7hp (Figure 17).
(B) *katG* transcript from Δ*gsrN* and *gsrN*$^{++}$ cells quantified as reads per kilobase per million mapped (RPKM). Data represent mean ± SD of 5 independent samples. Significance was evaluated with the Wald test.
(C) Label free quantification (LFQ) intensities of KatG peptides from Δ*gsrN* and *gsrN*$^{++}$ cells (mean ± SD, n=3; **** $p<0.0001$ Student's t-test).

Importantly, RNA-seq and proteomics experiments validated *katG* as a regulatory target of GsrN. *katG* transcript levels measured by RNA-seq were not significantly different between Δ*gsrN* and *gsrN*$^{++}$ strains (Figure 28B), consistent with our dot blot measurements of unstressed cultures (Figure 27A). Conversely, steady-state KatG protein levels estimated from our LC-MS/MS experiments were significantly reduced in Δ*gsrN*, consistent with our Western analysis of KatG protein (Figures 6.1.C and Figure 27B). Moreover, *katG* was the only gene that was significantly enriched in the pull-down and differentially expressed in the proteomic studies.

These results provide additional evidence that *katG* is a major target of GsrN, and that GsrN functions to enhance KatG expression at the post-transcriptional level.

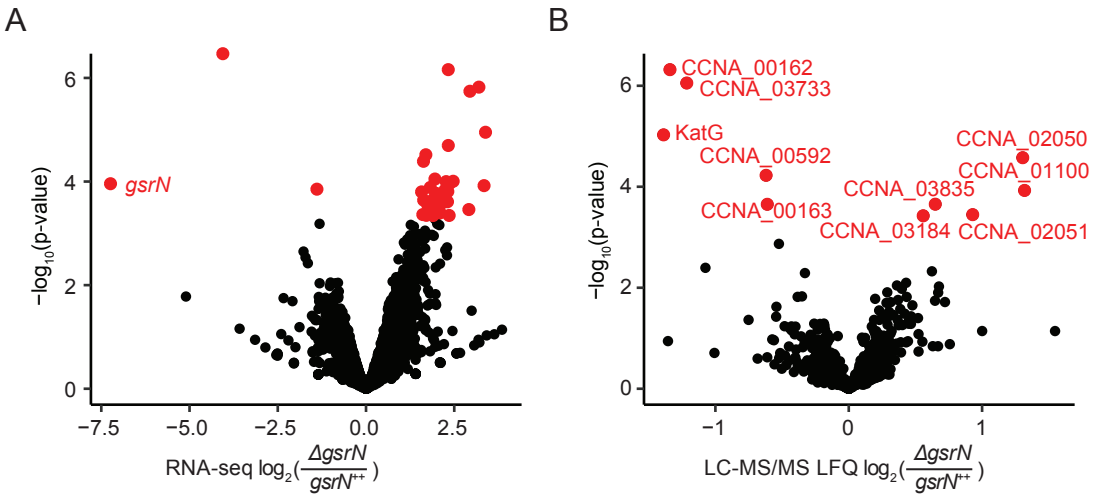


Figure 29. Transcriptomic and proteomic analysis of Δ*gsrN* and *gsrN*$^{++}$ show regulatory effects of several genes.
(A) RNA-seq analysis of Δ*gsrN* and *gsrN*$^{++}$ early stationary phase cultures ($OD_{660}$ ~0.85-0.90, which reflects the density of culture used to start a viability assay) represented as a volcano plot where expression changes are plotted as a function of p-value. Red indicates transcripts with a false discovery rate (FDR) corrected p-value < 0.05. Black indicates gene transcripts with a FDR p-value above the cut-off.
(B) LC-MS/MS total soluble protein signal from MaxQuant label free quantitation estimates from Δ*gsrN* and *gsrN*$^{++}$ cells grown to early stationary phase ($OD_{660}$ ~0.85-0.90, which reflects the density of culture used to start a viability assay). Log-2 transformed fold change in LFQ estimates from MaxQuant (Cox et al. 2014) are plotted as a function of p-values obtained from the multiple t-test analyses using GraphPad Prism version 6.04 for MacOS, GraphPad Software, La Jolla California USA, www.graphpad.com. Red indicates proteins with significant differences (false discovery corrected p-value < 0.05). Black represents proteins that do not meet the FDR cut-off

## 6.2. GsrN is necessary during osmotic stress in *Caulobacter crescentus*

Given our pull-down experiments and our transcriptomic and proteomic datasets, we reasoned that GsrN might contribute to other phenotypes associated with deletion of the GSR sigma factor, *sigT*. Indeed, the Δ*gsrN* mutant has a survival defect after exposure to hyperosmotic stress, similar to Δ*sigT* (Figure 30A). As we observed for peroxide stress, overexpression of *gsrN* protects cells under this physicochemically-distinct condition. Hyperosmotic stress survival does not require *katG* (Figure 30B), providing evidence that a separate GsrN regulatory target mediates this response. Unlike hydrogen peroxide (Figure 27A), hyperosmotic stress induces GsrN expression (Figure 31A). This is consistent with previous transcriptomic studies in *Caulobacter* in which hyperosmotic stress, but not peroxide stress, activated GSR transcription (Alvarez-Martinez et al. 2007). GsrN transcription is also significantly enhanced in stationary phase cultures relative to logarithmic phase cultures (Figure 31B). Though its functional role under this condition remains undefined, it has been reported that *katG* is a genetic determinant of stationary phase survival (Steinman, Fareed, and Weinstein 1997a).

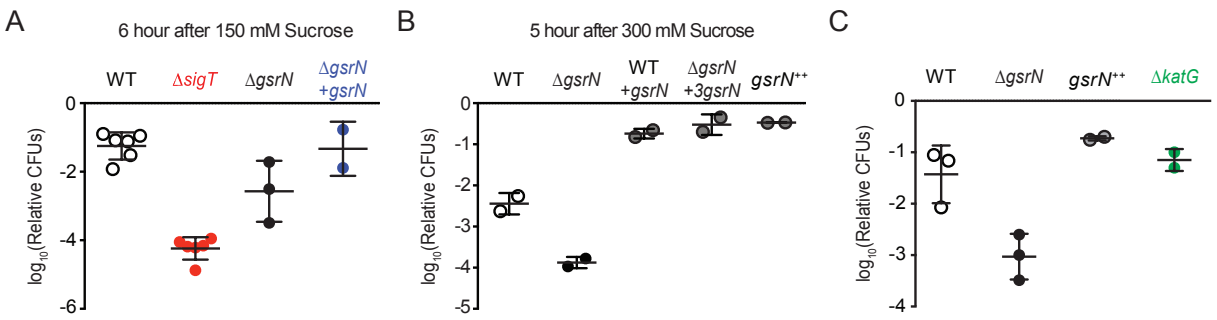


Figure 30. *gsrN* is necessary for hyper-osmotic stress survival.
(A) Relative CFUs of wild type, Δ*sigT,* Δ*gsrN*, and complementation strain (Δ*gsrN*+*gsrN)* of sucrose treated cultures to untreated cultures. Treatment was a 6-hour exposure to 150 mM sucrose. The sucrose was chosen to parallel past osmotic stress studies on the GSR (Alvarez-Martinez et al. 2007). Data represent mean ± SD from at least 2 biological replicates (points).
(B) Relative CFUs of wild type, Δ*gsrN*, and overexpression strains of *gsrN* of sucrose treated cultures to untreated cultures. Treatment was for 5-hour exposure to 300 mM sucrose. This was our first experimentation of increasing sucrose concentration to achieve a higher dynamic range between wild type, Δ*gsrN*, and overexpression strains of *gsrN*. Data represent mean ± SD from 2 independent experiments (points).
(C) Hyperosmotic stress survival of wild type, Δ*gsrN*, *gsrN*$^{++}$, and Δ*katG* cells relative to untreated cells. Stress was a 5-hour treatment with 300 mM sucrose. The increase of sucrose concentration from previous experiments in this study was chosen to display the dynamic range Δ*gsrN* susceptibility and *gsrN*$^{++}$ protection. Data represent mean ± SD from 2 independent experiments (points).

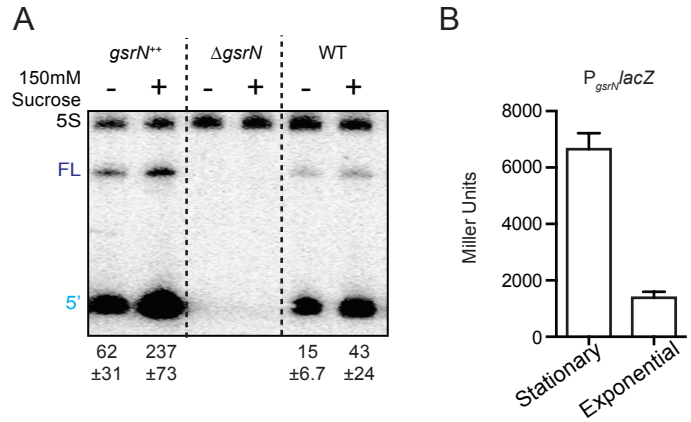


Figure 31. *gsrN* is induced under hyper-osmotic stress survival and under growth under stationary phase.
(A) Northern blot of total RNA from wild type, Δ*gsrN*, and *gsrN*$^{++}$ cultures with or without 150 mM sucrose stress. Blots were hybridized with GsrN and 5S rRNA probes. Normalized mean ± SD of total GsrN signal from 3 independent samples is quantified.
(B) β-galactosidase activity from the $P_{gsrN}$*lacZ* transcriptional fusion in exponentially growing ($OD_{660}$ ~0.25) and early stationary phase ($OD_{660}$ ~0.75) wild-type cells. Bars represent mean ± SD from 2 independent cultures.

## Chapter 7: Conclusions and Future Directions

Using cell cycle transcriptome data, we predicted a sRNA of unknown function, GsrN, might play a predominant role in the general stress response. Through genetic analysis, biochemical characterization, and molecular dissection of the mRNA leader of *katG*, we uncovered how the general stress response through GsrN confers adaptation to hydrogen peroxide exposure (Figure 32).

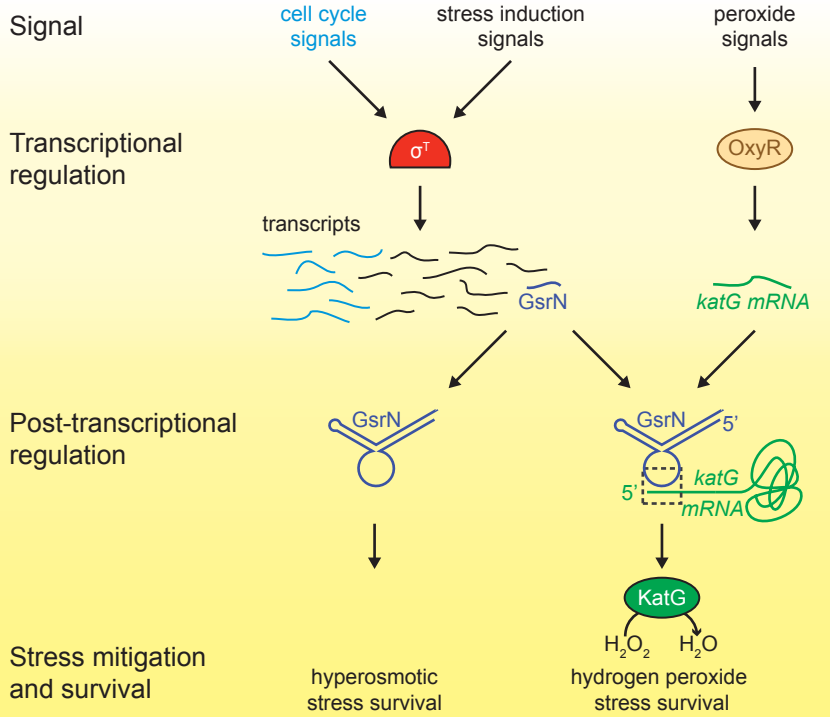


Figure 32. GsrN provides an additional layer of regulation to the general stress response. Expression of the GSR EcfG-sigma factor, *sigT* (SigT), and select genes in the GSR regulon is regulated as a function cell cycle phase. SigT-dependent transcription can be induced by certain signals (e.g. hyperosmotic stress), but is unaffected by hydrogen peroxide. Transcription of the sRNA, GsrN, is activated by SigT, and the cell cycle expression profile of *gsrN* is highly correlated with *sigT* and its upstream regulators. Transcription of the catalase/peroxidase *katG* is independent of $\sigma^T$. GsrN dependent activation of KatG protein expression is sufficient to rescue the peroxide survival defect of a Δ*sigT* null strain. GsrN convenes a post-transcriptional layer of gene regulation that confers resistance to peroxide and hyperosmotic stresses.

## 7.1. Role of GsrN in mitigating peroxide stress

The discovery and functional characterization of GsrN during peroxide stress echoes the principle of Jacob and Monod 1961: a cell's physiology is primarily controlled through gene regulation. In the case of GsrN, the catalase activity of *C. crescentus* cultures is tuned through the levels of GsrN, as measured through survival (Figure 9). This occurs through the base-pairing mechanisms of the 5' leader of *katG* mRNA to GsrN's C-rich 5' loop (Figure 20 and 5.7), which enables increased levels of KatG protein (Figure 27).

Hydrogen peroxide stress adaptation in *C. crescentus* works on two levels: transcriptional up-regulation through OxyR and post-transcriptional influence by GsrN on the sole catalase gene, *katG* (Italiani, Neto, et al. 2011, Steinman, Fareed, and Weinstein 1997b). Although transcriptional regulation by OxyR in *E. coli*'s *katG*, seems to be conserved, there is no analogous sRNA regulation of any of the catalase systems in *E. coli*. Interestingly *E. coli* contains two additional catalase systems, the alkyl hydroperoxide reductase (Ahp) and catalase E (KatE). Ahp is also under the regulation of OxyR and seems to be the more predominant scavenger of $H_2O_2$ (Seaver and Imlay 2001, Zheng et al. 2001); while KatE (as mentioned previously Chapter 1.2.II) is under the control of RpoS (Schellhorn and Hassan 1988). Catalase-production also seems to highly up-regulated during stationary growth as well in *E. coli* (Imlay 2013).

Although the mechanisms of regulating hydrogen peroxide stress vary between *E. coli* and *C. crescentus*, the birds eye view of these systems are fairly analogous. GsrN explains the role of GSR during peroxide stress akin to RpoS induction of *katE* transcription. Additionally, OxyR

transcriptionally up-regulates KatG production in a GSR-independent manner in both organisms. The three programmed catalase systems in *E. coli* and single catalase system in *C. crescentus* could be due to the ecological niches of each organism.

*C. crescentus* is an obligate aerobe that primarily grows in oxygen-rich environments, while *E. coli* is a facultative anaerobe and can switch from anaerobic to aerobic metabolism. The variety of metabolisms, perhaps, causes *E. coli* to have several systems to deal with oxidants, such as hydrogen peroxide, when switching to an aerobic-based metabolism where hydrogen peroxide can be a metabolic byproduct (Imlay 2013). *C. crescentus* might maintain KatG levels near its enzymatic capacity in the cell due to endogenous production of hydrogen peroxide in its aerobic lifestyle. Given the sensitivity of *C. crescentus* to exogenous hydrogen peroxide in *gsrN* deletion and overexpression, we believe that GsrN is playing a significant role in maintaining the catalase capacity of the cell.

KatG is one of the most abundant proteins in *C. crescentus* cultures during stationary phase and is required during this growth phase (Steinman, Fareed, and Weinstein 1997b). One role GsrN might be playing in stationary phase induction is to work in consortium with OxyR to maintain steady levels of KatG. Lastly, GsrN's regulation of *katG* also doesn't seem to be involved in GsrN's role in osmotic stress. The targets outlined in Table 1 would be promising candidates for future functional analysis of the GsrN's role in osmotic stress relief.

## 7.2. Prevalence of GsrN in Alphaproteobacteria.

The GSR system is broadly conserved in Alphaproteobacteria. Given the importance of GsrN as a post-transcriptional regulator of the *Caulobacter* GSR, we reasoned that a functionally related sRNAs might be a conserved feature of the GSR in this clade. To identify potential orthologs of *gsrN*, we surveyed the genomes of Alphaproteobacteria that encoded regulatory components of the GSR system and for which transcriptomic data were publically available.

We initially searched for GsrN-related sequences using BLASTn (Altschul et al. 1990). Hits to GsrN were limited to the Caulobacteraceae family (*Caulobacter*, *Brevundimonas*, and *Phenylobacterium*). The 5' C-rich loop of homologs identified in this family had the highest level of conservation compared to other regions of secondary structure (Figure 34A). Predicted *gsrN* homologs are often proximal to the genes encoding the core GSR regulators (*ecfG/sigT*, *nepR* and *phyR*) (Figure 33). *C. crescentus* is a notable exception where *gsrN* is positioned distal to the GSR locus. Therefore, we used genome position as a key parameter to identify additional GsrN or GsrN-like RNAs in Alphaproteobacteria outside of Caulobacteraceae.

Our search for GsrN focused on three parameters: evidence of intergenic transcription, identification of a near-consensus EcfG-binding site in the promoter region, and proximity to the *ecfG-phyR* chromosomal locus. Based on our criteria, we identified a set of putative GsrN homologs in the Rhizobiaceae family (Jans et al. 2013, Kim et al. 2014, Valverde et al. 2008) (Figure 34B). The predicted secondary structure of these putative GsrN homologues has features similar to GsrN from Caulobacteraceae. Specifically, there is an exposed cytosine-rich loop at the 5' end (Figure 34B).

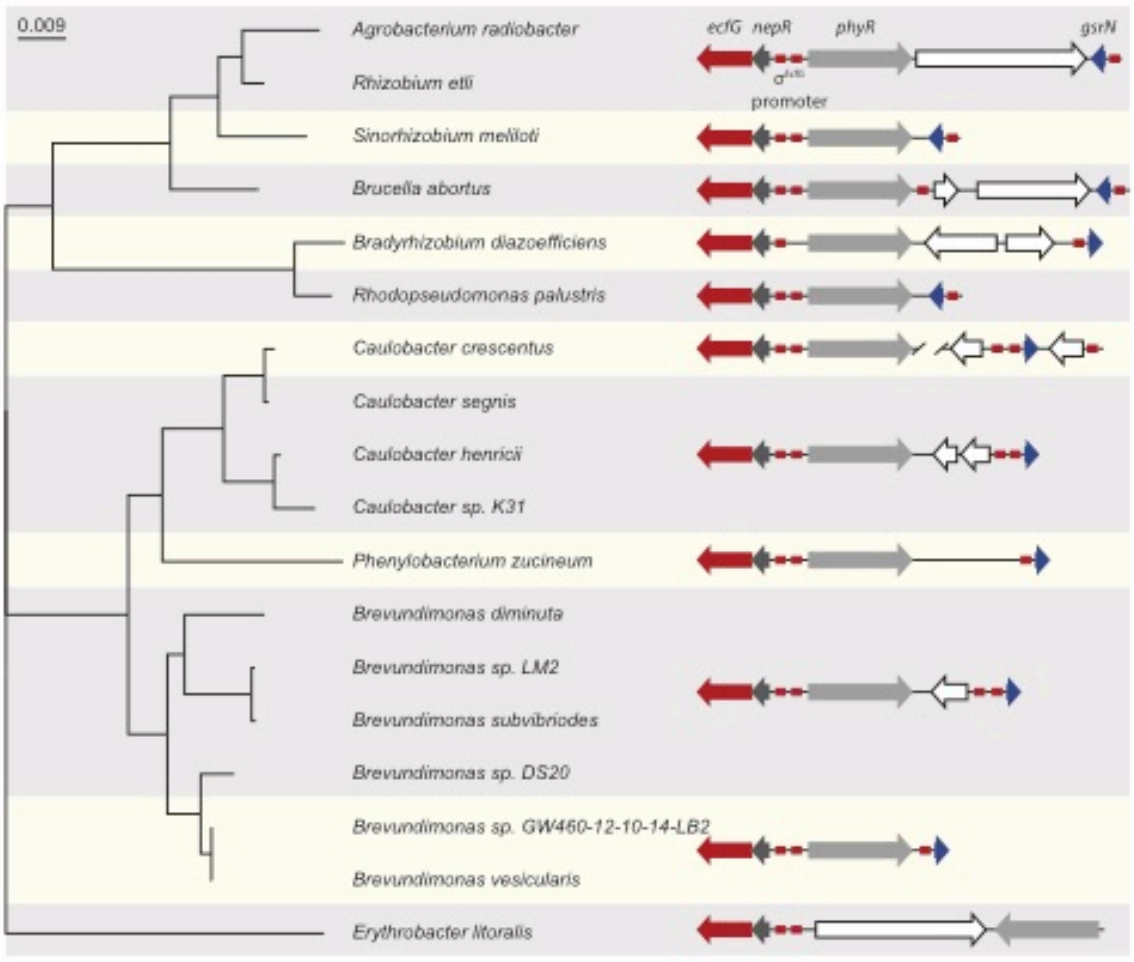


Figure 33. Synteny between the GSR locus and putative *gsrN* in Alphaproteobacteria
Locus diagrams showing predicted *gsrN* homologs in several Alphaproteobacteria. Tree was constructed from the 16s rRNA sequences of each strain where *Erythrobacter litoralis* (for which there is no apparent *gsrN*-like gene) was the out-group. Red arrows represent *ecfG*, dark gray arrows represent *nepR*, red boxes represent the conserved EcfG-binding site, light gray arrows represent *phyR*, and dark blue arrows represent *gsrN* (or its putative homologs). The prediction of GsrN orthologs in the Caulobacteraceae (*Caulobacter*, *Brevundimonas*, and *Phenylobacterium)* was based on a BLASTn search (Altschul et al. 1990). The prediction of GsrN in *Rhizobium etli*, *Sinorhizobium meliloti*, and *Brucella abortus* was based on evidence of expression in published transcriptome data, proximity to the GSR locus, and identification of a EcfG-binding site upstream of the gene. The prediction of *Agrobacterium radiobacter* was based on a BLASTn search of using the predicted GsrN sequence from *R. etli* as the query (Altschul et al. 1990). The prediction of *Rhodopseudomonas palustris* and *Bradyrhizobium diazoefficiens* is completely based on the proximity to the GSR locus and the presence of an upstream EcfG-binding site.

