

Ferrous iron sensing and responding in  
*Pseudomonas aeruginosa*

Thesis by  
Naomi Kreamer

In Partial Fulfillment of the Requirements for the  
degree of  
Doctor of Philosophy



CALIFORNIA INSTITUTE OF TECHNOLOGY  
Pasadena, California  
2015  
(Defended Mar 16, 2015)

©2015

Naomi Kreamer

All Rights Reserved

## Acknowledgements

Thank you to my advisor, Dianne Newman, and the entire Newman lab for their support, feedback, and guidance. My research was made possible by several fellowships: National Action Council for Minorities in Engineering Fellowship, National Institutes of Health Training Grant (GM07616), and the National Science Foundation Research Fellowship Program.

Collaborating with James Boedicker was a wonderful experience, both for useful discussions and exposure to a new and different way of thinking about scientific problems.

To my thesis committee members, Harry Gray, Doug Rees, Shu-ou Shan, and Rob Phillips thank you for your advise, guidance, and suggestions.

## Abstract

Controlling iron distribution is important for all organisms, and is key in bacterial pathogenesis. It has long been understood that cystic fibrosis (CF) patient sputum contains elevated iron concentrations. However, anaerobic bacteria have been isolated from CF sputum and hypoxic zones in sputum have been measured. Because ferrous iron [Fe(II)] is stable in reducing, acidic conditions, it could exist in the CF lung. I show that a two-component system, BqsRS, specifically responds to Fe(II) in the CF pathogen, *Pseudomonas aeruginosa*. Concurrently, a clinical study found that Fe(II) is present in CF sputum at all stages of lung function decline. Fe(II), not Fe(III), correlates with patients in the most severe disease state. Furthermore, transcripts of the newly identified BqsRS were detected in sputum. Two-component systems are the main method that bacteria use to interact with their extracellular environment. A typical two-component system contains a sensor histidine kinase, which upon activation phosphorylates a response regulator that then acts as a transcription factor to elicit a cellular response to stimuli. To explore the mechanism of BqsRS, I describe the Fe(II)-sensing RExxE motif in the sensor BqsS and determine the consensus DNA sequence BqsR binds. With the BqsR binding sequence, I identify novel regulon members through bioinformatic and molecular biology techniques. From the predicted function of new BqsR regulon members, I find that Fe(II) elicits a response that globally protects the cells against cationic stressors, including clinically relevant antibiotics. Subsequently, I use BqsR as a case study to determine if promoter outputs can accurately be predicted based only on a deep understanding of a



transcriptional activator's operator, or if a broader regulatory context is required for accurate predictions at all genomic loci. This work highlights the importance of Fe(II) as a (micro)environmental factor, even in conditions typically thought of as aerobic. Since the presence of Fe(II) can alter *P. aeruginosa's* antibiotic susceptibility, combining the current strategy of targeting Fe(III) with a new approach targeting Fe(II) may help eradicate infections in the CF lung in the future.

## TABLE OF CONTENTS

Acknowledgements.....	iii
Abstract .....	iv
Table of Contents .....	vi
List of Illustrations and/or Tables .....	vii
Chapter I: Introduction .....	1
Chapter summaries.....	4
References .....	7
Chapter II: BqsR/BqsS constitute a two-component system that senses extracellular Fe(II) in <i>Pseudomonas aeruginosa</i> .....	8
References .....	41
Chapter III: The ferrous iron responsive BqsRS two-component system activates genes that promote cationic stress tolerance.....	46
References .....	90
Chapter IV: Broader regulatory context outweighs operator occupation in predicting genome-wide regulation in bacteria .....	107
Supplemental Material .....	143
References .....	165
Chapter V: Future directions.....	170
References .....	181
Appendix A: $\Delta rhIR$ growth and $\Delta bqsR$ rhamnolipid phenotype .....	184
References .....	200
Appendix B: Phylogenetic analysis of BqsS.....	201
References .....	212

**List of figures and tables**

Chapter I	
Figure 1-1 .....	4
Chapter II	
Table 2-1 .....	14
Figure 2-1 .....	16
Figure 2-2 .....	17
Figure 2-3 .....	19
Figure 2-4 .....	20
Figure 2-5 .....	21
Figure 2-6 .....	23
Table 2-2 .....	23
Figure 2-7 .....	28
Table 2-3 .....	30
Table 2-4 .....	32
Chapter III	
Figure 3-1 .....	53
Figure 3-2 .....	55
Figure 3-3 .....	56
Figure 3-4 .....	60
Table 3-1 .....	60
Figure 3-5 .....	62
Figure 3-6 .....	64
Figure 3-7 .....	66
Figure 3-8 .....	69
Table 3-2 .....	70
Table 3-3 .....	76
Table 3-4 .....	78

Dataset S3-1 .....	96
Dataset S3-2 .....	100
Dataset S3-3 .....	104
Chapter IV	
Figure 4-1 .....	113
Figure 4-2 .....	117
Figure 4-3 .....	120
Figure 4-4 .....	122
Figure 4-5 .....	125
Figure 4-6 .....	128
Figure 4-7 .....	133
Figure 4-8 .....	135
Figure 4-S1 .....	146
Figure 4-S2 .....	147
Figure 4-S3 .....	148
Figure 4-S4 .....	150
Figure 4-S5 .....	153
Figure 4-S6 .....	156
Figure 4-S7 .....	156
Figure 4-S8 .....	157
Figure 4-S9 .....	159
Figure 4-S10 .....	160
Figure 4-S11 .....	161
Figure 4-S12 .....	162
Figure 4-S13 .....	162
Table 4-S1.....	164
Chapter V	
Figure 5-1 .....	176
Table 5-S1.....	183

Appendix A	
Figure A-1 .....	190
Figure A-2 .....	190
Figure A-3 .....	191
Figure A-4 .....	192
Figure A-5 .....	193
Figure A-6 .....	194
Table A-1 .....	195
Appendix B 201	
Figure B-1 .....	206
Figure B-2 .....	207
Figure B-3 .....	209 (212)

## Chapter I: Introduction

Iron (Fe) is a critical element for many cellular processes, including respiration, and amino acid and DNA biosynthesis (1). It is used in Fe-sulfur clusters, heme-containing enzymes, and as active centers in metalloenzymes. In biological systems, free iron is rarely available, because ferric iron [Fe(III)] is insoluble and ferrous iron [Fe(II)] is toxic. In aerobic environments, Fe(II)'s toxicity is mediated through the Fenton reaction, which generates hydroxyl radicals that attack biological macromolecules. Fe(II) has also been shown to have toxicity in anaerobic systems, though the mechanisms are not well understood (2, 3). Fe bioavailability is especially important for pathogenic bacteria, as the host and bacteria compete for Fe resources. Many bacteria can live in a biofilm growth mode, where the bacteria live in a community within an excreted extracellular matrix. To live in a biofilm the environmental Fe concentration must be just right, as too little Fe prevents biofilm formation (4, 5) and too much Fe causes biofilm dispersal (6). Bacteria living in biofilms are more resistant to antibiotic treatment than free-living cells.

Bacteria harness many methods to ensure they have the ideal Fe intracellular concentrations. This includes a remarkably diverse set of iron scavenging systems as well as regulatory networks that respond to the intracellular Fe concentrations (see Fig. 1-1). This includes siderophores, heme, and ferric citrate iron acquisition systems. In particular I will focus on those in *Pseudomonas aeruginosa*, an opportunistic pathogen best-known in the context of lung infections in cystic fibrosis (CF) patients. In addition, other patients most susceptible to *Pseudomonas* infections are burn victims, neutropenic patients, and patients with pneumonia (7). *Pseudomonas* also contains two

heme uptake systems whose outer membrane receptors are HasR and PhuR, which are energized by an ABC transporter and TonB complex, respectively (8). The TonB complex consists of the inner membrane protein TonB, which spans the periplasmic space and associates with outer membrane iron receptors, as well as two inner membrane accessory proteins ExbB and ExbD. It is driven by the proton motive force and allows active import through the outer membrane (9). Another method of iron acquisition is through the chelator citrate. Excreted citrate binds Fe(III) and is sensed by the outer membrane receptor FecA. Fe(III) is reduced to Fe(II) in the periplasm and the G-coupled Fe(II)-specific FeoB transporter allows iron into the cell (10).

*Pseudomonas* synthesizes two siderophores, pyoverdine and pyochelin, whose Fe(III) affinities are  $10^{24-27} \text{ M}^{-1}$  and  $10^5 \text{ M}^{-1}$ , respectively (11). The iron affinity for pyoverdine is so great that it can strip iron from CF host immune system associated Fe-carriers, lactoferrin (12) and serum Fe transporters, ferritin (13). Iron-loaded siderophores are sensed by outer membrane embedded receptors and iron is actively transported into the cell through the TonB complex. There are two receptor proteins associated with pyoverdine FpvA and FpvB. Both are required for Fe(III) uptake; it is thought that Fe(III) is reduced to Fe(II) in the periplasm allowing Fe to enter the cell and for Fe free pyoverdine recycling. Additionally, FpvA also interacts with an extracytoplasmic function (ECF) sigma factor/anti-sigma factor system FpvR/PvdS, allowing for Fe uptake feedback. Pyochelin uptake differs in that Fe(III)-pyochelin enters the periplasm through FptA then the complex is transported into the cytosol, where pyochelin is degraded to release Fe(III).

Most Fe studies have focused on Fe(III), the assumption being that Fe(III) is the iron species most commonly encountered because in the presence of oxygen Fe(II) is readily converted to Fe(III). However, Fe(II) is stable under reducing, acidic conditions. For example, the Fe(II) transporter FeoB is upregulated in low iron, low oxygen conditions (14). The master regulator responsible for the cell's global iron response is Fur. When the cell is iron-replete, Fe(II) binds Fur, which then acts as a repressor for all iron acquisition genes. Free cytosolic iron is always Fe(II); therefore, Fur represses its regulon regardless of Fe speciation.

However, some extracellular iron sensing systems exist in bacteria. Previously, a Fe(III) sensing two-component system (TCS), PmrAB, was described. TCSs are the main mechanism bacteria interact with their extracellular environment and the simplest type is composed of a sensor histidine kinase that phosphorylates its cognate response regulator, which then acts as a transcription factor. Here, I describe the first bacterial Fe(II)-specific TCS, BqsRS in *P. aeruginosa* that starts sensing Fe(II) at low micromolar Fe(II) concentrations. Fe(II) sensing is important for *P. aeruginosa* because Fe(II) has been found in CF sputum at concentrations where BqsRS is active (15). Furthermore, only Fe(II) correlates with severe lung function decline (15). The sickest patients have bacteria immersed in an environment not typically studied, which may change bacterial metabolism affecting the efficacy of typical antimicrobial therapies. Recently, another Fe(II)-responsive TCS FirRS has been described in another pathogen, *Haemophilus influenzae*.



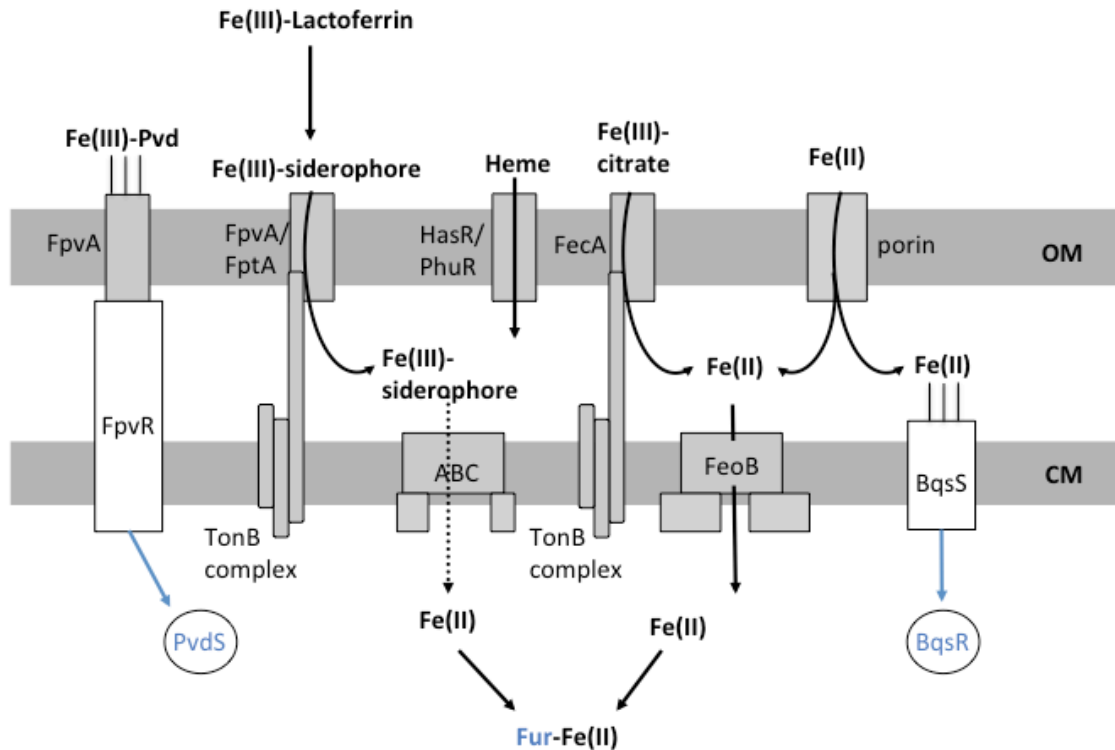


Figure 1-1. Iron acquisition and sensing in *P. aeruginosa*. Outer membrane iron receptors are energized, usually through the TonB complex. Ferric [Fe(III)]-siderophores are brought into the periplasm by FpvA for pyoverdine (Pvd) and FptA for pyochelin. Fe(III)-citrate uptake is performed by FecA. The two heme uptake receptors are HasR and PhuR. Because ferrous iron [Fe(II)] is soluble, it is thought to diffuse into the periplasm through outer membrane porins. Iron is actively transported into the cytosol, although all the proteins are not known (indicated by dashed line). FeoB brings Fe(II) into the cytosol. Individual extracellular Fe(III)-complexed Fe sources are sensed by receptors associated with ECF anti-sigma factor/sigma factor systems, such as the FpvR/PvdS pyoverdine sensing system. Fur, the master iron regulator, represses all iron acquisition systems (including PvdS function) in Fe replete conditions. The Fe(II)-specific two-component system, BqsRS is the first bacterial extracellular Fe(II)-sensing system characterized. Three lines indicate sensing and blue indicates a regulator.

## Overview

*Chapter 2. Can bacteria distinguish between extracellular Fe(II) and Fe(III)? I identify an operon containing a two-component system, BqsRS, through a microarray experiment comparing Fe(II) shocked cells to Fe(III) shocked cells. By comparing *bqsR* and *bqsS**

deletion mutants to WT and *feoB* mutant deficient in Fe(II) uptake, I show that BqsRS is responsible for extracellular Fe(II) sensing. BqsRS specifically responds to Fe(II) and not to other physiologically relevant divalent cations.

*Chapter 3.* How does Fe(II) affect *P. aeruginosa* through BqsRS? The periplasmic REXxE motif in BqsS is required for Fe(II) sensing and the consensus DNA sequence BqsR binds is shown. I observe that  $\Delta bqsR/S$  mutants have a severe growth defect in high Fe(II). Based on the predicted function of newly identified BqsR regulon members, I find two new phenotypes.  $\Delta bqsR/S$  mutants grown in high Fe(II) produce less spermidine than WT and are more sensitive than WT in high Fe(II) to aminoglycosides and polymyxins. I hypothesize this resistance is a pleiotropic effect from WT cells making their surface charge more positive, and thus providing resistance against only positively charged stressors, rather than acting as a general stress response mechanism.

*Chapter 4.* Are PWM missing critical information encoded in the operator or does the surrounding regulatory context dominate promoter outputs? I use the BqsR operator as a test case. Position weight matrices (PWM) are useful for generating hypotheses, but vastly over-predict transcription factor binding sites that have a regulatory consequence. Through bioinformatic genome analysis and generation of a promoter mutant library, I develop an energy binding based model of the BqsR operator. I find that while my model modestly outperforms PWM (generated in chapter 3), the surrounding regulatory parameters such as the presence of other transcription factor binding sites are more important in predicting promoter output.

*Chapter 5.* Where do we go from here? I suggest possible future experiments to address remaining questions.

*Appendix A.* How does Fe(II) impact the previously observed BqsRS-dependent rhamnolipid associated phenotypes (16)? I examine the effect of Fe(II) shock on rhamnolipid production and find that rhamnolipid concentration increases in both WT and  $\Delta bqsR$  strains. The quorum sensing regulator RhIR is responsible for rhamnolipid production. I examine how  $\Delta rhIR$ ,  $\Delta bqsR$ , and WT grow in minimal media in different carbon sources because  $\Delta rhIR$  grows much longer in exponential phase and to a higher maximum density than WT and  $\Delta bqsR$ .

*Appendix B.* Do other organism sense Fe(II) through the RExxE motif? A phylogenetic analysis reveals this motif is limited to *Pseudomonaceae* family; almost all strains in the *Pseudomonas* genus with a single *Cellvibrio* and *Azotobacter* strain. It would be interesting to test variants of the Fe(II)-sensing motif observed for their ability to bind Fe(II).

## References

1. **Wandersman C, Delepelaire P.** 2004. Bacterial iron sources: from siderophores to hemophores. *Annual Review of Microbiology* **58**:611-647.
2. **Dunning JC, Ma Y, Marquis RE.** 1998. Anaerobic killing of oral *streptococci* by reduced, transition metal cations. *Applied and Environmental Microbiology* **64**:27-33.
3. **Bird LJ, Coleman ML, Newman DK.** 2013. Iron and Copper Act Synergistically To Delay Anaerobic Growth of Bacteria. *Applied and Environmental Microbiology* **79**:3619-3627.
4. **Singh PK, Parsek MR, Greenberg EP, Welsh MJ.** 2002. A component of innate immunity prevents bacterial biofilm development. *Nature* **417**:552-555.
5. **Beinert H, Holm RH, Munck E.** 1997. Iron-Sulfur Clusters: Nature's Modular, Multipurpose Structures. *Science* **277**:653-659.
6. **Musk DJ, Banko DA, Hergenrother PJ.** 2005. Iron salts perturb biofilm formation and disrupt existing biofilms of *Pseudomonas aeruginosa*. *Chemistry & Biology* **12**:789-796.
7. **Cachia PJ, Hodges RS.** 2003. Synthetic peptide vaccine and antibody therapeutic development: Prevention and treatment of *Pseudomonas aeruginosa*. *Peptide Science* **71**:141-168.
8. **Ochsner UA, Johnson Z, Vasil ML.** 2000. Genetics and regulation of two distinct haem-uptake systems, *phu* and *has*, in *Pseudomonas aeruginosa*. *Microbiology* **146**:185-198.
9. **Vasil ML.** 2006. How we learnt about iron acquisition in *Pseudomonas aeruginosa*: a series of very fortunate events. *BioMetals* **20**:587-601.
10. **Marshall B, Stintzi A, Gilmour C, Meyer J-M, Poole K.** 2009. Citrate-mediated iron uptake in *Pseudomonas aeruginosa*: involvement of the citrate-inducible FecA receptor and the FeoB ferrous iron transporter. *Microbiology* **155**:305-315.
11. **Reid DW, Anderson GJ, Lamont IL.** 2009. Role of lung iron in determining the bacterial and host struggle in cystic fibrosis. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **297**:L795-L802.
12. **Skaar EP.** 2010. The Battle for Iron between Bacterial Pathogens and Their Vertebrate Hosts. *PLoS Pathog* **6**:e1000949.
13. **Sriyosachati S, Cox CD.** 1986. Siderophore-mediated iron acquisition from transferrin by *Pseudomonas aeruginosa*. *Infect. Immun.* **52**:885-891.
14. **Kim H, Lee H, Shin D.** 2013. The FeoC protein leads to high cellular levels of the Fe(II) transporter FeoB by preventing FtsH protease regulation of FeoB in *Salmonella enterica*. *Journal of Bacteriology* **195**:3364-3370.
15. **Hunter RC, Asfour F, Dingemans J, Osuna BL, Samad T, Malfroot A, Cornelis P, Newman DK.** 2013. Ferrous iron is a significant component of bioavailable iron in cystic fibrosis airways. *mBio* **4**:e00557-13.
16. **Dong Y-H, Zhang X-F, An S-W, Xu J-L, Zhang L-H.** 2008. A novel two-component system BqsS-BqsR modulates quorum sensing-dependent biofilm decay in *Pseudomonas aeruginosa*. *Commun Integr Biol* **1**:88-96.

**Chapter II: BqsR/BqsS constitute a two-component system that senses extracellular Fe(II) in *Pseudomonas aeruginosa***

Naomi N. K. Kreamer, Jessica C. Wilks, Jeffrey J. Marlow, Maureen L. Coleman, and Dianne K. Newman

This chapter was adapted from the manuscript: [BqsR/BqsS constitute a two-component system that senses extracellular Fe\(II\) in \*Pseudomonas aeruginosa\*](#). Kreamer NN, Wilks JC, Marlow JJ, Coleman ML, Newman DK. J Bacteriol. 2012 Mar;194(5):1195-204. doi: 10.1128/JB.05634-11.

## Abstract

*Pseudomonas aeruginosa* is a ubiquitous Gram-negative bacterium best known as the predominant opportunistic pathogen infecting the lungs of cystic fibrosis patients. In this context, it is thought to form biofilms, within which locally reducing and acidic conditions can develop that favor the stability of ferrous iron [Fe(II)]. Because iron is a signal that stimulates biofilm formation, we performed a microarray study to determine whether *P. aeruginosa* strain PA14 exhibits a specific transcriptional response to extracellular Fe(II). Among the genes that were most upregulated in response to Fe(II) were those encoding the two-component system BqsR/BqsS, previously identified for its role in *P. aeruginosa* strain PAO1 biofilm decay (1); here, we demonstrate its role in extracellular Fe(II) sensing. *bqsS* and *bqsR* form an operon together with two small upstream genes, *bqsP* and *bqsQ*, and one downstream gene, *bqsT*. BqsR/BqsS sense extracellular Fe(II) at physiologically relevant concentrations (>10  $\mu$ M) and elicit a specific transcriptional response, including its autoregulation. The sensor distinguishes between Fe(II), Fe(III), and other divalent cations [Ca(II), Cu(II), Mg(II), Mn(II), Zn(II)] under aerobic or anaerobic conditions. The gene that is most upregulated by BqsR/BqsS, as measured by quantitative reverse transcription-PCR (qRT-PCR), is PA14\_04180, which is predicted to encode a periplasmic oligonucleotide/oligosaccharide-binding domain (OB-fold) protein. Coincident with phenazine production during batch culture growth, Fe(II) becomes the majority of the total iron pool and *bqsS* is upregulated. The existence of a two-component system that senses Fe(II) indicates that extracellular Fe(II) is an important environmental signal for *P. aeruginosa*.

## Introduction

Iron is an essential element for nearly all forms of life. Iron generally presents itself to biological systems in one of two oxidation states: ferric [Fe(III)] or ferrous [Fe(II)] iron. Given that Fe(II) is oxidized by atmospheric oxygen, it is commonly assumed that organisms most frequently acquire iron in the Fe(III) form. In oxic environments at a circumneutral pH, Fe(II) rapidly oxidizes to Fe(III), precipitating as sparingly soluble ferric (hydr)oxide minerals (2). Within living systems, host proteins, such as members of the transferrin family, bind Fe(III) (3, 4). Given Fe(III)'s low solubility, considerable attention has been paid to Fe(III) sensing and acquisition by bacteria, and much is understood about these processes (5-8). Less is understood about Fe(II) sensing and uptake, despite the fact that many environments contain Fe(II) (9, 10) and that increased solubility of Fe(II) renders it more bioavailable than Fe(III) (2). In particular, Fe(II) is known to be important in acidic and/or reducing environments (2, 11, 12), and such conditions can easily arise at the microscale in diverse habitats (13, 14), including biofilms (15).

*Pseudomonas aeruginosa* is a Gram-negative bacterium capable of aerobic and anaerobic respiration. It inhabits many environments, ranging from soils (16) to marine sediments (17). It is best known as an opportunistic pathogen that infects immunocompromised individuals and, in particular, the mucus that accumulates on the surface of lung epithelial cells in individuals with cystic fibrosis (CF). Within this environment, *Pseudomonas* is thought to form biofilms—sessile, multicellular communities (18, 19) that are more resistant to conventional antimicrobial therapies

than are their free-living counterparts (20). *P. aeruginosa* requires iron to form cofactors of enzymes that play essential roles in electron transfer and other important cellular processes. Iron also signals biofilm formation (21, 22), but the level of iron necessary to promote biofilm formation exceeds that required for assimilatory purposes (23, 24). In addition, the production of several virulence determinants, such as exotoxin A, is regulated in response to iron (25).

*P. aeruginosa* can acquire iron in several ways. One strategy is to secrete high-affinity Fe(III)-binding molecules, called siderophores, which outcompete transferrin and other host proteins for Fe(III) (5). For example, pyoverdine is one of the dominant siderophores made by *P. aeruginosa*. When pyoverdine-bound Fe(III) enters the periplasm via TonB-dependent outer membrane transporters, iron is thought to be released from the pyoverdine complex via reduction facilitated by its binding to an inner membrane ABC transporter, allowing Fe(II) to enter the cytoplasm (7). In addition, *P. aeruginosa* can excrete citrate, which also chelates Fe(III), albeit with lower affinity than siderophores, causing the iron to enter the cytoplasm as Fe(II) through a different TonB-dependent porin and the G-protein-like transporter FeoB (26, 27). *Pseudomonas* also contains two heme uptake systems whose outer membrane receptors are HasR and PhuR, which are energized by an ABC transporter and TonB, respectively (28).

Limiting cells for Fe(III) has received considerable attention from a therapeutic perspective (22, 29, 30). This is because Fe(III) limitation has been shown to significantly reduce biofilm formation in various *in vitro* systems (22). However, these systems are commonly operated under aerobic conditions and may not reflect *in vivo* conditions.



Some evidence suggests anaerobic niches may exist in CF lung mucus where bacteria form dense biofilms (31), and biofilms themselves are known to harbor microdomains that are more acidic (15) or more reducing (32) than the bulk environment. Because steep chemical gradients can develop within biofilms over a small spatial scale (14), it is reasonable to expect that Fe(II) could exist in appreciable amounts in these microhabitats. Although the dominant iron oxidation state in the CF lung is unclear (33), it likely changes as infections progress and the local environment becomes increasingly reducing. On a related note, one reason to postulate that Fe(II) may be important for *P. aeruginosa* is that phenazines are known to be produced at micromolar concentrations in the CF lung (34). Phenazines can reduce host protein-bound Fe(III) (35), as well as Fe(III) bound up in mineral phases (36), to Fe(II). Fe(II) crosses the outer membrane via generic outer membrane porins and is actively transported across the cytoplasmic membrane via FeoB (26). Recently, it was shown that phenazine carboxylic acid can promote *P. aeruginosa* biofilm formation by stimulating Fe(II) acquisition in the presence of conalbumin (37). In addition, the presence of Fe(II) has been detected within a *P. aeruginosa* biofilm producing a gradient of the reduced phenazine pyocyanin (38).

Given these observations, we hypothesized that *P. aeruginosa* might respond differently to Fe(II) than to Fe(III). To test this, we performed a microarray experiment to measure the transcriptional response to Fe(II) or Fe(III) shock. Not only did we observe different responses depending on the iron oxidation state, we identified a two-component system that specifically senses extracellular Fe(II).

## Results

*P. aeruginosa* exhibits a transcriptional response that depends on the iron oxidation state. To determine if exogenous Fe(II) and Fe(III) are perceived differently at the cellular level by *P. aeruginosa* strain PA14, we used Affymetrix *P. aeruginosa* GeneChips to identify transcriptional changes in response to Fe(II) or Fe(III). To ensure that the iron oxidation state remained stable, strain PA14 was cultured anaerobically as described in Materials and Methods. Upon reaching exponential phase, triplicate cultures were exposed to approximately 100  $\mu$ M Fe(II), 100  $\mu$ M Fe(III), or water for 30 min, and total RNA was extracted, labeled, and hybridized to *P. aeruginosa* Affymetrix GeneChips.

In both the Fe(II) and Fe(III) shock conditions, relative to the no-iron control, iron-scavenging genes (e.g., pyochelin biosynthesis genes) were downregulated, as expected (see the supplemental material). To determine if any genes specifically responded to Fe(II), we compared the Fe(II) and Fe(III) shock treatments. Cultures exposed to extracellular Fe(II) displayed an expression profile that was distinct from that of Fe(III)-exposed cultures (Table 2-1). We found 3 genes that were downregulated and 18 genes that were upregulated at least 2-fold by Fe(II) with *q*values of  $\leq 0.01$ . Of the upregulated genes, a putative operon containing two short putative membrane proteins with PepSY domains and a putative two-component system showed the highest upregulation (Table 2-1).