Our analysis of the co-conservation of *katG* regulation by GsrN revealed that many Caulobacterales and Rhizobiales do not contain a *katG* gene. Of the species we examined, a majority of them had several genes capable of catalase activity (annotated in brown Figure 34A&B). *Caulobacter crescentus* and *Rhizobium etli* were the only organisms that have *katG* as the only gene responsible for catalase activity (Vargas Mdel et al. 2003, Steinman, Fareed, and Weinstein 1997a). Interestingly, these two organisms also show hydrogen peroxide sensitivity in a GSR compromised background (Alvarez-Martinez et al. 2007, Vercruysse et al. 2011). Inspection of the 5'UTR of *katG* in *R. etli* revealed a G-rich sequence that is complementary to seven nucleotides of the predicted exposed loop of the putative GsrN homologue in *R. etli*, echoing the mechanisms of GsrN and *katG* mRNA in *C. crescentus*.

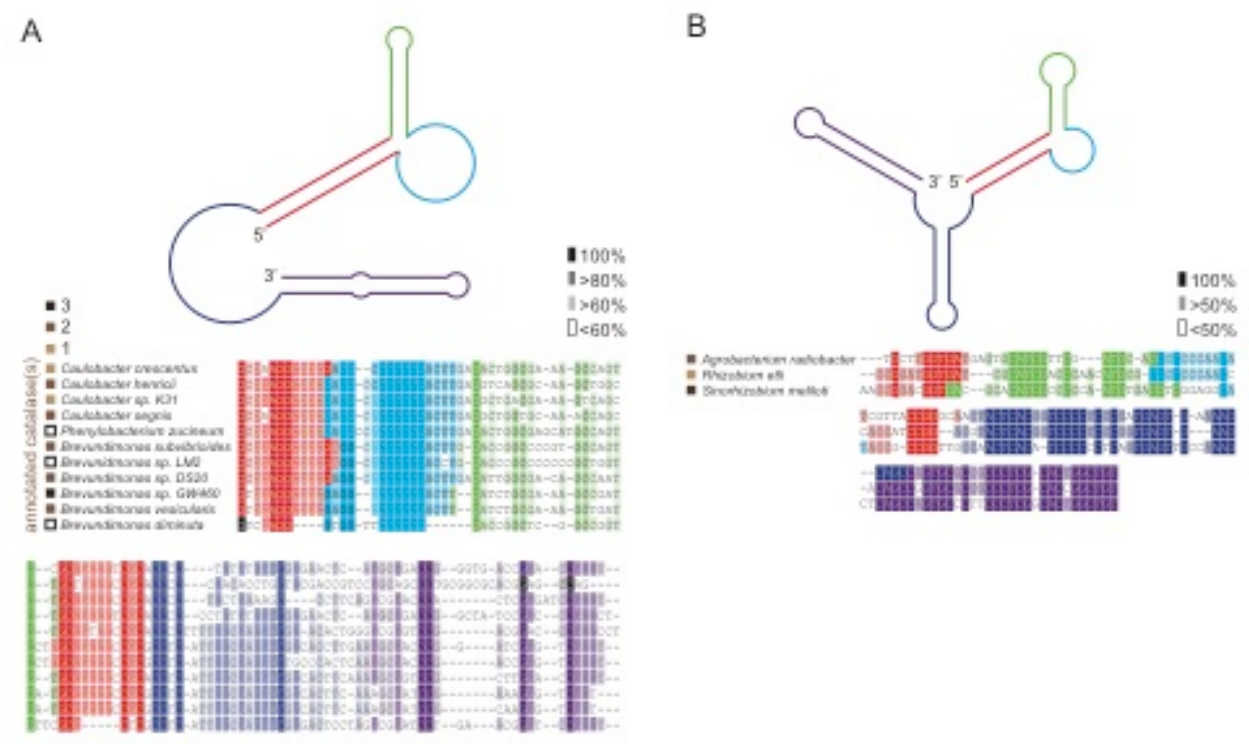


Figure 34 C-rich loop of GsrN seems to be the most conserved feature across other putative homologues.
(A) Diagram of predicted secondary structure of GsrN in other Caulobacteraceae is colored by secondary structure element. Colors highlighted in the sequence alignment correspond to the predicted secondary structure regions in the cartoon. Density of shading corresponds to conservation at that position. Brown boxes on the left of species name indicate the number of annotated catalase genes present in their genome annotations.
(B) Diagram of predicted secondary structure of predicted GsrN homologs in select Rhizobiaceae where the 5' portion contains an unpaired 5' G-rich loop (cyan) flanked by a small hairpin (green) and a stem loop involving the 5' terminus (red). Brown boxes are consistent with

## 7.3. GsrN stability and processing

The roles of sRNAs in stress adaptation have been investigated in many species, and a number of molecular mechanisms underlying sRNA-dependent gene regulation have been described. We have uncovered a connection between mRNA target site recognition and GsrN stability that presents challenges in the characterization of GsrN regulatory mechanisms. Specifically, mutations in the *katG* mRNA leader affect steady-state levels of GsrN (Figure 26). Given this result, one could envision scenarios in which changes in transcription of *katG* or some other direct GsrN target could broadly affect stress susceptibility by altering levels of GsrN and, in turn, the stability of other target mRNAs in the cell. In short, the concentrations of mRNA targets could affect each other via GsrN. Such effects should be considered when assessing mRNA target site mutations in this system and others.

GsrN is among a handful of sRNAs that are post-transcriptionally processed (Figure 12 and Figure 13) (Chapter 3.5). Select PP7hp insertions resulted in reduced 5' isoform formation; PP7hp insertion mutants with low 5' isoform levels did not complement the peroxide viability defect of Δ*gsrN*. Processing to a short 5' isoform may be necessary for GsrN to bind *katG* mRNA and regulate KatG expression. Alternatively, cleavage may not be required for function, and lack of complementation by certain hairpin insertion mutants may be due to PP7hp interfering with target recognition or simply reducing total levels of GsrN. Regardless, our data clearly show that GsrN is cleaved to yield a 5' isoform that is very stable in the cell (Figure 12) and is sufficient to protect *Caulobacter* from hydrogen peroxide treatment (Figure 11).

Lastly, the RNA-chaperone Hfq affects the stability of full-length GsrN and/or inhibits the process of endonucleolytic cleavage (Figure 35). Hfq is a part of the Sm-family of RNA-binding proteins and plays diverse roles in sRNA regulation. Some regulatory targets of sRNA require Hfq, while others depend on Hfq as a stability factor. Hfq is also known to associate with other RNA proteins such as RNase E and the transcriptional terminator Rho. In the case of GsrN, Hfq seems to be a strong regulator of isoform formation. It would be of interest to understand Hfq's role in sRNA-mRNA target stabilization or in controlling a potential RNase that might be causing the endonucleolytic cleavage event.

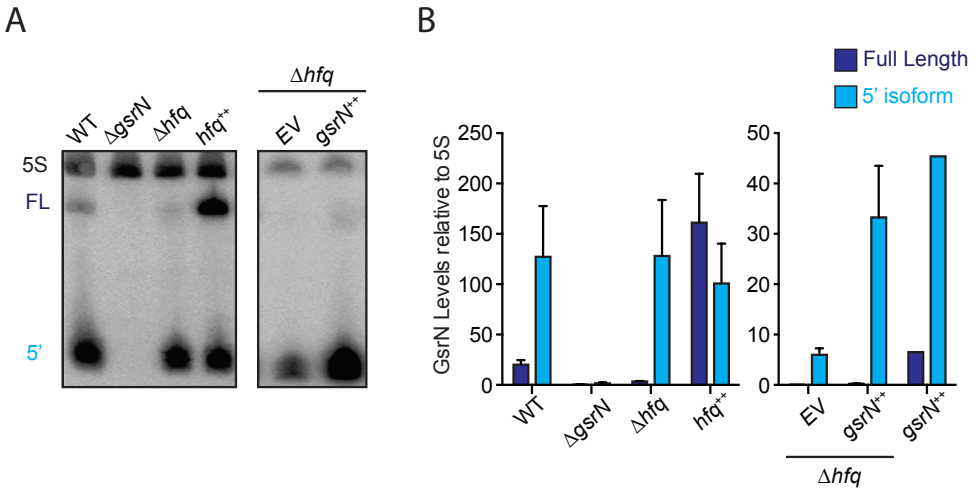


Figure 35. . Hfq affects GsrN isoform levels.
(A) Northern blot of total RNA from wild type, Δ*gsrN*, Δ*hfq*, and *hfq*$^{++}$ cultures. Blots were hybridized with GsrN and 5S rRNA probes. Blot on the right demonstrates that overexpression of GsrN cannot produce a high level of full-length GsrN.
(B) Blots were quantified by densitometry. GsrN signal from the full-length (FL; dark blue) and 5' isoform (5'; cyan) are normalized to 5S rRNA in each lane and multiplied by 100. Bars represent mean ± SD of triplicate extractions, each representing biologically independent samples.

Our understanding of the role of RNA metabolism in sRNA-dependent gene regulation is limited, and GsrN provides a good model to investigate mechanisms by which mRNA target levels and sRNA and mRNA processing control of gene expression.

## Chapter 8: Brief addendum on *Caulobacter crescentus* cell cycle regulation

An alternative perspective on gene regulation is cellular division, an intrinsic process inherent to all organisms. How are genes activated and repressed during the process of cellular division? Genes involved in DNA replication, peptidoglycan rearrangement, and cytokinesis machinery are produced and regulated at specific times during the each step of division (Hajduk, Rodrigues, and Harry 2016). However, what other genes might be regulated in actively dividing cells? How does appropriate timing of gene regulation affect cell physiology?

Studies on the dimorphic lifestyle of *Caulobacter crescentus* has provided a wonderful amount of information pertaining to cell cycle regulation. Since *Caulobacter crescentus* lives two distinct lifestyles: a motile non-reproductive swarmer cell and sessile reproductive stalked cell, understanding the difference in gene expression between these two cell types has elucidated several regulatory regimes that control gene expression during cellular replication (Collier 2016).

### 8.1. Proteomic circuits in *Caulobacter crescentus*

In *C. crescentus,* genome replication and cellular division are intimately coordinate so that DNA replication occurs once and only once per round of cellular division (Collier 2012). This coordination is different when considering the multiple rounds of DNA replication that can occur during cellular division in *E. coli* and *B. subtilis* (Mott and Berger 2007, Sharpe et al. 1998). The biological implications of the difference in coordination of DNA replication and cell division are not known; however, much like RNA transcription, DNA replication is largely regulated at its initiation phase.

Initiation of DNA replication in bacteria centers on the activity of DnaA, the initiating DNA binding factor that coordinates the melting of the origin of replication (*oriC*) and recruits DNA polymerase to the *oriC* (Kaguni 2006). In *C. crescentus*, DnaA is sequestered from its binding site at the *oriC* by the transcription factor and response regulator, CtrA; this CtrA binding event causes the swarmer cell to be non-replicative (Jonas, Chen, and Laub 2011). Although this competition drives DNA replication initiation, each DNA binding factor also moonlights as a transcription factor that dictates the activation and repression of several hundred genes (Hottes, Shapiro, and McAdams 2005, Laub et al. 2002).

The regulon of CtrA and DnaA have been defined in several studies (McGrath et al. 2007, Schrader et al. 2014). Moreover, it is the activity of these two factors that begins a time-coordinated molecular cascade and enables the production of several necessary components for cellular division. Some factors belonging to this downstream cascade include GcrA (global cell cycle regulator A), CcrM (a DNA methyltransferase), and SciP (small CtrA inhibitory protein). Approximately 20-30% of genes show strong cell cycle regulation and many of them can attribute their expression to the activities of these global regulators (Laub et al. 2000). For instance, the regulatory logic of DnaA, CtrA and CcrM allows for the cytokinesis machinery (*ftsZ*) to be expressed in *C. crescentus* stalked cell (Kelly et al. 1998, Gonzalez and Collier 2013). Moreover, the RpoN sigma factor, also known as sigma-54, is CtrA regulated and is essential for the transcription of late flagellum genes (*fla*) near the end of cellular replication (Wu, Benson, and Newton 1995, Brun and Shapiro 1992).

Studies in other alphaproteobacteria have shown that the homologues of the cell cycle regulators perform the same actions and time the replication of DNA intimately with the process of division (Collier 2016). Interestingly, the implications of why some genes are intimately tied to the cell cycle are left unknown. One such lingering mystery is the cell cycle regulation of the GSR sigma factors, *sigT*, which displays strong up-regulation in the swarmer-to-stalk transition in *C. crescentus'* life cycle (Laub et al. 2000).

## 8.2. Cell cycle regulation of *Caulobacter crescentus* under stress

Bacteria must balance gene regulatory decisions between an inherent cellular process such as cellular division and an external signaling cascade such as stress response. It is a common strategy for bacteria to arrest their growth under low nutrient conditions or under environmental perturbations (Bergkessel, Basta, and Newman 2016). Mechanisms in which bacteria stall growth are quite varied, however, in *C. crescentus* this regulatory decision is usually centered on DNA replication initiation, either through DnaA or CtrA.

There are several mechanisms in which DnaA activity is down regulated during stress. For instance, DnaA hyper-proteolysis by Lon is activated during proteomic toxicity and heat shock (Jonas et al. 2013). Under carbon starvation, DnaA levels drops dramatically. This mechanism seems to be post-transcriptional through some regulatory element of *dnaA's* 5' untranslated region in *C. crescentus* (Leslie et al. 2015). However, the alarmone, guanosine penta- or tetraphosphate ((p)ppGpp), which is synthesized and upregulated during stringent response causes the destabilization of DnaA and stabilization of CtrA (Boutte, Henry, and Crosson 2012, Lesley and Shapiro 2008), as well.

Although role of DnaA and CtrA have been tested under several conditions, many other stress responses seem to cause the same effect in *C. crescentus*. Using single-cell microscopy, *C. crescentus* strains exposed to sodium chloride, which induces osmotic up-shock, causes cellular arrest; interestingly cells undergoing active membrane growth and rearrangement were hypersensitive to osmotic up-shock (Mathis and Ackermann 2016). Converse to the previously mentioned mechanisms, CtrA seem to be rapidly degraded under this stress condition (Heinrich, Sobetzko, and Jonas 2016), thus leaving an unknown regulatory regime that is underlying this specific stress response.

# Chapter 9: Materials and Methods

## 9.1 Experimental model and subject details

### Growth Media and Conditions

*C. crescentus* was cultivated on peptone-yeast extract (PYE)-agar (0.2% peptone, 0.1% yeast extract, 1.5% agar, 1 mM $MgSO_4$, 0.5 mM $CaCl_2$) (Ely 1991) at 30°C. Antibiotics were used at the following concentrations on this solid medium: kanamycin 25 μg/ml; tetracycline 1 μg/ml; and chloramphenicol 2 μg/ml.

For liquid culture, *C. crescentus* was cultivated in either PYE or in M2X defined medium (Ely 1991). PYE liquid: 0.2%(w/v) peptone, 0.1%(w/v) yeast extract, 1 mM $MgSO_4$, and 0.5 mM $CaCl_2$, autoclaved before use. M2X defined medium: 0.15% (w/v) xylose, 0.5 mM $CaCl_2$, 0.5 mM $MgSO_4$, 0.01 mM Fe Chelate, and 1x M2 salts, filtered with a 0.22 micron bottle top filter. One liter of 20x M2 stock was prepared by mixing 17.4g $Na_2HPO_4$, 10.6 $KH_2PO_4$, and 10g $NH_4Cl$. To induce gene expression from the *vanA* promoter, 500 μM vanillate (final concentration) was added. Antibiotics were used at the following concentrations in liquid medium: kanamycin 5 μg/ml, tetracycline 1 μg/ml, nalidixic acid 20 μg/ml, and chloramphenicol 2 μg/ml.

For cultivation of *E. coli* in liquid medium, we used lysogeny broth (LB). Antibiotics were used at the following concentrations: ampicillin 100 μg/ml, kanamycin 50 μg/ml, tetracycline 12 μg/ml, and chloramphenicol 20 μg/ml.

**Strain construction**

All *C. crescentus* experiments were conducted using strain CB15 (Poindexter 1964) and derivatives thereof. Plasmids were conjugated into CB15 (Ely 1991) using the *E. coli* helper strain FC3 (Finan et al. 1986). Conjugations were performed by mixing the donor *E. coli* strain, FC3, and the CB15 recipient strain in a 1:1:5 ratio. Mixed cells were pelleted for 2 minutes at 15,000xg, resuspended in 100 µL, and spotted on a nonselective PYE-agar plate for 12-24 hours. Exconjugants containing the desired plasmid were spread on PYE agar containing the plasmid-specified antibiotic for selection. The antibiotic nalidixic acid (20 µg/ml) was used to counter-select against both *E. coli* strains (helper and plasmid donor).

Gene deletion and nucleotide substitution strains were generated using the integrating plasmid pNPTS138 (Ried and Collmer 1987). pNPTS138 transformation and integration occurs at a chromosomal site homologous to the insertion sequence in pNPTS138. Exconjugants with pNPTS138 plasmids were selected on PYE agar plates with 5 µg/ml kanamycin; 20 µg/ml nalidixic acid selected against the *E. coli* donor strain. Single colony exconjugants were inoculated into liquid PYE or M2X for 6-16 hours in a rolling 30°C incubator for non-selective growth. Nonselective liquid growth allows for a second recombination event to occur, which either restores the native locus or replaces the native locus with the insertion sequence that was engineered into pNPTS138. Counter-selection for the second recombination of pNPTS138 was carried out on PYE agar with 3% (w/v) sucrose. This selects for loss of the *sacB* gene during the second crossover event. Colonies were subjected to PCR genotyping and/or sequencing to identify to confirm the allele replacement.

Other strains utilized in this study originate from (Herrou et al. 2010), (Purcell et al. 2007), and (Foreman, Fiebig, and Crosson 2012). Complete strain list is listed in Table 2.

The Δ*gsrN* strains and Δ*sigT* strains were complemented by introducing the gene at an ectopic locus (either *vanA* or *xylX*) utilizing the integrating plasmids: pMT552, pMT674, and pMT680. pMT674 and pMT680 carry a chloramphenicol resistance marker gene (*cat*) and pMT552 carries a kanamycin resistance marker gene (*npt1*) (Thanbichler, Iniesta, and Shapiro 2007). pMT552 and pMT674 integrate into the *vanA* gene and pMT680 integrates into the *xylX* gene. Transformation of ectopic complementation plasmids conjugated (as described earlier). Introduction of *gsrN* complementation was done in the reverse direction of the inducible promoters. Introduction of *katG* was done in-frame in the same direction of the inducible promoters.

Replicating plasmids pPR9TT and pRKlac290 were conjugated as previously described earlier. pPR9TT and pRKlac290 were selected using tetracycline and chloramphenicol, respectively.

pMal-MBP-PP7CPHis was transformed into *E. coli* Rosetta by electroporation and plated on LB plates with ampicillin 100 μg/ml.

**Plasmid construction**

Plasmid pNPTS138 was used to perform allele replacements and to generate gene deletions (Ried and Collmer 1987, West, Yang, and Stephens 2002). Primers for in-frame deletions and GeneBlocks (Gblocks) are listed in Table 3. Gene fragments were created by splice-overlap-

extension and ligated into the digested pNPTS138 vector at restriction enzyme sites (HindIII, SpeI) or gene fragments were stitched together using Gibson assembly. pNPTS138 contains a *kan*$^{R}$ (*npt1*) antibiotic resistance marker and the counter-selectable marker gene *sacB*, which encodes levansucrase

Plasmids for *gsrN* genetic complementation experiments carried wild-type or mutant *gsrN* alleles cloned antisense into a vanillate inducible(*vanA*)-promoter. An in-frame stop codon was designed at a restriction enzyme site downstream of the *vanA* promoter to ensure translational read-through of the *vanA* transcript did not disrupt *gsrN* transcription. Tandem *gsrN* alleles (overexpression by multiple copies of *gsrN*) were constructed using Gblocks with unique ends for Gibson assembly into pMT552. Plasmids for genetic complementation of the *katG* mutant were constructed by cloning *katG* in-frame with the vanillate and xylose-inducible promoters of pMT674 and pMT680, respectively, at the NdeI and KpnI restriction sites. *katG* complementation plasmids did not include the 5' untranslated region (UTR) of *katG*.

Beta-galactosidase transcriptional and translational reporters utilized pRKlac290 (Ely 1991) and pPR9TT (Santos et al. 2001) replicating plasmids, respectively. Transcriptional reporters of *gsrN* contained upstream and promoter sequences of *gsrN* cloned into the EcoRI and HindIII sites of pPRKlac290. Translational reporters of *katG* contained the 191 nucleotides 3' of the annotated *katG* transcriptional start site (Zhou et al. 2015a) cloned into pPR9TT at HindIII and KpnI.

Protein expression plasmid pMal was used to express a maltose binding protein (MBP) fused to the N-terminus of a Pseudomonas Phage 7 coat protein fused to a His-tag at its C-terminus (to

generate MBP-PP7CP-His). The PP7CPHis protein sequence was amplified out of pET283xFlagPP7CPHis and inserted into pMal at SalI and EcoRI restriction sites. pET283xFlagPP7CPHis was a gift from Alex Ruthenburg and originates from Kathleen Collins (Addgene plasmid # 28174).

## 9.2. Experimental method details

**Hydrogen peroxide/osmotic stress assays**

Liquid cultures were passaged several times before stress treatment to insure that population growth rate and density was as consistent as possible prior to addition of hydrogen peroxide (oxidative stress) or sucrose (hyperosmotic stress). Briefly, starter cultures were inoculated in liquid M2X medium from colonies grown on PYE-agar plates. Cultures were grown overnight at 30°C in a rolling incubator. Overnight cultures were then diluted back to an optical density reading of 0.05 at 660 nm ($OD_{660}$=0.05) and grown in a rolling incubator at 30°C for 7-10 hours. After this period, cultures were re-diluted with M2X to $OD_{660}$=0.025 and grown for 16 hours at 30°C in a rolling incubator. After this period, $OD_{660}$ was consistently 0.85-0.90. These cultures were then diluted to $OD_{660}$=0.05 and grown for 1 hour and split into two tubes. One tube received stress treatment and the other tube was untreated. Treated cultures were subjected to either hydrogen peroxide or sucrose.

For stress treatment, we used a freshly prepared 10 mM $H_2O_2$ solution diluted from a 30% (w/w) stock bottle (stock never more than 3 months old) or a stock of 80% (w/v) sucrose. The amount of 10 mM $H_2O_2$ added for stress perturbation depended on the volume of the culture and the

desired final concentration of $H_2O_2$. Final volumes assessed in our studies are described for each experiment throughout this manuscript.

Treated cultures and untreated cultures were subsequently tittered (10 µL sample in 90 µL of PYE) by initially diluting into 96-well plates. 5 µL spots from each dilution were plated on PYE-agar. Once spots dried, plates were incubated at 30°C for 2 days. Clearly visible colonies begin to form after 36 hours in the incubator.

The difference in colony forming units (CFU) between treated and untreated cultures was calculated using the following formula:

$$Relative\ CFU = \frac{Treated\ CFU \times 10^{x}}{\mathrm{Un}treated\ CFU \times 10^{y}} \quad (1)$$

Where x represents the countable (resolvable) dilution in which colonies are found in the treated sample dilution series and y represents the untreated sample dilution.

**β-galactosidase gene expression reporter assays**

To assess reporter gene expression, liquid cultures were passaged several times as described in the hydrogen peroxide/osmotic stress assays section above. However, cultures were placed in a 30°C shaker instead of a 30°C rolling incubator. Exponential phase cultures were harvested when the last starter culture (i.e., the $OD_{660}$=0.05 culture at the 16 hour time point) reached an $OD_{660}$ of 0.2-0.25. Stationary growth cultures were harvested when the exponential phase culture reached an OD660 of 0.85-0.90. Reporter assays in which the effect of stress treatment was quantified were conducted on exponential phase cultures that were split immediately before treatment.

β-galactosidase activity from chloroform-permeabilized cells was measured using the colorimetric substrate o-nitrophenyl-β-D-galactopyranoside (ONPG). 1 mL enzymatic reactions contained 200-250 µL of chloroform-permeabilized cells, 550-600 µL of Z-buffer (60 mM $Na_2HPO_4$, 40 mM $NaH_2PO_4$, 10 mM KCl, 1 mM $MgSO_4$), and 200 µL of 4 mg/mL ONPG in 0.1 M $KPO_4$, pH 7.0. Chloroform-permeabilized cell samples were prepared from 100-150 µL of culture, 100 µL of PYE, and 50 µL of chloroform (chloroform volume is not included in the final calculation of the 1 mL reaction). Chloroform-treated cells were vortexed for 5-10 seconds to facilitate permeabilization. Z buffer and ONPG were added directly to chloroform-permeabilized cells. Reactions were incubated in the dark at room temperature and quenched with 1 mL of 1 M $Na_2CO_3$.

Each reporter construct was optimized with different reaction times and different volumes of cells. Reaction time and volume for each reporter was empirically determined by the development of the yellow pigment from chloroform-permeabilized *C. crescentus* CB15 cultures. Strains harboring the pRKlac290 transcriptional reporter plasmid containing the established GSR promoter reporter $P_{sigU}$ or $P_{gsrN}$ used 100 µL of cells and were quenched after 10 minutes and 18 minutes, respectively. Strains containing pRKlac290 with the *katG* promoter ($P_{katG}$) used 150 µL of cells and were quenched after 12 minutes. Strains with the translational reporter plasmid pPR9TT containing the 5'UTR of *katG* (wild-type and RS constructs) used 150 µL of cells and were quenched after 4 minutes.

Miller units were calculated as:

$$MU = \frac{A_{420} \times 1000}{A_{660} \times t \times v} \quad (2)$$

Where $A_{420}$ is the absorbance of the quenched reaction measured at 420 nm on a Spectronic Genesys 20 spectrophotometer (ThermoFisher Scientific, Waltham, MA). $A_{660}$ is the optical density of the culture of cells used for the assay. □ is time in minutes between the addition of ONPG to the time of quenching with $Na_2CO_3$. □ is the volume in milliliters of the culture added to the reaction.

**TRIzol RNA extractions**

Cultures used for the extraction of RNA were passaged in the same manner outlined in the hydrogen peroxide/osmotic stress assays section above. Exponential phase cultures were harvested from the last starter (i.e., the $OD_{660}$=0.05 culture at the 16 hour time point) when it reached an $OD_{660}$ of 0.20-0.25. Stationary cultures were harvested when the final culture diluted to $OD_{660}$=0.025 reached an $OD_{660}$ of 0.85-0.90.

Exponential phase cultures ($OD_{660}$ of 0.20-0.25) harvested for extraction of RNA were pelleted at 15000xg for 3 minutes at ≈23°C (i.e. room temperature). Early stationary cultures ($OD_{660}$ of 0.85-0.90) were also pelleted at 15000xg for 30 seconds at ≈23°C. All media were aspirated using a vacuum flask. Cell pellets were resuspended in 1 mL of TRIzol™. The TRIzol resuspension was heated for 10 minutes at 65°C, treated with 200 µL of chloroform and hand shaken. The chloroform mixture was allowed to stand for 5 minutes and then spun down at 15000xg for 15 minutes at 4°C. Approximately 500 µL of clear aqueous phase was extracted and mixed with 500 µL of 100% isopropanol. Samples were then incubated at -20°C overnight. Overnight isopropanol precipitation was then spun down at 15000xg for 15 minutes at 4°C.

Isopropanol was aspirated, the pellet was washed in 1mL of 75% ethanol, and sample was spun down at 15000xg for 15 minutes at 4°C. Ethanol was removed from pellet, and the pellet was left to dry for 15 minutes. The RNA pellet was resuspended in 25 µL of nuclease-free $H_2O$.

**Radiolabeled Oligonucleotides**

Oligonucleotides were radiolabeled with T4 Polynucleotide Kinase (PNK). 10µL labeling reactions were composed of 1µL of PNK, 1µL PNK 10x Buffer, 2µL of 5 µM oligonucleotides (1 µM final concentration), 4µL $H_2O$, and 2 µL ATP, [γ-32P]. Reactions were incubated for a minimum of 37°C for 30 minutes. Total reactions were loaded onto a BioRad P-6 column to clean the reaction. Radiolabeled samples were stored at 4°C.

**Northern Blots**

RNA samples were resolved on a 10% acrylamide:bisacrylamide (29:1), 7 M urea, 89 mM Tris Borate pH 8.3, 2 mM $Na_2EDTA$ (TBE) 17 by 15 cm gel, run for 1 hour and 50 minutes at 12 Watts constant power in TBE running buffer. The amount of sample loaded was between 1-5 µg of RNA, mixed in a 1:1 ratio with 2x RNA loading dye (9 M urea, 100 mM EDTA, 0.02% w/v xylene cyanol, 0.02% w/v bromophenol blue). Samples were heated for 8 minutes at 75°C and then subjected to an ice bath for 1 minute before loading. Acrylamide gels with immobilized samples were then soaked in TBE buffer with ethidium bromide and imaged. Samples immobilized on the gel were transferred onto Zeta-Probe Blotting Membrane with a Trans-Blot® SD Semi-Dry Transfer Cell. Transfer was done at 400 mA constant current with voltage not exceeding 25V for 2 hours. Membrane was then subjected to two doses of 120 mJ/$cm^2$ UV radiation, using a Stratalinker UV cross-linker. Membranes were subsequently prehybridized 2

times for 30 minutes in hybridization buffer at 65°C in a rotating hybridization oven. Hybridization buffer is a variation of the Church and Gilbert hybridization buffer (20 mM sodium phosphate, pH 7, 300 mM NaCl, 1% SDS). Blots were hybridized with hybridization buffer containing the radiolabeled oligonucleotide probes described above. Hybridization buffer was always prepared so that GsrN probe concentration was approximately 1 nM, 5S rRNA probe concentration was approximately 2 pM, and tRNA-Tyr probe was 500 pM. Hybridization took place over 16 hours at 65°C in a rotating hybridization oven. Membranes were then incubated with wash buffer three times for 20 minutes at 65°C in a rotating hybridization oven. Wash buffer contained 20 mM sodium phosphate (pH 7.2), 300 mM NaCl, 2 mM EDTA, and 0.1% SDS. Membranes were then wrapped in plastic wrap and placed directly against a Molecular Dynamics Phosphor Screen. Screens were imaged with Personal Molecular Imager™ (PMI™) System. Membrane exposure time was determined using a Geiger counter: 100x 2 minutes, 10x 30-60 minutes, 1.0x 8-16 hours, 0.1x 48-72 hours.

Intensity of GsrN bands or *katG* mRNA dots was calculated by dividing the probe signal specific to GsrN or *katG* mRNA over the probe signal specific to the 5S rRNA multiplied by 100. Normalization of *katG* mRNA specific probes in the dot blot was carried out in a manner similar to that described for Northern blot, in which the 5S rRNA probe signal was used for normalization.

$$Normalized\ volume_x = \frac{volume_x(CNT * mm^2)}{volume_{5s\ rRNA\ probe}(CNT * mm^2)} \quad (3)$$

**Rifampicin transcription inhibition assays**

Liquid *C. crescentus* CB15 cultures were passaged in the same manner outlined in the hydrogen peroxide/osmotic stress assays section. However, cells for transcription inhibition assays were grown to an $OD_{660}$ of 0.2-0.25 from the last starter culture (i.e., inoculated from the $OD_{660}$=0.05 culture from 16 hour growth) and split across 6 tubes and labeled: untreated, 30 second treatment, 2 minute treatment, 4 minute treatment, 8 minute treatment, and 16 minute treatment. Untreated cultures were the 0 time point where no rifampicin was added. Rifampicin treated cultures were subjected to a final concentration of 10 μg/mL (from a 10 mg/mL stock in methanol) and were grown in a rolling incubator at 30°C. The 30 second rifampicin treatment refers to the centrifugation time (15000xg for 30 seconds at room temperature) to pellet the cells. Thus, the 30 second sample was immediately pelleted after exposure to rifampicin. 2 minute, 4 minute, 8 minute, and 16 minute samples were placed into a rolling incubator after exposure and were removed 30 seconds prior to their indicated time point, (i.e. 2 minute culture was removed from the incubator at 1 minute and 30 seconds). Pellets were then subjected to TRIzol extraction as described earlier. RNA extracts were subjected to Northern Blot analysis as described earlier.

Intensity of full-length and 5'isoform of GsrN bands were first adjusted to the intensity of the 5S rRNA control, as described in Equation 3. To plot the GsrN decay curve, all adjusted bands were then divided by the intensity of the 0 time point (untreated culture) and plotted in Prism v6.04.

$$Normalized\ timepoint\ volume_t = \frac{Normalized\ volume_t}{Normalized\ volume_0} \quad (4)$$

**Primer extension**

Primer extension was carried out using the SuperScript™ IV Reverse Transcriptase standalone enzyme. Total RNA from *gsrN*$^{++}$ and Δ*gsrN* strains was extracted from stationary cultures ($OD_{660}$=0.95-1.0) as described in the TRIzol extraction section. Primers for extension were first HPLC purified (Integrated DNA technologies) and radiolabeled as described in the Radiolabeled Oligonucleotides section.

Briefly, 14 µL annealing reactions comprised of the following final concentrations/amounts: 0.1 µM of gene specific radiolabeled primer, 0.3-0.5 mM of dNTPs, 2 µg of total RNA, and when necessary 0.5 mM ddNTPs. ddNTP reactions had a 3 dNTP:5 ddNTP ratio and were conducted using total RNA from *gsrN*$^{++}$. Annealing reactions were incubated at 65°C for 5 minutes and subsequently incubated on ice for at least 1 minute.

Extension reactions contained 14 µL annealing reactions with 6 µL of SuperScript™ IV Reverse Transcriptase master mix (final concentrations/amount 5 mM DTT, 2.0 U/µL, 1x SSIV buffer). Reactions were incubated at 50–55°C for 10 minutes and then incubated at 80°C for 10 minutes to inactivate the reaction.

After the extension reaction, 1 µL of RNase H was added to the mixture. This was incubated at 37°C for 20 minutes and mixed with 20 µL of 2x RNA loading dye. Reactions were subsequently heated for 8 minutes at 80°C, subjected to an ice bath for 1 minute, and loaded onto a 33.8 by 19.7 cm 20% acrylamide:bisacrylamide gel (as outlined in the Northern Blot section). Reactions were loaded on the gel along with a labeled Low Molecular Weight Marker (10-100

nt; Affymetrix/USB). Final amounts loaded were estimated using a Geiger counter, such that 10 mR/hr was loaded for each sample. Primer extension samples were resolved on the gel at 10 Watts constant power until unextended primer reached the bottom of the gel. The acrylamide gel was wrapped in plastic, exposed, and imaged as outlined in the Northern Blot section.

**Affinity purification of GsrN using a PP7hp-PP7cp system**

GsrN constructs containing a Pseudomonas phage 7 RNA hairpin (PP7hp) sequence were affinity purified using a hairpin-binding phage coat protein (PP7cp) immobilized on agarose beads. To prepare the coat protein, a 50 mL culture of *E. coli* Rosetta carrying an expression plasmid for PP7cp fused to maltose binding protein (MBP) at its N-terminus and a His-tag at its C-terminus (pMal-PP7cp-HIS) was grown at 37°C in a shaking incubator overnight in LB-ampicillin broth. Overnight cultures were rediluted and grown to $OD_{600}$=0.6. Cells were then induced with 1mM IPTG for 5 hours and spun down at 8000g at 4°C for 10 minutes. The cell pellet was resuspended in 6 mL of ice-cold lysis buffer (125 mM NaCl, 25 mM Tris-HCl pH 7.5, 10 mM Imidazole) and mechanically lysed in a LV1 Microfluidizer. Lysate was immediately added to 500 μL of amylose resin slurry that was prewashed with ice-cold lysis buffer. After the sample was loaded, beads were washed in 50x bead volume (~10mL) of ice-cold lysis buffer.

A 50 mL culture of *C. crescentus* Δ*gsrN* carrying plasmid pMT552 expressing PP7hp-tagged alleles of *gsrN* was grown at 30°C in a shaking incubator overnight in M2X medium. The culture was prepared from a starter and passaged as outlined in the hydrogen peroxide/osmotic stress assays section. Cells were grown to an $OD_{660}$=0.85-0.90. Cells were spun down at 8000g at 4°C for 15 minutes, resuspended in 6 mL of ice-cold lysis buffer, and mechanically lysed in a LV1

Microfluidizer. Lysate was immediately loaded onto a column of amylose resin on which pMal-PP7cp-HIS had been immobilized. After the sample was loaded, beads were washed in 50x bead volume (~10mL) of ice-cold lysis buffer. Elution of MBP-PP7cp-HIS bound to GsrN-PP7hp and associated biomolecules was completed over three 0.5 mL elution steps using 500 mM maltose. Each 0.5 mL elution was then mixed with equal volumes of acid-phenol for RNA extraction for RNA analysis, or equal volumes of SDS-Loading Buffer (200 mM Tris-HCl pH 6.8, 400 mM DTT, 8% SDS, 0.4% bromophenol blue, 40% glycerol) for protein analysis. For the RNA analysis, the three elution fractions were combined in an isopropanol precipitation step. RNA samples were subjected to DNase treatment as outlined in the RNA-seq sequencing section.

**Acid-Phenol RNA extraction**

Samples for acid-phenol extractions were mixed with equal volumes of acid-phenol and vortexed intermittently at room temperature for 10 minutes. Phenol mixture was spun down for 15 minutes at maximum speed at 4°C. The aqueous phase was extracted, cleaned with an equal volume of chloroform, and spun down for 15 minutes at maximum speed at 4°C. The aqueous phase was extracted from the organic and equal volumes of 100% isopropanol were added. Linear acrylamide was added to the isopropanol precipitation to improve pelleting (1 µL per 100 µL of isopropanol sample). Samples were then incubated at -20°C overnight and spun down at 15000xg for 15 minutes at 4°C. The isopropanol was aspirated, the pellet washed in 1 mL of 75% ethanol, and sample spun again at 15000xg for 15 minutes at 4°C. Ethanol was removed from the RNA pellet, and pellet was left to dry for 15 minutes. Pellet was resuspended in 25 µL of nuclease-free $H_2O$.

**RNA dot blot analysis**

Samples (≈3 μg) for dot blot analysis were mixed with equal volumes of 2x RNA loading dye as in a Northern Blot, and heated for 8 minutes at 75°C. Samples were then spotted on a Zeta-Probe Blotting Membrane and left to dry for 30 minutes. Spotted membrane was then subjected to two doses of 120 mJ/cm$^2$ UV radiation (Stratalinker UV crosslinker). The membrane was then prehybridized 2 times for 30 minutes in hybridization buffer at 65°C in a rotating hybridization oven. After pre-hybridization, we added radiolabeled oligonucleotide probes. Hybridization buffer with probes was always prepared so that each probe's concentration was approximately 1 nM. *katG* mRNA was first hybridized for 16 hours at 65°C in a rotating hybridization oven. Membrane was then washed with wash buffer three times, 20 minutes each at 65°C in a rotating hybridization oven. The blot was exposed for 48 hours to a Molecular Dynamics Phosphor screen and imaged on a Personal Molecular Imager as described above. Membrane was subsequently stripped with two rounds of boiling in 0.1% SDS solution and incubated for 30 minutes at 65°C in a rotating hybridization oven. Following stripping, the membrane was subjected to two rounds of prehybridization and then hybridized for 16 hours at 65°C in a rotating hybridization oven with the probe specific to the 5' end of GsrN. Membrane was then washed again with wash buffer three times for 20 minutes each at 65°C in a rotating hybridization oven. This GsrN blot was exposed for 36 hours to the phosphor screen and imaged. The membrane was stripped four times after GsrN probe exposure. Following stripping, membrane was again subjected to two rounds of prehybridization and then hybridized for 16 hours at 65°C in a rotating hybridization oven with the probe specific to 5S rRNA. Membrane washed with Wash Buffer three times, 20 minutes each at 65°C in a rotating hybridization oven. This 5S RNA blot was exposed to the phosphor screen for 1 hour and imaged.