TABLE 2-1. Exogenous Fe (II) upregulates a specific subset of genes in PA14

Genes	PA14 ID	Locus	Fold		Description
			Change	q-value	
<i>bqsP</i>	PA14_29710*	PA2659	98.4	0.00E+00	probable PEPSY type peptidase
<i>bqsQ</i>	PA14_29720*	PA2658	41.1	0.00E+00	probable PEPSY type peptidase
<i>bqsR</i>	PA14_29730*	PA2657	34.1	0.00E+00	probable two-component response regulator conserved hypothetical protein; periplasmic localization
<i>bqsS</i>	PA14_04180*	PA0320	27.5	0.00E+00	probable two-component sensor
	PA14_29740*	PA2656	10.4	3.00E-05	probable ferric reductase
	PA14_07070*	PA0545	6.5	4.00E-03	probable carbonic anhydrase
	PA14_01240*	PA0102	6.5	2.00E-05	probable sulfate transporter
	PA14_01250*	PA0103	5.8	0.00E+00	probable transcription factor
<i>inaA</i>	PA14_04270*	PA0327	5.2	1.00E-05	heavy metal transport protein/chaperone
	PA14_18800	PA3520	4.4	5.14E-03	InaA protein
	PA14_56930	PA4378	3.9	5.60E-04	hypothetical protein
	PA14_52340	PA0921	3.4	2.39E-03	probable methyl-transferase
	PA14_56930	PA4379	3.2	5.60E-04	hypothetical protein
	PA14_29750	PA2655	3.1	2.33E-03	hypothetical protein
<i>dsbB</i>	PA14_34170	PA2358	3.0	5.52E-03	disulfide bond formation protein
	PA14_07000	PA0538	2.8	3.58E-03	probable iron-binding protein IscA
<i>iscA</i>	PA14_14750	PA3812	2.7	2.75E-03	hypothetical protein
	PA14_72370	PA5482	2.3	1.59E-03	hypothetical protein
	PA14_30410	PA2604	1.8	2.62E-03	pyridoxamine 5'-phosphate oxidase
<i>pdxH</i>	PA14_50800	PA1049	-1.6	5.61E-03	respiratory nitrate reductase beta chain
<i>narH</i>	PA14_13800	PA3874	-1.8	8.71E-03	azurin precursor
<i>azu</i>	PA14_65000	PA4922	-1.9	2.39E-03	molybdenum cofactor biosynthetic protein A1
<i>moeA1</i>	PA14_13280	PA3914	-2.1	1.86E-03	

	PA14_21630	PA3278	-2.1	4.00E-03	hypothetical protein
<i>rpmF</i>	PA14_25630	PA2970	-2.6	1.90E-03	50S ribosomal protein L32

\* Expression of these genes is controlled by the PA2657 (*bqsR*) response regulator as shown in Figure 2-3.

To determine whether these genes constitute an operon, we performed RT-PCR. All of the genes from PA14\_29710 to PA14\_29750 appear to be transcriptionally connected (Fig. 2-1), and thus we refer to this operon as the *bqsPQRST* operon, because homologs of BqsS and BqsR (exhibiting 100% amino acid identity) were previously identified in *P. aeruginosa* strain PAO1 for their role in biofilm development and quorum sensing (1). Whether *bqsT* is a bona fide member of this operon is less clear, as qRT-PCR results (Fig. 2-3) show lower fold change in this gene than in other members of the operon, possibly reflecting transcriptional bleed-through. Moreover, whereas all of the other members of the operon possess overlapping reading frames, there are 90 bp between *bqsS* and *bqsT*. The five-gene cluster is conserved in its entirety only among *P. aeruginosa* strains. Orthologs of the sensor BqsS are encoded in similar predicted operons with a response regulator and one or two PepSY domain-containing proteins across other *Pseudomonas* and *Azotobacter* species, but the fifth hypothetical gene is excluded.

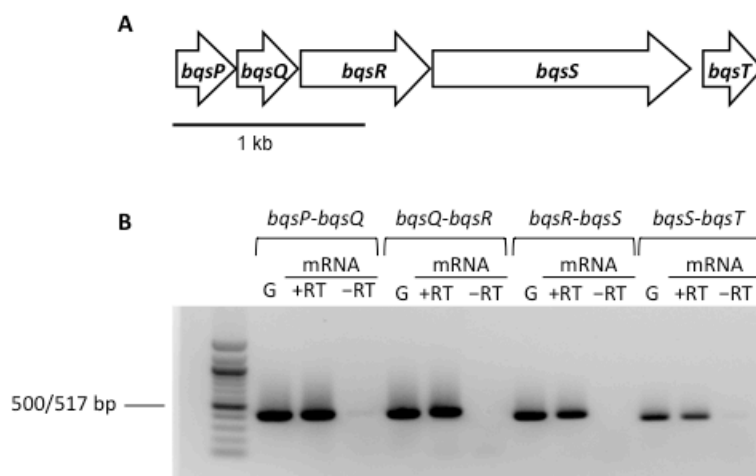


Figure 2-1. The *bqs* operon. (A) Genomic organization of *bqs* operon, which encodes two putative PepSY domain containing membrane proteins (*bqsP* [PA14\_29710] and *bqsQ* [PA14\_29720]), a response regulator (*bqsR* [PA14\_29730]), a histidine kinase (*bqsS* [PA14\_29740]), and a hypothetical protein (*bqsT* [PA14\_29750]) The length of the arrows represents the size of the gene relative to other genes in the operon. (B) A 1% agarose gel confirming that *bqs* is an operon. The notation *bqsP-bqsQ* indicates, for example, in lanes 2-4 PCR products were obtained using a forward primer within *bqsP* and a reverse primer within *bqsQ*. G denotes genomic DNA run as a positive control. mRNA -RT (reverse transcriptase) is a negative control to ensure there is no genomic DNA contamination in the mRNA. The presence of products in all of the cDNA lanes (mRNA +RT) reveals this is polycistronic mRNA that contains *bqsP*, *bqsQ*, *bqsR*, *bqsS*, and *bqsT*.

*Extracellular Fe(II) upregulates the bqs operon.*

It was striking that exogenous Fe(II) induced a two-component system, because such systems are responsible for transmitting extracellular signals (e.g., the presence of ferrous iron) to the intracellular environment (53). This suggested that our two-component system, comprising a cytoplasmic response regulator (*bqsR*) and a sensor histidine kinase (*bqsS*) and identified by its conserved kinase domain and its predicted localization to the inner membrane, was sensing extracellular Fe(II). To test this, we used qRT-PCR to measure the expression of *bqsR* and *bqsS* in response to Fe(II) shock in the

wild-type (WT) strain and in strain *feoB::MAR2xT7* [which has a severe defect in Fe(II) uptake across the cytoplasmic membrane (26, 27, 37)]. If the cells were sensing extracellular Fe(II), we anticipated that *bqsR* and *bqsS* expression would be the same in both the wild type and the *feoB* mutant. Alternatively, if the cells were sensing cytoplasmic Fe(II), we expected *bqsR* and *bqsS* to be upregulated in the wild type compared to their expression in the *feoB* mutant. *P. aeruginosa* strains PA14 and PA14 *feoB::MAR2xT7* were cultured under anaerobic conditions until early exponential phase, at which time 63  $\mu$ M Fe(II) was added to the medium. After 30 min, the RNA was stabilized and extracted for qRT-PCR. As seen in Fig. 2-2, *bqsR* and *bqsS* were upregulated in both the wild type and the *feoB* mutant upon Fe(II) shock, which indicates their expression is activated in response to extracellular Fe(II).

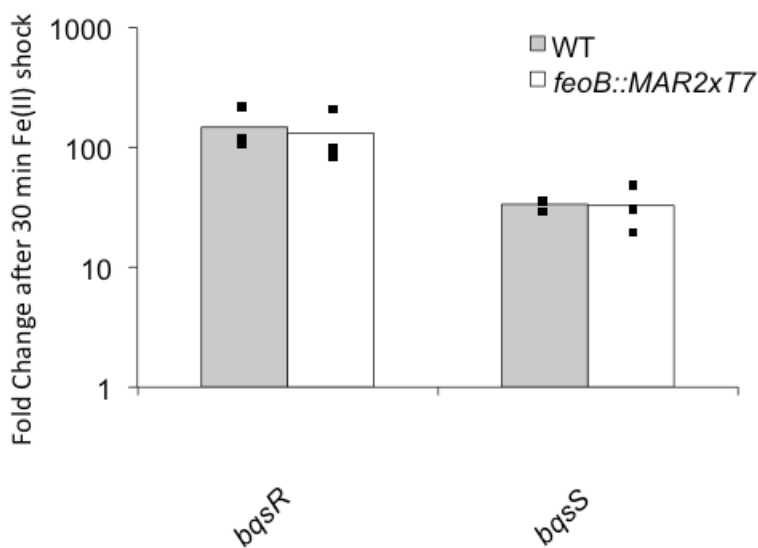


Figure 2-2. Extracellular Fe(II) upregulates *bqsR* and *bqsS*. Wild type strain PA14 (gray bars) and mutant strain *feoB::MAR2xT7* (white bars) were cultured anaerobically in MMM. The *feoB* mutant strain has a severe defect in Fe(II) uptake. A 30 min Fe(II) shock and qRT-PCR was performed to analyze expression of *bqsR* and *bqsS*. Since the wild type and the mutant show the same levels of transcriptional response, Fe(II) must be extracellular. The bars represent the average of biological triplicates and the individual points indicate the fold change of each replicate.

*Fe(II) response is due to the bqs operon.* To gain insight into the mechanism of the extracellular Fe(II) shock transcriptional response, we examined expression of the nine genes most upregulated in the microarray (each having a fold change of  $\geq 4$ ) in strains lacking either the regulator ( $\Delta bqsR$  mutants) or the sensor ( $\Delta bqsS$  mutants) (Fig. 2-3). Four of these genes are components of the *bqs* operon (*bqsPQRST*). Under anaerobic conditions, the  $\Delta bqsR$  mutant,  $\Delta bqsS$  mutant, and wild type were grown to early exponential phase and exposed to Fe(II). RNA was extracted for qRT-PCR analysis before and after 30 min of exposure. Our results reveal that BqsR/BqsS is autoregulated. When both BqsR and BqsS are present, extracellular Fe(II) induces the upregulation of the entire *bqs* operon (Fig. 2-3). However, if either *bqsR* or *bqsS* is deleted, the *bqs* operon no longer responds to extracellular Fe(II). Complementation with *bqsR* expressed on pMQ64 resulted in an Fe(II) response similar to that of the wild type (Fig. 2-3). The exception to this is expression of *bqsR* itself, which does not exhibit the same dramatic increase in expression with Fe(II) shock; this can be explained by the fact that before Fe(II) shock, *bqsR* is upregulated 75-fold compared to wild-type levels because its expression is being driven off a multicopy plasmid. No data are presented for *bqsR* in both the  $\Delta bqsR$  mutant and the vector-only control because no copies of *bqsR* are present and any values detected are within background error. The transcriptional responses of the five most highly upregulated genes outside this operon—PA14\_04180, PA14\_07070, PA14\_01240, PA14\_01250, and PA14\_04270—also depended on *bqsR* and *bqsS*. Of these genes, PA14\_04180 shows the highest upregulation by qRT-PCR (Fig. 2-3) and was the

fourth most upregulated in the microarray. In general, the strength of the transcriptional response in our qRT-PCR experiments was significantly higher than what we observed in the array study (Table 2-3), which is not unexpected because qRT-PCR is a more sensitive technique than microarrays.

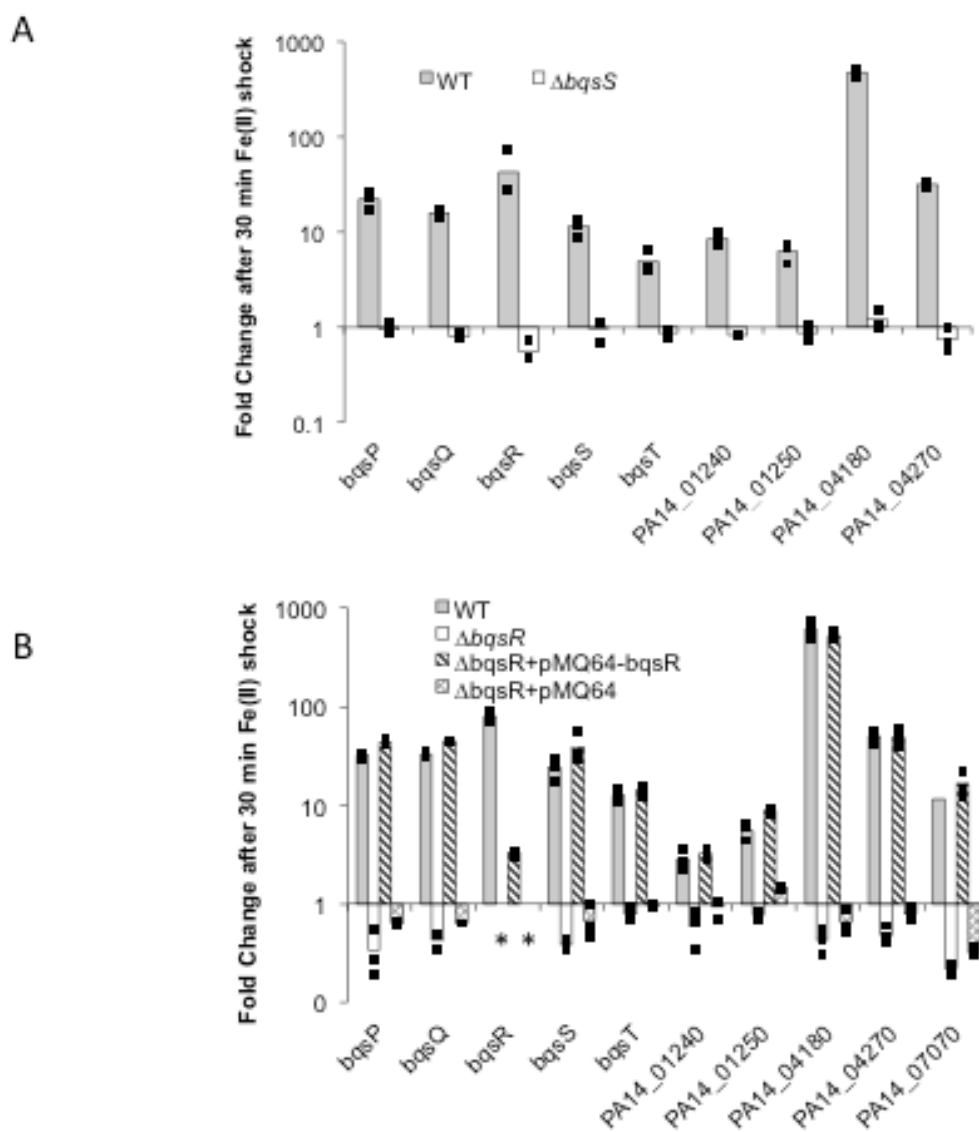


Figure 2-3. BqsS and BqsR are required for extracellular Fe(II) regulation. (A) WT (gray bars) and  $\Delta bqsS$  mutant (white bars) transcriptional fold change after Fe(II) shock, as measured by qRT-PCR. (B) WT (gray bars),  $\Delta bqsR$  mutant (white bars),  $\Delta bqsR$  mutant with pMQ64-*bqsR* (*bqsR* complement) (diagonal bars), and  $\Delta bqsR$  mutant with pMQ64 (vector-only control) (hashed bars) transcriptional fold change after Fe(II) shock, as measured by



qRT-PCR. Both *bqsR* and *bqsS* are required for the Fe(II)-induced upregulation of the *bqs* operon and upregulation of the genes in the *bqs* regulon.  $\Delta bqsR::pMQ64$ -*bqsR* shows WT levels of upregulation. The bars are the averages of biological triplicates, and the individual points indicate the fold changes of each replicate.

*The bqs operon responds to low concentrations of extracellular Fe(II).* Intrigued by the differential fold changes we observed in our qRT-PCR experiments, we hypothesized that the BqsS/BqsR-mediated transcriptional response was tunable by the Fe(II) concentration. To test this, we performed an Fe(II) titration experiment, varying Fe(II) over an order of magnitude from  $\sim 6 \mu\text{M}$  to  $\sim 60 \mu\text{M}$ . A concentration of  $\sim 20 \mu\text{M}$  was the lowest concentration of Fe(II) to elicit a response for *bqsR* or *bqsS*; the maximum response was seen for  $>60 \mu\text{M}$  Fe(II) (Fig. 2-4).

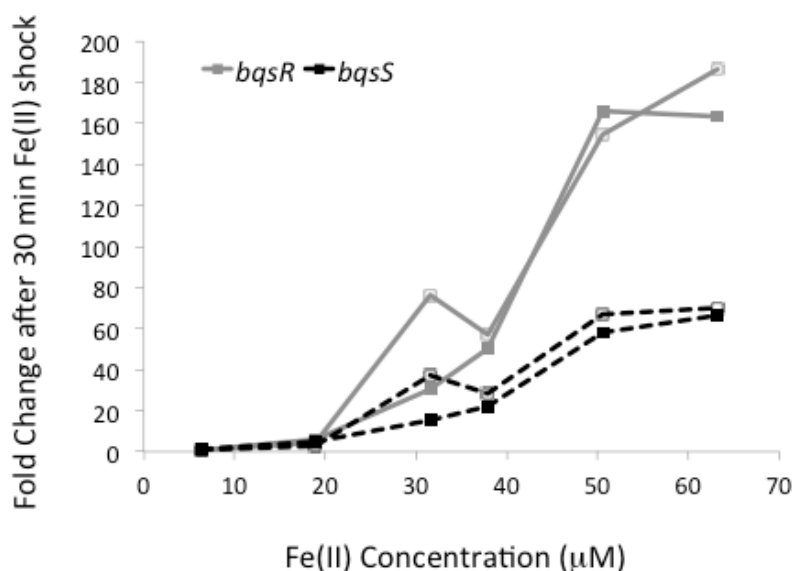


Figure 2-4. Fe(II) titration curve. Duplicate trials showing induction of a transcriptional response for *bqsR* (gray lines) and *bqsS* (dashed black lines) in response to increasing Fe(II) concentrations. The minimum Fe(II) concentration to elicit a response was  $\sim 20 \mu\text{M}$ , and the maximum fold change was seen with a concentration of  $\sim 60 \mu\text{M}$ . Additional replicates show similar trends.

The *bqs* operon is specific for Fe(II). Because some two-component systems are induced by multiple external signals, we questioned whether BqsR/BqsS was specific for Fe(II). To determine this, we performed transcriptional shock experiments with five other divalent cations that are commonly found in CF sputum (54). We also measured the transcriptional response to Fe(III) as a control. The wild type was cultured under either anaerobic or aerobic conditions and exposed to metal shock in early exponential phase. Fe(III), Ca(II), Cu(II), Mg(II), Mn(II), and Zn(II) do not significantly upregulate *bqsS* under either aerobic or anaerobic conditions (Fig. 2-5). From this, it appears that BqsR/BqsS is selective for ferrous iron. Notably, the extent of *bqsS* gene expression was attenuated in the aerobic Fe(II) shock experiment (Fig. 2-5). The lower degree of upregulation can be explained by rapid oxidation of Fe(II) to Fe(III): the initial concentration of Fe(II) was 100  $\mu$ M, but after completion of the 30-min shock, the concentration was 4.9  $\mu$ M.

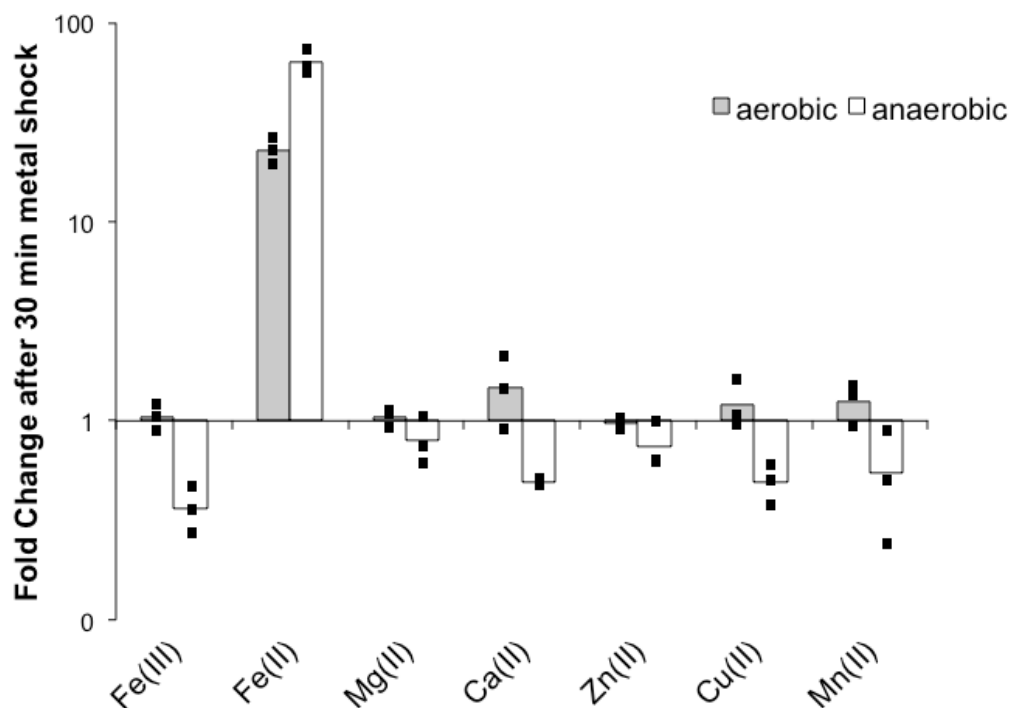


Figure 2-5. *bqsS* responds specifically to Fe(II). Expression of *bqsS* under both anaerobic (gray bars) and aerobic (white bars) conditions after shock with Fe(II) and other cations. Metal concentrations of 100  $\mu\text{M}$   $\text{FeCl}_3$ , 100  $\mu\text{M}$   $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ , 10 mM  $\text{MgCl}_2$ , 5 mM  $\text{CaCl}_2$ , 100  $\mu\text{M}$   $\text{ZnCl}_2$ , 100  $\mu\text{M}$   $\text{CuCl}_2$ , and 100  $\mu\text{M}$   $\text{MnCl}_2$  were used in the 30-min metal shocks for Fe(III), Fe(II), Mg(II), Ca(II), Zn(II), Cu(II), and Mn(II), respectively. The bars are the averages of biological triplicates, and the individual points indicate the fold changes of each replicate.

*The bqs operon is transcribed as Fe(II) is produced during growth.* To explore the connection between *bqs* operon transcriptional induction and biological iron reduction in a wild-type culture, a 27-h growth experiment was conducted in duplicate using PA14 grown in MOPS medium with an initial iron allocation of almost exclusively Fe(III) (Fig. 2-6).  $\text{OD}_{500}$  measurements show a classical growth curve exhibiting exponential growth between approximately 10 and 18 h after inoculation. Pyocyanin, a type of phenazine, was detected in the culture starting in mid-exponential phase and continuing through the end of the experiment. The culture's iron oxidation state shifted from entirely Fe(III) at the beginning of the experiment to >85% Fe(II) by the end; this shift coincided with the exponential phase of growth and the presence of phenazines. qRT-PCR showed upregulation of *bqsS* during exponential phase, when extracellular Fe(II) was at a concentration of  $\sim 125 \mu\text{M}$  (well above the concentration needed to induce the *bqs* operon), and a repression of transcription in stationary phase.

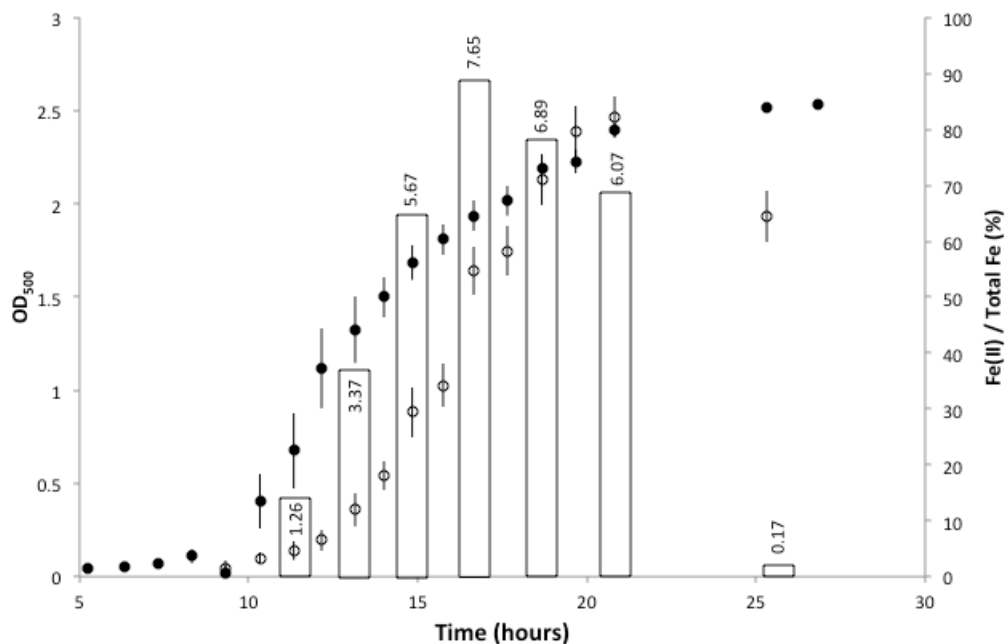


Figure 2-6. *bqsR* is transcribed over the course of growth as Fe(II) accumulates. Correlation between growth, phenazine production, iron reduction, and *bqsR* transcription in PA14. Growth curves (OD<sub>500</sub>) averaged for triplicate cultures of PA14 are plotted alongside the percentage of total Fe that were Fe(II). Bars indicate the average upregulation of *bqsS* (relative to *recA*) compared to cells grown without Fe and normalized to OD<sub>500</sub>. For the complete qPCR data set, see Table 2-2. Filled circles, PA14 OD<sub>500</sub>; open circles, percentage of total Fe as Fe(II).

Table 2-2. qRT-PCR data for Fig. 2-6.

Time (hours)	<i>bqsS</i>	SD
9.33	0.28	0.10
11.33	1.26	0.65
13.17	3.37	1.96
14.83	5.67	1.17
16.67	7.65	1.21
18.67	6.89	1.31
20.83	6.07	1.07
25.33	0.17	0.20

## Discussion

In this study, we identified a two-component system in *P. aeruginosa* strain PA14, BqsR/BqsS, that controls a specific transcriptional response to extracellular Fe(II). To our knowledge, this is the first extracellular Fe(II) sensor described in bacteria. BqsR/BqsS autoregulates its own expression, is turned on by Fe(II) concentrations of >10  $\mu$ M, and does not respond to other physiologically relevant dipositive ions. Previously, BqsR/BqsS was shown to play a role in the dispersal of *P. aeruginosa* strain PAO1 biofilms (1). This appears to be due to the fact that biofilms lacking BqsR or BqsS produce lower concentrations of rhamnolipids, which are known to promote biofilm dispersal (55).

Interestingly, *bqsR* and *bqsS* expression peaks during late exponential growth, in which cells are thought to experience conditions similar to those in biofilms (56, 57). The physiological context of *bqsR* and *bqsS* expression thus provides a satisfying link between our results and those of Dong *et al.*(1): as cell densities and phenazine concentrations rise, Fe(II) is generated, inducing *bqsR* and *bqsS* expression, which promotes biofilm dispersal. While several previous studies have indicated that moderate concentrations of iron are necessary to stimulate biofilm formation regardless of whether iron is in the ferric or ferrous form (21, 22), it has also been established that this phenomenon is concentration dependent: at sufficiently high concentrations, iron can promote biofilm dispersal (58). It thus seems likely that BqsR/BqsS acts as a “gatekeeper” to sense when Fe(II) concentrations have crossed the threshold where conditions are no longer favorable for biofilm maintenance.

That the BqsR/BqsS system responds specifically to extracellular Fe(II) but not Fe(III) was initially surprising, given that the PmrA/PmrB two-component system is specific for extracellular Fe(III) but not Fe(II) (8). In a detailed study by Wösten et al. (8), it was shown that PmrB directly interacts with Fe(III) through glutamate residues in two HExxE motifs and a serine outside these motifs in the periplasmic region. BqsS contains a similar motif that could be responsible for directly sensing Fe(II), a theory which we plan to test biochemically. Of the 23 BqsS orthologs we identified across currently sequenced pseudomonads, only 14 are predicted to contain the putative Fe(II) binding motif; 7 of these are strains of *P. aeruginosa*. Once the BqsR Fe(II) binding motif is biochemically confirmed, it will be interesting to determine how the BqsR/BqsS regulon differs among these species.