**Western Blot analysis**

Strains from which protein samples were prepared for Western blot analysis were grown and passaged as outlined in the hydrogen peroxide/osmotic stress assays section. However, cultures were taken from the overnight 16-hour growth when $OD_{660}$ reached 0.85-0.90. 1 mL of these cultures was then pelleted, resuspended in 125 µL of Western blot buffer (10 mM Tris pH 7.4, 1 mM $CaCl_2$, and 5 µg/mL of DNase), and mixed with 125 µL SDS-Loading buffer. Samples were boiled at 85°C for 10 minutes, and 10-20 µL of each sample was loaded onto a Mini-PROTEAN TGX Precast Gradient Gel (4-20%) with Precision Plus Protein™ Kaleidoscope™ Prestained Protein Standards. Samples were resolved at 35 mA constant current in SDS running buffer (0.3% Tris, 18.8% Glycine, 0.1% SDS). Gels were run until the 25 kDa marker reached the bottom of the gel. Gel was transferred to an Immobilon®-P PVDF Membrane using a Mini Trans-Blot® Cell after preincubation in Western transfer buffer (0.3% Tris, 18.8% Glycine, 20% methanol). Transfer was carried out at 4°C, 100 V for 1 hour and 20 minutes in Western transfer buffer. The membrane was then blocked in 5% (w/v) powdered milk in Tris-buffered Saline Tween (TBST: 137 mM NaCl, 2.3 mM KCl, 20 mM Tris pH 7.4, 0.1% (v/v) of Tween 20) overnight at room temperature on a rotating platform. Primary incubation with a DYKDDDDK(i.e. M2)-Tag Monoclonal Antibody (clone FG4R) was carried out for 3 hours in 5% powdered milk TBST at room temperature on a rotating platform (4 µL antibody in 12 mL). Membrane was then washed 3 times in TBST for 15 minutes each at room temperature on a rotating platform. Secondary incubation with Goat anti-Mouse IgG (H+L) Secondary Antibody, HRP was for 1 hour at room temperature on a rotating platform (3 µL antibody in 15 mL). Finally, membrane was washed 3 times in TBST for 15 minutes each at room temperature on a

rotating platform. Chemiluminescence was performed using the SuperSignal™ West Femto Maximum Sensitivity Substrate and was imaged using a ChemiDoc MP Imaging System version 6.0. Chemiluminescence was measured using the ChemSens program with an exposure time of ~2 minutes.

Western blot lane normalization of KatG-M2 specific bands was conducted by normalizing total signal from the doublet signal in the M2 specific background to that of the non-specific band (found in strains were there was no M2 tagged KatG). Samples extracted on the same day were run on the same gel. Lane normalized samples were then normalized to the levels of KatG-M2 signal in the wild-type untreated samples for that specific gel.

$$Lane\ Normalized\ volume_x = \frac{Volume\ of\ top_x + Volume\ of\ bottom_x}{Volume\ of\ non-specific_x} \quad (5)$$

$$WT\ Normalized\ volume_x = \frac{Lane\ Normalized\ volume_x}{Lane\ Normalized\ volume_{untreated\ WT}} \quad (6)$$

**RNA-seq preparation**

Total RNA was extracted from cultures passaged similarly to the hydrogen peroxide/osmotic stress assays section. However, cultures were harvested at $OD_{660}$=0.85-0.90 from the 16-hour overnight growth. Total RNA extraction followed the procedure outlined in the TRIzol extraction section. Resuspended RNA pellets after the 75% ethanol wash were loaded onto an RNeasy Mini Kit column (100 µL sample, 350 µL RLT, 250 µL 100% ethanol). Immobilized RNA was then subjected to an on-column DNase digestion with TURBO™ DNase. DNase

treatment was repeated twice on the same column; each incubation was 30 minutes at 30°C with 70 µL solutions of DNase Turbo (7 µL DNase, 7 µL 10x Buffer, 56 µL di$H_2O$). RNA was eluted from column, rRNA was depleted using Ribo-Zero rRNA Removal (Gram-negative bacteria) Kit (Epicentre). RNA-seq libraries were prepared with an Illumina TruSeq stranded RNA kit according to manufacturer's instructions. The libraries were sequenced on an Illumina HiSeq 4000 at the University of Chicago Functional Genomics Facility.

**Soluble protein extraction for LC-MS/MS proteomics**

Total soluble protein for proteomic measurements was extracted from cultures passaged similarly to the hydrogen peroxide/osmotic stress assays section. However, harvested cultures were grown to an $OD_{660}$=0.85-0.90 in 50 mL of M2X during the 16-hour overnight growth in a 30°C shaking incubator. Cells were spun down at 8000g at 4°C for 15 minutes. Cells were resuspended in 6 mL of ice-cold lysis buffer. Cells were mechanically lysed in LV1 Microfluidizer. Lysate was then spun down at 8000g at 4°C for 15 minutes. Protein samples were resolved on a 12% MOPS buffered 1D Gel (Thermo Scientific) for 10 minutes at 200V constant. Gel was stained with Imperial Protein stain (Thermo Scientific), and a ~2 cm plug was digested with trypsin. Detailed trypsin digestion and peptide extraction by the facility is published in (Truman et al. 2012).

**LC-MS/MS data collection and analysis**

Samples for analysis were run on an electrospray tandem mass spectrometer (Thermo Q-Exactive Orbitrap), using a 70,000 RP survey scan in profile mode, m/z 360-2000 Fa, with lockmasses, followed by 20 MSMS HCD fragmentation scans at 17,500 resolution on doubly

and triply charged precursors. Single charged ions were excluded, and ions selected for MS/MS were placed on an exclusion list for 60s (Truman et al. 2012).

## 9.3. Computational method details

**Network construction**

RNAseq data (15 read files) was obtained from the NCBI GEO database from (Fang et al. 2013). Read files are comprised of 3 biological replicates of total RNA extracted from *C. crescentus* cultures at 5 time points across the cell cycle (0, 30, 60, 90, and 120 minutes post synchrony). Reads were mapped and quantified with Rockhopper 2.0 (Tjaden 2015). The estimated expression levels of each gene across the 5 time points were extracted from the "Expression" column in the "_transcripts.txt" file, using the "verbose" output. Expression of each gene across the 5 time points was normalized using python scripts as follows: for a given gene, the normalized expression of the gene at a time point, t, is divided by the sum of the gene's expression across all the time points, Equation 7. Thus the sum of a gene's normalized expression across the 5 time points would equal 1.

$$Let\ t\ \in T, where\ T = \{0, 30, 60, 90, 120\}$$

$$Normalized\ Transcript_t = \frac{Expression_t}{\sum^T Expression_t} \quad (7)$$

We computed Pearson's correlation coefficient based on normalized expression between all pairwise combinations of genes. Correlation coefficients were organized into a numpy.matrix data structure where each row and column corresponds to the same gene order. Correlation

coefficients less than 0 were not considered for this analysis and were assigned the value 0. We refer to this matrix as the Rho-matrix. The Rho-matrix is symmetric and the product of its diagonal is 1. The Rho-matrix represents the weighted edges of the network, where the value of 0 demonstrates no edge is drawn between nodes.

A one-dimensional weight matrix that corresponds to the rows and columns of the Rho-matrix was constructed as a numpy.matrix data structure with all values initialized at 0. Lastly, a key array was constructed in conjunction with the Rho-matrix and weight-matrix for initializing the assignment of weight and obtaining the final weights of the algorithm. The weight-matrix represents the weight of the nodes of the network and the key matrix represents the gene name of the node.

**Iterative Ranking: Matrices and Algorithms**

Iterative ranking algorithms are a class of analytical tools used to understand relationships between nodes of a given network. The iterative ranking algorithm used to dissect the general stress response in the transcription-based network follows:

Given the Rho-matrix (P) and weight-matrix (*f*), the weight-matrix after *t*-iterations is Equation 8.

$$f^t = \propto f^0 + (1-\propto)\mathrm{P}f^{t-1} \quad (8)$$

For Equation 8, let $\propto$ represent a dampening factor applied to the initialize $(t) = 0$ weight of the nodes, $f^0$. The final weights of the weight-matrix as $t \to \infty$ converge to a stable solution, Equation 9.

$$f^{\infty} = \propto [I - (1-\propto)\mathrm{P}]^{-1} f^0 \quad (9)$$

Algorithm and solution information was adapted from (Wang and Marcotte 2010).

Initial weight-matrix, ($f^0$), was created by assigning the weight 1.0 to the corresponding positions of the seven genes known to regulate the General Stress Response (GSR) of *C. crescentus*: *sigT*, *phyR*, *phyK, sigU, nepR, lovR*, and *lovK*. Normalization of the values of the Rho-matrix, P, was performed by normalizing each column such that each column has a sum equal to 1 and then repeating the same normalization process by rows.

**Iterative rank parameter tuning**

As described in Chapter 2.3.I. Tuning script is available at https://github.com/mtien/IterativeRank.

**Identification of $\sigma^T$-promoter motifs**

Motif finder utilized a python script that scans 200 nucleotides upstream of annotated transcriptional start sites (Zhou et al. 2015a) or predicted translational start sites (TSS) (Marks et al. 2010).

We built a simple python library to take in genomic FASTA files, find specified regions of interest, and extract 200 nucleotides from a given strand. We used the *Caulobacter crescentus* NA1000 annotation (CP001340) from NCBI as the input genomic file and used the predicted TSS (when available) or annotated gene start sites as the region and strand specifier. After locating the position and strand within the file, we extracted the 200 nucleotides directly upstream of the site of interest and put the regions into a character-match calculator. Our simple calculator reported a list of positions for -35 (GGAAC) and for -10 elements (CGTT) of $\sigma^T$-dependent promoters within the 200-nucleotide input string. Only strict matches to these elements were reported. Spacers were calculated between all pairwise -35 and -10 matches. We identified potential $\sigma^T$-dependent promoters by identifying consensus -35 to -10 sequences with 15-17 base spacing. Sequence logos were generated from (Crooks et al. 2004)

**IntaRNA analysis**

IntaRNA version 2.0.2 is a program within the Freiberg RNA Tools collection (Mann, Wright, and Backofen 2017a). To predict likely RNA-RNA associations between predicted unstructured regions within GsrN and its RNA targets, we input the sequence of GsrN as the query ncRNA sequence and a FASTA file of either: 1) windows significantly enriched in the GsrN(37)-PP7hp purification from our sliding window analysis with an additional 100 base pairs (50 bp on each side of the window) or 2) entire gene windows that showed significant enrichment from our Rockhopper analysis (Figure 5 – source table 3).

Output from IntaRNA 2.0.2 comprised a csv file of target binding sites and the corresponding GsrN binding sites. We extract the predicted binding sites of the targets with a python script and

parsed the targets into those predicted to bind the first exposed loop and the second exposed loop. Sequence logos were generated from (Crooks et al. 2004)

**Phylogenetic tree construction**

A 16S rRNA phylogenetic tree of Alphaproteobacteria was constructed by extracting 16S rRNA sequences for all species listed in Figure S7A and using the tree building package in Geneious 11.0.2 (Kearse et al. 2012). The tree was constructed using a global alignment with free end gaps and a cost matrix of 65% similarity (5.0/~4.0). The genetic distance model was the Tamura-Nei and the tree building method employed was neighbor-joining. *E. litoralis* was the out-group for tree construction.

**Prediction of *gsrN* homologs**

A homology search based on the sequence of GsrN was conducted using BLASTn (Altschul et al. 1990). This simple search provided a list of clear GsrN homologs in the Caulobacteraceae family (*Caulobacter*, *Brevundimonas*, and *Phenylobacterium*).

Identification of homologs in other genera relied on analysis of published transcriptomic data, searching specifically for gene expression from intergenic regions. Analyzed data included *Rhizobium etli* (Jans et al. 2013), *Sinorhizobium meliloti* (Valverde et al. 2008) and *Brucella abortus* (Kim et al. 2014). The prediction of GsrN homologs in *Rhodopseudomonas palustris* and *Bradyrhizobium diazoefficiens* is completely based on the proximity of a GsrN-like sequence to the GSR locus and the presence of a $\sigma^{ecfG}$-binding site in the predicted promoters of these predicted genes.

**Mapping reads from RNA-seq data**

RNA-seq read files (fastQ) were aligned with sequence files (fastA) using bowtie 2.0 (Langmead and Salzberg 2012). SAMTools was then used to calculate the depth and coverage of each nucleotide in the hit output file from bowtie 2.0 (Li et al. 2009). Normalization of reads per nucleotide was computed by normalizing each count to the total number of reads mapped per million to all of the CP001340.1 genome. Normalized reads per nucleotide was then plotted in Prism v6.04 where standard error and mean were calculated.

**RNA-seq analysis of mRNAs that co-elute with GsrN**

RNA-seq read files (fastQ) from the three replicate GsrN(37)::PP7hp purifications and duplicate PP7hp-GsrN-3' purifications were quantified and analyzed with Rockhopper 2.0 (Tjaden 2015). Reads were mapped to modified *C. crescentus* genome files (fastA, PTT, RNT) where the wild-type *gsrN* locus was replaced with the sequence of *gsrN(37)-PP7hp*. Using the "verbose output" option and the resulting "transcripts.txt" file, we pruned the dataset to find genes that had low FDR values ("qValue" < .05), were significantly enriched in GsrN(37)::PP7 ("Expression GsrN(37)-PP7hp" > "Expression PP7hp-GsrN-3'"), and had a high total number of reads that mapped to GsrN(37)::PP7 ("Expression GsrN(37)-PP7hp" >1000). This analysis provided a list of 35 candidate genes (Figure 5-source data 1).

The Rockhopper analysis package organizes reads into IGV (integrative Genomic Viewer) files. Upon visual inspection and spot validation of the 35 candidates in IGV, we found 26 genes with consistently higher signal across the three GsrN(37)::PP7hp purifications relative to PP7hp-

GsrN-3' control fractions. In some cases, reads mapped outside coding sequences. Such reads mapped proximal to the 5' end of annotated genes and to intergenic regions. We observed uneven read distribution across some annotated genes. Cases in which reads were not evenly distributed across a gene were typically not classified as significantly different from the control samples in "Expression" or "qValue" by Rockhopper even when a clear bias in read density was visually evident (most often at the 5' end of the gene).

As a second approach, we performed a systematic window annotation analysis to capture the unaccounted read density differences between the two purified fractions (GsrN(37)::PP7hp and the PP7hp-GsrN-3' negative control). Windows were generated by *in silico* fragmentation of the *C. crescentus* NA1000 genome sequence, designating 25 base pair windows across the genome. We prepared new annotated window files (FASTA, PTT, RNT) for wild-type, *gsrN(37)-PP7hp,* and *PP7hp-gsrN-3'*. The window identification number corresponds to the same sequence across the three different FASTA sequences.

Mapping and quantification of reads to these windows was conducted using the EDGE-pro analysis pipeline (Magoc, Wood, and Salzberg 2013). A caveat of EDGE-pro quantification is the potential misattribution of reads to input windows. EDGE-pro quantification does not take strand information into account when mapping reads to input windows.

Read quantification of the *gsrN(37)::PP7hp* purifications showed consistent differences in one of the three samples. *gsrN(37)::PP7hp* sample 1 contained 2.69% reads mapped to *gsrN(37)-PP7hp* while sample 2 and 3 had 15.78% and 14.04% mapped to *gsrN(37)-PP7hp* respectively.

Additionally, we observed that sample 1 had several genes that were strongly enriched in sample 1 and not in sample 2 and 3. Thus we employed a metric to balance the discrepancies between the three separate purifications. To minimize potential false positives, we calculated the average of all three samples and the average of samples 2 and 3. If the total average was 1.5 times greater than the sample 2 and 3 average, we assumed that the sample 1 artificially raised the average RPKM value and did not consider any data from any of the purifications in that specific window. The total window population decreased from 161713 windows to 109648 windows after this correction. This process is reflected in the https://github.com/mtien/Sliding_window_analysis script "remove_high_variant_windows.py".

From the RPKM values calculated with EDGE-pro, we used the R-package, DESeq (Anders and Huber 2010), to assess statistically significant differences between windows of expression. Candidate windows enriched in the GsrN(37)::PP7 fractions were identified using metrics similar to what is applied to traditional RNA-seq data. Briefly, we identified windows that had a low p-values ( pvalue< .10), were enriched in the GsrN(37)::PP7 ("baseMean GsrN(37)-PP7hp" > "baseMean PP7hp"), and had a high level of reads mapped to the gene in the GsrN(37)::PP7 ("baseMean GsrN(37)-PP7hp" >1000) (Figure 5-source data 2). Since the read density of windows from the total RNA extracted from the PP7-purification did not converge when estimating dispersion with a general linear model, we added total RNA seq read density from wild-type strains grown in stationary phase to help model the dispersion for the negative binomial analysis by DESeq, GSE106168.

Adjacent significant windows were then combined and mapped onto the annotated genome of *C. crescentus*. In order to correct for strand information lost in EDGE-pro quantitation, bowtie file information was used to define the strand of reads mapped to combined significant windows (Table 1).

**RNA-seq processing of total RNA**

Analysis of whole genome RNA-seq data was conducted using the CLC Genomics Workbench (Qiagen). Reads were mapped to the *C. crescentus* NA1000 genome (accession CP001340.1) (Marks et al. 2010). Differential expression was determined using Wald test in the CLC Workbench suite (Figure 8- source table 2).

**LC-MS/MS processing of total soluble protein**

Raw files of LC-MS/MS data collected on wild-type, Δ*gsrN, and* $gsrN^{++}$ were processed using the MaxQuant software suitev1.5.1.2 (Cox et al. 2014). Samples were run against a FASTA file of proteins from the UniProt database (UP000001364) and standard contaminants. The label free quantitation (LFQ) option was turned on. Fixed modification included carbamidomethyl (C) and variable modifications were acetyl or formyl (N-term) and oxidation (M). Protein group files were created for three comparisons: wild-type versus Δ*gsrN*, Δ*gsrN* versus $gsrN^{++}$, and wild-type versus $gsrN^{++}$ samples.

LFQ values for each protein group were compiled across all three runs and used as estimated protein quantities in our analyses (Figure 8A). Each strain had a total of 6 LFQ values for every protein group, 2 from each of the comparisons. Average LFQ values were only calculated if 3 or

more LFQ values were found for a given protein group. This allowed for protein groups that had a sufficient amount of signal across all the samples and analyses to be considered for comparison. Once averages for each protein group were calculated, we calculated the fold change between samples from different backgrounds by dividing the averages and taking the log-2 transformation (log2Fold).

Multiple t-tests were conducted using all 6 LFQ value obtained across the three MaxQuant runs. We used the multiple t-test analysis from GraphPad Prism version 6.04 for MacOS, GraphPad Software, La Jolla California USA, www.graphpad.com. The false discovery rate (Q) value was set to 5.000% and each row was analyzed individually, without assuming a consistent SD.

Table 2. Strains used in this work.
Table 2 continued on page 106 and 107.

| Identifier | Genotype | Source |
|---|---|---|
| FC#19 | CB15 | Poindexter, 1964 |
| FC#674 | CB15 ΔsigT | Herrou et al., 2010 |
| FC#2838 | CB15 ΔCCNA_R0081( ΔgsrN ) | This work |
| FC#799 | CB15 ΔphyR | Herrou et al., 2010 |
| FC#1688 | CB15 ΔphyK | Foreman et al., 2012 |
| FC#2394 | CB15 ΔlovR | This work |
| FC#1188 | CB15 ΔlovK | Purcell et al., 2007 |
| FC#672 | CB15 ΔsigU | This work |
| FC#2839 | CB15 ΔCCNA_00535 | This work |
| FC#2840 | CB15 ΔCCNA_03669 | This work |
| FC#2841 | CB15 ΔCCNA_00201 | This work |
| FC#2842 | CB15 ΔCCNA_00987 | This work |
| FC#2843 | CB15 ΔCCNA_03668 | This work |
| FC#700 | CB15 ΔCCNA_01418 | This work |
| FC#670 | CB15 ΔCCNA_00282 | This work |
| FC#671 | CB15 ΔCCNA_01601 | This work |
| FC#2844 | CB15 ΔCCNA_03587 | This work |
| FC#675 | CB15 ΔCCNA_01237 | This work |
| FC#2845 | CB15 /pRKlac290- PgsrN | This work |
| FC#2846 | CB15 ΔsigT /pRKlac290- PgsrN | This work |
| FC#642 | CB15 /pRKlac290-PsigU | Foreman et al., 2012 |
| FC#820 | CB15 ΔsigT /pRKlac290-PsigU | Foreman et al., 2012 |
| FC#2847 | CB15 ΔgsrN/pRKlac290-PsigU | This work |
| FC#814 | CB15 ΔphyR/pRKlac290-PsigU | Foreman et al., 2012 |
| FC#1701 | CB15 ΔphyK/pRKlac290-PsigU | Foreman et al., 2012 |
| FC#2848 | CB15 vanA::pMT552-3gsrN /pRKlac290-PsigU | This work |
| FC#647 | CB15 vanA::pMT552-EV | This work |
| FC#2849 | CB15 ΔgsrN vanA::pMT552-EV | This work |
| FC#2850 | CB15 ΔgsrN vanA::pMT552-PgsrN-gsrN | This work |
| FC#2851 | CB15 vanA::pMT552-gsrN | This work |
| FC#2852 | CB15 ΔgsrN vanA::pMT552-3gsrN | This work |
| FC#2853 | CB15 vanA::pMT552-3gsrN (gsrN++) | This work |
| FC#2854 | CB15 ΔsigT vanA::pMT552-PvanA-gsrN | This work |
| FC#2855 | CB15 ΔsigT xylX::pMT585-PxylX-gsrN | This work |
| FC#451 | CB15 xylX::pMT674-EV | |
| FC#2856 | CB15 ΔgsrN xylX::pMT674-EV | This work |
| FC#2857 | CB15 ΔgsrN xylX::pMT674-gsrN | This work |
| FC#2858 | CB15 ΔgsrN xylX::pMT674-gsrNΔ(1-58) (gsrNΔ3') | This work |

| Table 2 continued. | | |
|---|---|---|
| FC#2859 | CB15 ΔgsrN xylX::[pMT674-2gsrNΔ(1-58) | This work |
| FC#2860 | CB15 gsrNΔ(10-50) (gsrNΔ5') | This work |
| FC#2861 | CB15 gsrNΔ(10-50) xylX::pMT674-gsrN | This work |
| FC#2862 | CB15 gsrNΔ(10-50) xylX::pMT674-gsrN(1-58) | This work |
| FC#2863 | CB15 ΔgsrN vanA::pMT552-5'PP7hp-gsrN | This work |
| FC#2864 | CB15 ΔgsrN vanA::pMT552-gsrN(37)PP7hp | This work |
| FC#2865 | CB15 ΔgsrN vanA::pMT552-PgsrN-gsrN(54)PP7hp | This work |
| FC#2866 | CB15 ΔgsrN vanA::pMT552-PgsrN-gsrN(59)PP7hp | This work |
| FC#2867 | CB15 ΔgsrN vanA::pMT552-PgsrN-gsrN(67)PP7hp | This work |
| FC#2868 | CB15 ΔgsrN vanA::pMT552-PgsrN-gsrN(93)PP7hp | This work |
| FC#2869 | CB15 ΔgsrN vanA::pMT552-PgsrN-gsrN(1-5)-PP7hp-gsrN(56-112) | This work |
| FC#2870 | CB15 /pRKlac290-promoter katG | This work |
| FC#2871 | CB15 ΔgsrN /pRKlac290-promoter katG | This work |
| FC#2872 | CB15 vanA::pMT552-3gsrN /pRKlac290-promoter katG | This work |
| FC#2873 | CB15 /pPR9TT-PrpoD(vanA)-katG(1-190) | This work |
| FC#2874 | CB15 ΔgsrN /pPR9TT-PrpoD(vanA)-katG(1-190) | This work |
| FC#2875 | CB15 vanA::pMT552-3gsrN /pPR9TT-PrpoD(vanA)-katG(1-190) | This work |
| FC#2876 | CB15 vanA::pMT552-EV/pPR9TT-PrpoD(vanA)-katG(RS)(1-190) | This work |
| FC#2877 | CB15 ΔgsrN vanA::pMT552-EV/pPR9TT-PrpoD(vanA)-katG(RS)(1-190) | This work |
| FC#2878 | CB15 vanA::pMT552-3gsrN /pPR9TT-PrpoD(vanA)-katG(RS)(1-190) | This work |
| FC#2879 | CB15 ΔgsrN vanA::pMT552-gsrN(RS) | This work |
| FC#2880 | CB15 ΔgsrN vanA::pMT552-3gsrN(RS) | This work |
| FC#2881 | CB15 ΔgsrN vanA::pMT552-3gsrN(RS) xylX::pMT680-3gsrN(RS) | This work |
| FC#2882 | CB15 ΔgsrN vanA::pMT552-3gsrN /pPR9TT-PrpoD(vanA)-katG(1-190) | This work |
| FC#2883 | CB15 ΔgsrN vanA::pMT552-3gsrN /pPR9TT-PrpoD(vanA)-katG(RS)(1-190) | This work |
| FC#2884 | CB15 ΔgsrN vanA::pMT552-3gsrN(RS) /pPR9TT-PrpoD(vanA)-katG(1-190) | This work |
| FC#2885 | CB15 ΔgsrN vanA::pMT552-3gsrN(RS) /pPR9TT-PrpoD(vanA)-katG(RS)(1-190) | This work |
| FC#2886 | CB15 katG(RS) ΔgsrN vanA::pMT552-3gsrN | This work |
| FC#2887 | CB15 katG(RS) ΔgsrN vanA::pMT552-3gsrN(RS) | This work |
| FC#2888 | CB15 ΔkatG /pRKlac290- gsrN promoter | This work |
| FC#2889 | CB15 katG(RS) /pRKlac290- gsrN promoter | This work |
| FC#2890 | CB15 katG-M2 | This work |
| FC#2891 | CB15 katG-M2 ΔgsrN | This work |
| FC#2892 | CB15 katG-M2 vanA::pMT552-3gsrN | This work |
| FC#2893 | Rosetta (DE3) pLysS / pMal-MBP-PP7cp-HIS | This work |

| Table 2 continued. | | |
|---|---|---|
| FC#2931 | CB15 ΔkatG | This work |
| FC#2932 | CB15 katG(RS) | This work |
| FC#2933 | CB15 ΔkatG ΔgsrN | This work |
| FC#2965 | CB15 xylX::pMT680-katG | This work |
| FC#2966 | CB15 ΔkatG xylX::pMT680-katG | This work |
| FC#2967 | CB15 ΔkatG vanA::pMT552-3gsrN | This work |
| FC#2972 | CB15 Δhfq | This work |
| FC#2973 | CB15 xylX::pMT612-hfq | This work |
| FC#2974 | CB15 Δhfq pMT552-3gsrN | This work |
| FC#2975 | CB15 Δhfq pMT552-EV | This work |

Identifier: Refers the Fiebig-Crosson collection of frozen strains.
Genotype: Refers to the genotype of the strain
Source: Refers to origin of the strain

Table 3. Plasmids used in this work

Table 3 continued on page 109.

| Identifier | Genotype | Source |
|---|---|---|
| FC#2894 | Top10 /pNPTS138-CCNA_R0081(10-95) | This work |
| FC#2383 | Top10 /pNPTS138-lovR | This work |
| FC#1209 | Top10 /pNPTS138-sigU | This work |
| FC#2895 | Top10 /pNPTS138-CCNA_00535 | This work |
| FC#2896 | Top10 /pNPTS138-CCNA_03669 | This work |
| FC#2897 | Top10 /pNPTS138-CCNA_00201 | This work |
| FC#2898 | Top10 /pNPTS138-CCNA_00987 | This work |
| FC#2899 | Top10 /pNPTS138-CCNA_03668 | This work |
| FC#2900 | Top10 /pNPTS138-CCNA_03587 | This work |
| FC#2901 | Top10 /pNPTS138-CCNA_01418 | This work |
| FC#2902 | Top10 /pNPTS138-CCNA_00282 | This work |
| FC#2903 | Top10 /pNPTS138-CCNA_01601 | This work |
| FC#2904 | Top10 /pNPTS138-CCNA_01237 | This work |
| FC#2905 | Top10 /pRKlac290-PgsrN promoter | This work |
| FC#634 | Top10 /pRKlac290-PsigU promoter | Foreman et al., 2012 |
| FC#2906 | Top10 /pMT552-PgsrN-gsrN | This work |
| FC#2907 | Top10 /pMT552-(PgsrN-gsrN)x3 | This work |
| FC#2908 | Top10 /pMT552-PrpoD(vanA)-gsrN | This work |
| FC#2909 | Top10 /pMT585-PrpoD(xylX)-gsrN | This work |
| FC#2910 | Top10 /pMT674-PgsrN-gsrN(1-58) | This work |
| FC#2911 | Top10 /pMT674-PgsrN-gsrN | This work |
| FC#2912 | Top10 /pNPTS138-CCNA_R0081(10-50) | This work |
| FC#2913 | Top10 /pMT552-PgsrN-5'PP7hp-gsrN | This work |
| FC#2914 | Top10 /pMT552-PgsrN-gsrN(37)PP7hp | This work |
| FC#2915 | Top10 /pMT552-PgsrN-gsrN(54)PP7hp | This work |
| FC#2916 | Top10 /pMT552-PgsrN-gsrN(59)PP7hp | This work |
| FC#2917 | Top10 /pMT552-PgsrN-gsrN(67)PP7hp | This work |
| FC#2918 | Top10 /pMT552-PgsrN-gsrN(93)PP7hp | This work |
| FC#2919 | Top10 /pMT552-PgsrN-gsrN(1-5)-PP7hp-gsrN(56-112) | This work |
| FC#2920 | Top10/ pMal-MBP-PP7cp-HIS | This work |
| FC#2921 | Top10/ pET28_3xFLAG_PP7CP_His | Gift from Alex Ruthenburg, originally from Kathleen Collins (Addgene Plasmid #21339) |
| FC#2922 | Top10 /pRKlac290-promoter katG | This work |
| FC#2923 | Top10 /pPR9TT-PrpoD(vanA)-katG(1-190) | This work |
| FC#2924 | Top10 /pPR9TT-PrpoD(vanA)-katG(RS)(1-190) | This work |
| FC#2925 | Top10 /pMT552-PgsrN-gsrN(RS) | This work |
| FC#2926 | Top10 /pMT552-(PgsrN-gsrN(RS))x3 | This work |

| Table 3 continued. | | |
|---|---|---|
| FC#2927 | Top10 /pMT680-(PgsrN-gsrN(RS))x3 | This work |
| FC#2928 | Top10 /pNPTS138-CCNA_03138 (katG) | This work |
| FC#2929 | Top10 /pNPTS138-CCNA_03138 (katG(RS)) | This work |
| FC#2930 | Top10 /pNPTS138-CCNA_03138 (katG::M2) | This work |
| MTLS# 4266 | Top10/pMT552-EV | Thanbichler et al., 2007 |
| MTLS# 4396 | Top10/pMT585-EV | Thanbichler et al., 2007 |
| MTLS# 4259 | Top10/pMT674-EV | Thanbichler et al., 2007 |
| MTLS# 4389 | Top10/pMT680-EV | Thanbichler et al., 2007 |
| FC#55 | DH10B/pNPTS138 | Ried and Collmer, 1987 |
| FC#54 | S17-1:E.coli 294::RP4-2(Tc::Mu)(Km::T7)/pRKlac290 | Ely, 1991 |
| FC#1101 | Top10/pPR9TT | Santos et al., 2001 |
| FC#118 | Top10/pMal | New England Biolabs (E8200S) |
| FC#2968 | Top10/ pMT680-katG | This work |
| FC#2976 | Top10/ pNPTS138-CCNA_01819(hfq) | This work |
| FC#2977 | Top10/ pMT612-hfq | This work |

Identifier: Refers the Fiebig-Crosson collection of frozen strains carrying the plasmids.
Genotype: Refers to the genotype of the strain carrying the plasmid
Source: Refers to origin of the plasmid.

| Table 4. Nucleic acids used in this work<br>Table 4 continued on pages 111-118. | | | | | |
|---|---|---|---|---|---|
| Identifier | Name | Sequence | Source | Plasmid | Restriction Enzyme Sites |
| Primers | | | | | |
| 1016 | R0081 KO A Up-F | ATCGACTAGTAAGCAAGGAGGACCGACATG | This work | FC#2894 | SpeI |
| 1017 | R0081 KO B Up-R | GTCCGGCGCCTTCGCACTCACAAAA | This work | FC#2894 | |
| 1018 | R0081 KO C Down-F | AAGGCGCCGGACCGCCTTTCCTTATG | This work | FC#2894 | |
| 1019 | R0081 KO D Down-R | ATCGAAGCTTAACATCGATGCGCTTCATGC | This work | FC#2894 | HindIII |
| 1082 | R0081 (10-50) KO B Up-R | GAAGGCCGCCTTCGCACTCACAAAA | This work | FC#2915 | |
| 1083 | R0081 (10-50) KO C Down-F | AAGGCGGCCTTCGCACCCCTTTGTTG | This work | FC#2915 | |
| 1022 | PgsrN F | CATAGAATTCGGCAACGCCTCTTCAAAAA | This work | FC#2905 | |
| 1023 | PgsrN R | CATAAAGCTTCCGCCTTCGCACTCACAA | This work | FC#2905 | |
| 1044 | 03668_KO A Up-F | ATCGACTAGTCTAAGCAAGGAGGACCGACA | This work | FC#2899 | SpeI |
| 1045 | 03668_KO B Up-R | ATGATGATGTGGACGCCATAGCCACG | This work | FC#2899 | |
| 1046 | 03668_KO C Down-F | CCACATCATCATGACGTCGCTCCCCT | This work | FC#2899 | |
| 1047 | 03668_KO D Down-R | ATCGAAGCTTGTCGCCGAAGTTCTGCACGA | This work | FC#2899 | HindIII |
| 1048 | 00987_KO A Up-F | ATCGACTAGTCTCCTATCCGCTGACGCCGT | This work | FC#2898 | SpeI |
| 1049 | 00987_KO B Up-R | TAGCCTGCTCATGGCCAAATCCTCTG | This work | FC#2898 | |
| 1050 | 00987_KO C Down-F | ATGAGCAGGCTATGACTGACAAGGCG | This work | FC#2898 | |
| 1051 | 00987_KO D Down-R | ATCGAAGCTTGAGCTGGCGAATGTCGAGAT | This work | FC#2898 | HindIII |
| 1052 | 03587_KO A Up-F | ATCGACTAGTCGATCCGCATGCTGATCAAG | This work | FC#2900 | SpeI |
| 1053 | 03587_KO B Up-R | ATGCGTTAAGTCTTAGGCCAAGGTCC | This work | FC#2900 | |

| Table 4 continued. | | | | | |
|---|---|---|---|---|---|
| 1054 | 03587_KO C Down-F | GACTTAACGCATCATTTTCTCCGCTG | This work | FC#2900 | |
| 1055 | 03587_KO D Down-R | ATCGAAGCTTTCATTGATCCGGATCGCTTG | This work | FC#2900 | HindIII |
| 1056 | 03669_KO A Up-F | ATCGACTAGTATCATGACGTCGCTCCCCTG | This work | FC#2896 | SpeI |
| 1057 | 03669_KO B Up-R | ACGACCTATTACTGGGAAGACGGCGC | This work | FC#2896 | |
| 1058 | 03669_KO C Down-F | GTAATAGGTCGTCTTCATCTTGGCTC | This work | FC#2896 | |
| 1059 | 03669_KO D Down-R | ATCGAAGCTTTGCGGTAGCATGTTCATGGT | This work | FC#2896 | HindIII |
| 1060 | 00201_KO A Up-F | ATCGACTAGTCCTTCAGGTCGTAGTGGTCT | This work | FC#2897 | SpeI |
| 1061 | 00201_KO B Up-R | GGTCAGGAATTTCGGCATGGTCGCAA | This work | FC#2898 | |
| 1062 | 00201_KO C Down-F | AAATTCCTGACCTGATGGCTCTACGG | This work | FC#2899 | |
| 1063 | 00201_KO D Down-R | ATCGAAGCTTTGTGCACGCCCTCGTACATC | This work | FC#2900 | HindIII |
| 1009 | 00535 KO A Up-F | CATACTAGTGCTCACCTTCGTCGATGACA | This work | FC#2895 | SpeI |
| 1010 | 00535 KO B Up-R | GCGTAGGACGCCGGCGAGCTTAAG | This work | FC#2895 | |
| 1011 | 00535 KO C Down-F | GGCGTCCTACGCAACGGCTGATCG | This work | FC#2895 | |
| 1012 | 00535 KO D Down-R | CATAAAGCTTCTTGGGCTCTTCGAACTCGT | This work | FC#2895 | HindIII |
| | lovR KO A Up-F | AAAAGAATTCGTCGCTGAGAGGTATGCTGG | This work | FC#2383 | EcoRI |
| | lovR KO B Up-R | CAGGGTGTCAATCAGAACATCGAGCGCTGG | This work | FC#2383 | |
| | lovR KO C Down-F | ATGTTCTGATTGACACCCTGACCTGATACGC | This work | FC#2383 | |
| | lovR KO D Down-R | AAAAGGATCCATTTCCTTGTTGATCGCGCC | This work | FC#2383 | BamI |
| | sigU KO A Up-F | GAATTCCGGGATGACATGGGATTTT | This work | FC#1209 | EcoRI |
| | sigU KO B Up-R | CTCGAGTTCGAAGCTCTTCTGAGTCTG | This work | FC#1209 | XhoI |
| | sigU KO C Down-F | CTCGAGATGCTGCGCGATCTGGAAC | This work | FC#1209 | XhoI |
| | sigU KO D Down-R | AAGCTTCCCAGACACACCAGTGACAG | This work | FC#1209 | HindIII |

| Table 4 continued. | | | | | |
|---|---|---|---|---|---|
| | 01418 KO A Up-F | AAGCTTAAATCGGCTTGAACCACTTG | This work | FC#2901 | HindIII |
| | 01418 KO B Up-R | CTCGAGTGTTATTGTCGTCATCGTTCG | This work | FC#2901 | XhoI |
| | 01418 KO C Down-F | CTCGAGCGAGGCAGCTGAGCCATCAT | This work | FC#2901 | XhoI |
| | 01418 KO D Down-R | GAATTCGACATCATCGAAGGCACGTT | This work | FC#2901 | EcoRI |
| | 00282 KO A Up-F | GAATTCATCCGCATGACCTGGAC | This work | FC#2902 | EcoRI |
| | 00282 KO B Up-R | GGATCCGATCAGCATCAGCGTGGAACG | This work | FC#2902 | BamI |
| | 00282 KO C Down-F | GGATCCCACGAGGCGACAAAGTAA | This work | FC#2902 | BamI |
| | 00282 KO D Down-R | AAGCTTCAAGAACGTGATCCTGATGC | This work | FC#2902 | HindIII |
| | 01237 KO A Up-F | GAATTCTCGATGCTTTCGAGGTCTCT | This work | FC#2904 | EcoRI |
| | 01237 KO B Up-R | CTCGAGAAGCCGCTTTCGGTTCAC | This work | FC#2904 | XhoI |
| | 01237 KO C Down-F | CTCGAGTTCCGCCTGTGACCAGTC | This work | FC#2904 | XhoI |
| | 01237 KO D Down-R | AAGCTTCGAGGTTCTGCGCCATAG | This work | FC#2904 | HindIII |
| | 01601 KO A Up-F | GAATTCACCGTCACCTTGGTGGTTT | This work | FC#2903 | EcoRI |
| | 01601 KO B Up-R | TCTAGAATGTTCCAACTTCGTTTGAGG | This work | FC#2903 | XbaI |
| | 01601 KO C Down-F | TCTAGAGACGACGCCAAGTCCAAGT | This work | FC#2903 | XbaI |
| | 01601 KO D Down-R | AAGCTTGCTCGTGCGAATAGAACCA | This work | FC#2903 | HindIII |
| 1090 | gsrN F | CCTCGAGCTCCTCTTCAAAAAAATCGCGGTG | This work | FC#2906 | SacI |
| 1091 | gsrN R | CATCGGTACCTCGCCGAAGTTCTGCACGAA | This work | FC#2906 | KpnI |
| 1069 | pET28_3xFLAG_PP7CP_HIS F | CAATGAATTCATGGCCAAAACCATCGTTCT | This work | FC#2927 | EcoRI |
| 1070 | pET28_3xFLAG_PP7CP_HIS R | CAATGTCGACTCAGTGGTGGTGGTGGTGGT | This work | FC#2927 | SalI |
| 1191 | PkatG F | CATAGAATTCGAGCCTTGGTCAAGGCCGG | This work | FC#2922 | EcoRI |
| 1192 | PkatG R | CATACTGCAGGCGCCTAGAAGCCGCTCTT | This work | FC#2922 | PstI |