Not only does BqsS differentiate Fe(II) from Fe(III), it also distinguishes Fe(II) from other divalent ions that *Pseudomonas* might encounter: when challenged with Ca(II), Cu(II), Mg(II), Mn(II), and Zn(II), the wild type did not upregulate *bqsS*. This differs from the Fe(III)-sensing PmrA/PmrB system, which can also be activated by low extracellular Mg(II) (59) and mild acid (60). In addition, BqsR/BqsS are highly sensitive to Fe(II). The minimum Fe(II) level found to elicit a transcriptional response was  $\sim 20 \mu\text{M}$ ; the response rapidly increased over a small concentration range, with the maximum response achieved by a concentration of  $\sim 60 \mu\text{M}$ . These levels of Fe(II) are well within physiologically relevant Fe(II) concentrations, as the Fe(II) concentration has been measured up to  $110 \mu\text{M}$  in CF sputum samples (R.C. Hunter and D. K. Newman,

unpublished data). Other studies have obtained sputum samples and measured total Fe concentrations, which range from approximately 20  $\mu\text{M}$  to 140  $\mu\text{M}$  (61).

Both BqsP/BqsQ contain PepSY domains (Pfam accession number [PF03413](#)) (62) and are predicted to be single-pass transmembrane proteins by TMpred (63) and DAS (64). BqsP is predicted to share structural homology with the protein Yycl from *Bacillus subtilis* (75% estimated precision using the Phyre homology modeling program) (65). Yycl, along with another auxiliary protein, is known to regulate the kinase activity of a two-component system, YycFG, through its transmembrane domain (66). Deleting Yycl caused a 10-fold increase in expression of a *lacZ* kinase promoter fusion (32). Given the similarity between BqsP and Yycl, it seems likely that BqsP/BqsQ play a similar role in modulating the kinase activity of BqsS. *bqsT* encodes a hypothetical protein that lacks conserved domains and has only very weak similarity to any known proteins. Orthologs to *bqsT* are only present in other *P. aeruginosa* strains, implying that, if it is a true member of the *bqs* operon, it has been integrated into the operon recently and may not serve a critical function.

Strikingly, only a small number of genes in our microarray were upregulated in the Fe(II) shock condition compared to their levels in the Fe(III) shock condition, and we did not observe upregulation of *pqsA*, *phnA* (PQS biosynthesis), and *rhIA* (rhamnolipid production), as reported by Dong et al. (1). This is likely because we performed our shock experiment on cells harvested from early exponential phase. If BqsR/BqsS are involved in the regulation of genes that are also quorum regulated (i.e., *pqsA*, *phnA*, and *rhIA*) we would not have detected them. Another possibility is that the BqsR/BqsS

regulon is different in PA14 than in PAO1. Regardless, the “core” Fe(II)-sensitive BqsR/BqsS regulon is quite limited. This may be due to the fact that only Fe(II) enters and is sensed in the cytoplasm, no matter whether Fe(III) or Fe(II) is extracellular (Fig. 2-7). Thus, the primary network of iron-responsive genes would not need to distinguish between the two forms. Perhaps instead, extracellular Fe(II) is sensed to assess environmental conditions that foster Fe(II) formation. Fe(II) exists in reducing and acidic environments. Consistent with this, genes that respond to Fe(II) shock include PA14\_01240, encoding carbonic anhydrase, and *iscA*, an acid response gene involved in iron-sulfur cluster biogenesis (67). Their protein products can help the cell adjust to an external decrease in pH. If the extracellular environment becomes more reducing, the disulfide bonds of periplasmic proteins are reduced to free sulfhydryl groups, greatly decreasing their stability. The upregulated *dsbB* encodes a disulfide bond-forming protein with predicted periplasmic localization, which could ameliorate these effects.



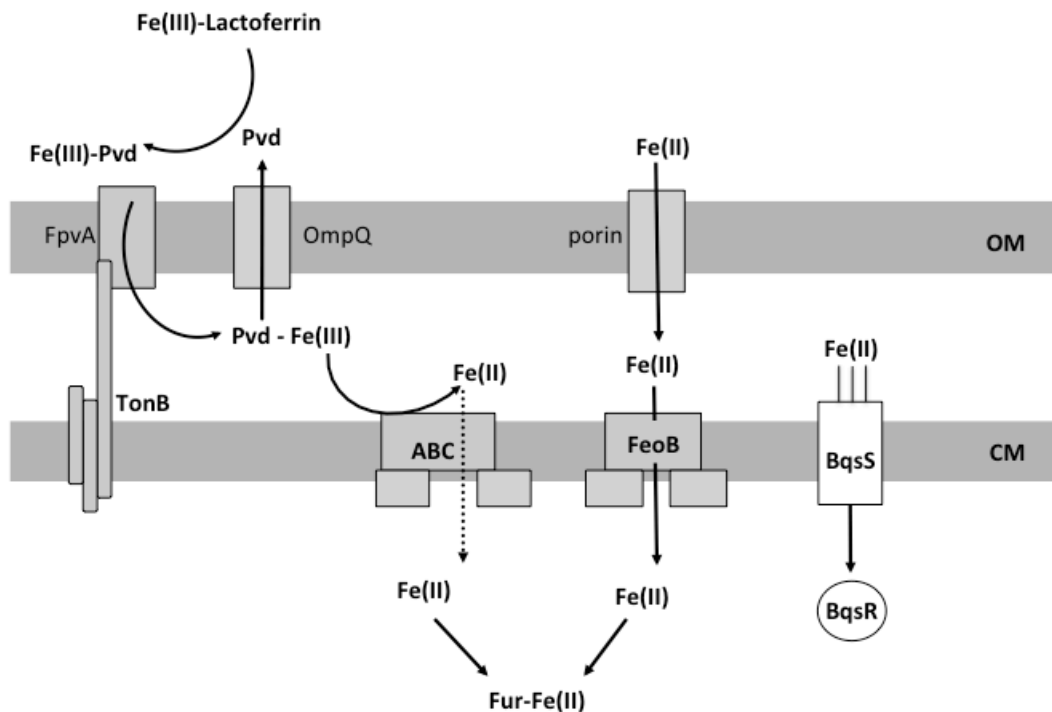


Figure 2-7. Iron acquisition model. **Fe(III)**: Outside of the cell Fe(III) binds to the siderophore pyoverdine and this complex enters the periplasm via the TonB-dependent outer membrane transporter FpvA (7). The reduction of Fe(III) to Fe(II) frees the iron from pyoverdine. A pyoverdine efflux pump, OmpQ, recycles free pyoverdine by transporting it out of the cell (68). The Fe(II) then enters the cytoplasm through an unidentified ABC transporter. Upon entering the cytoplasm, Fe(II) binds Fur, which subsequently represses multiple iron acquisition and iron storage genes (69). **Fe(II)**: Fe(II) enters the periplasm through a nonspecific porin in the outer membrane. While in the periplasm, Fe(II) activates the histidine kinase BqsS, which subsequently activates the response regulator BqsR. BqsR then goes on to induce its regulon. Fe(II) enters the cytoplasm through the ferrous iron transporter FeoB (26). Once in the cytoplasm, Fe(II) binds Fur, which results in the repression of iron acquisition genes. Solid arrows indicate known pathways, dashed arrows indicate putative pathways, and three lines indicate binding. Pvd, pyoverdine; OM, outer membrane; CM, cytoplasmic membrane.

One clear result is that the PA14\_04180 gene is highly upregulated by BqsR/BqsS. Bioinformatic predictions suggest it is periplasmically localized and has a bacterial oligonucleotide/oligosaccharide-binding domain (OB-fold). The distinguishing feature of bacterial OB-fold proteins compared to other OB-fold proteins is that the bacterial ligand-binding domain binds cationic ligands rather than nucleotides (70).

Proteins with such domains can serve a variety of functions and have been implicated in bacterial pathogenesis (71). In fact, the Fe(III)-binding PmrA/PmrB regulates a similar periplasmic OB-fold protein (OmdA) that confers resistance to polymyxin B, a cationic antimicrobial peptide (72). Whether PA14\_04180 confers a similar resistance in *Pseudomonas* species is currently unknown.

The identification of an extracellular Fe(II)-sensing system that is specific for Fe(II), can distinguish between Fe(II) and Fe(III), and is insensitive to other divalent ions opens up several avenues for future research. It reminds us that the concentration of Fe(II) can be sufficiently high in some habitats for bacteria to have evolved specific mechanisms to sense it. What their transcriptional response is and how it helps them adapt to these conditions remain to be determined. Once the molecular basis of Fe(II) specificity is understood, it may be possible to identify other proteins that have a similar specificity for Fe(II). Finally, monitoring the transcription of genes regulated by BqsR/BqsS may provide an opportunity to indirectly gauge whether *P. aeruginosa* is experiencing Fe(III) or Fe(II) in its environment. For example, knowing the iron oxidation state in CF sputum or biofilm microenvironments would inform the proper administration of antimicrobial therapies targeting iron uptake.

### **Materials and Methods**

*Bacterial strains and growth conditions.* Bacterial strains used in this study are listed in Table 2-3. Stationary-phase cultures of *P. aeruginosa* strains were grown aerobically at 37°C in 3 g/liter Bacto tryptic soy broth containing 100 mM KNO<sub>3</sub>. The overnight cultures were diluted to an optical density at 500 nm (OD<sub>500</sub>) of 0.01 in minimal metal medium

(MMM). MMM consists of 0.3 g/liter Bacto-tryptic soy broth, 50 mM glutamate, 1% glycerol, 100 mM KNO<sub>3</sub>, and 50 mM MOPS (morpholinepropanesulfonic acid); the pH was adjusted to 7.0 with NaOH, and the medium was treated with 1% (wt/vol) Amberlite chelating resin (Sigma) for 1 h and filtered into acid-washed medium bottles prior to autoclaving. After autoclaving, MMM was immediately sparged to maintain anaerobic conditions and brought into a vinyl anaerobic chamber containing a 5% hydrogen, 15% carbon dioxide, and 80% nitrogen gas mixture (Coy Laboratory Products). Ten milliliters of medium was added to 25-ml anaerobic Balch tubes, which were acid-washed in 1 M HCl to remove residual metals. Unless otherwise specified, inoculations were performed anaerobically in the anaerobic chamber with the previously stated gas mixture and incubated and shaken at 37°C. Spectrophotometric measurements were performed on a Beckman Spec20 unless otherwise specified.

TABLE 2-3. Strains and plasmids used in this study

Strains/plasmids	Characteristics	Source or reference
<i>P. aeruginosa</i>		
PA14		
WT	Clinical isolate UCBPP-PA14	(39)
$\Delta bqsR$	PA14 with deletion in <i>bqsR</i>	This study
$\Delta bqsR$ + pMQ64	PA14 with deletion in <i>bqsR</i> containing plasmid pMQ64	This study
$\Delta bqsR$ + pMQ64- <i>bqsR</i>	PA14 with deletion in <i>bqsR</i> containing plasmid pMQ64- <i>bqsR</i>	This study
$\Delta bqsS$	PA14 with deletion in <i>bqsS</i>	This study
<i>feoB</i>	PA14 <i>feoB</i> :: MAR2xT7; transposon mutant containing Gm <sup>R</sup> cassette	(40)
<i>E. coli</i>		
UQ950	<i>E. coli</i> DH5 $\alpha$ $\lambda$ (pir) host for cloning; F- $\Delta$ ( <i>argF-lac</i> )169 $\theta$ 80 <i>dlacZ</i> 58( $\Delta$ M15) <i>glnV</i> 44(AS) <i>rfbD</i> 1 <i>gyrA</i> 96(Nal <sup>R</sup> ) <i>recA</i> 1 <i>endA</i> 1 <i>spoT</i> 1 <i>thi</i> -1 <i>hsdR</i> 17	D. Lies, Caltech

BW29427	<i>deoR</i> $\lambda$ <i>pir+</i> Donor strain for conjugation: <i>thrB1004 pro thi rpsL hsdS lacZ</i> $\Delta$ M15RP4-1360 $\Delta$ ( <i>araBAD</i> )567 $\Delta$ <i>dapA1341::[erm pir(wt)]</i>	W. Metcalf, University of Illinois
<i>S. cerevisiae</i> InvSc1	<i>MATa/MAT<math>\alpha</math> leu2/leu2 trp1-289/trp1-289 ura3-52/ura3-52 his3-<math>\Delta</math>1/his3-<math>\Delta</math>1</i>	Invitrogen
Plasmids		
pMQ30	pEX18-Gm <sup>R</sup> , <i>URA3</i> , <i>CEN6</i> , <i>ARSH4</i> ; Allelic replacement vector	(41)
p $\Delta$ <i>bqsR</i>	2kb fusion PCR fragment containing $\Delta$ <i>bqsR</i> cloned into the BamHI/EcoRI site of pMQ30; used to make $\Delta$ <i>bqsR</i>	This study
p $\Delta$ <i>bqsS</i>	2kb fusion PCR fragment containing $\Delta$ <i>bqsS</i> cloned into the BamHI/EcoRI site of pMQ30; used to make $\Delta$ <i>bqsS</i>	This study
pMQ64	pMQ30, deletion of <i>SacB</i>	(41)
pMQ64- <i>bqsR</i>	<i>bqsR</i> cloned into pMQ64	This study

Gm, Gentamicin

*Construction of bqsR and bqsS unmarked deletions.* The *bqsR* unmarked deletion was created using a pMQ30-based construct, which was made by applying the *Saccharomyces cerevisiae*-based molecular tool kit (41) (Table 2-3). This procedure takes advantage of yeast homologous recombination to combine the vector with PCR-amplified regions upstream and downstream from the gene of interest. The 5' region (~1 kb in length) of the sequence flanking *bqsR* was amplified using primers JW1\_F and JW1\_R, and the 3' region (~1 kb in length) of the sequence flanking *bqsR* was amplified using primers JW2\_F and JW2\_R (Table 2-4). Both primer pairs contained regions that were homologous to each other, and the 5' ends of the JW1\_F and the JW2\_R primers contained 40-bp regions of homology to pMQ30.

TABLE 2-4. Oligonucleotide primers

Gene/PA14 ORF	Primer	Primer Sequence (5'-3')	Application
01240	01240_F	GTGCTCACCGAAGAGAACG	Real-time PCR
	01240_R	AGGAAGCGACCCTGCTT	Real-time PCR
01250	01250_F	GTCTTCACCGACCTGCTGAC	Real-time PCR
	01250_R	GCATGGTCGACATAGCTGAG	Real-time PCR
04180	04180_F	CAAGGGCGACATCTACGAGT	Real-time PCR
	04180_R	CTTGTCGTTGATTTCCATCG	Real-time PCR
04270	04270_F	TCTACGAGATCCACGGTTTCC	Real-time PCR
	04270_R	GTGTCCAGCTCGACTACCAGA	Real-time PCR
07070	07070_F	TATAGCGACGAGCAGGGCAAG	Real-time PCR
	07070_R	TCAGCGCATACGGAAGG	Real-time PCR
<i>bqsP</i>	PA2659_F	GCCGCTACATCTACCAGGTC	Real-time PCR
	PA2659_R	AGTCGTCCTGGTGGTCCTT	Real-time PCR
<i>bqsQ</i>	PA2658_F	ATGTCTACGAGGTGGAAGTCT	Real-time PCR
	PA2658_R	GTCGTCCTCCTCGTCCTTCA	Real-time PCR
<i>bqsR</i>	JW1_F	GGAATTGTGAGCGGATAACAATTTACACA GGAAACAGCTGCCGAGATCTCCAGCAGGT	Construction of <i>bqsR</i> clean deletion
	JW1_R	GTTTCGTCGGCCAGGGGAACGAGGTGATCGA GACCCGCCGT	Construction of <i>bqsR</i> clean deletion
	JW2_F	AGGTGATCGAGACCCGCCGTGTTTCGTCGGC CAGGGGAACG	Construction of <i>bqsR</i> clean deletion
	JW2_R	AGGCAAATTCTGTTTTATCAGACCGCTTCTG CGTTCTGATCCTGGATGCCAGCGCCTTT	Construction of <i>bqsR</i> clean deletion
<i>bqsR</i> _F	CAGTGCCTGTGCCTGAAC	Real-time PCR	
<i>bqsR</i> _R	GCGGGTCTCGATCACCT	Real-time PCR	
NK29	ACCCGTTTTTTTTGGGCTAGCGAATTCGAGCT CGGTACCCGAGGAGATATACATATGCGGTT GCTGCTGGTTG	Construction of <i>bqsR</i> complementati on plasmid	
NK30	AATCTTCTCTCATCCGCCAAAACAGCCAAGC TTGCATGCCTGTGGTCGAACAGCCACAGG	Construction of <i>bqsR</i> complementati	

<i>bqsS</i>	<i>dbqsS</i> 1	GGAATTGTGAGCGGATAACAATTTACACA GGAAACAGCTAAGGACCACCAGGACGACT	on plasmid Construction of <i>bqsS</i> clean deletion
	<i>dbqsS</i> 2	GAAGAACTGTTAAGCCCTGGCCCGCTGGAT CGACCTCAT	Construction of <i>bqsS</i> clean deletion
	<i>dbqsS</i> 3	ATGAGGTCGATCCAGCGGGCCAGGGCTTAA CAGTTCTTC	Construction of <i>bqsS</i> clean deletion
	<i>dbqsS</i> 4	AGGCAAATTCTGTTTTATCAGACCGCTTCTG CGTTCTGATGGTTCGGTGCTGGAAGTG	Construction of <i>bqsS</i> clean deletion
<i>bqsT</i>	<i>bqsS</i> _F	CTGGCGCACATTTTCGATT	Real-time PCR
	<i>bqsS</i> _R	ACTTCGCTATCGGCCCACTT	Real-time PCR
	29750 F	TGGTGTTCTGTTGGGCGAGT	Real-time PCR
	29750 R	TTCCACGGCCTGCTCGTA	Real-time PCR
	<i>feoB</i>	<i>feoB</i> _F	ACCCGATCCAGGTCGAGA
	<i>feoB</i> _R	CGGATCTGCATGAGTGTCG	Confirm tn mutant

*S. cerevisiae* strain InvSc1 (Table 2-3) was cotransformed with plasmid linearized with EcoRI and BamHI as well as with the upstream and downstream PCR fragments; recombinant plasmids were generated by yeast homologous recombination (41). Recombinants were selected on synthetic defined medium lacking uracil (SD-Ura). Plasmids were liberated from uracil-auxotrophic yeast using the QIAprep Spin miniprep kit (Qiagen) yeast protocol and electroporated into *Escherichia coli* DH5 $\alpha$ . Plasmids obtained from DH5 $\alpha$  were electroporated into *E. coli* BW29427. The deletion plasmid was mobilized from BW29427 into PA14 using biparental conjugation (42). PA14 single recombinants (merodiploid containing the intact *bqsR* gene and a deleted gene) were

selected on LB agar containing 100 µg/ml gentamicin. These colonies were then grown in the absence of selection to resolve merodiploids. Potential *bqsR* deletion mutants (resolved merodiploids) were identified by selecting for colonies that grew in the presence of 10% sucrose. Strains with properties of double recombination were further analyzed by PCR and sequenced to verify *bqsR* deletion ( $\Delta bqsR$ ). The above protocol was performed for the *bqsS* deletion ( $\Delta bqsS$ ), substituting primers *dbqsS* 1 and *dbqsS* 2 for primers JW1\_F and JW1\_R and *dbqsS* 3 and *dbqsS* 4 for primers JW2\_F and JW2\_R (see Table 2-4 for primer sequences). Because the reading frames in the *bqs* operon overlap, we were careful to construct our deletions in this region so as to both render them nonpolar and preserve the integrity of the remaining genes in the operon.

*Analysis of global gene expression using P. aeruginosa Affymetrix GeneChips.* The GeneChip *P. aeruginosa* genome array (Affymetrix) contains probe sets for more than 5,500 genes from *P. aeruginosa* PAO1 and 117 additional genes from strains other than PAO1. Due to the strong similarity between strains PAO1 and PA14, the Affymetrix array has been used for gene expression analysis of both strains (43-45). *P. aeruginosa* PA14 was grown in anaerobic MMM as described in “Bacterial strains and growth conditions.” The cultures were grown anaerobically at 37°C until they reached exponential phase (OD<sub>500</sub> of 0.3), at which time they were removed from the incubator and were taken into the anaerobic chamber containing a gas mixture of 5% H<sub>2</sub> and 95% N<sub>2</sub> (Coy Laboratory Products). Water, 100 µM FeCl<sub>3</sub>, or 100 µM FeCl<sub>2</sub> was added to the cultures. Iron concentrations were calculated by weight, assuming 100% purity of Fe(III) and Fe(II)

stocks. The cultures were vortexed every 5 min over a 30-min period, after which 5 ml of culture was removed and mixed with 10 ml of bacterial RNAprotect (Qiagen). The cultures were incubated for 5 min at room temperature before cells were pelleted (10 min,  $5,000 \times g$ ). DNA-free RNA was isolated from the cell pellet using the RNeasy minikit (Qiagen) that includes on-column DNase treatment. The preparation of the cDNA, as well as the processing of the *P. aeruginosa* GeneChip arrays, was performed by the BioMicro Center at the Massachusetts Institute of Technology (MIT) using an Affymetrix fluidics station. GeneChip arrays were performed in biological triplicate. Signal intensities were background corrected, normalized, and summarized using the robust multiarray averaging (RMA) method in the Bioconductor package affy (46-48). Linear models for each gene were fit using the program limma, and moderated *t* statistics were calculated with limma's empirical Bayes approach (49). *P* values were corrected for multiple testing using the Benjamini-Hochberg method to control the false-discovery rate (50); these adjusted values are referred to as *q* values. A threshold *q* value of 0.01 was used to identify differentially expressed genes, resulting in a false-discovery rate of 1% among genes called significant. Verification of GeneChip data was performed with quantitative reverse transcriptase PCR (qRT-PCR).

*Validating iron concentrations.* Fe(II) readily oxidizes in the presence of trace levels of oxygen. Different Fe(II) salts have different stabilities in water. For most experiments, the Fe(II) source was ferrous chloride; it was used because the anion, chloride, was expected to have little effect on the metabolism of PA14. However, this form is less



stable in water than ferrous ammonium sulfate. To be certain of the total iron concentration and to distinguish between Fe(II) and Fe(III), the ferrozine assay (51) was used. Ferrozine binds to Fe(II) and provides a colorimetric readout of Fe(II) concentration compared to that of a standard. Ferrous ammonium sulfate in acid was used as a standard because it is the most stable form of Fe(II) and, under acidic conditions, all the iron is expected to remain in the Fe(II) state. Iron-bound ferrozine was measured at 562 nm. Total iron concentration was determined by reducing the sample with hydroxylamine hydrochloride to convert all Fe(III) to Fe(II).

*Fe(II) shock experiments.* Genes found to be strongly induced in the array experiment were examined using qRT-PCR (primers are listed in Table 2-4). *P. aeruginosa* strains listed in Table 2-3 were grown anaerobically in 10-ml cultures in triplicate as described in “Bacterial strains and growth conditions.” The Fe(II) shock was performed in the anaerobic chamber after the cultures grown in MMM reached early exponential phase. Five-milliliter samples of each culture were removed immediately before the Fe(II) shock and mixed with 2 volumes of bacterial RNAprotect (Qiagen), incubated for 5 min at room temperature, and centrifuged for 10 min at  $5,000 \times g$ . These served as the Fe-free control. A 50  $\mu\text{M}$  concentration of  $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$  was added to the cultures, mixed, and incubated for 30 min. After the 30-min shock, the remaining 5-ml culture was immediately added to 10 ml RNAprotect, incubated for 5 min at room temperature, and centrifuged for 10 min at  $5,000 \times g$ . The pellets were stored at  $-80^\circ\text{C}$  until RNA isolation, and qRT-PCR analysis were performed.

*Anaerobic metal shock experiments.* PA14 was grown in MMM to early exponential phase and treated as described in “Fe(II) shock experiments,” substituting Fe(III) and various dipositive cations for Fe(II). The metals tested were 100  $\mu\text{M}$   $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$  [containing 100  $\mu\text{M}$  Fe(II), as determined by weight and the ferrozine assay], 100  $\mu\text{M}$   $\text{FeCl}_3$ , 5 mM  $\text{CaCl}_2$ , 100  $\mu\text{M}$   $\text{MnCl}_2$ , 10 mM  $\text{MgCl}_2$ , 100  $\mu\text{M}$   $\text{ZnCl}_2$ , and 100  $\mu\text{M}$   $\text{CuCl}_2$ . The concentrations of calcium and magnesium were selected to approximate physiologically relevant concentrations (52).

*Aerobic metal shock experiments.* PA14 was grown in MMM aerobically to early exponential phase and shocked with the metal concentrations given above. The Fe(II) concentration in the  $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$  was determined by the ferrozine assay before and after the aerobic shock to determine if any iron remained in the Fe(II) state after the 30-min exposure to oxygen. The initial Fe(II) concentration was 100  $\mu\text{M}$ , and the concentration after the 30-min shock was 4.9  $\mu\text{M}$ .

*Fe(II) titration.* Triplicate samples of 50 ml MMM in 125-ml acid-washed serum vials were inoculated from cultures grown as described in “Bacterial stains and growth conditions” and sealed. When the cells reached early exponential phase, the Fe(II) shock was performed. Five milliliters of culture was treated with RNAprotect before the iron shock and used as the no-iron control. Subsequently, 5 ml of culture was added to tubes containing iron and incubated for 30 min. The Fe(II) concentrations assayed were 6  $\mu\text{M}$ ,

19  $\mu\text{M}$ , 31.6  $\mu\text{M}$ , 38  $\mu\text{M}$ , 51  $\mu\text{M}$ , and 63  $\mu\text{M}$  (10  $\mu\text{M}$ , 30  $\mu\text{M}$ , 50  $\mu\text{M}$ , 60  $\mu\text{M}$ , 80  $\mu\text{M}$ , and 100  $\mu\text{M}$   $\text{FeCl}_2$  by weight, respectively). After the shock, the cultures were added to 10 ml RNAprotect for RNA stabilization.

*RNA isolation and qRT-PCR analysis.* Total RNA was extracted from the cell pellet using the RNeasy minikit (Qiagen), including the optional DNase treatment step, according to the manufacturer's instructions. cDNA was generated using the extracted RNA as a template for an iScript (Bio-Rad) random-primed reverse transcriptase reaction by following the manufacturer's protocol. The cDNA was used as a template for quantitative PCR (Real Time 7500 PCR machine; Applied Biosystems) using SYBR green with the ROX detection system (Bio-Rad). Samples were assayed in biological triplicate. The threshold cycle ( $C_T$ ) values of *recA* and *clpX* were used as endogenous controls (45). Fold changes were calculated using the  $\Delta\Delta C_T$  method. Briefly, all samples were normalized to each other by subtracting the  $C_T$  value for the control gene *recA* in the shock condition from the value in the control condition ( $\Delta C_T \text{ recA}$ ). This value was then converted from the  $\log_2$  using the following formula, which accounts for machine error:  $a = 2^{\Delta C_T \text{ recA}}$ . The change in the genes of interest was calculated by subtracting the  $C_T$  value of the gene in the shock condition from the  $C_T$  value for the gene in the control condition ( $\Delta C_T \text{ gene}$ ). This value was then converted from the  $\log_2$  using the following formula:  $b = 2^{\Delta C_T \text{ gene}}$ . The final relative fold change is calculated by the following formula: fold change =  $b/a$ . To ensure *recA* was constant under all conditions tested, the relative fold change for the internal control *clpX*, whose expression was also

expected to remain constant across all our treatments, was calculated as described above. Only those samples with a *clpX* fold change between 0.5 and 2 were used. Primers (Integrated DNA Technologies) were designed using Primer3, with settings modified for qRT-PCR primers.

*Growth, Fe(III) reduction, and bqsR expression experiment.* PA14 was grown aerobically on LB agar plates at 37°C. Two single colonies were picked, and each was transferred to a fresh tube containing 3 ml of sterilized MOPS minimal medium (50 mM MOPS, 2.2 mM  $\text{KH}_2\text{PO}_4$ , 43 mM NaCl, 93 mM  $\text{NH}_4\text{Cl}$ , 40 mM succinate, 1 mM  $\text{MgSO}_4$ , and 3.6  $\mu\text{M}$   $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ ; pH 7.1) and grown overnight at 37°C. One milliliter of turbid inoculum was then added to 50 ml of sterilized MOPS medium with 500  $\mu\text{M}$   $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$  and incubated at 37°C in triplicate, and a low-iron control culture was inoculated in MOPS minimal medium without additional Fe(III). At given time points, aliquots were removed from the culture for OD measurements, pyocyanin (a phenazine) concentration measurements, tests to distinguish the iron oxidation state, and qRT-PCR measurements of the *bqs* operon. A sample of 500  $\mu\text{l}$  was taken for OD measurements, 1 ml was frozen for subsequent high-performance liquid chromatography (HPLC) analysis, 1.6 ml was taken to distinguish between iron oxidation states, and 2 to 5 ml was taken for RNA isolation. The  $\text{OD}_{500}$  was measured in triplicate using the Beckman DU 800 spectrophotometer. Pyocyanin HPLC peaks were identified from experimental samples and quantified with known concentrations of a pyocyanin reference on a Beckman System Gold instrument with an acetonitrile gradient, a Waters Symmetry  $\text{C}_{18}$  reverse-phase column (5- $\mu\text{m}$

particle size; 4.6 by 250 mm), and a 168 diode array detector. The ferrozine assay was used to determine the iron concentrations and the oxidation states of the samples. All iron oxidation state distinctions were conducted in an anaerobic chamber to avoid reoxidation of ferrous ions, and measurements were made with a Biotek Synergy 4 plate reader. Relative fold changes for qRT-PCR were calculated by comparing samples from MOPS minimal medium plus Fe to samples with the same OD<sub>500</sub> from MOPS minimal medium without Fe.