| Table 4 continued. | | | | | |
|---|---|---|---|---|---|
| 1194 | tandem mutant A F | GTTCGAATTCTCCGGAGCTCCTCTTCAAAAAATCGCGGTG | This work | FC#2930 | SacI |
| 1195 | tandem mutant A R | ACTGGCCGTCGTTTTACTCGCCGAAGTTCTGCACGAA | This work | FC#2930 | |
| 1196 | tandem mutant B F | CGAGTAAAACGACGGCCAGTCTCTTCAAAAAAATCGCGGTG | This work | FC#2930 | |
| 1197 | tandem mutant B R | GTCATAGCTGTTTCCTGTCGCCGAAGTTCTGCACGAA | This work | FC#2930 | |
| 1198 | tandem mutant C F | CGACAGGAAACAGCTATGACCTCTTCAAAAAAATCGCGGTG | This work | FC#2930 | |
| 1199 | tandem mutant C R | TTAATTAATATGCATGGTACCTGATCGCCGAAGTTCTGCACGAA | This work | FC#2930 | KpnI |
| 1138 | katG_KO A | ATCGACTAGTGCAGCGGTTCGCCTTCAC | This work | FC#2928 | SpeI |
| 1139 | katG_KO B | GTCCGCGTTGGCGGGACCACGACC | This work | FC#2928 | |
| 1140 | katG_KO C | GCCAACGCGGACTCGCAGGAGAAG | This work | FC#2928 | |
| 1141 | katG_KO D | ATCGAAGCTTGATCGGCCACGTCTCGCC | This work | FC#2928 | HindIII |
| 1204 | katG KI B | CATCAGGGGCACGCGCCTAGAAGCCGCTCTTT | This work | FC#2929 | |
| 1205 | katG KI C | GCGTGCCCCTGATGGAAACACCTGCGCGGAG | This work | FC#2929 | |
| 1173 | katG_Cterm_M2 A | TACGTAATACGACTCACTAGGGCCGACTTGATCGTGCTG | This work | FC#2930 | |
| 1174 | katG_Cterm_M2 B | CAGGTCGACGCGGGTTGC | This work | FC#2930 | |
| 1175 | katG_Cterm_M2 C | CGAGGCTTCCGGGATGACAC | This work | FC#2930 | |
| 1176 | katG_Cterm_M2 D | AGATATCCTGCAGAGAAGCTGTCGATCGGCCACGTCTCG | This work | FC#2930 | |
| 1179 | katG_Cterm_M2 E | GACGACGACGACAAGTGATAGGATCTGGCGGCTTAGGATCTG | This work | FC#2930 | |
| 1178 | katG_Cterm_M2 F | GTGTCATCCCGGAAGCCTCG | This work | FC#2930 | |
| 1142 | katG_OE F | CATCCATATGATGGAAACACCTGCGCGG | This work | FC#2968 | NdeI |
| 1143 | katG_OE R | CATCGGTACCGTGTCATCCCGGAAGCCTC | This work | FC#2968 | KpnI |
| 1037 | hfq_KO A Up-F | ATCGACTAGTGGTTTGTCGACGACATGGGC | This work | FC#2976 | SpeI |
| 1038 | hfq_KO B Up-R | GTCGTCCTTCTTTTCGGCGGACATACG | This work | FC#2976 | |
| 1039 | hfq_KO C Down-F | AAGAAGGACGACTGACCCTAGAGCCGT | This work | FC#2976 | |

| Table 4 continued. | | | | | |
|---|---|---|---|---|---|
| 1040 | hfq_KO D Down-R | ATCGAAGCTTTCAGCCGATTGAACAGCGT | This work | FC#2976 | HindIII |
| 1041 | hfq Xyl F | CATCCATATGTCCGCCGAAAAGAAGCAAAAT | This work | FC#2977 | NdeI |
| 1042 | hfq Xyl R | CATCGGTACCCTCTAGGGTCAGTCGTCGGC | This work | FC#2977 | KpnI |
| Northern Blot probes | | | | | |
| NB02 | Landt Probe (62-89) | CCGGTCCGCCTGAGTTCGCATCAACAAA | Landt et al., 2008 | | |
| NB05 | R0081 Probe (10-57) | TGCGAAGGCGGCGGACTGGCTTTCGCCAGTCTCAAGTCGGGGGGCGTC | This work | | |
| NB13 | 5S rRNA Probe | CTTGAGACGAAGTACCATTGGCCCAGGGGGACTTAACGAC | This work | | |
| NB24 | tRNA-Tyr probe | CCTTTGCCACTCGGGACATTCC | This work | | |
| NB21 | katG probe | TCAGGCCCTCCAGGCGCAGGCTCTGGGGCCACCAGTCCCG | This work | | |
| 1077 | T7 PP7hp Template Strand | GGAGCGACGCCATATCGTCTGCTCCTATAGTGAGTCGTATTAGTGATC | This work | | |
| Primer extention | | | | | |
| 1186 | anti-sense to GsrN 18-36 | CGCCAGTCTCAAGTCGGGG | This work | | |
| 1189 | anti-sense to GsrN 72-89 | GGTCCGCCTGAGTTCGCA | This work | | |
| GBlocks | | | | | |
| G003 | R0081 tandem A | GTTCGAATTCTCCGGAGCTCCGCGCGGCAACGCCTCTTCAAAAAAATCGCGGTGCGCGGGAACCAAGGGACGGGGCGGTCGTTTTGTGAGTGCGAAGGCGGCGACGCCCCCCGACTTGAGACTGGCGAAAGCCAGTCCGCCGCCTTCGCACCCCTTTGTTGATGCGAACTCAGGCGGACCGGGTGACCGGACCGCCTTTCCTTATGCGCCGAGGCCAACCGCCGCGACCTGCACCGTTCAAGCCTCTGGGCTCGCAAGGCTCGCATGGCAAGCCCTGCTTCTGCGATCGCAGACACACCAAAGGCGACCGGCCGGCCCTATCGCCTTCGTGCAGAACTTCGGCGAGTAAAACGACGGCCAGT | IDT | FC#2907 | |
| G004 | R0081 tandem B | CGAGTAAAACGACGGCCAGTCTCTTCAAAAAATCGCGGTGCGCGGGAACCAAGGGACGGGGCGGTCGTTTTGTGAGTGCGAAGGCGGCGACGCCCCCCGACTTGAGACTGGCGAAAGCCAGTCCGCCGCCTTCGCACCCCTTTGTTGATGCGAACTCAGGCGGACCGGGTGACCGGACCGCCTTTCCTTATGCGCCGAGGCCAACCGCCGCGACCTGCACCGTTCAAGCCTCTGGGCTCGCAAGGCTCGCATGGCAAGCCCTGCTTCTGCGATCGCAGACACACCAAAGGCGACCGGCC | IDT | FC#2907 | |

| | | | | | |
|---|---|---|---|---|---|
| | | GGCCCTATCGCCTTCGTGCAGAACTTCGGC GACAGGAAACAGCTATGAC | | | |
| Table 4 continued. | | | | | |
| G005 | R0081 tandem C | CGACAGGAAACAGCTATGACCTCTTCAAAA AAATCGCGGTGCGCGGGAACCAAGGGACG GGGCGGTCGTTTTGTGAGTGCGAAGGCGGC GACGCCCCCCGACTTGAGACTGGCGAAAGC CAGTCCGCCGCCTTCGCACCCCTTTGTTGAT GCGAACTCAGGCGGACCGGGTGACCGGACC GCCTTTCCTTATGCGCCGAGGCCAACCGCC GCGACCTGCACCGTTCAAGCCTCTGGGCTC GCAAGGCTCGCATGGCAAGCCCTGCTTCTG CGATCGCAGACACACCAAAGGCGACCGGCC GGCCCTATCGCCTTCGTGCAGAACTTCGGC GACGCGCTCAGGAGCGGGTACCATGCATAT TAATTAA | IDT | FC#2907 | |
| G006 | Pvan(rpoD)::R0081 complementation | CCTCGAGCTCCTCTTCAAAAAAATCGCGGT GCGCGTTGACGTCCGTTTGATTACGATCAA GATTGTGAGTGCGAAGGCGGCGACGCCCCC CGACTTGAGACTGGCGAAAGCCAGTCCGCC GCCTTCGCACCCCTTTGTTGATGCGAACTCA GGCGGACCGGGTGACCGGACCGCCTTTCCT TATGCGCCGAGGCCAACCGCCGCGACCTGC ACCGTTCAAGCCTCTGGGCTCGCAAGGCTC GCATGGCAAGCCCTGCTTCTGCGATCGCAG ACACACCAAAGGCGACCGGCCGGCCCTATC GCCTTCGTGCAGAACTTCGGCGAGGTACCG ATG | IDT | FC#2908 | |
| G007 | Pxyl(rpoD)::R0081 complementation | CCTCGAGCTCCTCTTCAAAAAAATCGCGGT GCGCGTTGCCGTCCCCACATGTTAGCGCTAC CAATGTGAGTGCGAAGGCGGCGACGCCCCC CGACTTGAGACTGGCGAAAGCCAGTCCGCC GCCTTCGCACCCCTTTGTTGATGCGAACTCA GGCGGACCGGGTGACCGGACCGCCTTTCCT TATGCGCCGAGGCCAACCGCCGCGACCTGC ACCGTTCAAGCCTCTGGGCTCGCAAGGCTC GCATGGCAAGCCCTGCTTCTGCGATCGCAG ACACACCAAAGGCGACCGGCCGGCCCTATC GCCTTCGTGCAGAACTTCGGCGAGGTACCG ATG | IDT | FC#2909 | |
| G002 | R0081 1-58nt complementation | CATCGGTACCCTCTTCAAAAAAATCGCGGT GCGCGGGAACCAAGGGACGGGGCGGTCGTT TTGTGAGTGCGAAGGCGGCGACGCCCCCCG ACTTGAGACTGGCGAAAGCCAGTCCGCCGC CTTCGCGCCTTTCCTTATGCGCCGAGGCCAA CCGCCGCGACCTGCACCGTTCAAGCCTCTG GGCTCGCAAGGCTCGCATGGCAAGCCCTGC TTCTGCGATCGCAGACACACCAAAGGCGAC CGGCCGGCCCTATCGCCTTCGTGCAGAACTT CGGCGACATATGGATG | IDT | FC#2912 | |
| G009 | 5'PP7hp R0081 | CCTCGAGCTCCTCTTCAAAAAAATCGCGGT GCGCGGGAACCAAGGGACGGGGCGGTCGTT TTGTGAGTGCGAACTAGAAAGGAGCAGACG ATATGGCGTCGCTCCCTGCAGGGCGAAGGC GGCGACGCCCCCCGACTTGAGACTGGCGAA | IDT | FC#2917 | |

| | | | | | |
|---|---|---|---|---|---|
| | | AGCCAGTCCGCCGCCTTCGCACCCCTTTGTT<br>GATGCGAACTCAGGCGGACCGGGTGACCGG<br>ACCGCCTTTCCTTATGCGCCGAGGCCAACC<br>GCCGCGACCTGCACCGTTCAAGCCTCTGGG<br>CTCGCAAGGCTCGCATGGCAAGCCCTGCTT<br>CTGCGATCGCAGACACACCAAAGGCGACCG<br>GCCGGCCCTATCGCCTTCGTGCAGAACTTCG<br>GCGAGGTACCGATG | | | |
| Table 4 continued. | | | | | |
| G012 | R0081-PP7hp@37nt | CCTCGAGCTCCTCTTCAAAAAAATCGCGGT<br>GCGCGGGAACCAAGGGACGGGGCGGTCGTT<br>TTGTGAGTGCGAAGGCGGCGACGCCCCCCG<br>ACTTGAGACTGGCGAGGAGCAGACGATATG<br>GCGTCGCTCCAAGCCAGTCCGCCGCCTTCG<br>CACCCCTTTGTTGATGCGAACTCAGGCGGA<br>CCGGGTGACCGGACCGCCTTTCCTTATGCGC<br>CGAGGCCAACCGCCGCGACCTGCACCGTTC<br>AAGCCTCTGGGCTCGCAAGGCTCGCATGGC<br>AAGCCCTGCTTCTGCGATCGCAGACACACC<br>AAAGGCGACCGGCCGGCCCTATCGCCTTCG<br>TGCAGAACTTCGGCGAGGTACCGATG | IDT | FC#2918 | |
| G013 | R0081-PP7hp@54nt | CCTCGAGCTCCTCTTCAAAAAAATCGCGGT<br>GCGCGGGAACCAAGGGACGGGGCGGTCGTT<br>TTGTGAGTGCGAAGGCGGCGACGCCCCCCG<br>ACTTGAGACTGGCGAAAGCCAGTCCGCCGC<br>CTGGAGCAGACGATATGGCGTCGCTCCTCG<br>CACCCCTTTGTTGATGCGAACTCAGGCGGA<br>CCGGGTGACCGGACCGCCTTTCCTTATGCGC<br>CGAGGCCAACCGCCGCGACCTGCACCGTTC<br>AAGCCTCTGGGCTCGCAAGGCTCGCATGGC<br>AAGCCCTGCTTCTGCGATCGCAGACACACC<br>AAAGGCGACCGGCCGGCCCTATCGCCTTCG<br>TGCAGAACTTCGGCGAGGTACCGATG | IDT | FC#2919 | |
| G014 | R0081-PP7hp@59nt | CCTCGAGCTCCTCTTCAAAAAAATCGCGGT<br>GCGCGGGAACCAAGGGACGGGGCGGTCGTT<br>TTGTGAGTGCGAAGGCGGCGACGCCCCCCG<br>ACTTGAGACTGGCGAAAGCCAGTCCGCCGC<br>CTTCGCGGAGCAGACGATATGGCGTCGCTC<br>CACCCCTTTGTTGATGCGAACTCAGGCGGA<br>CCGGGTGACCGGACCGCCTTTCCTTATGCGC<br>CGAGGCCAACCGCCGCGACCTGCACCGTTC<br>AAGCCTCTGGGCTCGCAAGGCTCGCATGGC<br>AAGCCCTGCTTCTGCGATCGCAGACACACC<br>AAAGGCGACCGGCCGGCCCTATCGCCTTCG<br>TGCAGAACTTCGGCGAGGTACCGATG | IDT | FC#2920 | |
| G015 | R0081-PP7hp@67nt | CCTCGAGCTCCTCTTCAAAAAAATCGCGGT<br>GCGCGGGAACCAAGGGACGGGGCGGTCGTT<br>TTGTGAGTGCGAAGGCGGCGACGCCCCCCG<br>ACTTGAGACTGGCGAAAGCCAGTCCGCCGC<br>CTTCGCACCCCTTTGGAGCAGACGATATGG<br>CGTCGCTCCTTGATGCGAACTCAGGCGGAC<br>CGGGTGACCGGACCGCCTTTCCTTATGCGCC<br>GAGGCCAACCGCCGCGACCTGCACCGTTCA<br>AGCCTCTGGGCTCGCAAGGCTCGCATGGCA<br>AGCCCTGCTTCTGCGATCGCAGACACACCA | IDT | FC#2921 | |

| | | | | | |
|---|---|---|---|---|---|
| | | AAGGCGACCGGCCGGCCCTATCGCCTTCGT GCAGAACTTCGGCGAGGTACCGATG | | | |
| Table 4 continued. | | | | | |
| G016 | R0081-PP7hp@93nt | CCTCGAGCTCCTCTTCAAAAAAATCGCGGT GCGCGGGAACCAAGGGACGGGGCGGTCGTT TTGTGAGTGCGAAGGCGGCGACGCCCCCCG ACTTGAGACTGGCGAAAGCCAGTCCGCCGC CTTCGCACCCCTTTGTTGATGCGAACTCAGG CGGACCGGGTGGAGCAGACGATATGGCGTC GCTCCGACCGGACCGCCTTTCCTTATGCGCC GAGGCCAACCGCCGCGACCTGCACCGTTCA AGCCTCTGGGCTCGCAAGGCTCGCATGGCA AGCCCTGCTTCTGCGATCGCAGACACACCA AAGGCGACCGGCCGGCCCTATCGCCTTCGT GCAGAACTTCGGCGAGGTACCGATG | IDT | FC#2923 | |
| G017 | PP7hp+R0081(TERMhp) | CCTCGAGCTCCTCTTCAAAAAAATCGCGGT GCGCGGGAACCAAGGGACGGGGCGGTCGTT TTGTGAGTGCGAACTAGAAAGGAGCAGACG ATATGGCGTCGCTCCCTGCAGGGAGGCGGA CCGGGTGACCGGACCGCCTTTCCTTATGCGC CGAGGCCAACCGCCGCGACCTGCACCGTTC AAGCCTCTGGGCTCGCAAGGCTCGCATGGC AAGCCCTGCTTCTGCGATCGCAGACACACC AAAGGCGACCGGCCGGCCCTATCGCCTTCG TGCAGAACTTCGGCGAGGTACCGATG | IDT | FC#2924 | |
| G020 | rpoD-katG-5'UTR | ATCGGGTACCCAATCAATTTGACGTCCGTTT GATTACGATCAAGATTAGGCGCGTCGGGGT GATGGAAACACCTGCGCGGAGGAACACGCC AATGGACGCCAAGGTCGAAGACAATATCGC GGGCAAATGCCCCATGGGCCACGGTCGTGG TCCCGCCAACCGGGACTGGTGGCCCCAGAG CCTGCGCCTGGAGGGCCTGAACCAGCACGC CCCGCGCTCCAATCCGATGGGCGAGGCGTT CGAAAGCTTATTG | IDT | FC#2923 | |
| G024 | rpoD-katG-5'UTR (RS) | ATCGGGTACCCAATCAATTTGACGTCCGTTT GATTACGATCAAGATTAGGCGCGTGCCCCT GATGGAAACACCTGCGCGGAGGAACACGCC AATGGACGCCAAGGTCGAAGACAATATCGC GGGCAAATGCCCCATGGGCCACGGTCGTGG TCCCGCCAACCGGGACTGGTGGCCCCAGAG CCTGCGCCTGGAGGGCCTGAACCAGCACGC CCCGCGCTCCAATCCGATGGGCGAGGCGTT CGAAAGCTTATTG | IDT | FC#2924 | |
| G023 | R0081 binding site swap katG | CCTCGAGCTCCTCTTCAAAAAAATCGCGGT GCGCGGGAACCAAGGGACGGGGCGGTCGTT TTGTGAGTGCGAAGGCGGCGACGCCGGGGC ACTTGAGACTGGCGAAAGCCAGTCCGCCGC CTTCGCACCCCTTTGTTGATGCGAACTCAGG CGGACCGGGTGACCGGACCGCCTTTCCTTA TGCGCCGAGGCCAACCGCCGCGACCTGCAC CGTTCAAGCCTCTGGGCTCGCAAGGCTCGC ATGGCAAGCCCTGCTTCTGCGATCGCAGAC ACACCAAAGGCGACCGGCCGGCCCTATCGC CTTCGTGCAGAACTTCGGCGAGGTACCGAT G | IDT | FC#2925 | |

| Table 4 continued. | | | | | |
|---|---|---|---|---|---|
| G019 | katG Cterm M2 | GCAACCCGCGTCGACCTGATCTTCGGCTCCC ACGCCGAACTGCGGGCCTTCGCCGAGGTCT ATGCCTGCGCGGACTCGCAGGAGAAGTTCG TCTGCGACTTCGTCACCGCCTGGAACAAGG TGATGAACGCCGACCGCCTGGATCTGGCGG CTGGATCGGGGAGCATGGACTATAAGGACC ATGACGGGGATTATAAGGACCACGACATCG ACTATAAGGACGACGACGACAAGTGA | IDT | FC#2930 | |

Identifier: Refers to the Mattthew Tien number
Name: Convienient name
Sequence: Sequence from 5' to 3'
Source: Refers to the origin
Plasmid: Refers to the plasmid strain that was constructed from the primers or Gblock
Restriction Enzyme Site: refers to any restriction enzyme sites that might have been used for cloning.