*Microarray data accession number.* Our microarray data have been deposited in EBI Array Express (accession number E-MEXP-3459).

### **Acknowledgements**

We thank the Howard Hughes Medical Institute (HHMI) for supporting this work. D.K.N. is an HHMI investigator. N.K. was supported by an NIH training grant (GM07616). J.C.W. and J.J.M. were supported by NSF graduate fellowships.

### **References**

1. **Dong Y-H, Zhang X-F, An S-W, Xu J-L, Zhang L-H.** 2008. A novel two-component system BqsS-BqsR modulates quorum sensing-dependent biofilm decay in *Pseudomonas aeruginosa*. *Commun Integr Biol* **1**:88-96.
2. **Morel F, Hering JG.** 1993. Principles and applications of aquatic chemistry. Wiley, New York.
3. **Huebers HA, Josephson B, Huebers E, Csiba E, Finch CA.** 1984. Occupancy of the iron binding sites of human transferrin. *Proceedings of the National Academy of Sciences* **81**:4326-4330.

4. **Aisen P, Leibman A, Zweier J.** 1978. Stoichiometric and site characteristics of the binding of iron to human transferrin. *The Journal of Biological Chemistry* **253**:1930-1937.
5. **Ratledge C, Dover LG.** 2000. Iron metabolism in pathogenic bacteria. *Annual Review of Microbiology* **54**:881-941.
6. **Stearman R, Yuan DS, Yamaguchilwai Y, Klausner RD, Dancis A.** 1996. A permease-oxidase complex involved in high-affinity iron uptake in yeast. *Science* **271**:1552-1557.
7. **Wandersman C, Delepelaire P.** 2004. Bacterial iron sources: from siderophores to hemophores. *Annual Review of Microbiology* **58**:611-647.
8. **Wösten MMSM, Kox LFF, Chamnongpol S, Soncini FC, Groisman EA.** 2000. A signal transduction system that responds to extracellular iron. *Cell* **103**:113-125.
9. **Baker BJ, Banfield JF.** 2003. Microbial communities in acid mine drainage. *FEMS Microbiology Ecology* **44**:139-152.
10. **Liang L, Korte N, Gu B, Puls R, Reeter C.** 2000. Geochemical and microbial reactions affecting the long-term performance of in situ 'iron barriers'. *Advances in Environmental Research* **4**:273-286.
11. **Naikare H, Palyada K, Panciera R, Marlow D, Stintzi A.** 2006. Major role for FeoB in *Campylobacter jejuni* ferrous iron acquisition, gut colonization, and intracellular survival. *Infect. Immun.* **74**:5433-5444.
12. **Velayudhan J, Hughes NJ, McColm AA, Bagshaw J, Clayton CL, Andrews SC, Kelly DJ.** 2000. Iron acquisition and virulence in *Helicobacter pylori*: a major role for FeoB, a high-affinity ferrous iron transporter. *Molecular Microbiology* **37**:274-286.
13. **Satoh H, Ono H, Rulin B, Kamo J, Okabe S, Fukushi K-I.** 2004. Macroscale and microscale analyses of nitrification and denitrification in biofilms attached on membrane aerated biofilm reactors. *Water Research* **38**:1633-1641.
14. **Li J, Bishop PL.** 2003. Monitoring the influence of toxic compounds on microbial denitrifying biofilm processes. *Water Sci Technol* **47**:211-216.
15. **Hunter RC, Beveridge TJ.** 2005. Application of a pH-sensitive fluoroprobe (C-SNARF-4) for pH microenvironment analysis in *Pseudomonas aeruginosa* biofilms. *Applied and Environmental Microbiology* **71**:2501-2510.
16. **Hugo WB, Turner M.** 1957. A soil bacterium producing an unusual blue pigment. *Journal of Bacteriology* **73**:154-157.
17. **Kimata N, Nishino T, Suzuki S, Kogure K.** 2004. *Pseudomonas aeruginosa* isolated from marine environments in Tokyo Bay. *Microb Ecol* **47**:41-47.
18. **Lam J, Chan R, Lam K, Costerton JW.** 1980. Production of mucoid microcolonies by *Pseudomonas aeruginosa* within infected lungs in cystic fibrosis. *Infect. Immun.* **28**:546-556.
19. **Singh PK, Schaefer AL, Parsek MR, Moninger TO, Welsh MJ, Greenberg EP.** 2000. Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms. *Nature* **407**:762-764.
20. **Costerton JW, Stewart PS, Greenberg EP.** 1999. Bacterial Biofilms: A Common Cause of Persistent Infections. *Science* **284**:1318-1322.

21. **Beinert H, Holm RH, Munck E.** 1997. Iron-Sulfur Clusters: Nature's Modular, Multipurpose Structures. *Science* **277**:653-659.
22. **Singh PK, Parsek MR, Greenberg EP, Welsh MJ.** 2002. A component of innate immunity prevents bacterial biofilm development. *Nature* **417**:552-555.
23. **Patriquin GM, Banin E, Gilmour C, Tuchman R, Greenberg EP, Poole K.** 2008. Influence of quorum sensing and iron on twitching motility and biofilm formation in *Pseudomonas aeruginosa*. *Journal of Bacteriology* **190**:662-671.
24. **Banin E, Vasil ML, Greenberg EP.** 2005. Iron and *Pseudomonas aeruginosa* biofilm formation. *Proceedings of the National Academy of Sciences of the United States of America* **102**:11076-11081.
25. **Litwin CM, Calderwood SB.** 1993. Role of iron in regulation of virulence genes. *Clinical Microbiology Reviews* **6**:137-149.
26. **Marlovits TC, Haase W, Herrmann C, Aller SG, Unger VM.** 2002. The membrane protein FeoB contains an intramolecular G protein essential for Fe(II) uptake in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **99**:16243-16248.
27. **Marshall B, Stintzi A, Gilmour C, Meyer J-M, Poole K.** 2009. Citrate-mediated iron uptake in *Pseudomonas aeruginosa*: involvement of the citrate-inducible FecA receptor and the FeoB ferrous iron transporter. *Microbiology* **155**:305-315.
28. **Ochsner UA, Johnson Z, Vasil ML.** 2000. Genetics and regulation of two distinct haem-uptake systems, *phu* and *has*, in *Pseudomonas aeruginosa*. *Microbiology* **146**:185-198.
29. **Moreau-Marquis S, O'Toole GA, Stanton BA.** 2009. Tobramycin and FDA-approved iron chelators eliminate *P. aeruginosa* biofilms on cystic fibrosis cells. *American Journal of Respiratory Cell Molecular Biology* **41**:2008-0299OC.
30. **Banin E, Lozinski A, Brady KM, Berenshtein E, Butterfield PW, Moshe M, Chevion M, Greenberg EP, Banin E.** 2008. The potential of desferrioxamine-gallium as an anti-*Pseudomonas* therapeutic agent. *Proceedings of the National Academy of Sciences* **105**:16761-16766.
31. **Worlitzsch D, Tarran R, Ulrich M, Schwab U, Cekici A, Meyer KC, Birrer P, Bellon G, Berger J, Weiss T, Botzenhart K, Yankaskas JR, Randell S, Boucher RC, Döring G.** 2002. Effects of reduced mucus oxygen concentration in airway *Pseudomonas* infections of cystic fibrosis patients. *The Journal of Clinical Investigation* **109**:317-325.
32. **Stewart PS, Franklin MJ.** 2008. Physiological heterogeneity in biofilms. *Nature Reviews Microbiology* **6**:199-210.
33. **Reid DW, Anderson GJ, Lamont IL.** 2008. Cystic fibrosis: ironing out the problem of infection? *AJP: Lung Cellular and Molecular Physiology* **295**:L23-L24.
34. **Wilson R, Sykes DA, Watson D, Rutman A, Taylor GW, Cole PJ.** 1988. Measurement of *Pseudomonas aeruginosa* phenazine pigments in sputum and assessment of their contribution to sputum sol toxicity for respiratory epithelium. *Infect. Immun.* **56**:2515-2517.
35. **Cox CD.** 1986. Role of pyocyanin in the acquisition of iron from transferrin. *Infect. Immun.* **52**:263-270.

36. **Wang Y, Newman DK.** 2008. Redox reactions of phenazine antibiotics with ferric (hydr)oxides and molecular oxygen. *Environmental Science & Technology* **42**:2380-2386.
37. **Wang Y, Wilks JC, Danhorn T, Ramos I, Croal L, Newman DK.** 2011. Phenazine-1-carboxylic acid promotes bacterial biofilm development via ferrous iron acquisition. *Journal of Bacteriology* **193**:JB.00396-00311.
38. **Koley D, Ramsey MM, Bard AJ, Whiteley M.** 2011. Discovery of a biofilm electrocline using real-time 3D metabolite analysis. *Proceedings of the National Academy of Sciences*. **108**:19996-20001.
39. **Rahme L, Stevens E, Wolfort S, Shao J, Tompkins R, Ausubel F.** 1995. Common virulence factors for bacterial pathogenicity in plants and animals. *Science* **268**:1899-1902.
40. **Liberati NT, Urbach JM, Miyata S, Lee DG, Drenkard E, Wu G, Villanueva J, Wei T, Ausubel FM.** 2006. An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proceedings of the National Academy of Sciences of the United States of America* **103**:2833-2838.
41. **Shanks RMQ, Caiazza NC, Hinsa SM, Toutain CM, O'Toole GA.** 2006. *Saccharomyces cerevisiae*-based molecular tool kit for manipulation of genes from gram-negative bacteria. *Applied and Environmental Microbiology* **72**:5027-5036.
42. **Whiteley M, Lee KM, Greenberg EP.** 1999. Identification of genes controlled by quorum sensing in *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences* **96**:13904-13909.
43. **Whiteley M, Banger MG, Bumgarner RE, Parsek MR, Teitzel GM, Lory S, Greenberg EP.** 2001. Gene expression in *Pseudomonas aeruginosa* biofilms. *Nature* **413**:860-864.
44. **Mashburn LM, Jett AM, Akins DR, Whiteley M.** 2005. *Staphylococcus aureus* Serves as an Iron Source for *Pseudomonas aeruginosa* during In Vivo Coculture. *Journal of Bacteriology* **187**:554-566.
45. **Dietrich LEP, Price-Whelan A, Petersen A, Whiteley M, Newman DK.** 2006. The phenazine pyocyanin is a terminal signalling factor in the quorum sensing network of *Pseudomonas aeruginosa*. *Molecular Microbiology* **61**:1308-1321.
46. **Bolstad BM, Irizarry RA, Astrand M, Speed TP.** 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**:185-193.
47. **Gentleman R, Carey V, Bates D, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Lacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini A, Sawitzki G, Smith C, Smyth G, Tierney L, Yang J, Zhang J.** 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**:R80.
48. **Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP.** 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**:249-264.



49. **Smyth G.** 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology* **3**.
50. **Benjamini Y, Hochberg Y.** 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**:289-300.
51. **Stookey LL.** 1970. Ferrozine---a new spectrophotometric reagent for iron. *Analytical Chemistry* **42**:779-781.
52. **Levy J, Smith AL, Kenny MA, Ramsey B, Schoenknecht FD.** 1983. Bioactivity of gentamicin in purulent sputum from patients with cystic-fibrosis or bronchiectasis - comparison with activity in serum. *Journal of Infectious Diseases* **148**:1069-1076.
53. **Stock AM, Robinson VL, Goudreau PN.** 2000. Two-component signal transduction Annual Review of Biochemistry **69**:183-215.
54. **Gray RD, Duncan A, Noble D, Imrie M, O'Reilly DSJ, Innes JA, Porteous DJ, Greening AP, Boyd AC.** 2010. Sputum trace metals are biomarkers of inflammatory and suppurative lung disease. *Chest* **137**:635-641.
55. **Boles BR, Thoendel M, Singh PK.** 2005. Rhamnolipids mediate detachment of *Pseudomonas aeruginosa* from biofilms. *Molecular Microbiology* **57**:1210-1223.
56. **Spoering AL, Lewis K.** 2001. Biofilms and planktonic cells of *Pseudomonas aeruginosa* have similar resistance to killing by antimicrobials. *Journal of Bacteriology* **183**:6746-6751.
57. **Waite RD, Papakonstantinou A, Littler E, Curtis MA.** 2005. Transcriptome analysis of *Pseudomonas aeruginosa* growth: comparison of gene expression in planktonic cultures and developing and mature biofilms. *Journal of Bacteriology* **187**:6571-6576.
58. **Musk DJ, Banko DA, Hergenrother PJ.** 2005. Iron salts perturb biofilm formation and disrupt existing biofilms of *Pseudomonas aeruginosa*. *Chemistry & Biology* **12**:789-796.
59. **Soncini FC, Vescovi EG, Solomon F, Groisman EA.** 1996. Molecular basis of the magnesium deprivation response in *Salmonella typhimurium*: Identification of PhoP-regulated genes. *Journal of Bacteriology* **178**:5092-5099.
60. **Perez JC, Groisman EA.** 2007. Acid pH activation of the PmrA/PmrB two-component regulatory system of *Salmonella enterica*. *Molecular Microbiology* **63**:283-293.
61. **Reid DW, Withers NJ, Francis L, Wilson JW, Kotsimbos TC.** 2002. Iron deficiency in cystic fibrosis. *Chest* **121**:48-54.
62. **Yeats C, Rawlings ND, Bateman A.** 2004. The PepSY domain: a regulator of peptidase activity in the microbial environment? *Trends Biochem.Sci.* **29**:169-172.
63. **Hofmann K, Stoffel W.** 1993. TMbase - A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* **374**.

64. **Cserzo M, Wallin E, Simon I, von Heijne G, Elofsson A.** 1997. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Engineering* **10**:673-676.
65. **Kelley LA, Sternberg MJE.** 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protocols* **4**:363-371.
66. **Szurmant H, Mohan MA, Imus PM, Hoch JA.** 2007. YycH and YycI interact to regulate the essential YycFG two-component system in *Bacillus subtilis*. *J. Bacteriol.* **189**:3280-3289.
67. **White S, Tuttle FE, Blankenhorn D, Dosch DC, Slonczewski JL.** 1992. pH-dependence and gene structure of InaA in *Escherichia coli*. *Journal of Bacteriology* **174**:1537-1543.
68. **Yeterian E, Martin LW, Lamont IL, Schalk IJ.** 2010. An efflux pump is required for siderophore recycling by *Pseudomonas aeruginosa*. *Environmental Microbiology Reports* **2**:412-418.
69. **Hassett DJ, Sokol PA, Howell ML, Ma JF, Schweizer HT, Ochsner U, Vasil ML.** 1996. Ferric uptake regulator (Fur) mutants of *Pseudomonas aeruginosa* demonstrate defective siderophore-mediated iron uptake, altered aerobic growth, and decreased superoxide dismutase and catalase activities. *Journal of Bacteriology* **178**:3996-4003.
70. **Ginalski K.** 2004. BOF: a novel family of bacterial OB-fold proteins. *FEBS Letters* **567**:297-301.
71. **Murzin AG.** 1993. OB (oligonucleotide oligosaccharide binding)-fold - common structural and functional solution for nonhomologous sequences. *Embo J.* **12**:861-867.
72. **Pilonieta MC, Erickson KD, Ernst RK, Detweiler CS.** 2009. A protein important for antimicrobial peptide resistance, Ydel/OmdA, is in the periplasm and interacts with OmpD/NmpC. *Journal of Bacteriology* **191**:7243-7252.

**Chapter III: The ferrous iron responsive BqsRS two-component system activates genes that promote cationic stress tolerance**

Naomi N. Kreamer, Flavia Costa, and Dianne K. Newman

This chapter was adapted from the manuscript: The ferrous iron-responsive BqsRS two-component system activates genes that promote cationic stress tolerance. Kreamer NN, Costa F, Newman DK. *Mbio*. 2015 Feb 24;6(2). E02549-14. DOI: 10.1128/mbio.02549-14.

**Abstract**

The physiological resistance of pathogens to antimicrobial treatment is a severe problem in the context of chronic infections. For example, the mucus-filled lungs of cystic fibrosis patients are readily colonized by diverse antibiotic-resistant microorganisms, including *Pseudomonas aeruginosa*. Previously, we showed that bioavailable ferrous iron [Fe(II)] is present in CF sputum at all stages of infection, and constitutes a significant portion of the iron pool at advanced stages of lung function decline (1). *P. aeruginosa*, a dominant CF pathogen, senses Fe(II) using a two-component signal transduction system BqsRS, which is transcriptionally active in CF sputum (1, 2). Here, we show that an RExxE motif in BqsS is required for BqsRS activation. Once Fe(II) is sensed, BqsR binds a tandem repeat DNA sequence, activating transcription. The BqsR regulon—defined through iterative bioinformatic predictions and experimental validation—includes several genes whose products are known to drive antibiotic resistance to aminoglycosides and polymyxins. Among them are genes encoding predicted determinants of polyamine transport and biosynthesis. Compared to the wild type, *bqsS* and *bqsR* deletion mutants are sensitive to high levels of Fe(II), produce less spermidine in high Fe(II), and are more sensitive to tobramycin and polymyxin B but not arsenate, chromate, or cefsolodin. BqsRS thus mediates a physiological response to Fe(II) that guards the cell against positively charged molecules, but not negatively-charged stressors. These results suggest Fe(II) is an important environmental signal that, via BqsRS, bolsters tolerance of a variety of cationic stressors, including clinically important antimicrobial agents.

## Significance Statement

Clearing chronic infections is challenging due to the physiological resistance of opportunistic pathogens to antibiotics. Effective treatments are hindered by a lack of understanding of how these organisms survive *in situ*. Fe(II) is typically present at micromolar levels in soils and sedimentary habitats, as well as in CF sputum. All *P. aeruginosa* strains possess a two-component system, BqsRS, that specifically senses extracellular Fe(II) at low micromolar concentrations. Our work shows that BqsRS protects the cell against cationic perturbations to the cell envelope as well as low pH and Eh, conditions under which Fe<sup>2+</sup> is stable. Fe(II) can thus be understood as a proxy for a broader environmental state; the cellular response to its detection may help rationalize the resistance of *P. aeruginosa* to clinically-important cationic antibiotics. This finding demonstrates the importance of considering environmental chemistry when exploring mechanisms of microbial survival in habitats that include the human body.

## Introduction

Iron is an essential trace element for most life forms, occupying key positions in diverse metalloenzymes spanning those involved in energy-yielding reactions to those critical for cellular defense (3-5). In the context of infections, the iron requirement of pathogens has resulted in an evolutionary arms race between host and microbe to sequester and obtain iron, respectively (6). The outcome of the infection depends on who wins this race: for example, iron can promote the formation of biofilms, complex

microbial structures that are physiologically resistant to common antibiotics (7-9). Iron commonly exists in one of two oxidation states, the oxidized form—ferric iron [Fe(III)]—or the reduced form, ferrous iron [Fe(II)]. Fe(III) predominates in oxic environments, whereas Fe(II) is more stable under hypoxic and/or reducing conditions, with its ionic form ( $\text{Fe}^{2+}$ ) predominating at low pH and reduction potential (Eh) (10). While Fe(III) uptake systems have been the focus of most mechanistically-oriented studies, growing attention has been placed on the importance of Fe(II) sensing and uptake by pathogens. For example, *Pseudomonas aeruginosa*, the opportunistic pathogen renowned for causing chronic pulmonary infections in CF patients, contains an Fe(II)-specific two-component system, BqsRS (2); *Haemophilus influenzae*, another respiratory-tract pathogen, contains a different Fe(II)-responsive two-component system, FirRS (11).

Recently, we reported average per patient micromolar concentrations of Fe(II) in the mucus-filled environment of the lungs of cystic fibrosis patients (1). While the average value of sputum Fe(II) from patients at advanced stages of disease progression was  $\sim 40$   $\mu\text{M}$ , considerable variation existed between samples, with many presenting values greater than 100  $\mu\text{M}$  and some over 200  $\mu\text{M}$ . Fe(II), not Fe(III), correlated with severe lung function decline, with statistical significance comparable to or greater than other inflammatory indices (12, 13). While the presence of high Fe(II) in the lung may seem paradoxical given it is the site for oxygen exchange into the bloodstream (14), this finding is consistent with several studies that have found the CF mucus environment to contain hypoxic, if not anoxic niches (15, 16). It is perhaps not coincidental that BqsRS-containing

members of the Pseudomonadaceae are commonly found in soils and sediments (17), where Fe(II) can reach low micromolar concentrations (18). This environmental reservoir, together with the high Fe(II) concentrations found in sputum, may help contextualize why BqsRS is present in all sequenced strains of *P. aeruginosa* and why *bqsS* transcripts were detected in the majority of CF sputum samples tested (1).

With Fe(II) as a significant component of the CF lung environment and extracellular Fe(II)-sensing predicted to be a conserved trait amongst diverse members of the Pseudomonadaceae, it is important to understand how *P. aeruginosa* responds to Fe(II). Other members of the Pseudomonadaceae (*e.g.*, *Azotobacter vinelandii*) also contain BqsRS homologs, further motivating studies to better characterize this sensory system. *P. aeruginosa* has one of the largest regulatory networks of any bacterial species, which allows it to precisely control its response to a variety of environmental stimuli (19). The two-component system BqsRS was originally discovered as an effector of biofilm dissolution and quorum sensing (20); in a study using zebrafish embryos to screen two-component systems in *P. aeruginosa* for phenotypes, BqsS was shown to significantly impact the establishment of infection in the zebrafish host (21). When the trace metal concentration in the medium was controlled by chelation, we showed that BqsRS responds specifically to extracellular Fe(II) at concentrations exceeding 10 nM (2, 20). However, while our previous study identified the *bqs* operon and a small number of BqsR regulated genes, it did not define the BqsR regulon.

Because *bqsS* transcripts are detectable in CF sputum from patients with chronic *P. aeruginosa* infections, we wondered whether the BqsRS regulon might control factors that affect its adaptation to the Fe(II)-replete environment of CF sputum. We are broadly interested in understanding how *P. aeruginosa* survives under conditions relevant to the microenvironment of chronic infections, including anoxia and antimicrobial stressors. Here, we sought to gain insight into how BqsS senses Fe(II) and how BqsR controls gene transcription. We identified the BqsRS regulon through bioinformatic, genetic, and biochemical approaches and tested predicted phenotypes. This study expands our knowledge of how *P. aeruginosa* responds to Fe(II), and suggests a common mechanism helps protect the cell against Fe(II) and other cationic stressors, including certain clinically-relevant antibiotics.

## Results

### *Fe(II) sensing residues*

Under standard laboratory conditions, the BqsRS system is specifically sensitive to extracellular Fe(II) and not other dipositive cations (2). Multiple bacterial protein prediction programs [TMpred (22) and DAS (23)] predict the sensor kinase BqsS to be a two-pass transmembrane protein with a periplasmic region containing a RExxE motif (Fig. 3-1A). A similar motif is present in a variety of iron-binding proteins; a HExxE motif is found in the Fe(III) sensing *Salmonella* PmrAB two-component system (24) and a RExxE is present in the yeast Ftr iron permease (25). To determine whether the RExxE motif is



involved in Fe(II) sensing, we generated strains with substitutions in these residues in a  $\Delta bqsS$  mutant background. Alleles with substitutions in the RExxE residues were created and expressed from a chromosomal insertion at the *glmS* locus comprised of the *bqs* promoter and the *bqsS* gene (see Methods). Activation of the BqsS variants in the presence of Fe(II) was measured by following the transcriptional response of *bqsP*, the first gene in the *bqs* operon, to 200  $\mu$ M Fe(II) (hereafter referred to as “Fe(II) shock”).

When the glutamates at positions 45 and 48 in BqsS were replaced with alanine (RAXxA) or serine (RSxxS) the transcriptional Fe(II) response was completely abolished. No change was seen in the levels of *bqsP* in these mutants, as was for the  $\Delta bqsS$  negative control (Fig. 3-1B). The arginine residue at position 44 is also critical: when this amino acid was replaced with alanine (AExxE) the Fe(II) shock response decreased by 98% and when replaced with histidine (HExxE), the response was abolished. None of the site-directed mutants responded to Fe(III) shock (Fig. 3-1C). Fe(II) sensing thus appears to be mediated through the periplasmic RExxE motif in BqsS.

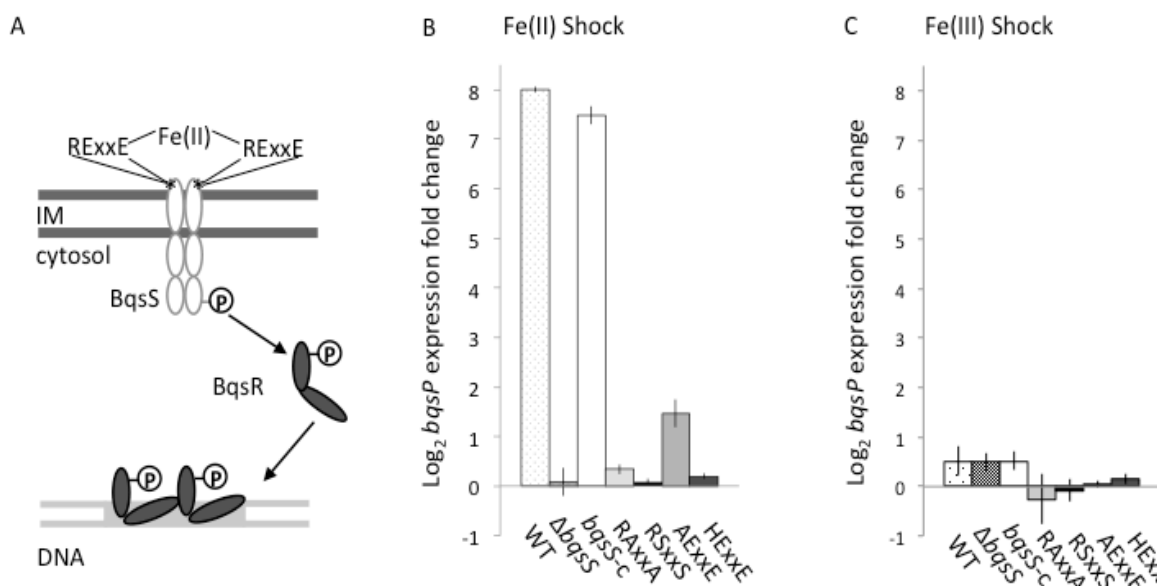


Figure 3-1. BqsS senses Fe(II) through the RExxE motif. (A) Model of BqsS activation by Fe(II) through the periplasmic RExxE motif and signal transduction to BqsR. This model is based on analogy to similar two-component systems (54). (B) Point mutations of the RExxE motif in BqsS show a decreased transcriptional response to Fe(II), suggesting that this may be the binding site of ferrous iron. From left to right the bars show the log<sub>2</sub> fold change in *bqsP* (the first gene in the *bqs* operon) WT,  $\Delta bqsS$ , *bqsS-c* ( $\Delta bqsS::pbqs-bqsS$ ), RAxxA ( $\Delta bqsS::pbqs-RAxxA$ ), RSxxS ( $\Delta bqsS::pbqs-RSxxS$ ), AExxE ( $\Delta bqsS::pbqs-AExxE$ ), and HExxE ( $\Delta bqsS::pbqs-HExxE$ ). (C) Point mutations of the RExxE motif in BqsS show no transcriptional response to Fe(III). From left to right the bars show the log<sub>2</sub> fold change in *bqsP* (the first gene in the *bqs* operon) WT,  $\Delta bqsS$ , *bqsS-C* ( $\Delta bqsS::pbqs-bqsS$ ), RAxxA ( $\Delta bqsS::pbqs-RAxxA$ ), RSxxS ( $\Delta bqsS::pbqs-RSxxS$ ), AExxE ( $\Delta bqsS::pbqs-AExxE$ ), and HExxE ( $\Delta bqsS::pbqs-HExxE$ ).