Table 5. Key reagents and resources.
Table 5 continued on page 120.

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| Goat anti-Mouse IgG (H+L) Secondary Antibody, HRP | ThermoFisher | 32430 |
| DYKDDDDK Tag Monoclonal Antibody (FG4R) | ThermoFisher | MA1-91878-1MG |
| Bacterial and Virus Strains | | |
| See Supplementary File 1 | | |
| Chemicals, Peptides, and Recombinant Proteins | | |
| Agar | Lab Scientific | A466 |
| 30% Hydrogen Peroxide | ThermoFisher | H325-100 |
| substrate o-nitrophenyl-β-D-galactopyranoside (ONPG) | GoldBio | N-275-100 |
| acrylamide:bisacrylamide (29:1) | BioRad | 1610156 |
| Acid-Phenol | Ambion | Am9722 |
| TRIzol | ThermoFisher | 15596026 |
| T4 Polynucleotide Kinase | New England Biolabs | M0201L |
| ATP, [γ-32P]- 3000Ci/mmol 10mCi/ml EasyTide | PerkinElmer | BLU502A500UC |
| SuperScript™ IV Reverse Transcriptase | ThermoFisher | 18090010 |
| RNase H | New England Biolabs | M0297S |
| TURBO™ DNase | ThermoFisher | AM2238 |
| Critical Commercial Assays | | |
| Micro Bio-Spin Columns With Bio-Gel P-6 in Tris Buffer | BioRad | 7326221 |
| Amylose Resin | New England Biolabs | E8021L |
| RNeasy Mini Kit | Qiagen | 74106 |
| SuperSignal™ West Femto Maximum Sensitivity Substrate | ThermoFisher | 34095 |
| Deposited Data | | |
| Raw and analyzed RNA-seq data | This paper | GEO: GSE106168 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE106168 |
| Raw and analyzed LC-MS/MS data | This paper | PRIDE: PXD008128 |
| Raw and analyzed RNA-seq data for GsrN-PP7hp purification | This paper | GEO: GSE106171 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE106171 |
| Raw and analyzed RNA-seq data for Network construction | (Fang et al. 2013) | GEO: GSE46915 |
| Gel and Blotting equipment | | |
| Zeta-Probe Blotting Membranes | BioRad | 162-0165 |
| Low Molecular Weight Marker, 10-100 nt | Alfa Aesar | J76410 |
| Mini-PROTEAN TGX Precast Gel, 4-20% | BioRad | 456-1094 |
| Precision Plus Protein™ Kaleidoscope™ Prestained Protein Standards | BioRad | 1610375 |
| Software and Algorithms | | |
| Bowtie2 | (Langmead and Salzberg 2012) | http://bowtie-bio.sourceforge.net/bowtie2/index.shtml |

| Table 5 continued. | | |
|---|---|---|
| SAMTools | (Li et al. 2009) | http://samtools.sourceforge.net/ |
| IntaRNA 2.0.2 | (Mann, Wright, and Backofen 2017a) | http://rna.informatik.uni-freiburg.de/IntaRNA/Input.jsp |
| Prism v6.04 | GraphPad Software, Inc. | https://www.graphpad.com/scientific-software/prism/ |
| WebLogo | (Crooks et al. 2004) | http://weblogo.berkeley.edu/logo.cgi |
| Geneious 11.0.2 | (Kearse et al. 2012) | https://www.geneious.com/ |
| R v 3.3.3 | | https://www.r-project.org/ |
| Python v2.7 | | https://www.python.org/download/releases/2.7/ |
| Rockhopper 2.0 | (Tjaden 2015) | https://cs.wellesley.edu/~btjaden/Rockhopper/ |
| Edge-pro | (Magoc, Wood, and Salzberg 2013) | http://ccb.jhu.edu/software/EDGE-pro/index.shtml |
| DESeq | (Anders and Huber 2010) | http://bioconductor.org/packages/release/bioc/html/DESeq.html |
| CLC Genomics Workbench 10 | (Qiagen) | https://www.qiagenbioinformatics.com/products/clc-genomics-workbench/ |
| MaxQuant | (Cox et al. 2014) | http://www.coxdocs.org/doku.php?id=maxquant:start |
| IterativeRank | This paper | https://github.com/mtien/IterativeRank |
| Sliding_window_analysis | This paper | https://github.com/mtien/Sliding_window_analysis |

Reagent or Resource: Refers to important experimental tools needed for this study
Souce: Refers to where this resource originated.
Indentifiers: Refers to product information to find the resource.

## References


Abromaitis, S., and J. E. Koehler. 2013. "The Bartonella quintana extracytoplasmic function sigma factor RpoE has a role in bacterial adaptation to the arthropod vector environment." *J Bacteriol* 195 (11):2662-74. doi: 10.1128/JB.01972-12.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. "Basic local alignment search tool." *J Mol Biol* 215 (3):403-10. doi: 10.1016/S0022-2836(05)80360-2.

Alvarez-Martinez, C. E., R. F. Lourenco, R. L. Baldini, M. T. Laub, and S. L. Gomes. 2007. "The ECF sigma factor sigma(T) is involved in osmotic and oxidative stress responses in Caulobacter crescentus." *Mol Microbiol* 66 (5):1240-55. doi: 10.1111/j.1365-2958.2007.06005.x.

Anders, S., and W. Huber. 2010. "Differential expression analysis for sequence count data." *Genome Biol* 11 (10):R106. doi: 10.1186/gb-2010-11-10-r106.

Andronescu, M., A. Condon, H. H. Hoos, D. H. Mathews, and K. P. Murphy. 2007. "Efficient parameter estimation for RNA secondary structure prediction." *Bioinformatics* 23 (13):i19-28. doi: 10.1093/bioinformatics/btm223.

Arora, S., C. Lund, R. Motwani, M. Sudan, and M. Szegedy. 1998. "Proof verification and the hardness of approximation problems." *J. ACM* 45 (3):501-555. doi: Doi 10.1145/278298.278306.

Badger, J. L., and V. L. Miller. 1995. "Role of RpoS in survival of Yersinia enterocolitica to a variety of environmental stresses." *J Bacteriol* 177 (18):5370-3.

Bak, G., K. Han, D. Kim, and Y. Lee. 2014. "Roles of rpoS-activating small RNAs in pathways leading to acid resistance of Escherichia coli." *Microbiologyopen* 3 (1):15-28. doi: 10.1002/mbo3.143.

Battesti, A., J. R. Hoskins, S. Tong, P. Milanesio, J. M. Mann, A. Kravats, Y. M. Tsegaye, A. Bougdour, S. Wickner, and S. Gottesman. 2013. "Anti-adaptors provide multiple modes for regulation of the RssB adaptor protein." *Genes Dev* 27 (24):2722-35. doi: 10.1101/gad.229617.113.

Battesti, A., N. Majdalani, and S. Gottesman. 2011. "The RpoS-mediated general stress response in Escherichia coli." *Annu Rev Microbiol* 65:189-213. doi: 10.1146/annurev-micro-090110-102946.

Bergkessel, M., D. W. Basta, and D. K. Newman. 2016. "The physiology of growth arrest: uniting molecular and environmental microbiology." *Nat Rev Microbiol* 14 (9):549-62. doi: 10.1038/nrmicro.2016.107.

Boutte, C. C., J. T. Henry, and S. Crosson. 2012. "ppGpp and polyphosphate modulate cell cycle progression in Caulobacter crescentus." *J Bacteriol* 194 (1):28-35. doi: 10.1128/JB.05932-11.

Bouvier, M., C. M. Sharma, F. Mika, K. H. Nierhaus, and J. Vogel. 2008. "Small RNA binding to 5' mRNA coding region inhibits translational initiation." *Mol Cell* 32 (6):827-37. doi: 10.1016/j.molcel.2008.10.027.

Brin, S., and L. Page. 1998. "The anatomy of a large-scale hypertextual Web search engine." *Comput Networks ISDN* 30 (1-7):107-117. doi: Doi 10.1016/S0169-7552(98)00110-X.

Britos, L., E. Abeliuk, T. Taverner, M. Lipton, H. McAdams, and L. Shapiro. 2011. "Regulatory response to carbon starvation in Caulobacter crescentus." *PLoS One* 6 (4):e18179. doi: 10.1371/journal.pone.0018179.

Brun, Y. V., and L. Shapiro. 1992. "A temporally controlled sigma-factor is required for polar morphogenesis and normal cell division in Caulobacter." *Genes Dev* 6 (12A):2395-408.

Burgess, R. R., and A. A. Travers. 1970. "Escherichia coli RNA polymerase: purification, subunit structure, and factor requirements." *Fed Proc* 29 (3):1164-9.

Burgess, R. R., A. A. Travers, J. J. Dunn, and E. K. Bautz. 1969. "Factor stimulating transcription by RNA polymerase." *Nature* 221 (5175):43-6.

Caimano, M. J., C. H. Eggers, K. R. Hazlett, and J. D. Radolf. 2004. "RpoS is not central to the general stress response in Borrelia burgdorferi but does control expression of one or more essential virulence determinants." *Infect Immun* 72 (11):6433-45. doi: 10.1128/IAI.72.11.6433-6445.2004.

Campanacci, V., D. Nurizzo, S. Spinelli, C. Valencia, M. Tegoni, and C. Cambillau. 2004. "The crystal structure of the Escherichia coli lipocalin Blc suggests a possible role in phospholipid binding." *FEBS Lett* 562 (1-3):183-8. doi: 10.1016/S0014-5793(04)00199-1.

Campbell, E. A., O. Muzzin, M. Chlenov, J. L. Sun, C. A. Olson, O. Weinman, M. L. Trester-Zedlitz, and S. A. Darst. 2002. "Structure of the bacterial RNA polymerase promoter specificity sigma subunit." *Mol Cell* 9 (3):527-39.

Carey, J., P. T. Lowary, and O. C. Uhlenbeck. 1983. "Interaction of R17 coat protein with synthetic variants of its ribonucleic acid binding site." *Biochemistry* 22 (20):4723-30.

Christman, M. F., G. Storz, and B. N. Ames. 1989. "OxyR, a positive regulator of hydrogen peroxide-inducible genes in Escherichia coli and Salmonella typhimurium, is

homologous to a family of bacterial regulatory proteins." *Proc Natl Acad Sci U S A* 86 (10):3484-8.

Collier, J. 2012. "Regulation of chromosomal replication in Caulobacter crescentus." *Plasmid* 67 (2):76-87. doi: 10.1016/j.plasmid.2011.12.007.

Collier, J. 2016. "Cell cycle control in Alphaproteobacteria." *Curr Opin Microbiol* 30:107-113. doi: 10.1016/j.mib.2016.01.010.

Corcoran, C. P., R. Rieder, D. Podkaminski, B. Hofmann, and J. Vogel. 2012. "Use of aptamer tagging to identify in vivo protein binding partners of small regulatory RNAs." *Methods Mol Biol* 905:177-200. doi: 10.1007/978-1-61779-949-5_11.

Cox, J., M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj, and M. Mann. 2014. "Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ." *Mol Cell Proteomics* 13 (9):2513-26. doi: 10.1074/mcp.M113.031591.

Crooks, G. E., G. Hon, J. M. Chandonia, and S. E. Brenner. 2004. "WebLogo: a sequence logo generator." *Genome Res* 14 (6):1188-90. doi: 10.1101/gr.849004.

de Castro Ferreira, I. G., M. M. Rodrigues, J. F. da Silva Neto, R. R. Mazzon, and M. do Valle Marques. 2016. "Role and regulation of ferritin-like proteins in iron homeostasis and oxidative stress survival of Caulobacter crescentus." *Biometals* 29 (5):851-62. doi: 10.1007/s10534-016-9956-y.

Dong, T., and H. E. Schellhorn. 2010. "Role of RpoS in virulence of pathogens." *Infect Immun* 78 (3):887-97. doi: 10.1128/IAI.00882-09.

Ely, B. 1991. "Genetics of Caulobacter crescentus." *Methods Enzymol* 204:372-84.

Fang, F. C., S. J. Libby, N. A. Buchmeier, P. C. Loewen, J. Switala, J. Harwood, and D. G. Guiney. 1992. "The alternative sigma factor katF (rpoS) regulates Salmonella virulence." *Proc Natl Acad Sci U S A* 89 (24):11978-82.

Fang, G., K. D. Passalacqua, J. Hocking, P. M. Llopis, M. Gerstein, N. H. Bergman, and C. Jacobs-Wagner. 2013. "Transcriptomic and phylogenetic analysis of a bacterial cell cycle reveals strong associations between gene co-expression and evolution." *BMC Genomics* 14:450. doi: 10.1186/1471-2164-14-450.

Feklistov, A., B. D. Sharon, S. A. Darst, and C. A. Gross. 2014. "Bacterial sigma factors: a historical, structural, and genomic perspective." *Annu Rev Microbiol* 68:357-76. doi: 10.1146/annurev-micro-092412-155737.

Fiebig, A., J. Herrou, J. Willett, and S. Crosson. 2015. "General Stress Signaling in the Alphaproteobacteria." *Annu Rev Genet* 49:603-25. doi: 10.1146/annurev-genet-112414-054813.

Finan, T. M., B. Kunkel, G. F. De Vos, and E. R. Signer. 1986. "Second symbiotic megaplasmid in Rhizobium meliloti carrying exopolysaccharide and thiamine synthesis genes." *J Bacteriol* 167 (1):66-72.

Foreman, R., A. Fiebig, and S. Crosson. 2012. "The LovK-LovR two-component system is a regulator of the general stress pathway in Caulobacter crescentus." *J Bacteriol* 194 (12):3038-49. doi: 10.1128/JB.00182-12.

Francez-Charlot, A., J. Frunzke, C. Reichen, J. Z. Ebneter, B. Gourion, and J. A. Vorholt. 2009. "Sigma factor mimicry involved in regulation of general stress response." *Proc Natl Acad Sci U S A* 106 (9):3467-72. doi: 10.1073/pnas.0810291106.

Francez-Charlot, A., A. Kaczmarczyk, H. M. Fischer, and J. A. Vorholt. 2015. "The general stress response in Alphaproteobacteria." *Trends Microbiol* 23 (3):164-71. doi: 10.1016/j.tim.2014.12.006.

Frohlich, K. S., K. Papenfort, A. A. Berger, and J. Vogel. 2012. "A conserved RpoS-dependent small RNA controls the synthesis of major porin OmpD." *Nucleic Acids Res* 40 (8):3623-40. doi: 10.1093/nar/gkr1156.

Gonzalez, D., and J. Collier. 2013. "DNA methylation by CcrM activates the transcription of two genes required for the division of Caulobacter crescentus." *Mol Microbiol* 88 (1):203-18. doi: 10.1111/mmi.12180.

Gourion, B., M. Rossignol, and J. A. Vorholt. 2006. "A proteomic study of Methylobacterium extorquens reveals a response regulator essential for epiphytic growth." *Proc Natl Acad Sci U S A* 103 (35):13186-91. doi: 10.1073/pnas.0603530103.

Gourion, B., S. Sulser, J. Frunzke, A. Francez-Charlot, P. Stiefel, G. Pessi, J. A. Vorholt, and H. M. Fischer. 2009. "The PhyR-sigma(EcfG) signalling cascade is involved in stress response and symbiotic efficiency in Bradyrhizobium japonicum." *Mol Microbiol* 73 (2):291-305. doi: 10.1111/j.1365-2958.2009.06769.x.

Groat, R. G., J. E. Schultz, E. Zychlinsky, A. Bockman, and A. Matin. 1986. "Starvation proteins in Escherichia coli: kinetics of synthesis and role in starvation survival." *J Bacteriol* 168 (2):486-93.

Gruber, T. M., and C. A. Gross. 2003. "Multiple sigma subunits and the partitioning of bacterial transcription space." *Annu Rev Microbiol* 57:441-66. doi: 10.1146/annurev.micro.57.030502.090913.

Hajduk, I. V., C. D. Rodrigues, and E. J. Harry. 2016. "Connecting the dots of the bacterial cell cycle: Coordinating chromosome replication and segregation with cell division." *Semin Cell Dev Biol* 53:2-9. doi: 10.1016/j.semcdb.2015.11.012.

Hao, Y., T. B. Updegrove, N. N. Livingston, and G. Storz. 2016. "Protection against deleterious nitrogen compounds: role of sigmaS-dependent small RNAs encoded adjacent to sdiA." *Nucleic Acids Res* 44 (14):6935-48. doi: 10.1093/nar/gkw404.

Hartuv, Erez, and Ron Shamir. 2000. "A clustering algorithm based on graph connectivity." *Information Processing Letters* 76 (4):175-181. doi: https://doi.org/10.1016/S0020-0190(00)00142-3.

Hecker, M., J. Pane-Farre, and U. Volker. 2007. "SigB-dependent general stress response in Bacillus subtilis and related gram-positive bacteria." *Annu Rev Microbiol* 61:215-36. doi: 10.1146/annurev.micro.61.080706.093445.

Heinrich, K., P. Sobetzko, and K. Jonas. 2016. "A Kinase-Phosphatase Switch Transduces Environmental Information into a Bacterial Cell Cycle Circuit." *PLoS Genet* 12 (12):e1006522. doi: 10.1371/journal.pgen.1006522.

Hengge-Aronis, R. 2002. "Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase." *Microbiol Mol Biol Rev* 66 (3):373-95, table of contents.

Herrou, J., and S. Crosson. 2011. "Function, structure and mechanism of bacterial photosensory LOV proteins." *Nat Rev Microbiol* 9 (10):713-23. doi: 10.1038/nrmicro2622.

Herrou, J., R. Foreman, A. Fiebig, and S. Crosson. 2010. "A structural model of anti-anti-sigma inhibition by a two-component receiver domain: the PhyR stress response regulator." *Mol Microbiol* 78 (2):290-304. doi: 10.1111/j.1365-2958.2010.07323.x.

Hogg, J. R., and K. Collins. 2007. "RNA-based affinity purification reveals 7SK RNPs with distinct composition and regulation." *RNA* 13 (6):868-80. doi: 10.1261/rna.565207.

Hottes, A. K., L. Shapiro, and H. H. McAdams. 2005. "DnaA coordinates replication initiation and cell cycle transcription in Caulobacter crescentus." *Mol Microbiol* 58 (5):1340-53. doi: 10.1111/j.1365-2958.2005.04912.x.

Imlay, James A. 2013. "The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium." *Nature Reviews Microbiology* 11:443. doi: 10.1038/nrmicro3032.

Italiani, V. C., J. F. da Silva Neto, V. S. Braz, and M. V. Marques. 2011. "Regulation of catalase-peroxidase KatG is OxyR dependent and Fur independent in Caulobacter crescentus." *J Bacteriol* 193 (7):1734-44. doi: 10.1128/JB.01339-10.

Italiani, V. C. S., J. F. D. Neto, V. S. Braz, and M. V. Marques. 2011. "Regulation of Catalase-Peroxidase KatG Is OxyR Dependent and Fur Independent in Caulobacter crescentus." *Journal of Bacteriology* 193 (7):1734-1744. doi: 10.1128/Jb.01339-10.

Jacob, F., and J. Monod. 1961. "Genetic regulatory mechanisms in the synthesis of proteins." *J Mol Biol* 3:318-56.

Jans, A., M. Vercruysse, S. Gao, K. Engelen, I. Lambrichts, M. Fauvart, and J. Michiels. 2013. "Canonical and non-canonical EcfG sigma factors control the general stress response in Rhizobium etli." *Microbiologyopen* 2 (6):976-87. doi: 10.1002/mbo3.137.

Jishage, M., and A. Ishihama. 1995. "Regulation of RNA polymerase sigma subunit synthesis in Escherichia coli: intracellular levels of sigma 70 and sigma 38." *J Bacteriol* 177 (23):6832-5.

Jonas, K., Y. E. Chen, and M. T. Laub. 2011. "Modularity of the bacterial cell cycle enables independent spatial and temporal control of DNA replication." *Curr Biol* 21 (13):1092-101. doi: 10.1016/j.cub.2011.05.040.

Jonas, K., J. Liu, P. Chien, and M. T. Laub. 2013. "Proteotoxic stress induces a cell-cycle arrest by stimulating Lon to degrade the replication initiator DnaA." *Cell* 154 (3):623-36. doi: 10.1016/j.cell.2013.06.034.