#### *BqsR* consensus DNA binding site

Having identified some of the key BqsS residues responsible for Fe(II) sensing, we sought to elucidate the BqsR DNA binding sequence. Previously, we performed a microarray experiment to determine whether there were genes uniquely upregulated in response to Fe(II) rather than Fe(III) or other divalent cations (2). Not only did this experiment reveal the existence of the BqsRS system, it also alluded to genes that might

be under its control. By inputting the 500 base pair region upstream from the five most upregulated genes in the microarray (2) into the motif search program MEME (26), we generated putative consensus DNA binding motifs. BqsR belongs to the OmpR-like class of two-component systems, the most common type (27), in which the response regulator contains a phosphoreceiver domain and an effector domain. Members of this class of response regulators, such as OmpR and PhoP, are known to bind tandem repeat DNA sequences (28, 29). Accordingly, we reasoned that the most likely predicted BqsR DNA motif consisted of a 5 bp tandem repeat with 6 intervening nucleotides (Fig. 3-2A); we went on to test this biochemically and through iterative bioinformatic and transcriptional measurements.

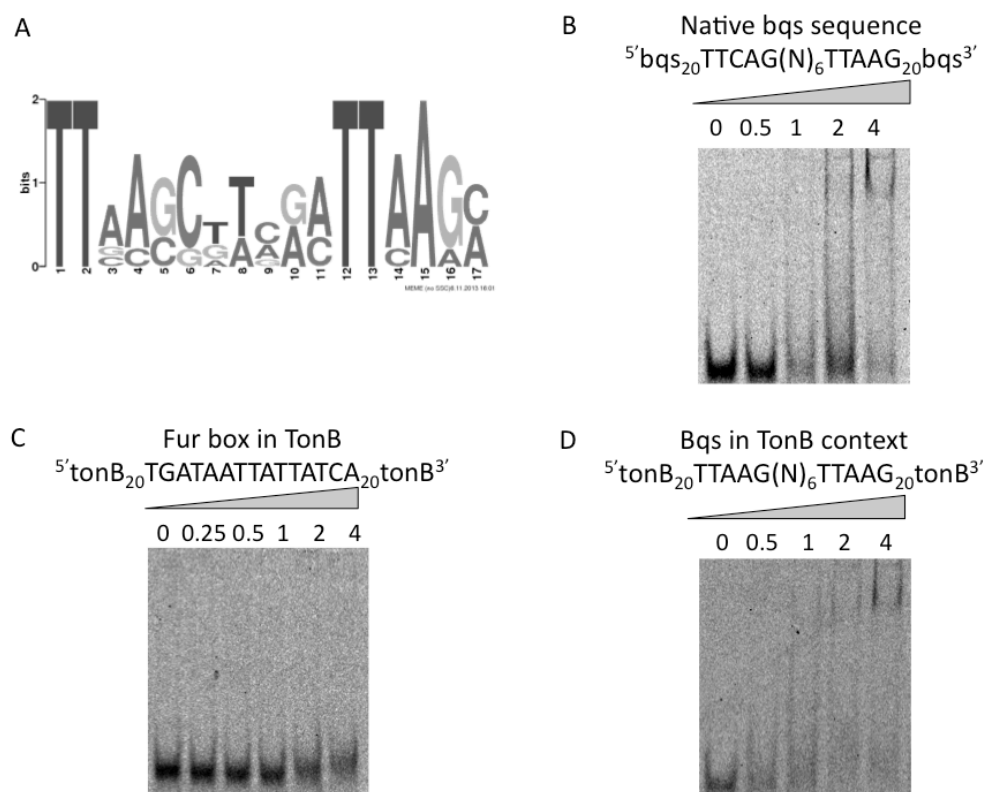


Figure 3-2. A) Predicted sequence used to find TTAAG(N)6TTAAG used in the gel shift assays. B) Native *bqs* sequence shows the exact sequence in front of the *bqs* operon shows a gel shift when incubated with BqsR. C) Fur box in *tonB* is the negative control showing no non-specific DNA binding for BqsR. D) Bqs in the TonB context shows the perfect BqsR tandem repeat exchanging the *bqs* flanking regions with the flanking 20 bp from *tonB* and the gel shift shows that the BqsR sequence is sufficient to induce the gel shift.

Gel shift assays proved BqsR binds specifically to the MEME-predicted tandem repeat DNA sequence. Because the *bqs* operon is autoregulated (2), we incubated purified BqsR with a synthetic 60 bp DNA sequence from the *bqs* promoter region containing the tandem repeat. The native sequence contains one nucleotide substitution in the perfect tandem repeat consensus DNA sequence. Both the perfect tandem repeat (Fig. 3-3A) and the native tandem repeat (Fig. 3-2B) showed decreased mobility with increasing protein concentrations on a native polyacrylamide gel, indicating BqsR-bound DNA. To ensure this was a sequence specific shift, BqsR was incubated with the 56 bp sequence in the *tonB* promoter containing a Fur box (a palindromic sequence with no homology to the predicted BqsR box) with flanking 20 bp DNA regions. No shift was observed for the *tonB* sequence with the same concentrations of BqsR (Fig. 3-2C). To determine whether the BqsR box was sufficient to bind BqsR, we placed the *bqs* tandem repeat into a *tonB* context by exchanging the flanking *bqs* 20 bp regions for those flanking the Fur box from the *tonB* promoter (Fig. 3-2D). This hybrid sequence also showed a similar BqsR concentration dependent decrease in mobility, signifying the *bqs* tandem repeat is sufficient for BqsR binding. As a final control, we randomized the *bqs* tandem repeat sequence while preserving its GC content (Fig. 3-3A); there was no shift with increasing

BqsR concentration, demonstrating that the 5 bp bqs tandem repeat is necessary and sufficient for BqsR binding. While these experiments were performed without chemically phosphorylating BqsR (acetyl phosphate failed to generate stable phosphorylated BqsR), we would expect phosphorylated BqsR would increase its affinity for DNA containing the BqsR box, as has been observed for other response regulators (30).

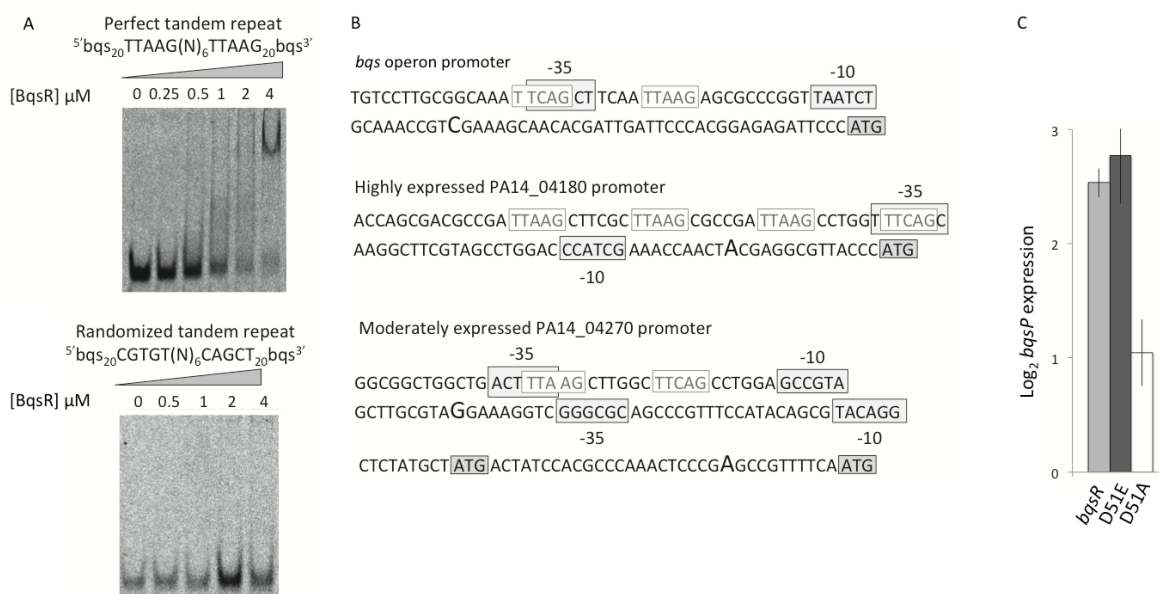


Figure 3-3. BqsR binds to a bioinformatically predicted tandem repeat that overlaps with the RNA polymerase binding site and appears to be activated by phosphorylation. (A) Gel shift assays demonstrate that BqsR binds to the specific tandem repeat TTAAG(N)<sub>6</sub>TTAAG, and within the same flanking DNA sequence does not bind to a randomized tandem repeat with the same binding energy. (B) By determining the transcription start site of three genes, *bqsP*, PA14\_04180, and PA14\_04270, that are upregulated to different degrees by cRACE, it is evident that the BqsR binding site overlaps with the RNA polymerase -35 binding site. The box shaded dark grey highlights the translation start site; the light grey box indicates the -10 and -35 RNA polymerase binding sites; the white box with grey letters shows the BqsR binding sites. (C) In the  $\Delta bqsR$  background, *bqsP*, *bqsP*-D51E (a phosphomimic), or *bqsP*-D51A (cannot be phosphorylated) were inserted into the *gImS* locus under a constitutively active promoter. Cells were grown in 100  $\mu\text{M}$  Fe(II) and 1  $\mu\text{M}$  Fe(II) and transcription of *bqsP* was measured by qRT-PCR. The light grey bar shows native *bqsP* activation by comparing expression between 100  $\mu\text{M}$  Fe(II) and 1  $\mu\text{M}$  Fe(II). The dark grey bar shows expression of *bqsP* in *bqsP*-D51E in 1  $\mu\text{M}$  Fe(II) to

native *bqsR* grown in 1  $\mu\text{M}$  Fe(II). The white bar compares expression of *bqsP* in *bqsR*-D51A in 100  $\mu\text{M}$  Fe(II) to native *bqsR* grown in 1  $\mu\text{M}$  Fe(II). Because the D51E phosphomimic mutant is similar to WT activation and D51A shows no upregulation, the phosphorylated form of BqsR likely binds DNA and activates transcription.

#### *Promoter architecture and BqsR activation mechanism*

To obtain a more finely resolved picture of how BqsR interacts with promoter regions—specifically, to identify where BqsR binds with respect to the RNA polymerase binding site—we performed Rapid Amplification of cDNA Ends (cRACE) on three representative BqsR-regulated genes (PA14\_04180, *bqsS* and PA14\_04270) with a range of induced expression levels from high to low (Fig. 3-3B). Using the transcription start sites identified by cRACE, we could identify the -10 and -35 RNA polymerase binding sites. For both PA14\_04180 and the *bqs* promoter, the -35 RNA polymerase binding site overlaps with one of the BqsR tandem repeat sites. PA14\_04270 contains two transcription start sites, of which the most upstream site's -35 position also overlaps with the BqsR binding site.

Because the BqsR DNA binding site overlaps with the RNA polymerase binding site, we wondered whether BqsR acts as a classical transcriptional activator. Previously, it has been shown that transcription factor binding sites that overlap with the -35 RNA polymerase sigma binding site can activate transcription by helping recruit RNA polymerase to the promoter (31). To test whether BqsR serves as an activator *in vivo*, we replaced the conserved phospho-acceptor aspartate with a glutamate to mimic a constitutively phosphorylated response regulator, or we changed the aspartate to an

alanine to prevent the response regulator from being phosphorylated; this experimental design assumes functional conservation between BqsR and other two-component regulators at this residue (32-34). *bqsR* alleles containing one of these mutations were placed under the control of a constitutive weak promoter at a neutral site (35), in the  $\Delta bqsR$  mutant background. The expression level of a BqsRS-regulated gene (*bqsP*) in the strain carrying the phosphomimic mutant should indicate the function of the activated BqsR, and its expression level in the non-phosphorylatable mutant should indicate the function of the unphosphorylated form. The D51E mutant is constitutively activated, and when grown in 1  $\mu$ M Fe(II) shows similar transcript levels to WT cells grown in 100  $\mu$ M Fe(II); the D51A mutant does not enhance *bqsP* expression under any condition (Fig. 3-3C). This suggests that BqsR is an activator when phosphorylated. Consistent with this interpretation, the  $\Delta bqsR$  strain does not induce the BqsR regulon. Together, these *in vivo* data suggest that BqsRS is a classical two-component system, with the Fe(II) signal presumably causing BqsS to phosphorylate BqsR, and phosphorylated BqsR binding DNA to activate a transcriptional response; this remains to be proven biochemically.

#### *Predicted BqsR regulon*

Having validated the predicted BqsR binding motif, we searched the genome to identify candidate BqsR-regulated genes using two strategies. First, we used a regular expression approach. We selected the most common variants of the *bqs* tandem repeat consensus sequences by eye from our initial position weight matrix (PWM) (Fig. 3-2). A

representative regular expression sequence was  $TTAMG(N)_{4-8}TTAMG$ , which assumed no preference for the nucleotides between the tandem repeat sequences (Fig. 3-4A). We then searched the intergenic regions of the PA14 genome for matches to these regular expressions using the Motif Search tool available on the Pseudomonas Genome Database (36). The final PWM (Fig. 3-4B) was reached through multiple rounds of generating new lists of potential BqsR regulated genes, verifying new members of the regulon by q-reverse transcriptase-PCR (qRT-PCR) and adding the upstream regions to the MEME input to generate a refined PWM. This PWM identified approximately 75 genes, while the regular expression approach identified 62 gene targets. Genes identified through both methods were annotated using information from a variety of databases (Dataset S1) and classified by DAVID (37) (Table 3-1). A subset of these were tested by qRT-PCR if they met one of three criteria: 1) found in both the regular expression approach and the PWM-based approach (Fig. 3-4C); 2) present in a cluster of genes with similar function (Table 3-1); or 3) had an E-value of less than 50 from the PWM approach (Dataset S1). This high E-value was chosen to obtain the broadest possible set of putative BqsR-regulated genes.



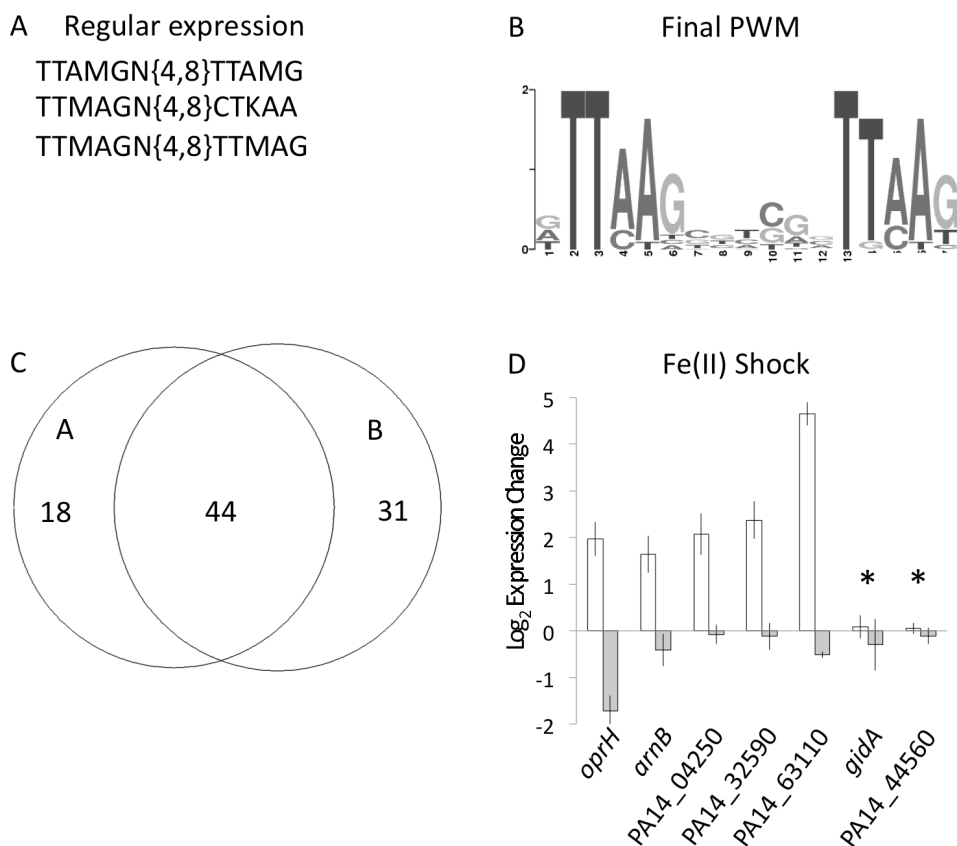


Figure 3-4. Bioinformatic predictions of the BqsR regulon. (A) A motif reflecting the search parameters used to manually search (using regular expressions) the intergenic regions of PA14 genome. (B) Final position-weight matrix based on genes with BqsR-dependent qRT-PCR-verified upregulation in high Fe(II). (C) A Venn diagram depicting in circle A the number of genes predicted by the regular expression motif (4a) and in circle B the number of genes predicted through the qRT-PCR generated PWM (4b) given in Dataset S1. There was substantial overlap between the proposed regulons of the regular expression PWM and the qRT-PCR generated PWM. (D) qRT-PCR validation of 5 genes predicted to be BqsR-regulated: *oprH*, *arnB*, PA14\_04250, PA14\_32590, and PA14\_63110, with 2 negative controls (marked by \*) without the tandem repeat: *gidA* and PA14\_44560. White bars correspond to WT, grey bars to the  $\Delta bqsR$  strain.

Table 3-1. Broad predicted functions of genes identified by either regular expression or the PWM approach based on homology of their gene products to known proteins and categorized by DAVID. Some genes fit into two categories.

Function	# Genes	% Annotation
protease/hydrolase activity	15	8.6

transposition, DNA metabolic process	3	1.7
(poly)amine metabolism	4	2.3
metal binding/hydrolase	14	8.0
protein complex assembly	3	1.7
outer membrane/envelope	6	3.4
metabolism/regulation	50	28.7
RNA metabolic process	3	1.7
signal transducer/two-component system	5	2.9
membrane/transport	18	10.3
redox/cofactor binding	11	6.3
ATPase/nucleotide-binding	6	3.4
transferase activity	7	4.0
intracellular	3	1.7
23S RNA	4	2.3
hypothetical	22	12.6

All of the previously identified Fe(II)-responsive genes (2) showed BqsR-dependent upregulation with Fe(II) shock when tested by qRT-PCR. While some of these predicted genes appeared in our earlier microarray study (2), that study was performed only in the WT in exponential phase and the transcriptional response observed aggregated all direct and indirect effects of Fe(II) shock, not only those specific to BqsR. 5 out of 12 newly predicted genes showed a BqsR-dependent Fe(II) shock response (Fig. 3-4D) in early exponential phase. Several of these genes have known or predicted functions involving cell stress resistance, polyamine biosynthesis/transport, or polymyxin resistance in *P. aeruginosa* strains. For example, *oprH*, a small outer membrane porin, is part of an operon containing the two-component system *phoPQ*, and all three genes are linked to polymyxin resistance (38). *oprH* is thought to stabilize the outer membrane, while *phoPQ* is upregulated by phosphate and Mg<sup>2+</sup> limitation and contributes to resistance to polymyxins among other cationic antimicrobials (38, 39). *arnB* is the first gene of the

operon *arnBCADTEF*, which together synthesize an aminoarabinose derivative covalently attached to lipid A, preventing electrostatic interaction of the polymyxin amine groups with the phosphates in the LPS (40). PA14\_04220-PA14\_04250 forms an operon with similar structure and functional homology to the *potABCD* operon, a known spermidine/putrescine transporter (41). Finally, PA14\_63110 is in an operon with PA14\_63120, PA14\_63130, and *pmrAB*, as confirmed by RT-PCR (Fig. 3-5). PA14\_63110 and PA14\_63120 are predicted spermidine synthesis genes (42) and *pmrAB*, while unresponsive to Fe(III) in *P. aeruginosa* (unlike *Salmonella typhimurium*) (43), has been shown to regulate genes that partially overlap the *phoPQ* regulon (44). PmrAB has also been shown to be involved in polymyxin resistance (44).

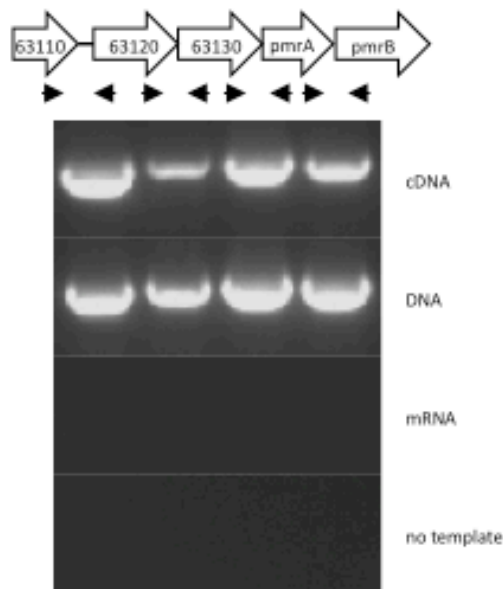


Figure 3-5. RT-PCR showing that PA14\_63110, PA14\_63120, PA14\_63130, *pmrA*, and *pmrB* are in an operon. From top to bottom, the gel shows PCR products spanning the

junctions between genes in cDNA, PCR products with genomic DNA, no bands observed for the mRNA, and no template.

### *RNA-Seq*

Because our qRT-PCR analysis and previous microarray study (2) were limited in scope (*e.g.*, focused on particular genes under a single condition), we performed an RNA-Seq experiment to identify a broader set of BqsR-regulated genes and validate our bioinformatics predictions. For this experiment, we compared WT to  $\Delta bqsR$  grown anaerobically in Fe-limited conditions until deep stationary phase, when quorum sensing is active [validated by measuring rhamnolipid production (45) and Appendix A] and subsequently shocked with 200  $\mu$ M Fe(II). We expected that some genes under this condition might be co-regulated by quorum sensing regulators, which we would have missed in our earlier experiments. We analyzed the RNAseq data in multiple ways: by calculating the reads per kilobase per million reads mapped (RPKM) for each gene and sRNA in the genome (Dataset S2), and by calculating this value for each predicted transcription unit as defined by Wurtzel *et. al* (46) (Dataset S3). Genes or transcriptional units that showed 2-fold higher expression in the WT strain compared to the mutant strain under the iron shock condition, and had an FDR value less than 0.01 (47), were considered to be differentially expressed in a BqsR-dependent manner.

The regular expression approach predicted 10/29 of the transcriptional units upregulated by BqsR in the RNA-Seq data (Fig. 3-4A) and the PWM predicted 11/29 (Fig. 3-4B), together accounting for 38% of the transcriptional units upregulated by BqsR in the

RNA-Seq data. The functional annotation tool of the DAVID suite identified several categories among the genes upregulated in our RNA-Seq data set: metabolism/regulation (30%), membrane (14%), unclassified (11%), protease/protein metabolism (10%), (poly)amine metabolism (8%), hypothetical (8%), metal-binding (6%), localization (5%), outer membrane (4%), and redox (4%). We note that growth conditions impact the expression of the BqsR regulon. For example, qRT-PCR data showed that PA14\_04250 has BqsR-dependent upregulation in early exponential phase but not in deep stationary phase; in contrast, *napE* is expressed in deep stationary but not in early exponential phase. Thus, the list of BqsR-regulated genes presented in Dataset S2 and S3 likely represents a subset of the genes under its control.

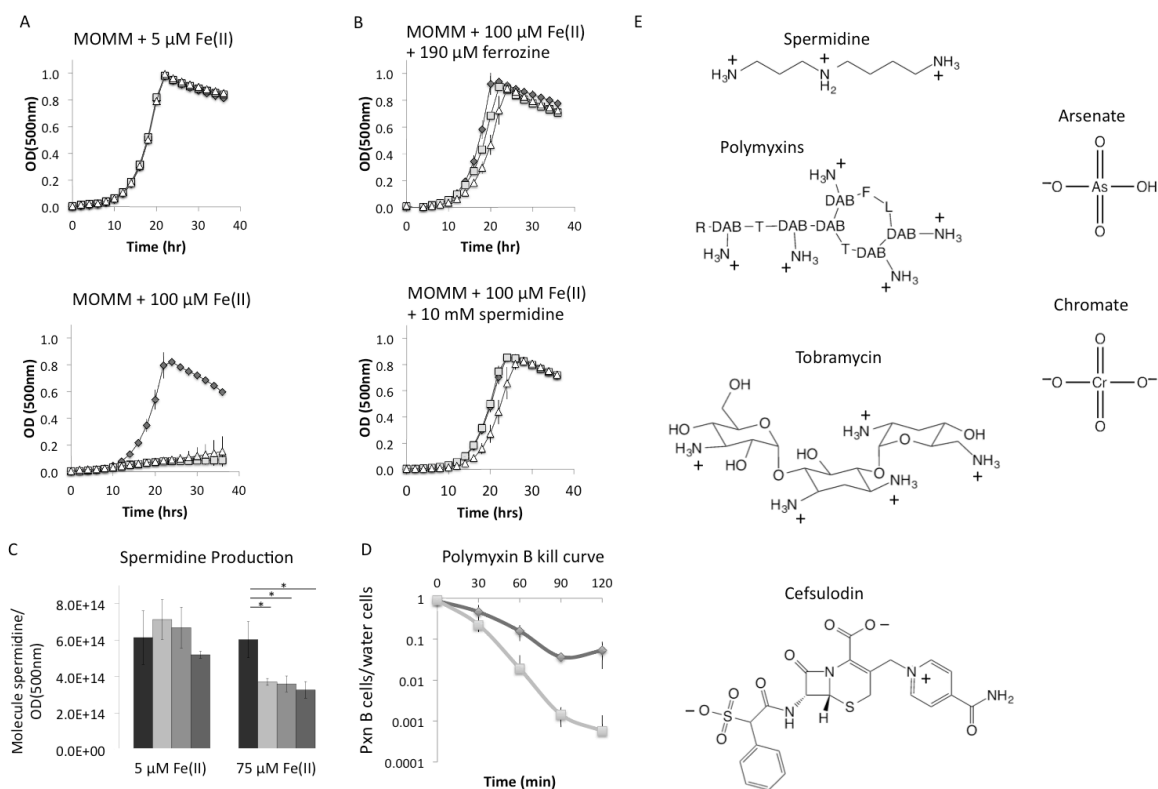


Figure 3-6. Physiological defects in deletion mutants in BqsRS both in growth deficiencies and increased antibiotic susceptibility. For growth in MOMM shown in part A and B, WT are diamonds,  $\Delta bqsR$  are squares, and  $\Delta bqsS$  are triangles. (A) At 5  $\mu\text{M}$  Fe(II) all strains grow like WT. At 100  $\mu\text{M}$  Fe(II), a condition where BqsRS is normally active, the deletion mutants show a deficiency in growth in both lag phase and final density. (B) When 190  $\mu\text{M}$  ferrozine, which chelates 2 ferrozine:1 Fe(II), is added to 100  $\mu\text{M}$  Fe(II), leaving  $\sim 5$   $\mu\text{M}$  free Fe, the growth defect is abolished. Exogenous addition of 10 mM spermidine can also rescue of the growth defect. (C) BqsR upregulates genes in both spermidine biosynthesis and transport. Spermidine production was measured by HPLC and reported in molecules spermidine/OD<sub>500</sub>. The bars show spermidine production from left to right WT,  $\Delta bqsR$ ,  $\Delta bqsS$ , and  $\Delta bqsRS$  in 5  $\mu\text{M}$  and 75  $\mu\text{M}$  Fe(II). In high Fe(II) the deletion mutants produce significantly less spermidine. (D) Polymyxin B kill curves show % survival of cells grown in 75  $\mu\text{M}$  Fe(II) treated with polymyxin B over 2 hours.  $\Delta bqsR$  shown in light gray survived an order of magnitude less than WT is shown in dark gray. (E) The structure of compounds used in this study at pH 7.2. Cationic structures include spermidine, the general polymyxin structure, and tobramycin (an aminoglycoside). Anionic compounds shown are cefulodin (a  $\beta$ -lactam), arsenate, and chromate.

#### *Phenotypic response to Fe(II)*

Because much of the BqsR regulon is consistently upregulated upon Fe(II) exposure, this allowed us to identify phenotypes potentially under BqsRS control. Our first prediction was that BqsRS promotes cell tolerance of high Fe(II) levels. To test this, we compared the growth of WT,  $\Delta bqsR$ , and  $\Delta bqsS$  strains in minimal medium under conditions in which BqsRS is inactive [5  $\mu\text{M}$  Fe(II)] and conditions in which BqsRS is active [100  $\mu\text{M}$  Fe(II)] (2). In the presence of low iron [5  $\mu\text{M}$  Fe(II)], the strains grew similarly, whereas at high iron [100  $\mu\text{M}$  Fe(II)], the  $\Delta bqsR$ , and  $\Delta bqsS$  mutants showed significant growth defects compared to WT (Fig. 3-6A). Complementation restored WT growth (Fig. 3-7A). The growth defect was also relieved by the addition of ferrozine, a Fe(II) chelator (Fig. 3-6B).

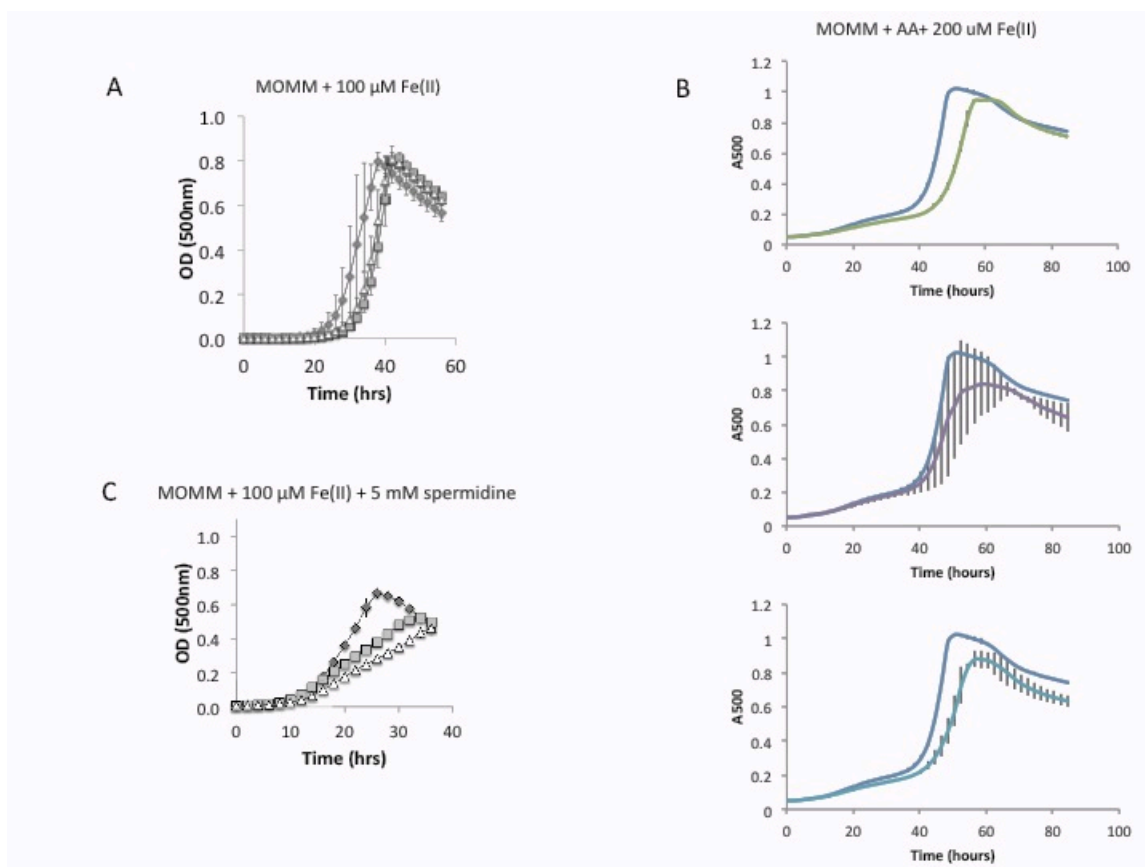


Figure 3-7. A) Chromosomal complements for deletion strains in 100  $\mu\text{M}$  Fe(II) shows  $\Delta bqsR::bqsR$  (squares) and  $\Delta bqsS::bqsS$  (triangles) have the same growth phenotype as WT. B) Growth curves performed in transposon mutants of BqsR-responsive genes and grown anaerobically in MOMM + AA + 200  $\mu\text{M}$  Fe(II). WT is plotted in dark blue in all graphs and all transposon mutants show a moderate growth defect. The top graph shows WT and PA14\_63110 transposon mutant, the middle graph plots WT and the PA14\_63120 mutant, and the bottom graph shows WT and the *pmrA* mutant. C) A growth curve showing 5 mM spermidine partially rescues  $\Delta bqsR$  (squares) and  $\Delta bqsS$  (triangles) growth deficiency in 100  $\mu\text{M}$  Fe(II) when compared to WT (diamonds).

A striking aspect of the BqsR regulon is that many of its genes are predicted to play a role in polyamine biosynthesis and transport. For instance, decarboxylation of S-adenosylmethionine is predicted to be catalyzed by PA14\_63110, and, subsequently, an amine group from S-adenosyl-L-methioninamine is transferred to putrescine by

PA14\_63120 to form spermidine. Transposon mutants of members of this operon, PA14\_63110, PA14\_63120, and *pmrA*, also show a moderate growth defect in high Fe(II) (see Fig. 3-7B). As mentioned previously, PA14\_04220-PA14\_04250 are predicted to encode a spermidine/putrescine-specific transporter (48) and the operon PA14\_04190 – PA14\_04210 (induced in the RNA-Seq experiment) is also predicted by COG (49) to encode a polyamine transporter. Given that PA14\_63110-63120 homologues in *P. aeruginosa* PAO1 have been shown to promote outer membrane stability (50), we hypothesized that spermidine might protect the cell from anaerobic Fe(II) toxicity by electrostatically repelling the ferrous ion from the cell surface and intracellular components. Consistent with this hypothesis, the mutants' growth defect was partially rescued by exogenous addition of 5 mM spermidine (Fig. 3-7C) and fully rescued by 10 mM spermidine (Fig. 3-6B). Whether other polyamines can provide similar protection remains to be determined.

Because exogenous spermidine protected the mutants, we reasoned that the *bqsR/S* deletion mutants were deficient in spermidine production in the presence of high Fe(II). HPLC analysis of polyamine extracts taken from cell lysate showed that the mutants produced approximately 50-66% as much spermidine as the WT when grown in the presence of 75  $\mu$ M Fe(II) (Fig. 3-6C), whereas no significant difference in spermidine production was observed between the WT and the  $\Delta bqsR$ ,  $\Delta bqsS$ , and  $\Delta bqsRS$  strains in 5  $\mu$ M Fe(II). Interestingly, for the WT, the total amount of spermidine in the culture did not increase from the 5  $\mu$ M to the 75  $\mu$ M Fe(II) growth condition.



Previous studies have indicated that spermidine can protect *P. aeruginosa* from several classes of cationic antimicrobials (42, 50). Additionally, BqsR-upregulated genes such as *pmrAB*, *arnB*, *oprH* and *phoPQ* have been demonstrated to be involved in polymyxin resistance (19, 44). Accordingly, we tested susceptibility to polymyxin B by comparing killing curves of WT and  $\Delta bqsR$  grown in 75  $\mu$ M Fe(II) to late exponential phase (Fig. 3-6D). Typically polymyxin survival assays are performed in aerobic conditions in rich medium, whereas our assays were performed anaerobically in minimal medium where the doubling time is 4-10 times longer. This accounts for the increased time scale in the killing curve. After an hour of incubation at 37 °C with 25  $\mu$ g/ml polymyxin B, differences in WT and  $\Delta bqsR$  survival were observed. After two hours, 3.7% of WT survived, whereas only 0.14% of  $\Delta bqsR$  survived. When complemented, the  $\Delta bqsR$  strain behaved like the WT. Additionally, we observed a distinct morphological difference between WT and  $\Delta bqsR$  colonies. The  $\Delta bqsR$  colonies treated with polymyxin B varied greatly in size and several satellite colonies could be seen (Fig. 3-8A). We therefore counted the total number of colonies, but kept track of the percentage of large and small. After 24 hours of incubation, half of the  $\Delta bqsR$  colonies visible after 48 hours were unobservable; counting solely colonies larger than a pinprick at this time point, we observed 0.07% survival for  $\Delta bqsR$  compared to 3.0% survival in WT (Fig. 3-8B). Figure 4D shows survival calculated from final cell counts after 48 hours of incubation.  $\Delta bqsR$  water-treated control and WT polymyxin B-treated cells showed consistent colony morphology after 24 hours of incubation.

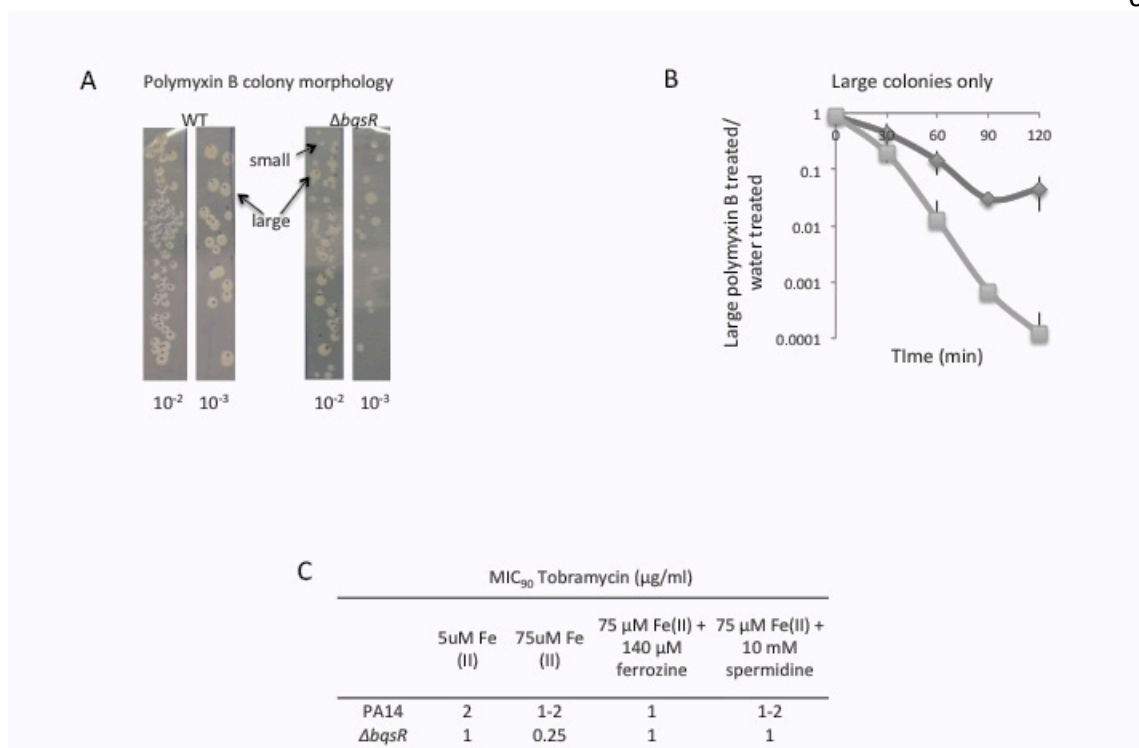


Figure 3-8. A) After 48 hours of growth the colonies were visible as shown; marked colonies on these images correspond to colonies visible after 24 hours. The marked colonies after 48 hours are the "large" colonies, and the unmarked colonies are the "small" colonies. For the *ΔbqsR* 10<sup>-3</sup> dilution, none of the colonies were present at 24 hour, so all the colonies seen here were counted as small colonies. B) A 25 μg/ml polymyxin B anaerobic kill curve in 75 μM Fe(II) that excludes the small colony variants observed in *ΔbqsR* (light gray) percent survival compared to the percent survival of WT (dark gray). C) When ferrozine or spermidine was added, the high Fe(II) tobramycin MIC was restored to WT levels in the mutant.

To test the hypothesis that BqsRS provides specific protection against diverse cationic stressors when activated by Fe(II), we performed MIC<sub>90</sub> assays of both cationic and anionic stressors (Table 3-2). Due to differences in growth rate and lag phase between WT and *ΔbqsR* in 75 μM Fe(II), we analyzed the MIC<sub>90</sub> experiments at the time point where each strain achieved its maximum optical density in MOMEM without the stressor. Under these conditions, there was no MIC difference between WT and *ΔbqsR* at

5  $\mu\text{M}$  Fe(II) for any of the stressors, as expected, because this concentration is below the activation threshold for BqsRS. Only in the case of Tobramycin, a cationic aminoglycoside (Fig. 4E), a significant difference could be seen at 75  $\mu\text{M}$  Fe(II) (Table 3-2). When treated with tobramycin, 75  $\mu\text{M}$  Fe(II), and 140  $\mu\text{M}$  ferrozine (two ferrozines chelate one iron, leaving  $\sim 5$   $\mu\text{M}$  free Fe), both  $\Delta bqsR$  and WT had the same MIC (Fig. 3-8C). Negatively charged stressors tested included cefsulodin, 1 negative charge (51), potassium chromate (2 negative charges), and sodium arsenate (1 or 2 negative charges; pKa 7.03). For both chromate and arsenate the MIC was the same for both strains regardless of the [Fe(II)]. For cefsulodin, the MIC differed depending on the [Fe(II)]—a phenomenon that has been observed elsewhere (52)—however, consistent with our hypothesis, BqsR did not affect the sensitivity level.

Table 3-2. BqsRS guards against cationic but not anionic stressors. MIC assays were performed for the WT and  $\Delta bqsR$  under anaerobic conditions at two different iron concentrations for tobramycin, cefsulodin, arsenate, and chromate. Numbers in bold highlight where the absence of BqsR makes a difference.

	MIC <sub>90</sub> 5 $\mu\text{M}$ Fe(II)	MIC <sub>90</sub> 75 $\mu\text{M}$ Fe(II)	Stressor	Stressor's charge
PA14	2 ( $\mu\text{g}/\text{ml}$ )	<b>1-2 (<math>\mu\text{g}/\text{ml}</math>)</b>	Tobramycin	+
$\Delta bqsR$	1 ( $\mu\text{g}/\text{ml}$ )	<b>0.25 (<math>\mu\text{g}/\text{ml}</math>)</b>	Tobramycin	+
PA14	256-512 ( $\mu\text{g}/\text{ml}$ )	8-16 ( $\mu\text{g}/\text{ml}$ )	Cefsulodin	-
$\Delta bqsR$	256 ( $\mu\text{g}/\text{ml}$ )	8 ( $\mu\text{g}/\text{ml}$ )	Cefsulodin	-
PA14	2-4 mM	2-4 mM	$\text{HArO}_4^-$	-
$\Delta bqsR$	2-4 mM	2-4 mM	$\text{HArO}_4^-$	-
PA14	250 $\mu\text{M}$	250 - 500 $\mu\text{M}$	$\text{CrO}_4^-$	-
$\Delta bqsR$	250 $\mu\text{M}$	250 - 500 $\mu\text{M}$	$\text{CrO}_4^-$	-

## Discussion

Characterizing the environment that pathogens encounter within the human host can help us understand their behavior. This principle is illustrated by connecting the output of the BqsRS regulatory system to the chemical context of CF sputum. Our research suggests that Fe(II) is an important environmental variable that *P. aeruginosa* senses through BqsRS, leading to a response that helps it cope not only with elevated Fe(II) concentrations but multiple cationic stressors within the CF lung environment.

The ability of BqsRS to specifically respond to Fe(II) is an interesting example of metal selectivity (2). Our results indicate that BqsS recognizes Fe(II) via the RExxE motif in its periplasmic domain, based on analogy to other Fe-sensing proteins. Although it is unusual for glutamates to prefer Fe(II) to Fe(III), the strongly positive arginine may tune the ligand environment to prefer the less positively charged Fe(II) over Fe(III). While PmrB from *Salmonella typhimurium* senses Fe(III) rather than Fe(II) through a similar motif [HExxE], *Pseudomonas* PmrB does not bind Fe(III) (43). The *Pseudomonas* PmrB has a 32% amino acid similarity to *Salmonella*, yet the Fe(III)-binding domain in *Salmonella* strains is missing in *P. aeruginosa*'s PmrB. The *Salmonella* PmrB requires a distal serine in addition to the HExxE motif for Fe(III) binding (24). This suggests that other, less proximal residues may contribute to Fe(II) recognition by BqsS, particularly as arginine replacement by histidine or alanine is insufficient to convert BqsS to an Fe(III) sensor. HbpS from *Streptomyces reticuli* contains 8 ExxE Fe(II) binding sites, each with a required adjacent Arg or Lys (single point mutations of the R/K resulted in reduced activity even with 7

other intact binding sites) (53). This is consistent with the hypothesis that RExxE may directly bind Fe(II) and the R is a critical part of the Fe(II) binding event. The RExxE motif is just one strategy by which cells sense Fe(II): the FirRS system from *H. influenzae* utilizes a different motif, DYRED (11). Upon activation, two-component sensors can act as a kinase or a phosphatase to their cognate response regulators (54). BqsR prefers DNA binding sites that overlap RNA polymerase binding sites, suggesting that BqsR may help recruit RNA polymerase to promoter regions. As one of the few two-component systems for which the effector is known, BqsRS presents an attractive system for future biochemical studies to validate these predictions and explore the mechanisms of signal transduction.

Our primary goal in this study was to predict the biological response to extracellular Fe(II) mediated by BqsRS. Using iterative bioinformatic and experimental approaches to define BqsR regulon, we identified a set of approximately 100 genes potentially under BqsR control. Approximately a third of these genes encode proteins involved in stress tolerance, such as cation-binding and transport determinants and lipopolysaccharide (LPS)-modulating proteins. Notably, several experimentally validated genes under BqsR control were annotated as playing a role in polyamine synthesis and transport. Polyamines, including spermidine, are present in all cells and environments (55), and are known to stabilize DNA and RNA, enhance translation and transcription, control porin-mediated transport, and increase the stability of LPS (41). A variety of metals are known to bind LPS (56) and divalent ions modulate polymyxin sensitivity

through interactions with LPS (57). Growing evidence has revealed that Fe(II) can be toxic under anaerobic conditions (58-60).

Because Fe(II) is thought to enter the cell through outer membrane porins (61, 62), that *bqs* mutants are particularly sensitive to high levels of Fe(II) yet can be rescued by exogenous spermidine is not surprising. The *bqs* mutants synthesize less spermidine than the WT, and while over 5 mM exogenous spermidine [a concentration higher than the reported intracellular spermidine concentration of 1-3 mM in bacteria (41)] is necessary to fully rescue these mutants, this can be rationalized by the fact that exogenous supplementation is different than endogenous production and that spermidine production is only one of several responses mediated by BqsRS that potentially helps confer resistance to Fe(II). *P. aeruginosa* possesses two different spermidine synthesis operons: *speDE* (bacterial in origin) and PA14\_63110/PA14\_63120 (archaeal in origin). The bacterial S-adenosylmethionine decarboxylase, *speD*, requires a dipositive cation (most commonly Mg<sup>2+</sup>) for activity, whereas the archaeal homolog to PA14\_63110 does not require a cation for activity (63). Because other metals may substitute for Mg<sup>2+</sup> with reduced enzymatic activity (63, 64), we speculate that high Fe(II) perturbs intracellular metal homeostasis enough to reduce SpeD activity, accounting for the lower spermidine concentrations in the  $\Delta bqsR/S$  mutants. At high Fe(II), BqsR induction of PA14\_63110/PA14\_63120 may compensate for SpeD's reduced activity, increasing the concentration of spermidine to homeostatic levels.

Given the plausible electrostatic mechanism underpinning polyamine-mediated resistance to Fe(II), we reasoned that BqsRS might also influence resistance to diverse cationic antibiotics but not anionic stressors. Previous studies have established that polyamines can mediate antibiotic resistance to cationic peptides, aminoglycosides, quinolones, and oxidative stress (42, 50, 65, 66). We show that *bqs* mutants exhibit increased sensitivity to tobramycin compared to the WT when stressed by high Fe(II). In addition to the genes involved in spermidine production and transport, BqsRS upregulates other genes known to be involved in polymyxin resistance, such as *arnB* and *oprH*. Moreover, BqsRS activates *pmrAB* and *phoP*, which mediate polymyxin resistance in response to other environmental signals (limiting Mg/Ca, or phosphate starvation, respectively (50, 67)). The polymyxin B killing curve demonstrates that  $\Delta bqsR$  is more sensitive than WT in high Fe(II). Thus, BqsRS may be a master regulator under anoxic, Fe(II)-replete conditions, exerting indirect effects via modulating the expression of other regulatory systems. That BqsRS mediates a specific response to cationic stressors in the presence of high Fe(II) is supported by its lack of an effect on resistance to the negatively charged stressors cefsulodin, arsenate, and chromate.

Beyond mediating resistance to cationic antibiotics, BqsRS elicits a cellular response that is broadly relevant to survival in an environment where Fe(II) is a dominant parameter. Fe(II) is stable in a well-defined subset of environmental conditions spanning a range of acidic and reducing conditions (68). It is well known that such microhabitats are present in soil environments, which are also replete with diverse natural products that

have antibiotic activities (69). While the evolutionary history of the BqsRS system is unknown, we note that pseudomonads (including *P. aeruginosa* isolates) are commonly found in soil environments (17). Sensing Fe(II) as proxy for these conditions may allow the cell to modulate its behavior accordingly. Consistent with this hypothesis, the BqsR regulon includes genes that protect the cell against pH and redox stress. For example, BqsRS upregulates *dsbB*, which helps control disulfide bond formation in the periplasm (70) and carbonic anhydrase, which promotes acid tolerance (71). Intriguingly, the most upregulated gene in the BqsR regulon, PA14\_04180 (upregulated ~4000x), is a predicted periplasmic bacterial OB-fold protein that likely binds positively charged molecules. This suggests that the cell may perceive a significant proportion of the Fe(II) pool in cationic form, consistent with the dominance of this species at low pH and Eh.

In summary, the connection between Fe(II) sensing and a response that broadly protects the cell against diverse cationic molecules—including clinically relevant antibiotics—reminds us of the importance of considering environmental chemistry when exploring mechanisms of microbial survival in habitats that include the human body. Typically, CF patients are first treated with aminoglycosides, beta-lactams, or fluoroquinolones. As lung function declines, polymyxins are employed as a last line of defense due to their harmful side-effects (72). It is worth exploring whether a combination of aminoglycosides and polymyxins in conjunction with Fe(II) chelators or novel molecules targeting BqsRS could potentially be administered at lower concentrations, reducing side-effects and increasing drug efficacy.



## Materials and Methods

*Growth media.* *E. coli* was grown aerobically in LB/ampicillin  $100 \mu\text{g ml}^{-1}$  at  $37^\circ\text{C}$ . *P.*

*aeruginosa* PA14 was grown both aerobically and anaerobically at  $37^\circ\text{C}$  in MOPS minimal medium (MOMM) in acid washed glassware to ensure cells were Fe limited. The basic MOMM is composed of 40 mM succinate ( $\text{C}_4\text{H}_4\text{Na}_2\text{O}_4 \cdot 6\text{H}_2\text{O}$ ), 9.3 mM  $\text{NH}_4\text{Cl}$ , 2.2 mM  $\text{KH}_2\text{PO}_4$ , 25 mM  $\text{KNO}_3$ , 25 mM  $\text{NaNO}_3$ , 25 mM MOPS, and 25 mM NaMOPS pH 7.2.

Additionally, immediately prior to inoculation  $100 \mu\text{M}$   $\text{CaCl}_2$ ,  $1 \mu\text{M}$   $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ ,  $1 \text{ mM}$   $\text{MgSO}_4$ , and trace metals were added (73). All PA14 cultures were prepared by inoculation of MOMM media with the desired strains overnight shaking aerobically at  $37^\circ\text{C}$ . Cultures were grown aerobically into exponential phase ( $\text{OD}_{500} \approx 0.4-0.6$ ) in fresh MOMM then grown anaerobically with 50 mM nitrate as the electron acceptor in a Coy chamber with an atmosphere of 80%  $\text{N}_2$ , 15%  $\text{CO}_2$ , and 5%  $\text{H}_2$ . For anaerobic growth curves, biological triplicates of WT and deletion mutants were incubated in MOMM with 5  $\mu\text{M}$  or 100  $\mu\text{M}$  Fe(II). Anaerobic cultures were also grown with MOMM + 100  $\mu\text{M}$  Fe(II) supplemented with 190  $\mu\text{M}$  ferrozine, 5 mM spermidine, or 10 mM spermidine.

Table 3-3. Strain table

Strain	genotype	plasmid
<i>E. coli</i> BL21	wt	pET15b- <i>bqsR</i>
<i>P. aeruginosa</i> PA14	wt	
<i>P. aeruginosa</i> PA14	wt	pMQ72
<i>P. aeruginosa</i> PA14	$\Delta bqsR$	pMQ72

<i>P. aeruginosa</i> PA14	wt
<i>P. aeruginosa</i> PA14	$\Delta bqsR$
<i>P. aeruginosa</i> PA14	$\Delta bqsS$
<i>P. aeruginosa</i> PA14	$\Delta bqsRS$
<i>P. aeruginosa</i> PA14	$\Delta bqsR\Omega glmS-pbqs-bqsR$
<i>P. aeruginosa</i> PA14	$\Delta bqsS\Omega glmS-pbqs-bqsS$
<i>P. aeruginosa</i> PA14	$\Delta bqsRS\Omega glmS-pbqs-bqsRS$
<i>P. aeruginosa</i> PA14	$\Delta bqsR\Omega glmS-ptrc-bqsR$
<i>P. aeruginosa</i> PA14	$\Delta bqsR\Omega glmS-ptrc-bqsR-D51E$
<i>P. aeruginosa</i> PA14	$\Delta bqsR\Omega glmS-ptrc-bqsR-D51A$
<i>P. aeruginosa</i> PA14	$\Delta bqsS\Omega glmS-pbqs-bqsS-$ $E45A/E48A$
<i>P. aeruginosa</i> PA14	$\Delta bqsS\Omega glmS-pbqs-bqsS-E45S/E48S$
<i>P. aeruginosa</i> PA14	$\Delta bqsS\Omega glmS-pbqs-bqsS-R44A$
<i>P. aeruginosa</i> PA14	$\Delta bqsS\Omega glmS-pbqs-bqsS-R44H$

*Ferrozine assay.* The Stookey method (74) was modified for 96-well plate format to detect [Fe(II)]. All measured Fe concentrations were within 5% of reported value.

*Cloning methods.* The genes of interest were cloned into vectors using the Gibson assembly technique. Briefly, genes were amplified with PCR primers with 20 bp overlap with the vector insertion site sequence with KAPA-HiFi HotStart (Kapa-biosystems) and stitched together with the Gibson assembly kit (New England Biolabs). The resulting plasmids were transformed into *E. coli* TOP10. For *E. coli* BqsR expression, the pET15b-BqsR plasmid was transferred to *E. coli* BL21 expression strain. The pMQ72 was electroporated into electrocompetent *P. aeruginosa* PA14 (1) and into *P. aeruginosa* PA14  $\Delta bqsR$ . The point mutants were generated through site-directed mutagenesis. For the point mutants expressed from the *glmS* locus in the *P. aeruginosa*  $\Delta bqsR$  strain,

tetraparental conjugation was used to integrate the pUC18T-mini-Tn7T-BqsR plasmid and the point mutation plasmids (separately) into *P. aeruginosa* PA14  $\Delta bqsR$ .

Complementation of the mutants  $\Delta bqsR$ ,  $\Delta bqsS$ , and  $\Delta bqsRS$  were performed using the pUC18T-mini-Tn7 plasmid that included the 1 Kb upstream sequence of the bqs operon, and bqsR, bqsS, or bqsRS. These constructs were electroporated into *E. coli* TOP10 and tetraparental conjugation used to conjugate *P. aeruginosa*  $\Delta bqsR$ ,  $\Delta bqsS$ , and  $\Delta bqsRS$  mutants.