Kaczmarczyk, A., S. Campagne, F. Danza, L. C. Metzger, J. A. Vorholt, and A. Francez-Charlot. 2011. "Role of Sphingomonas sp. strain Fr1 PhyR-NepR-sigmaEcfG cascade in general stress response and identification of a negative regulator of PhyR." *J Bacteriol* 193 (23):6629-38. doi: 10.1128/JB.06006-11.

Kaguni, J. M. 2006. "DnaA: controlling the initiation of bacterial DNA replication and more." *Annu Rev Microbiol* 60:351-75. doi: 10.1146/annurev.micro.60.080805.142111.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, and A. Drummond. 2012. "Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data." *Bioinformatics* 28 (12):1647-9. doi: 10.1093/bioinformatics/bts199.

Kelly, A. J., M. J. Sackett, N. Din, E. Quardokus, and Y. V. Brun. 1998. "Cell cycle-dependent transcriptional and proteolytic regulation of FtsZ in Caulobacter." *Genes Dev* 12 (6):880-93.

Kim, H. S., C. C. Caswell, R. Foreman, R. M. Roop, 2nd, and S. Crosson. 2013. "The Brucella abortus general stress response system regulates chronic mammalian infection and is controlled by phosphorylation and proteolysis." *J Biol Chem* 288 (19):13906-16. doi: 10.1074/jbc.M113.459305.

Kim, H. S., J. W. Willett, N. Jain-Gupta, A. Fiebig, and S. Crosson. 2014. "The Brucella abortus virulence regulator, LovhK, is a sensor kinase in the general stress response signalling pathway." *Mol Microbiol* 94 (4):913-25. doi: 10.1111/mmi.12809.

Kulkarni, G., C. H. Wu, and D. K. Newman. 2013. "The general stress response factor EcfG regulates expression of the C-2 hopanoid methylase HpnP in Rhodopseudomonas palustris TIE-1." *J Bacteriol* 195 (11):2490-8. doi: 10.1128/JB.00186-13.

Lalaouna, D., and E. Masse. 2015. "Identification of sRNA interacting with a transcript of interest using MS2-affinity purification coupled with RNA sequencing (MAPS) technology." *Genom Data* 5:136-8. doi: 10.1016/j.gdata.2015.05.033.

Landt, S. G., E. Abeliuk, P. T. McGrath, J. A. Lesley, H. H. McAdams, and L. Shapiro. 2008. "Small non-coding RNAs in Caulobacter crescentus." *Mol Microbiol* 68 (3):600-14. doi: 10.1111/j.1365-2958.2008.06172.x.

Landt, S. G., J. A. Lesley, L. Britos, and L. Shapiro. 2010. "CrfA, a small noncoding RNA regulator of adaptation to carbon starvation in Caulobacter crescentus." *J Bacteriol* 192 (18):4763-75. doi: 10.1128/JB.00343-10.

Lane, W. J., and S. A. Darst. 2006. "The structural basis for promoter -35 element recognition by the group IV sigma factors." *PLoS Biol* 4 (9):e269. doi: 10.1371/journal.pbio.0040269.

Lange, R., and R. Hengge-Aronis. 1991. "Identification of a central regulator of stationary-phase gene expression in Escherichia coli." *Mol Microbiol* 5 (1):49-59. doi: 10.1111/j.1365-2958.1991.tb01825.x.

Lange, R., and R. Hengge-Aronis. 1994. "The cellular concentration of the sigma S subunit of RNA polymerase in Escherichia coli is controlled at the levels of transcription, translation, and protein stability." *Genes Dev* 8 (13):1600-12.

Langmead, B., and S. L. Salzberg. 2012. "Fast gapped-read alignment with Bowtie 2." *Nat Methods* 9 (4):357-9. doi: 10.1038/nmeth.1923.

Laub, M. T., S. L. Chen, L. Shapiro, and H. H. McAdams. 2002. "Genes directly controlled by CtrA, a master regulator of the Caulobacter cell cycle." *Proc Natl Acad Sci U S A* 99 (7):4632-7. doi: 10.1073/pnas.062065699.

Laub, M. T., H. H. McAdams, T. Feldblyum, C. M. Fraser, and L. Shapiro. 2000. "Global analysis of the genetic network controlling a bacterial cell cycle." *Science* 290 (5499):2144-8.

Lemaux, P. G., S. L. Herendeen, P. L. Bloch, and F. C. Neidhardt. 1978. "Transient rates of synthesis of individual polypeptides in E. coli following temperature shifts." *Cell* 13 (3):427-34.

Lesley, J. A., and L. Shapiro. 2008. "SpoT regulates DnaA stability and initiation of DNA replication in carbon-starved Caulobacter crescentus." *J Bacteriol* 190 (20):6867-80. doi: 10.1128/JB.00700-08.

Leslie, D. J., C. Heinen, F. D. Schramm, M. Thuring, C. D. Aakre, S. M. Murray, M. T. Laub, and K. Jonas. 2015. "Nutritional Control of DNA Replication Initiation through the Proteolysis and Regulated Translation of DnaA." *PLoS Genet* 11 (7):e1005342. doi: 10.1371/journal.pgen.1005342.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and Subgroup Genome Project Data Processing. 2009. "The Sequence Alignment/Map format and SAMtools." *Bioinformatics* 25 (16):2078-9. doi: 10.1093/bioinformatics/btp352.

Lim, F., T. P. Downey, and D. S. Peabody. 2001. "Translational repression and specific RNA binding by the coat protein of the Pseudomonas phage PP7." *J Biol Chem* 276 (25):22507-13. doi: 10.1074/jbc.M102411200.

Lim, F., and D. S. Peabody. 2002. "RNA recognition site of PP7 coat protein." *Nucleic Acids Res* 30 (19):4138-44.

Loewen, P. C., and B. L. Triggs. 1984. "Genetic mapping of katF, a locus that with katE affects the synthesis of a second catalase species in Escherichia coli." *J Bacteriol* 160 (2):668-75.

Lourenco, R. F., C. Kohler, and S. L. Gomes. 2011. "A two-component system, an anti-sigma factor and two paralogous ECF sigma factors are involved in the control of general stress response in Caulobacter crescentus." *Mol Microbiol* 80 (6):1598-612. doi: 10.1111/j.1365-2958.2011.07668.x.

Magoc, T., D. Wood, and S. L. Salzberg. 2013. "EDGE-pro: Estimated Degree of Gene Expression in Prokaryotic Genomes." *Evol Bioinform Online* 9:127-36. doi: 10.4137/EBO.S11250.

Majdalani, N., C. Cunning, D. Sledjeski, T. Elliott, and S. Gottesman. 1998. "DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription." *Proc Natl Acad Sci U S A* 95 (21):12462-7.

Majdalani, N., D. Hernandez, and S. Gottesman. 2002. "Regulation and mode of action of the second small RNA activator of RpoS translation, RprA." *Mol Microbiol* 46 (3):813-26.

Mandin, P., and S. Gottesman. 2010. "Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA." *EMBO J* 29 (18):3094-107. doi: 10.1038/emboj.2010.179.

Mann, M., P. R. Wright, and R. Backofen. 2017a. "IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions." *Nucleic Acids Res*. doi: 10.1093/nar/gkx279.

Mann, M., P. R. Wright, and R. Backofen. 2017b. "IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions." *Nucleic Acids Res* 45 (W1):W435-W439. doi: 10.1093/nar/gkx279.

Marks, M. E., C. M. Castro-Rojas, C. Teiling, L. Du, V. Kapatral, T. L. Walunas, and S. Crosson. 2010. "The genetic basis of laboratory adaptation in Caulobacter crescentus." *J Bacteriol* 192 (14):3678-88. doi: 10.1128/JB.00255-10.

Martinez, A., and R. Kolter. 1997. "Protection of DNA during oxidative stress by the nonspecific DNA-binding protein Dps." *J Bacteriol* 179 (16):5188-94.

Mathis, R., and M. Ackermann. 2016. "Response of single bacterial cells to stress gives rise to complex history dependence at the population level." *Proc Natl Acad Sci U S A* 113 (15):4224-9. doi: 10.1073/pnas.1511509113.

McCullen, C. A., J. N. Benhammou, N. Majdalani, and S. Gottesman. 2010. "Mechanism of positive regulation by DsrA and RprA small noncoding RNAs: pairing increases translation and protects rpoS mRNA from degradation." *J Bacteriol* 192 (21):5559-71. doi: 10.1128/JB.00464-10.

McGrath, P. T., H. Lee, L. Zhang, A. A. Iniesta, A. K. Hottes, M. H. Tan, N. J. Hillson, P. Hu, L. Shapiro, and H. H. McAdams. 2007. "High-throughput identification of transcription start sites, conserved promoter motifs and predicted regulons." *Nat Biotechnol* 25 (5):584-92. doi: 10.1038/nbt1294.

Meisenzahl, A. C., L. Shapiro, and U. Jenal. 1997. "Isolation and characterization of a xylose-dependent promoter from Caulobacter crescentus." *J Bacteriol* 179 (3):592-600.

Mogull, S. A., L. J. Runyen-Janecky, M. Hong, and S. M. Payne. 2001. "dksA is required for intercellular spread of Shigella flexneri via an RpoS-independent mechanism." *Infect Immun* 69 (9):5742-51.

Mott, M. L., and J. M. Berger. 2007. "DNA replication initiation: mechanisms and regulation in bacteria." *Nat Rev Microbiol* 5 (5):343-54. doi: 10.1038/nrmicro1640.

Murakami, K. S., S. Masuda, E. A. Campbell, O. Muzzin, and S. A. Darst. 2002. "Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex." *Science* 296 (5571):1285-90. doi: 10.1126/science.1069595.

Murakami, K. S., S. Masuda, and S. A. Darst. 2002. "Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution." *Science* 296 (5571):1280-4. doi: 10.1126/science.1069594.

Nair, S., and S. E. Finkel. 2004. "Dps protects cells against multiple stresses during stationary phase." *J Bacteriol* 186 (13):4192-8. doi: 10.1128/JB.186.13.4192-4198.2004.

Papenfort, K., E. Espinosa, J. Casadesus, and J. Vogel. 2015. "Small RNA-based feedforward loop with AND-gate logic regulates extrachromosomal DNA transfer in Salmonella." *Proc Natl Acad Sci U S A* 112 (34):E4772-81. doi: 10.1073/pnas.1507825112.

Papenfort, K., N. Said, T. Welsink, S. Lucchini, J. C. Hinton, and J. Vogel. 2009. "Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA." *Mol Microbiol* 74 (1):139-58. doi: 10.1111/j.1365-2958.2009.06857.x.

Papenfort, K., Y. Sun, M. Miyakoshi, C. K. Vanderpool, and J. Vogel. 2013. "Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis." *Cell* 153 (2):426-37. doi: 10.1016/j.cell.2013.03.003.

Poindexter, J. S. 1964. "Biological Properties and Classification of the Caulobacter Group." *Bacteriol Rev* 28:231-95.

Purcell, E. B., D. Siegal-Gaskins, D. C. Rawling, A. Fiebig, and S. Crosson. 2007. "A photosensory two-component system regulates bacterial cell attachment." *Proc Natl Acad Sci U S A* 104 (46):18241-6. doi: 10.1073/pnas.0705887104.

CLC Genomics Workbench.

Ried, J. L., and A. Collmer. 1987. "An nptI-sacB-sacR cartridge for constructing directed, unmarked mutations in gram-negative bacteria by marker exchange-eviction mutagenesis." *Gene* 57 (2-3):239-46.

Said, N., R. Rieder, R. Hurwitz, J. Deckert, H. Urlaub, and J. Vogel. 2009. "In vivo expression and purification of aptamer-tagged small RNA regulators." *Nucleic Acids Res* 37 (20):e133. doi: 10.1093/nar/gkp719.

Santos, P. M., I. Di Bartolo, J. M. Blatny, E. Zennaro, and S. Valla. 2001. "New broad-host-range promoter probe vectors based on the plasmid RK2 replicon." *FEMS Microbiol Lett* 195 (1):91-6.

Sauviac, L., H. Philippe, K. Phok, and C. Bruand. 2007. "An extracytoplasmic function sigma factor acts as a general stress response regulator in Sinorhizobium meliloti." *J Bacteriol* 189 (11):4204-16. doi: 10.1128/JB.00175-07.

Schellhorn, H. E., and H. M. Hassan. 1988. "Transcriptional regulation of katE in Escherichia coli K-12." *J Bacteriol* 170 (9):4286-92.

Schrader, J. M., B. Zhou, G. W. Li, K. Lasker, W. S. Childers, B. Williams, T. Long, S. Crosson, H. H. McAdams, J. S. Weissman, and L. Shapiro. 2014. "The coding and noncoding architecture of the Caulobacter crescentus genome." *PLoS Genet* 10 (7):e1004463. doi: 10.1371/journal.pgen.1004463.

Seaver, L. C., and J. A. Imlay. 2001. "Alkyl hydroperoxide reductase is the primary scavenger of endogenous hydrogen peroxide in Escherichia coli." *J Bacteriol* 183 (24):7173-81. doi: 10.1128/JB.183.24.7173-7181.2001.

Sharma, C. M., and J. Vogel. 2009. "Experimental approaches for the discovery and characterization of regulatory small RNA." *Curr Opin Microbiol* 12 (5):536-46. doi: 10.1016/j.mib.2009.07.006.

Sharpe, M. E., P. M. Hauser, R. G. Sharpe, and J. Errington. 1998. "Bacillus subtilis cell cycle as studied by fluorescence microscopy: constancy of cell length at initiation of DNA replication and evidence for active nucleoid partitioning." *J Bacteriol* 180 (3):547-55.

Sledjeski, D. D., A. Gupta, and S. Gottesman. 1996. "The small RNA, DsrA, is essential for the low temperature expression of RpoS during exponential growth in Escherichia coli." *EMBO J* 15 (15):3993-4000.

Smulevich, Giulietta, Christa Jakopitsch, Enrica Droghetti, and Christian Obinger. 2006. "Probing the structure and bifunctionality of catalase-peroxidase (KatG)." *Journal of Inorganic Biochemistry* 100 (4):568-585. doi: https://doi.org/10.1016/j.jinorgbio.2006.01.033.

Staron, A., H. J. Sofia, S. Dietrich, L. E. Ulrich, H. Liesegang, and T. Mascher. 2009. "The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) sigma factor protein family." *Mol Microbiol* 74 (3):557-81. doi: 10.1111/j.1365-2958.2009.06870.x.

Steinman, H. M., F. Fareed, and L. Weinstein. 1997a. "Catalase-peroxidase of Caulobacter crescentus: function and role in stationary-phase survival." *J Bacteriol* 179 (21):6831-6.

Steinman, H. M., F. Fareed, and L. Weinstein. 1997b. "Catalase-peroxidase of Caulobacter crescentus: function and role in stationary-phase survival." *Journal of Bacteriology* 179 (21):6831-6836.

Struhl, K. 1999. "Fundamentally different logic of gene regulation in eukaryotes and prokaryotes." *Cell* 98 (1):1-4. doi: 10.1016/S0092-8674(00)80599-1.

Thanbichler, M., A. A. Iniesta, and L. Shapiro. 2007. "A comprehensive set of plasmids for vanillate- and xylose-inducible gene expression in Caulobacter crescentus." *Nucleic Acids Res* 35 (20):e137. doi: 10.1093/nar/gkm818.

Tjaden, B. 2015. "De novo assembly of bacterial transcriptomes from RNA-seq data." *Genome Biol* 16:1. doi: 10.1186/s13059-014-0572-2.

Touati, E., E. Dassa, and P. L. Boquet. 1986. "Pleiotropic mutations in appR reduce pH 2.5 acid phosphatase expression and restore succinate utilisation in CRP-deficient strains of Escherichia coli." *Mol Gen Genet* 202 (2):257-64.

Truman, A. W., K. Kristjansdottir, D. Wolfgeher, N. Hasin, S. Polier, H. Zhang, S. Perrett, C. Prodromou, G. W. Jones, and S. J. Kron. 2012. "CDK-dependent Hsp70 Phosphorylation controls G1 cyclin abundance and cell-cycle progression." *Cell* 151 (6):1308-18. doi: 10.1016/j.cell.2012.10.051.

Tu, N., A. Lima, Z. Bandeali, and B. Anderson. 2016. "Characterization of the general stress response in Bartonella henselae." *Microb Pathog* 92:1-10. doi: 10.1016/j.micpath.2015.12.010.

Valverde, C., J. Livny, J. P. Schluter, J. Reinkensmeier, A. Becker, and G. Parisi. 2008. "Prediction of Sinorhizobium meliloti sRNA genes and experimental detection in strain 2011." *BMC Genomics* 9:416. doi: 10.1186/1471-2164-9-416.

Vargas Mdel, C., S. Encarnacion, A. Davalos, A. Reyes-Perez, Y. Mora, A. Garcia-de los Santos, S. Brom, and J. Mora. 2003. "Only one catalase, katG, is detectable in Rhizobium etli, and is encoded along with the regulator OxyR on a plasmid replicon." *Microbiology* 149 (Pt 5):1165-76. doi: 10.1099/mic.0.25909-0.

Vercruysse, M., M. Fauvart, A. Jans, S. Beullens, K. Braeken, L. Cloots, K. Engelen, K. Marchal, and J. Michiels. 2011. "Stress response regulators identified through genome-wide transcriptome analysis of the (p)ppGpp-dependent response in Rhizobium etli." *Genome Biol* 12 (2):R17. doi: 10.1186/gb-2011-12-2-r17.

Vogel, J., V. Bartels, T. H. Tang, G. Churakov, J. G. Slagter-Jager, A. Huttenhofer, and E. G. Wagner. 2003. "RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria." *Nucleic Acids Res* 31 (22):6435-43.

Wagner, E. G., and P. Romby. 2015. "Small RNAs in bacteria and archaea: who they are, what they do, and how they do it." *Adv Genet* 90:133-208. doi: 10.1016/bs.adgen.2015.05.001.

Wang, P. I., and E. M. Marcotte. 2010. "It's the machine that matters: Predicting gene function and phenotype from protein networks." *J Proteomics* 73 (11):2277-89. doi: 10.1016/j.jprot.2010.07.005.

Wang, Y. 2002. "The function of OmpA in Escherichia coli." *Biochem Biophys Res Commun* 292 (2):396-401. doi: 10.1006/bbrc.2002.6657.

Weber, H., T. Polen, J. Heuveling, V. F. Wendisch, and R. Hengge. 2005. "Genome-wide analysis of the general stress response network in Escherichia coli: sigmaS-dependent genes, promoters, and sigma factor selectivity." *J Bacteriol* 187 (5):1591-603. doi: 10.1128/JB.187.5.1591-1603.2005.

Wehrli, W. 1983. "Rifampin: mechanisms of action and resistance." *Rev Infect Dis* 5 Suppl 3:S407-11.

West, L., D. Yang, and C. Stephens. 2002. "Use of the Caulobacter crescentus genome sequence to develop a method for systematic genetic mapping." *J Bacteriol* 184 (8):2155-66.

Wu, J., A. K. Benson, and A. Newton. 1995. "Global regulation of a sigma 54-dependent flagellar gene family in Caulobacter crescentus by the transcriptional activator FlbD." *J Bacteriol* 177 (11):3241-50.

Zambrano, M. M., D. A. Siegele, M. Almiron, A. Tormo, and R. Kolter. 1993. "Microbial competition: Escherichia coli mutants that take over stationary phase cultures." *Science* 259 (5102):1757-60.

Zheng, M., X. Wang, L. J. Templeton, D. R. Smulski, R. A. LaRossa, and G. Storz. 2001. "DNA microarray-mediated transcriptional profiling of the Escherichia coli response to hydrogen peroxide." *J Bacteriol* 183 (15):4562-70. doi: 10.1128/JB.183.15.4562-4570.2001.

Zhou, B., J. M. Schrader, V. S. Kalogeraki, E. Abeliuk, C. B. Dinh, J. Q. Pham, Z. Z. Cui, D. L. Dill, H. H. McAdams, and L. Shapiro. 2015a. "The global regulatory architecture of transcription during the Caulobacter cell cycle." *PLoS Genet* 11 (1):e1004831. doi: 10.1371/journal.pgen.1004831.

Zhou, Bo, Jared M. Schrader, Virginia S. Kalogeraki, Eduardo Abeliuk, Cong B. Dinh, James Q. Pham, Zhongying Z. Cui, David L. Dill, Harley H. McAdams, and Lucy Shapiro. 2015b. "The Global Regulatory Architecture of Transcription during the Caulobacter Cell Cycle." *PLOS Genetics* 11 (1):e1004831. doi: 10.1371/journal.pgen.1004831.

Zuker, M. 2003. "Mfold web server for nucleic acid folding and hybridization prediction." *Nucleic Acids Res* 31 (13):3406-15.