Table 3-4. Primer table

Primer Name	Sequence	Function
cloning		
$\Delta bqsRS$ P1	GGAATTGTGAGCGGATAACAATTTACACAGGAAAC AGCTTCGCCAGGTCCAGTCCCTGG CGGCAAATGGTGAAGAACTGTCAGTCGTCTTCCTCG	$\Delta bqsRS$ construction
$\Delta bqsRS$ P2	TCCT AGGACGAGGAAGACGACTGACAGTTCTTACCATT	$\Delta bqsRS$ construction
$\Delta bqsRS$ P3	GCCG AGGCAAATTCTGTTTTATCAGACCGCTTCTGCGTTCT	$\Delta bqsRS$ construction
$\Delta bqsRS$ P4	GATCCCAGAGCATCGACAGCCTG	$\Delta bqsRS$ construction
$\Delta bqsRS\Omega glmS$ -pbqs- <i>bqsRS</i> g1	catgagctcactagtgatccGGCAGGTCGAGATGGTAG	$\Delta bqsRS$ complementation
$\Delta bqsRS\Omega glmS$ -pbqs- <i>bqsRS</i> g2	aaccgcatGGGAATCTCTCCGTGGGA	$\Delta bqsRS$ complementation
$\Delta bqsRS\Omega glmS$ -pbqs- <i>bqsRS</i> g3	agattcccATGCGGTTGCTGCTGGTT	$\Delta bqsRS$ complementation
$\Delta bqsRS\Omega glmS$ -pbqs- <i>bqsRS</i> g4	gaggtaccgggccaagcttAACTGTTAAGCCCTGGCG	$\Delta bqsRS$ complementation $\Delta bqsS$ complementation (PCR from plasmid complement)
$\Delta bqsS$ comp P1	aaaaagagctcGGCAGGTCGAGATGGTAGTC	$\Delta bqsS$ complementation
$\Delta bqsS$ comp P2	aaaaActagtAACTGTTAAGCCCTGGCGGC	site directed mutagenesis
<i>bqsS</i> -R <sub>AxxA</sub> P1	AACCTGCGCGCGGAGGCGCGAACCTGCTG	R <sub>AxxA</sub>
<i>bqsS</i> -R <sub>AxxA</sub> P2	CAGCAGGTTCCGCCCTCCGCGCGCAGGTT	site directed mutagenesis
<i>bqsS</i> -A <sub>ExxE</sub> P1	CGCCGGCAACCTGGCCGAGGAGGCGGAG	R <sub>AxxA</sub>
<i>bqsS</i> -A <sub>ExxE</sub> P2	CTCCGCCTCTCGGCCAGGTTGCCGGCG	site directed mutagenesis
<i>bqsS</i> -H <sub>ExxE</sub> P1	CGCCGGCAACCTGCAAGGAGGCGGAG	A <sub>ExxE</sub>
		site directed mutagenesis
		H <sub>ExxE</sub>

<i>bqsS</i> -HExxE P2	CTCCGCCTCCTCGTGCAGGTTGCCGGCG	site directed mutagenesis HExxE
<i>bqsS</i> -RSxxS P1	AACCTGCGTCAGGAGGCTCAGAACCTGCTG	site directed mutagenesis RSxxS
<i>bqsS</i> -RSxxS P2	CAGCAGGTTCTGAGCCTCTGACGCAGGTT	site directed mutagenesis RSxxS
$\Delta$ <i>bqsR</i> $\Omega$ <i>glmS</i> -ptrc- <i>bqsR</i> g1	TCAACCAGCAGCAACCGCAgagctcgaattccatggtct	$\Delta$ <i>bqsR</i> constitutive expression complementation
$\Delta$ <i>bqsR</i> $\Omega$ <i>glmS</i> -ptrc- <i>bqsR</i> g2	CCGGGGTGGCCCGGATGAggtacctcggaaggccttg	$\Delta$ <i>bqsR</i> constitutive expression complementation
$\Delta$ <i>bqsR</i> $\Omega$ <i>glmS</i> -ptrc- <i>bqsR</i> g3	agaccatggaattcgagctcATGCGGTTGCTGCTGGTTGA	$\Delta$ <i>bqsR</i> constitutive expression complementation
$\Delta$ <i>bqsR</i> $\Omega$ <i>glmS</i> -ptrc- <i>bqsR</i> g4	caaggccttcgaggtaccT CATCCGGCGGCCACCCCGG	$\Delta$ <i>bqsR</i> constitutive expression complementation
<i>bqsR</i> -D51A P1	CTGATCATTCTCGCGCTCGGCCTGCCGGG	site directed mutagenesis D51A
<i>bqsR</i> -D51A P2	CCCGGGCAGGCCGAGCGCGAGAATGATCAG	site directed mutagenesis D51A
<i>bqsR</i> -D51E P1	CTGATCATTCTCGAGCTCGGCCTGCCGGG	site directed mutagenesis D51E
<i>bqsR</i> -D51E P2	CCCGGGCAGGCCGAGCTCGAGAATGATCAGS	site directed mutagenesis D51E
$\Delta$ <i>bqsR</i> $\Omega$ <i>glmS</i> -pbqs- <i>bqsR</i> g1	TGATTCCCACGGAGAGATT CATGCGGTTGCTGCTGG TTGA	$\Delta$ <i>bqsR</i> complementation
$\Delta$ <i>bqsR</i> $\Omega$ <i>glmS</i> -pbqs- <i>bqsR</i> g2	caaggccttcgaggtaccT CATCCGGCGGCCACCCCGG	$\Delta$ <i>bqsR</i> complementation
$\Delta$ <i>bqsR</i> $\Omega$ <i>glmS</i> -pbqs- <i>bqsR</i> g3	TCAACCAGCAGCAACCGCATGAATCTCTCCGTGGGA ATCA	$\Delta$ <i>bqsR</i> complementation
$\Delta$ <i>bqsR</i> $\Omega$ <i>glmS</i> -pbqs- <i>bqsR</i> g4	CCGGGGTGGCCCGGATGAggtacctcggaaggccttg	$\Delta$ <i>bqsR</i> complementation
gel shift		
F corrected <i>bqs</i> BqsR binding	GGCCATGTCCTTGCGGCAAATTAAGCTTCAATTAAG AGCGCCCGTTAATCTGCAAACCG	corrected sequence in front of <i>bqs</i> operon for potential BqsR binding site and flanking 20 bp (60 bp): with 5' labelled cy5
R corrected <i>bqs</i> BqsR binding	CGGTTTGCAGATTAACCGGGCGCTCTTAATTGAAGC TTAATTTGCCGAAGGACATGGCC	corrected sequence in front of <i>bqs</i> operon for potential BqsR binding site and flanking 20 bp (60 bp): fur box in front of TonB (as negative control) 55 bp: for BqsR binding with 5' labelled cy5
Fur box F	GGTAAGGAAGTAGAGTCTTCTGATAATTATTATCATT CAGTCGCGTCTCAGGGC	
Fur box R	GCCCTGAGGACGCGACTGAATGATAATAATTATCAG AAGACTCTACTTCTTACC	fur box in front of TonB (as negative control) 55 bp:

bqsR BS in Fur F	GGTAAGGAAGTAGAGTCTTCTTAAGCTTCAATTAAG TTCAGTCGCGTCTCAGGGC	BqsR binding site tandem repeats in the Fur box context (56 bp)
bqsR BS in Fur R	GCCCTGAGGACGCGACTGAACTTAATTGAAGCTTAA GAAGACTCTACTTCCTTACC GCC ATG TCC TTG CGG CAA ACG TGT CTT CAA	BqsR binding site tandem repeats in the Fur box context (56 bp)
FW Bqs BS delete all	CAG CTA GCG CCC GGT TAA TCT GCA AAC GTT TGC AGA TTA ACC GGG CGC TAG CTG TTG	
RV Bqs BS delete all	AAG ACA CGT TTG CCG CAA GGA CAT GGC	

cRACE		
WNp213	GTCTCGTTAGCTCGCTGGATCCTA 3' Inverted T	adaptor
WNp210	TAGGATCCAGCGAGCTAACGAGAC	prime off adaptor
bqs1	CGGCGTCACGCAGCTTCA	gene specific primer 1
bqs2	AGC TTC AGC GCC TCG T	gene specific primer 2
PA14_04180 RACE 1	TCGTCGATCTCCACCTTCA	gene specific primer 1
PA14_04180 RACE 2	ACTCGTAGATGTCGCCCTTGA	gene specific primer 2
PA14_04270 RACE 1	GCGGAAACGTCGTCGTTCAA	gene specific primer 1
PA14_04270 RACE 2	GTAGTCCGGCAGCCATATC	gene specific primer 2

**BqsR purification.** BqsR was expressed heterologously in *E. coli* with an N-terminal His-tag by inoculating LB/carbenicillin 50 ( $\mu\text{g ml}^{-1}$ ) with 1% inoculum from an overnight culture. Cells were grown to  $A_{600} = 0.4 - 0.5$  and BqsR expression induced with 1 mM IPTG and shifted to 16 °C overnight. Cells were pelleted at 5000 x g for 30 min and resuspended twice with wash buffer (500 mM KCl, 20 mM imidazole, 10% glycerol, and 20 mM Tris pH 8). DNase ( $2.5 \text{ units ml}^{-1}$ ) lysate, 5 mM  $\text{MgSO}_4$ , and 130  $\mu\text{M}$   $\text{CaCl}_2$  were added to the suspension. Cell lysate was generated with 14,000 psi on the Avestin Emulsiflex C3 and protease inhibitors added. BqsR was purified on Akta FPLC using 5 ml HisTrap HP (GE Healthcare). Column-bound BqsR was washed with 100 mM imidazole to remove contaminating proteins and eluted with 200 mM imidazole. After loading 1 mg protein (determined by the Bradford assay) per lane, the 200 mM imidazole fraction contained

one band, confirmed by MALDI-TOF to be BqsR. BqsR was dialyzed overnight with 3 changes of dialysis buffer (50% glycerol, 300 mM KCl, 50 mM KPO<sub>4</sub> pH 7.4), concentrated, and flash-frozen BqsR aliquots stored at -80 °C.

*Gel shift assays.* Double stranded DNA was generated by boiling equimolar concentrations of single stranded Cy5 labeled reverse complementary DNA sequences (approximately 60 bp in length) for 10 min and allowing to slowly come to room temperature over an hour. Various concentrations of BqsR were incubated with 5 nM double stranded DNA (synthesized and purified by IDT; both strands 5'-labeled with Cy5) in reaction buffer (15 mM KCl, 10% glycerol, 1 mM MgSO<sub>4</sub>, 50 mM , 40.1 mM K<sub>2</sub>HPO<sub>4</sub>, 9.9 mM KH<sub>2</sub>PO<sub>4</sub> pH 7.4) for 20 min at room temperature and run on an 8% acrylamide gel (0.5x TBE, 29 acrylamide: 1 bis-acrylamide [Bio-Rad]). Each of the above parameters was optimized to yield the cleanest gel shift, yet some smearing remained despite these efforts, possibly influenced by the instability of BqsR. Attempts to chemically phosphorylate BqsR with acetyl-P to enhance binding affinity were unsuccessful, which is not unprecedented for two-component transcription factors (75). DNA was visualized with Storm 860 in fluorescence mode using the 635 nm excitation laser. Through the duration of the experiment, the reaction and gel were shielded from light with aluminum foil.

*Transcription start site determination.* . cDNA was generated using gene specific primer 1 and Superscript II (Invitrogen) including the RNase inhibitors, RNasin (Promega) in the

reaction. The mRNA template was degraded with RNase H. The cDNA ligated to adaptor DNA (WNp210) with an inverted 3' T. A nested PCR reaction was performed with gene specific primer 2 and a primer complementary to the adaptor sequence (WNp213). The PCR reaction was TOPO cloned and sent for sequencing. The transcription start site is the junction between the adaptor sequence and the gene specific sequence.

*Point mutation growth conditions.* Aerobic cultures of *P. aeruginosa* PA14

complementation strains and point mutants were grown overnight in 5 ml MOMM. BqsR strains were inoculated into anaerobic 1  $\mu$ M Fe(II) MOMM and 100  $\mu$ M Fe(II) MoMM in triplicate. Anaerobic cultures were grown at 37 °C until early exponential phase (Beckman Spec20 A500 = 0.2) and harvested with 10 ml RNAprotect. BqsS strains were inoculated into 1  $\mu$ M Fe(II) MOMM and shocked with 200  $\mu$ M Fe(II) anaerobically and harvested with 10 ml RNAprotect.

*Consensus sequence generation.* To make an initial BqsR consensus DNA binding site prediction, the 500 base pairs upstream from the translation start site of the 5 most upregulated genes (*bqsP*, PA14\_04180, PA14\_04270, PA14\_01240, and PA14\_07070) discovered in an Fe(II)/Fe(III) shock microarray (2) were used as input sequences for the motif finding program, MEME (version 4.9.0) (26). The MEME input parameters allowed for zero or one sequence repetition and 3 motifs returned with 6 – 50 nucleotides per motif. This generated the position weight matrix (PWM) used as the basis for the gel shift

assays (Fig. 3-2). Several PWM were generated by adding the 500 bp upstream regions of newly qRT-PCR verified BqsR-regulated genes to the list of MEME input sequences as the genes were identified. The final version of the PWM was generated with MEME (version 4.9.1). The MEME input parameters allowed for any number of sequence repetitions, 3 motifs returned with 6 – 50 nucleotides per motif, and the program searched the given strand only. For the final version of the PWM the seeding sequences were from *Pseudomonas aeruginosa* PA14 and were validated by Fe(II) shock qRT-PCR experiments that showed upregulation in WT but not in a  $\Delta bqsR$  mutant. The 500 bp upstream from the following gene translation start sites used for the final PWM were *bqsP*, PA14\_04180, PA14\_04270, PA14\_01240, PA14\_07070, PA14\_01250, *dsbB*, *arnB*, PA14\_32270, PA14\_063110, PA14\_32590, and *oprH*.

*BqsR regulon prediction.* For the regular expression searching, we identified variants by selecting the most common sequences by eye from our initial PWM (shown in Fig. S1A). For example, in Fig. S1A, position four in the first pentamer shows the most common nucleotides are A and C. Accordingly, our first regular expression was designed with this preference in a tandem repeat orientation, including a variable length linker region (assuming no nucleotide preference in this region). The second regular expression kept the same sequence preferences for the first pentamer, but considered a palindromic structure. The last regular expression was designed as a tandem repeat, only in this case we varied the sequence to reflect the preference shown in the second pentamer in the



initial PWM (e.g., position 12-16). Using these regular expressions, we searched the intergenic regions of the *P. aeruginosa* UCBPP-PA14 genome to identify potential BqsR-regulated genes (Dataset S1).

The PWMs were uploaded into the MAST (version 4.9.1) tool (26) with the upstream regions database selected for the *P. aeruginosa* UCBPP-PA14 genome. The E-value cutoff was set to 50 and each strand was treated separately since the motif is a tandem repeat (rather than a palindrome). Also, by searching the intergenic regions of the PA14 genome for the most common variants of the tandem repeat consensus sequence, assuming no preference for the 6 nucleotides between the tandem repeat sequences (regular expression), a list of potential BqsR-regulated genes were identified. These upstream regions were then analyzed by MEME and MAST to generate the most broadly predicted BqsR regulon. We deliberately cast a wide net in our search process in order to identify a large set of genes to consider in follow-up tests, recognizing that some may be false-positives (Dataset S1).

*Fe(II) shock conditions.* Triplicate aerobic cultures of *P. aeruginosa* WT-pMQ72 and  $\Delta bqsR$ -pMQ72 were grown in MOMM + 100  $\mu\text{g ml}^{-1}$  gentamycin at 37 °C for 36 hours. Anaerobic cultures were inoculated with 1% inoculum in MOMM + 100  $\mu\text{g ml}^{-1}$  gentamycin + 1% arabinose. For qRT-PCR, when the cells reached early exponential phase ( $\text{OD}_{500} = 0.2$ ), RNAprotect was added before and after a 30 min 200  $\mu\text{M}$   $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$  shock to sample the “unshocked” and “shocked” states, respectively. For the RNA-Seq

experiments, when the cells reached deep stationary phase ( $OD_{500} = 0.8$ ), RNAprotect was added before and after a 30 min 200  $\mu\text{M}$   $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$  shock. We are confident that Fe(II) induces transcriptional changes, not the counter ion, sulfate, because MOMM contains a background of 1 mM sulfate. Additionally, previous 100  $\mu\text{M}$  shock experiments were done with either  $\text{FeCl}_2$  or  $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$  and no difference in expression was seen (2).

*BqsR regulon gene annotation and classification.* Genes seen in bioinformatic predictions and in the RNA-Seq data were manually annotated from the database with the lowest e-value over the largest portion of the gene as viewed on the *Pseudomonas* Genome Database (36). Databases used include COG (77), TIGRFAM (78), Pfam (79), CD (80), prk (81), and SMART (82). To cluster genes with similar annotations, DAVID (37) was set to medium stringency with all possible input databases and terms selected.

*RNA-Seq processing.* RNeasy kit (Qiagen) was used to isolate RNA and the Ribo-Zero Magnetic kit (epicentre) was used to deplete rRNA. The NEBNext mRNA Library Prep kit for Illumina (NEB) was used to prepare cDNA libraries for sequencing. Libraries were sequenced on an Illumina HiSeq2500 with 15 million read sequencing depth, at the Caltech Millard and Muriel Jacobs Genetics and Genomics Laboratory. To analyze the data, Trimmomatic -0.32 (83) was used to trim the low quality bases from the reads with parameters set to LEADING:27 TRAILING:27 SLIDINGWINDOW:4:20 MINLEN:35. The

trimmed reads were mapped to the genome using Bowtie 1.0.1 (84). Samtools 0.1.19 (85) was used to sort the mapped reads. Read counts per gene or transcriptional unit were calculated using Easy RNASeq (86), using .gff gene description files generated from the curated genome hosted by NCBI (NC\_008463.1) (Dataset S2), modified by the results of the single nucleotide resolution sequencing published by Wurtzel et al (46) (Dataset S3). Significance values for differential expression were determined (47) using Degust. The full dataset is publicly available through the NCBI gene expression omnibus (GEO accession number GSE65393).

*mRNA isolation and qRT-PCR data analysis.* mRNA was isolated using RNeasy kit mini (Qiagen) with optional on-column DNA digestion according to the manufacturer's instructions. Subsequently, the RNA was treated with TURBO DNA-free (Applied Biosystems), the rigorous treatment protocol. cDNA was generated using the extracted RNA as a template for an iScript (Bio-Rad) random-primed reverse transcriptase reaction following the manufacturer's protocol. The cDNA was used as template for quantitative PCR (Real Time 7500 PCR Machine, Applied Biosystems) using iTaq Universal SYBR Green Supermix (Bio-Rad). The elimination of genomic contamination was confirmed by running a qRT-PCR control plate with mRNA. Samples were assayed with 3-5 biological replicates. *recA* and *clpX* were used as endogenous controls (87). Fold changes were calculated using the  $\Delta\Delta C_t$  method (2).

*Spermidine quantification by HPLC analysis.* In order to ensure sufficient spermidine production for visual quantification, 5 mM arginine and 5 mM methionine were supplemented to the MOMM media to provide the precursors for spermidine synthesis; this medium will be subsequently referred to as MOMM+AA. Biological triplicates of PA14,  $\Delta bqsR$ ,  $\Delta bqsS$ , and  $\Delta bqsRS$  were in 5 mL MOMM+AA with either 5  $\mu$ M Fe or 75  $\mu$ M Fe and diluted to an  $OD_{500}=0.01$  in batch tubes containing 20 mL MOMM+AA with either 5  $\mu$ M or 75  $\mu$ M Fe(II). Cultures were stoppered and grown shaking at 37°C until  $OD \approx 0.7$ . Cells were pelleted in two centrifugations in 15 mL falcon tubes at 6800 rpm for 15 min at 4°C, and supernatant was removed. Pellets were then suspended in 1 mL 1M NaCl and 0.1M HEPES pH=7.2 and incubated for 10 min at 37°C shaking. Suspension was transferred to a 2 mL microfuge tube and centrifuged at 8000 rpm for 10 min. Supernatant was removed, and pellet was suspended in 1 mL 10% trichloroacetic acid and incubated at 4°C for 3 hours. Suspension was again centrifuged at 8000 rpm for 10 min and supernatant was siphoned off and stored in a glass vial at 4°C. 1,3-diaminopropane was added to the supernatants at a concentration of 30  $\mu$ g/ml as an internal standard for subsequent dabsylation and HPLC analysis of polyamines. 4  $\mu$ L of each sample was dabsylated in triplicate and analyzed by HPLC. Results were normalized by  $OD_{500}$ .

*Polymyxin B killing curve.* Triplicate WT and  $\Delta bqsR$  cultures were grown anaerobically in MOMM + AA + 75  $\mu$ M Fe(II) to an  $OD_{500} = 0.7$ . Aliquots were incubated anaerobically with

either water or  $25 \mu\text{g ml}^{-1}$  polymyxin B for 2 hours. Samples were plated onto LB every 30 min. CFU counts were taken after 24 and 48 hours, incubating aerobically at  $37^\circ\text{C}$ .

*MIC assays.* Cells were inoculated anaerobically in MOMM pH 7.2 with tobramycin (0-8  $\mu\text{g/ml}$ ), cefsulodin (0-128  $\mu\text{g/ml}$ ), sodium arsenate (0-16  $\mu\text{g/ml}$ ), and potassium chromate (0-2  $\mu\text{g/ml}$ ) in either  $5 \mu\text{M Fe(II)}$  or  $75 \mu\text{M Fe(II)}$ .  $\text{MIC}_{90}$  was calculated at the time each strain achieved maximum  $\text{OD}_{500}$  in wells with no antibiotic added. Each assay was performed with biological triplicates and results were confirmed on two separate occasions.

*Tobramycin MIC rescue experiments.* Ferrozine and spermidine were tested for their ability to rescue tobramycin sensitivity. Cultures were prepared and analyzed as usual, but with the addition of  $140 \mu\text{M}$  ferrozine or  $10 \text{ mM}$  spermidine to MOMM + AA +  $75 \mu\text{M Fe(II)}$ , with tobramycin concentrations ranging from 0-8  $\mu\text{g/ml}$ .

### **Acknowledgments**

We thank Megan Bergkessel, other members of the Newman laboratory, and anonymous reviewers for constructive comments on the manuscript. This work was supported by grants to DKN from the Howard Hughes Medical Institute (HHMI) and the National Heart, Lung, and Blood Institute of the National Institutes of Health (R01HL117328) and by the

Millard and Muriel Jacobs Genetics and Genomics Laboratory at California Institute of Technology. DKN is an HHMI Investigator.

## References

1. **Hunter RC, Asfour F, Dingemans J, Osuna BL, Samad T, Malfroot A, Cornelis P, Newman DK.** 2013. Ferrous iron is a significant component of bioavailable iron in cystic fibrosis airways. *mBio* **4**:e00557-13.
2. **Kreamer N NK, Wilks JC, Marlow JJ, Coleman ML, Newman DK.** 2012. BqsR/BqsS constitute a two-component system that senses extracellular Fe(II) in *Pseudomonas aeruginosa*. *J Bacteriol* **194**:1195-1204.
3. **Kiley PJ, Beinert H.** 2003. The role of Fe-S proteins in sensing and regulation in bacteria. *Curr Opin Microbiol* **6**:181-185.
4. **Fontecave M.** 2006. Iron-sulfur clusters: ever-expanding roles. *Nat Chem Biol* **2**:171-174.
5. **Lill R.** 2009. Function and biogenesis of iron-sulphur proteins. *Nature* **460**:831-838.
6. **Schalk I.** 2006. New insights on iron acquisition mechanisms in pathogenic *Pseudomonas*, p. 1-34. In Ramos J-L, Levesque R (ed.), *Pseudomonas*. Springer US.
7. **Banin E, Vasil ML, Greenberg EP.** 2005. Iron and *Pseudomonas aeruginosa* biofilm formation. *Proc Natl Acad Sci U S A* **102**:11076-11081.
8. **Patriquin GM, Banin E, Gilmour C, Tuchman R, Greenberg EP, Poole K.** 2008. Influence of quorum sensing and iron on twitching motility and biofilm formation in *Pseudomonas aeruginosa*. *J Bacteriol* **190**:662-671.
9. **Mah T-FC, O'Toole GA.** Mechanisms of biofilm resistance to antimicrobial agents. *Trends Microbiol* **9**:34-39.
10. **Stumm W, Lee GF.** 1961. Oxygenation of Ferrous Iron. *Ind Eng Chem* **53**:143-146.
11. **Steele KH, O'Connor LH, Burpo N, Kohler K, Johnston JW.** 2012. Characterization of a ferrous iron-responsive two-component system in nontypeable *Haemophilus influenzae*. *J Bacteriol* **194**:6162-6173.
12. **Mayer-Hamblett N, Aitken ML, Accurso FJ, Kronmal RA, Konstan MW, Burns JL, Sagel SD, Ramsey BW.** 2007. Association between pulmonary function and sputum biomarkers in cystic fibrosis. *Am J Respir Crit Care Med* **175**:822-828.
13. **Kim J-S, Okamoto K, Rubin BK.** 2006. Pulmonary function is negatively correlated with sputum inflammatory markers and cough clearability in subjects with cystic fibrosis but not those with chronic bronchitis. *Chest* **129**:1148-1154.
14. **Zeng A-P, Kim E-J.** 2004. Iron availability, oxygen limitation, *Pseudomonas aeruginosa* and cystic fibrosis. *Microbiol* **150**:516-518.
15. **Worlitzsch D, et. al.** 2002. Effects of reduced mucus oxygen concentration in airway *Pseudomonas* infections of cystic fibrosis patients. *J Clin Invest* **109**:317-325.
16. **Kolpen M, Kühl M, Bjarnsholt T, Moser C, Hansen CR, Lienggaard L, Kharazmi A, Pressler T, Høiby N, Jensen PØ.** 2014. Nitrous Oxide production in sputum from

- cystic fibrosis patients with chronic *Pseudomonas aeruginosa* lung infection. PLoS ONE **9**:e84353.
17. **Green SK, Schroth MN, Cho JJ, Kominos SD, Vitanza-Jack VB.** 1974. Agricultural Plants and Soil as a Reservoir for *Pseudomonas aeruginosa*. Appl Microbiol **28**:987-991.
  18. **Gotoh S, Patrick WH.** 1974. Transformation of Iron in a Waterlogged Soil as Influenced by Redox Potential and pH. Soil Sci Soc Am J **38**:66-71.
  19. **Gooderham WJ, Hancock REW.** 2009. Regulation of virulence and antibiotic resistance by two-component regulatory systems in *Pseudomonas aeruginosa*. FEMS Microbiol Rev **33**:279-294.
  20. **Dong Y-H, Zhang X-F, An S-W, Xu J-L, Zhang L-H.** 2008. A novel two-component system BqsS-BqsR modulates quorum sensing-dependent biofilm decay in *Pseudomonas aeruginosa*. Commun Integr Biol **1**:88-96.
  21. **Chand NS, Lee JS-W, Clatworthy AE, Golas AJ, Smith RS, Hung DT.** 2011. The sensor kinase KinB regulates virulence in acute *Pseudomonas aeruginosa* infection. J Bacteriol **193**:2989-2999.
  22. **Hofmann.** 1993. TMbase - A database of membrane spanning proteins segments. Biol Chem Hoppe-Seyler **374**.
  23. **Cserző M, Wallin E, Simon I, von Heijne G, Elofsson A.** 1997. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. Protein Eng **10**:673-676.
  24. **Wösten MMSM, Kox LFF, Chamnongpol S, Soncini FC, Groisman EA.** 2000. A signal transduction system that responds to extracellular iron. Cell **103**:113-125.
  25. **Stearman R, Yuan DS, Yamaguchiwai Y, Klausner RD, Dancis A.** 1996. A permease-oxidase complex involved in high-affinity iron uptake in yeast. Science **271**:1552-1557.
  26. **Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS.** 2009. MEME Suite: tools for motif discovery and searching. Nucleic Acids Res **37**:W202-W208.
  27. **Galperin MY.** 2010. Diversity of structure and function of response regulator output domains. Curr Opin Microbiol **13**:150-159.
  28. **Rampersaud A, Norioka S, Inouye M.** 1989. Characterization of OmpR binding sequences in the upstream region of the *ompF* promoter essential for transcriptional activation. J Biol Chem **264**:18693-18700.
  29. **Blanco AG, Sola M, Gomis-Rüth FX, Coll M.** 2002. Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator. Structure **10**:701-713.
  30. **Schaaf S, Bott M.** 2007. Target genes and DNA-binding sites of the response regulator PhoR from *Corynebacterium glutamicum*. J Bacteriol **189**:5002-5011.
  31. **Ghosh T, Bose D, Zhang X.** 2010. Mechanisms for activating bacterial RNA polymerase. FEMS Microbiol Rev **34**:611-627.



32. **Gupte G, Woodward C, Stout V.** 1997. Isolation and characterization of *rcsB* mutations that affect colanic acid capsule synthesis in *Escherichia coli* K-12. *J Bacteriol* **179**:4328-4335.
33. **Lan C-Y, Igo MM.** 1998. Differential expression of the OmpF and OmpC porin proteins in *Escherichia coli* K-12 Depends upon the level of active OmpR. *J Bacteriol* **180**:171-174.
34. **Klose KE, Weiss DS, Kustu S.** 1993. Glutamate at the site of phosphorylation of nitrogen-regulatory protein NTRC mimics aspartyl-phosphate and activates the protein. *J Mol Biol* **232**:67-78.
35. **Choi K-H, Schweizer HP.** 2006. mini-Tn7 insertion in bacteria with single attTn7 sites: example *Pseudomonas aeruginosa*. *Nat Protocols* **1**:153-161.
36. **Winsor GL, Lam DKW, Fleming L, Lo R, Whiteside MD, Yu NY, Hancock REW, Brinkman FSL.** 2011. *Pseudomonas* Genome Database: improved comparative analysis and population genomics capability for *Pseudomonas* genomes. *Nucleic Acids Res* **39**:D596-D600.
37. **Huang DW, Sherman BT, Lempicki RA.** 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protocols* **4**:44-57.
38. **Macfarlane ELA, Kwasnicka A, Ochs MM, Hancock REW.** 1999. PhoP–PhoQ homologues in *Pseudomonas aeruginosa* regulate expression of the outer-membrane protein OprH and polymyxin B resistance. *Mol Microbiol* **34**:305-316.
39. **Poole K.** 2005. Aminoglycoside Resistance in *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* **49**:479-487.
40. **Miller AK, Brannon MK, Stevens L, Johansen HK, Selgrade SE, Miller SI, Høiby N, Moskowitz SM.** 2011. PhoQ mutations promote Lipid A modification and polymyxin resistance of *Pseudomonas aeruginosa* found in colistin-treated cystic fibrosis patients. *Antimicrob Agents Chemother* **55**:5761-5769.
41. **Shah P, Swiatlo E.** 2008. A multifaceted role for polyamines in bacterial pathogens. *Mol Microbiol* **68**:4-16.
42. **Johnson L, Mulcahy H, Kanevets U, Shi Y, Lewenza S.** 2012. Surface-localized spermidine protects the *Pseudomonas aeruginosa* outer membrane from antibiotic treatment and oxidative stress. *J Bacteriol* **194**:813-826.
43. **McPhee JB, Lewenza S, Hancock REW.** 2003. Cationic antimicrobial peptides activate a two-component regulatory system, PmrA-PmrB, that regulates resistance to polymyxin B and cationic antimicrobial peptides in *Pseudomonas aeruginosa*. *Mol Microbiol* **50**:205-217.
44. **Moskowitz SM, Ernst RK, Miller SI.** 2004. PmrAB, a two-component regulatory system of *Pseudomonas aeruginosa* that modulates resistance to cationic antimicrobial peptides and addition of aminoarabinose to Lipid A. *J Bacteriol* **186**:575-579.
45. **Parsek MR, Greenberg EP.** 2005. Sociomicrobiology: the connections between quorum sensing and biofilms. *Trends Microbiol* **13**:27-33.

46. **Wurtzel O, Yoder-Himes DR, Han K, Dandekar AA, Edelheit S, Greenberg EP, Sorek R, Lory S.** 2012. The single-nucleotide resolution transcriptome of *Pseudomonas aeruginosa* grown in body temperature. *PLoS Pathog* **8**:e1002945.
47. **Law C, Chen Y, Shi W, Smyth G.** 2014. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**:R29.
48. **Yao X, Lu C-D.** 2014. Functional characterization of the potRABCD operon for spermine and spermidine uptake and regulation in *Staphylococcus aureus*. *Curr Microbiol* **69**:75-81.
49. **Tatusov RL, Koonin EV, Lipman DJ.** 1997. A genomic perspective on protein families. *Science* **278**:631-637.
50. **Kwon DH, Lu C-D.** 2006. Polyamines induce resistance to cationic peptide, aminoglycoside, and quinolone antibiotics in *Pseudomonas aeruginosa* PAO1. *Antimicrob Agents Chemother* **50**:1615-1622.
51. **Yotsuji A, Mitsuyama J, Hori R, Yasuda T, Saikawa I, Inoue M, Mitsuhashi S.** 1988. Outer membrane permeation of *Bacteroides fragilis* by cephalosporins. *Antimicrob Agents Chemother* **32**:1097-1099.
52. **Avery AM, Goddard HJ, Sumner ER, Avery SV.** 2004. Iron blocks the accumulation and activity of tetracyclines in bacteria. *Antimicrob Agents Chemother* **48**:1892-1894.
53. **Wedderhoff I, Kursula I, Groves MR, Ortiz de Oru  Lucana Do.** 2013. Iron binding at specific sites within the octameric HbpS protects *Streptomyces* from iron-mediated oxidative stress. *PLoS ONE* **8**:e71579.
54. **Laub MT, Goulian M.** 2007. Specificity in two-component signal transduction pathways. *Ann Rev Genet* **41**:121-145.
55. **Minguet EG, Vera-Sirera F, Marina A, Carbonell J, Bl zquez MA.** 2008. Evolutionary diversification in polyamine biosynthesis. *Mol Biol Evol* **25**:2119-2128.
56. **Langley S, Beveridge TJ.** 1999. Effect of O-side-chain-lipopolysaccharide chemistry on metal binding. *Appl Environ Microbiol* **65**:489-498.
57. **Boggis W, Kenward MA, Brown MRW.** 1979. Effects of divalent metal cations in the growth medium upon sensitivity of batch-grown *Pseudomonas aeruginosa* to EDTA or polymyxin B. *J Appl Bacteriol* **47**:477-488.
58. **Bird LJ, Coleman ML, Newman DK.** 2013. Iron and copper act synergistically to delay anaerobic growth of bacteria. *Appl Environ Microbiol* **79**:3619-3627.
59. **Dunning JC, Ma Y, Marquis RE.** 1998. Anaerobic killing of oral *Streptococci* by reduced, transition metal cations. *Appl Environ Microbiol* **64**:27-33.
60. **Poulain AJ, Newman DK.** 2009. *Rhodobacter capsulatus* catalyzes light-dependent Fe(II) oxidation under anaerobic conditions as a potential detoxification mechanism. *Appl Environ Microbiol* **75**:6639-6646.
61. **Nikaido H, Vaara M.** 1985. Molecular basis of bacterial outer membrane permeability. *Microbiol Rev* **49**:1-32.

62. **Craig SA, Carpenter CD, Mey AR, Wyckoff EE, Payne SM.** 2011. Positive regulation of the *Vibrio cholerae* porin OmpT by iron and fur. *J Bacteriol* **193**:6505-6511.
63. **Kim AD, Graham DE, Seeholzer SH, Markham GD.** 2000. S-adenosylmethionine decarboxylase from the archaeon *Methanococcus jannaschii*: identification of a novel family of pyruvoyl enzymes. *J Bacteriol* **182**:6667-6672.
64. **Lu ZJ, Markham GD.** 2007. Metal ion activation of S-adenosylmethionine decarboxylase reflects cation charge density. *Biochem* **46**:8172-8180.
65. **Jung IL, Oh TJ, Kim IG.** 2003. Abnormal growth of polyamine-deficient *Escherichia coli* mutant is partially caused by oxidative stress-induced damage. *Arch Biochem Biophys* **418**:125-132.
66. **Tkachenko AG, Nesterova LY.** 2003. Polyamines as modulators of gene expression under oxidative stress in *Escherichia coli*. *Biochem (Moscow)* **68**:850-856.
67. **McPhee JB, Bains M, Winsor G, Lewenza S, Kwasnicka A, Brazas MD, Brinkman FSL, Hancock REW.** 2006. Contribution of the PhoP-PhoQ and PmrA-PmrB two-component regulatory systems to Mg<sup>2+</sup>-Induced gene regulation in *Pseudomonas aeruginosa*. *J Bacteriol* **188**:3995-4006.
68. **Morel F, Hering JG.** 1993. Principles and applications of aquatic chemistry. Wiley, New York.
69. **Castro AP, Fernandes GDR, Franco OL.** 2014. Insights into novel antimicrobial compounds and antibiotic resistance genes from soil metagenomes. *Front Microbiol* **5**.
70. **Regeimbal J, Bardwell JCA.** 2002. DsbB catalyzes disulfide bond formation de novo. *J Biol Chem* **277**:32706-32713.
71. **Wen Y, Feng J, Scott DR, Marcus EA, Sachs G.** 2007. The HP0165-HP0166 two-component system (ArsRS) regulates acid-induced expression of HP1186  $\alpha$ -carbonic anhydrase in *Helicobacter pylori* by activating the pH-dependent promoter. *J Bacteriol* **189**:2426-2434.
72. **Beringer P.** 2001. The clinical use of colistin in patients with cystic fibrosis. *Curr Opin Pulm Med* **7**:434-440.
73. **Kopf SH, Henny C, Newman DK.** 2013. Ligand-enhanced abiotic iron oxidation and the effects of chemical versus biological iron cycling in anoxic environments. *Environ Sci Technol* **47**:2602-2611.
74. **Stookey LL.** 1970. Ferrozine---a new spectrophotometric reagent for iron. *Anal Chem* **42**:779-781.
75. **Lukat GS, McCleary WR, Stock AM, Stock JB.** 1992. Phosphorylation of bacterial response regulator proteins by low molecular weight phospho-donors. *Proc Natl Acad Sci U S A* **89**:718-722.
76. **Bose A, Newman DK.** 2011. Regulation of the phototrophic iron oxidation (pio) genes in *Rhodospseudomonas palustris* TIE-1 is mediated by the global regulator, FixK. *Mol Microbiol* **79**:63-75.

77. **Tatusov RL, Galperin MY, Natale DA, Koonin EV.** 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28**:33-36.
78. **Haft DH, Selengut JD, White O.** 2003. The TIGRFAMs database of protein families. *Nucleic Acids Res* **31**:371-373.
79. **Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M.** 2014. Pfam: the protein families database. *Nucleic Acids Res* **42**:D222-D230.
80. **Marchler-Bauer A, et. al.** 2013. CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* **41**:D348-D352.
81. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. *J Mol Biol* **215**:403-410.
82. **Schultz J, Milpetz F, Bork P, Ponting CP.** 1998. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc Natl Acad Sci U S A* **95**:5857-5864.
83. **Bolger AM, Lohse M, Usadel B.** 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics.* **30**:2114-2120.
84. **Langmead B, Trapnell C, Pop M, Salzberg S.** 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**:R25.
85. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP.** 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078-2079.
86. **Delhomme N, Padioleau I, Furlong EE, Steinmetz LM.** 2012. easyRNASeq: a bioconductor package for processing RNA-Seq data. *Bioinformatics* **28**:2532-2533.
87. **Dietrich LEP, Price-Whelan A, Petersen A, Whiteley M, Newman DK.** 2006. The phenazine pyocyanin is a terminal signalling factor in the quorum sensing network of *Pseudomonas aeruginosa*. *Mol Microbiol* **61**:1308-1321.
88. **Koski P, Helander IM, Sarvas M, Vaara M.** 1987. Analysis of polyamines as their dabsyl derivatives by reversed-phase high-performance liquid chromatography. *Analytical Biochemistry* **164**:261-266.

Dataset S3-1. Bioinformatic Predictions. A “+” in the “regular expression” column indicates the gene in this row was identified by this technique. E-values are for genes identified by the PWM.

Gene	Regular expression (3A)	e-value	Function
PA14_29690	+	0.047	hypothetical
PA14_29710	+	0.047	bqsP
PA14_07060	+	0.07	Predicted phosphoesterase (cog)
PA14_07070	+	0.07	Predicted ferric reductase (cog)
PA14_04250	+	0.092	ABC-type spermidine/putrescine transport systems, ATPase components (cog, prk)
PA14_04270	+	0.092	transcription factor (PseudoCAP Function Class)
PA14_32250	+	0.18	hypothetical
PA14_32270	+	0.18	outer membrane porin, OprD family (pfam)
PA14_04180	+	0.25	Bacterial OB-Fold protein (pfam)
PA14_63110	+	0.53	S-adenosylmethionine decarboxylase (cog, pfam, tigrfam); Spermidine synthase (cog, pfam, tigrfam) operon
PA14_32580	+	0.65	Response regulator (cog, pfam, cd)
PA14_32590	+	0.65	Thiol:disulfide interchange protein (cog, prk)
PA14_35790	+	1	Homospermidine synthase (cog)
PA14_57880	+	1	ABC-type transport system involved in resistance to organic solvents, ATPase component (cog)
PA14_57890	+	1	Predicted sugar phosphate isomerase involved in capsule formation (cog)
PA14_01240	+	1.1	Zn-binding carbonic anhydrase (doi: 10.1099/mic.0.066357-0)
PA14_18370	+	1.3	UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase (cog)
PA14_54120	+	1.9	ACP phosphodiesterase (cog)
PA14_54130	+	1.9	Transcriptional regulator (cog, pfam)
PA14_58390		2.3	ABC-type dipeptide transport system, periplasmic component (cog, prk, cd)
aguB	+	3.5	N-carbamoylputrescine amidase (tigrfam, cd)
PA14_52340	+	3.5	hypothetical
PA14_52350	+	3.5	hypothetical
PA14_35300	+	3.8	Gluconate 2-dehydrogenase subunit 3 (pfam)
PA14_31630		3.8	LysR substrate binding domain (pfam, prk)
PA14_31640		3.8	This conserved domain belongs to a superfamily including the bleomycin (cd) resistance protein, glyoxalase I, and

## type I ring-cleaving dioxygenases

oprH		3.8	outer membrane porin stabilizes outer membrane by LPS binding and increases antibiotic resistance (doi: 10.1074/jbc.M111.280933)
napE	+	3.8	nitrate reductase periplasmic (PMID: 9560320)
PA14_24350	+	4.4	cprR - phosphorelay response regulator activity antimicrobial peptide activated (PMID: 23006746)
PA14_24360	+	4.4	Peptidase S8 family domain in Autotransporter serine proteases (cd)
PA14_46900	+	4.5	hypothetical
PA14_46910	+	4.5	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain (cog)
PA14_37990	+	8.2	RNA polymerase sigma factor (cog, prk, tigrfam)
PA14_38000	+	8.2	hypothetical
algD	+	8.2	GDP-mannose 6-dehydrogenase AlgD (PMID: 12705829, PMID: 7521247)
PA14_61840	+	8.6	Plasmid maintenance system antidote protein (cog)
PA14_61850	+	8.6	Outer membrane receptor proteins, mostly Fe transport (cog)
lasR	+	10	quorum sensing transcription factor
PA14_34150	+	12	Transcriptional regulators containing an AAA-type ATPase domain and a DNA-binding domain (cog)
PA14_02520		14	hypothetical
PA14_02530		14	Predicted esterase of the alpha/beta hydrolase fold (cog)
PA14_24690	+	16	D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein 4) (cog)
PA14_24700		16	hypothetical
PA14_40170		19	Membrane transporters of cations and cationic drugs (cog)
PA14_40180		19	Aerobic-type carbon monoxide dehydrogenase, small subunit CoxS/CutS homologs (cog)
feoA	+	23	Fe <sup>2+</sup> transport system protein A
PA14_63470		25	S-adenosylmethionine-dependent methyltransferases (CD), Methyltransferase domain (pfam)
PA14_02760		25	cyl CoA:acetate/3-ketoacid CoA transferase, alpha subunit (cog)
PA14_47380		27	TonB-dependent heme/hemoglobin receptor family protein (tigrfam)
PA14_54830		29	Predicted hydrolases or acyltransferases (cog)
PA14_27070		32	Predicted membrane protein (cog)
PA14_21020	+	32	Non-ribosomal peptide synthetase modules and related proteins (cog)
PA14_21030	+	32	ATP-dependent Clp protease proteolytic subunit (prk, cog)

metN		32	DL-methionine transporter ATP-binding subunit (prk)
PA14_38380	+	35	Transcriptional regulator (cog)
PA14_38395	+	35	periplasmic multidrug efflux lipoprotein precursor (prk)
PA14_08600		35	23S ribosomal RNA
PA14_55631		35	23S ribosomal RNA
PA14_62060		35	23S ribosomal RNA
PA14_70880		35	23S ribosomal RNA
PA14_07430	+	36	hypothetical
PA14_17380		38	LysR family transcriptional regulator (prk)
adhC		38	Zn-dependent alcohol dehydrogenases, class III( cog)
PA14_44820		38	Transthyretin-like protein (cog)
PA14_44830		38	Catalytic domain of bacterial PuuE allantoinases, Schizosaccharomyces pombe chitin deacetylase 1 (SpCDA1), and similar proteins (cd)
PA14_58375	+	43	Xaa-Pro aminopeptidase (cog)
PA14_58380	+	43	Predicted transcriptional regulators (cog)
PA14_59550		43	hypothetical
PA14_52960		43	hypothetical
PA14_10080		45	hypothetical
PA14_10090		45	Transcriptional regulator (cog)
PA14_01490		45	hemolysin
PA14_25840		47	Dehydrogenases (flavoproteins) [Energy production and conversion] (cog)
etfB		47	Electron transfer flavoprotein, beta subunit (cog)
PA14_55380		47	Hemolysin activation/secretion protein (cog)
PA14_63220	+		hypothetical
PA14_35740	+		Transposase and inactivated derivatives, TnpA family (cog)
mexG	+		hypothetical part of operon MexGHI-OpmD efflux pump confers resistance to vanadium and is probably involved in acyl-homoserine lactone homeostasis. (cog)
PA14_51610	+		hypothetical
PA14_51620	+		Transposase and inactivated derivatives (cog)
PA14_51310	+		Predicted redox protein, regulator of disulfide bond formation (cog)
PA14_51320	+		Putative Zn-dependent protease, contains TPR repeats [General function prediction only] (cog)
PA14_13890	+		Integrase core domain (pfam)
PA14_35210	+		Transcriptional regulator [Transcription] (cog)
PA14_35230	+		Sugar phosphate permease (cog)
PA14_31130	+		hypothetical
PA14_28200	+		hypothetical

PA14_31350	+	Uncharacterized NAD(FAD)-dependent dehydrogenases [General function prediction only] (cog)
PA14_56790	+	Predicted signal transduction protein containing a membrane domain, an EAL and a GGDEF domain (cog)
PA14_14160	+	Xenobiotic acyltransferase (XAT): The XAT class of hexapeptide acyltransferases is composed of a large number of microbial enzymes that catalyze the CoA-dependent acetylation of a variety of hydroxyl-bearing acceptors such as chloramphenicol and streptogramin, among others (CD)
aprA	+	alkaline metalloproteinase (pdb 1XAP, 1AKL)
PA14_32710	+	RNA polymerase sigma factor (cog, prk,tigfam, pfam)
PA14_18870	+	hypothetical



Dataset S3-2. Gene analysis for RNA-seq data for 200  $\mu$ M Fe(II) shocked late stationary phase cells.

locusID	Gene name	log2 fold change WT after 30 min Fe(II) shock	log2 fold change $\Delta bqsR$ after 30 min Fe(II) shock	FDR	1st - 2nd column (WT vs $\Delta bqsR$ )	Predicted function
PA14_04180		11.61	-0.90	5.68E-08	12.51	Bacterial OB-Fold protein (pfam)
PA14_29710	<i>bqsP</i>	9.03	-0.14	6.68E-12	9.17	Predicted membrane protein (COG)
PA14_04270		8.55	-0.28	1.04E-11	8.84	transcription factor (PseudoCAP Function Class)
PA14_04190		8.43	-0.27	1.78E-12	8.70	acytyl polyamine deacetylase (COG)
PA14_29720	<i>bqsQ</i>	8.61	0.06	1.78E-12	8.54	Predicted membrane protein (COG)
PA14_29730	<i>bqsR</i>	7.79	-0.29	6.83E-12	8.08	Fe(II) responsive response regulator
PA14_29740	<i>bqsS</i>	6.99	0.05	1.78E-12	6.94	Fe(II) responsive sensor histidine kinase
PA14_04290		6.74	-0.19	3.58E-12	6.92	arginine-specific autotransporter (doi: 10.1371/journal.ppat.1002854)
PA14_07070		7.09	0.65	1.05E-11	6.44	Predicted ferric reductase (cog)
PA14_04210		5.38	-0.50	2.75E-11	5.88	amino acid transporter, potE homologue (polyamine transporter) (cog)
PA14_29750	<i>bqsT</i>	5.95	0.20	7.42E-10	5.75	hypothetical
PA14_52340		5.31	-0.41	1.61E-06	5.71	hypothetical

PA14_63110		4.78	-0.18	2.07E-08	4.96	S-adenosylmethionine decarboxylase (cog, pfam, tigrfam)
PA14_63120		3.86	-0.16	1.88E-09	4.02	Spermidine synthase (cog, pfam, tigrfam)
PA14_01240		2.97	-1.04	9.80E-09	4.01	Zn-binding carbonic anhydrase (doi: 10.1099/mic.0.066357-0)
PA14_01250		2.44	-0.41	2.66E-08	2.85	sulfate transporter (cog, pfam)
PA14_57880		3.08	0.31	1.62E-09	2.77	ABC-type transport system involved in resistance to organic solvents, ATPase component (cog)
PA14_57870		2.59	0.28	1.67E-08	2.32	ABC-type transport system involved in resistance to organic solvents, permease component (cog)
PA14_63150	<i>pmrA</i>	2.19	-0.08	7.34E-08	2.26	<i>pmrA</i> two-component regulator
PA14_63130		1.97	-0.13	8.73E-07	2.10	hypothetical
PA14_57840		2.41	0.32	1.09E-04	2.09	Anti-anti-sigma regulatory factor (antagonist of anti-sigma factor) (cog)
PA14_49100		1.40	-0.59	3.11E-06	1.99	Glutathione S-transferase (COG)
PA14_57850		2.33	0.47	2.33E-06	1.86	ABC-type transport system involved in resistance to organic solvents, periplasmic component (cog)
PA14_52330		1.82	0.02	1.16E-04	1.80	hypothetical

PA14_57830		2.18	0.43	1.17E-05	1.75	ABC-type transport system involved in resistance to organic solvents, auxiliary component (cog)
PA14_49180	<i>phoP</i>	1.37	-0.37	1.59E-06	1.74	two-component system
PA14sr_013		1.94	0.22	4.62E-06	1.72	regulatory RNA
PA14_63160	<i>pmrB</i>	1.48	-0.17	1.34E-07	1.65	<i>pmrB</i> two-component sensor
PA14_03710		3.47	1.86	4.52E-04	1.61	hypothetical
PA14_28070		3.20	1.62	5.23E-03	1.59	hypothetical
PA14_07000	<i>dsbB</i>	1.42	-0.12	4.56E-07	1.54	Disulfide bond formation protein DsbB (cog)
PA14_29760		2.27	0.73	1.95E-07	1.54	Methyl-accepting chemotaxis protein (cog, pfam, SMART)
PA14_46900		1.44	-0.09	6.09E-04	1.53	hypothetical
PA14_19970		-2.99	-4.45	1.88E-06	1.46	hypothetical
PA14_49200	<i>oprH</i>	1.16	-0.26	1.53E-05	1.42	outer membrane porin stabilizes outer membrane by LPS binding and increases antibiotic resistance (doi: 10.1074/jbc.M111.280933)
PA14_70860		1.43	0.06	1.73E-06	1.37	ABC-type phosphate transport system, periplasmic component (cog)
PA14_37830		3.60	2.31	3.17E-04	1.30	cysteine desulfurase (cog, tigrfam, prk)
PA14_13010		3.20	1.95	6.41E-05	1.25	ABC-type metal ion transport system, periplasmic component/surface antigen (COG)
PA14_03700		3.19	2.00	6.15E-04	1.19	ABC-type sulfate

PA14_04220	1.09	-0.04	3.35E-04	1.13	transport system, periplasmic component (COG) Spermidine/putresci ne-binding periplasmic protein (COG)
PA14sr_109	1.18	0.08	2.45E-05	1.10	regulatory RNA
PA14_50740	1.10	0.04	2.59E-06	1.07	hypothetical
PA14_34740	2.80	1.76	2.34E-03	1.05	hypothetical
P35	1.52	0.48	1.64E-05	1.05	Non-coding RNA gene
PA14_57720	2.91	1.90	3.13E-04	1.01	sulfate adenylyltransferase subunit 2 (PRK)

Dataset S3. Transcription unit analysis for RNA-seq data for 200  $\mu$ M Fe(II) shocked late stationary phase cells.

region	log2 fold change wt after 30min Fe(II) shock	log2 fold change $\Delta$ bqsR after 30min Fe(II) shock	1 <sup>st</sup> -2 <sup>nd</sup> column (WT vs $\Delta$ bqsR)	FDR	Predicted function
regionPA14_04180	11.62	-0.91	12.53	1.36E-08	Bacterial OB-Fold protein (pfam)
regionPA14_04270	8.54	-0.29	8.83	8.23E-12	transcription factor (PseudoCAP Function Class)
regionPA14_29710 -PA14_29740	8.27	-0.03	8.30	4.74E-13	bqs operon; Two-component system (PseudoCAP Function Class)
regionPA14_04190  PA14_04210	7.44	-0.41	7.85	1.43E-12	deacetylase (cog); amino acid transporter, potE homologue (polyamine transporter) (cog)
regionPA14_04290	6.75	-0.19	6.94	4.53E-12	arginine-specific autotransporter (doi: 10.1371/journal.ppat.1002854)
regionPA14_07070	7.10	0.65	6.46	8.57E-12	Predicted ferric reductase (cog)
regionPA14_29750	5.96	0.19	5.77	1.11E-09	hypothetical
regionPA14_52340	5.32	-0.41	5.73	2.77E-06	hypothetical
regionPA14_01240	2.98	-1.04	4.02	1.07E-08	Zn-binding carbonic anhydrase (doi: 10.1099/mic.0.066357-0)
regionPA14_63110 -PA14_63160	2.99	-0.15	3.14	2.57E-09	S-adenosylmethionine decarboxylase (cog, pfam, tigrfam); Spermidine synthase (cog, pfam, tigrfam); hypothetical; pmrA; pmrB
regionPA14_01250	2.45	-0.42	2.87	2.75E-08	sulfate transporter (cog, pfam)

regionPA14_57850 -PA14_57880	2.82	0.33	2.49	7.17E-09	57830/57840 in operon (DOOR operon predictor) ABC-type transport system involved in resistance to organic solvents, periplasmic component (cog); ABC-type transport system involved in resistance to organic solvents, permease component (cog); ABC-type transport system involved in resistance to organic solvents, ATPase component (cog)
regionPA14_57830  PA14_57840	2.35	0.34	2.01	2.22E-05	ABC-type transport system involved in resistance to organic solvents, auxiliary component (cog); Anti-anti-sigma regulatory factor (antagonist of anti-sigma factor) (cog)
regionPA14_52330	1.83	0.02	1.81	1.12E-04	hypothetical
regionPA14_03710	3.48	1.85	1.63	5.11E-04	hypothetical
regionPA14_28070	3.21	1.61	1.60	6.11E-03	FOG: GAF domain (cog)
regionPA14_07000	1.43	-0.13	1.56	4.23E-07	Disulfide bond formation protein DsbB (cog)
regionPA14_29760	2.28	0.73	1.55	1.98E-07	Methyl-accepting chemotaxis protein (cog, pfam, SMART)
regionPA14_49170  PA14_49180	1.17	-0.38	1.55	2.41E-06	PhoP/PhoQ (predicted to be part of the oprH operon)
regionPA14_46900	1.45	-0.10	1.55	5.24E-04	hypothetical
regionPA14_19970	-2.98	-4.45	1.48	2.10E-06	hypothetical
regionPA14_49200	1.17	-0.26	1.44	1.15E-05	oprH- outer membrane porin stabilizes outer membrane by LPS binding and increases antibiotic resistance (doi: 10.1074/jbc.M111.280933)

regionPA14_70860	1.44	0.05	1.39	1.65E-06	ABC-type phosphate transport system, periplasmic component (cog)
regionPA14_37830	3.60	2.30	1.30	3.67E-04	cysteine desulfurase (cog, tigrfam, prk)
regionPA14_50740	1.12	0.03	1.09	1.58E-06	Outer membrane lipoprotein (cog)
regionPA14_17930	2.40	1.39	1.01	6.80E-06	glycerol-3-phosphate dehydrogenase (cog, tigrfam)
regionPA14_63470	0.91	-0.09	1.00	2.14E-04	S-adenosylmethionine-dependent methyltransferases (CD), Methyltransferase domain (pfam)
regionPA14_39590	0.31	1.73	-1.42	4.15E-04	metE, cobalamine independent methione synthetase (PMID: 1339288 )
regionPA14_39570  PA14_39580	-0.63	1.02	-1.66	8.30E-03	rimJ Acetyltransferase, including N-acetylases of ribosomal proteins (cog), This enzyme acetylates the N-terminal alanine of ribosomal protein S5. Plays also a role in the temperature regulation of pap pilin transcription (uniprot) ; hypothetical

**Chapter IV: Broader regulatory context outweighs operator occupation in predicting  
genome-wide regulation in bacteria**

Naomi Kreamer, Rob Phillips, Dianne Newman, and James Boedicker

The following is modified from a submitted manuscript.



**Abstract**

Predicting global gene expression patterns directly from genome sequence remains a daunting challenge. Quantitative models that could parlay genomic information into predicted global gene expression levels would enable us to take full advantage of the increased availability of whole genome sequences. Here, we examine the relative contributions of an individual regulatory input, the ferrous iron-responsive regulatory element, BqsR, on global patterns of gene expression in the bacterium *Pseudomonas aeruginosa*. The position weight matrix (PWM) derived for BqsR uncovered thousands of likely binding sites throughout the genome. Only a subset of these potential binding sites had a regulatory consequence, suggesting that BqsR/DNA interactions were not captured within the PWM or that the broader regulatory context at each promoter played a greater role in setting promoter outputs. The architecture of the BqsR operator was systematically varied to understand how the sequence, length, and orientation of the binding site influences expression. We found that BqsR operator affinity was predicted by the PWM well. However, for most promoter containing BqsR operators, the surrounding regulatory context and the specific combination and arrangement of the transcription factors at each promoter were more influential in setting promoter outputs than the binding energy of individual BqsR binding sites. These results indicate that models derived from PWMs can accurately describe operator affinity, but more comprehensive models that include local regulatory contexts are needed to develop a predictive understanding of global regulatory outputs.

## Introduction

It is well appreciated that the rate of generation of new genome sequencing data is far outpacing our ability to make sense of it. For example, although considerable progress has been made in recent years to understand the roles of noncoding genomic regions (1, 2), our ability to apply this understanding in a predictive fashion is still quite limited. More fundamentally, it is not clear to what extent we can accurately predict genome-wide regulatory outputs from any DNA nucleotide sequence, though many complementary approaches are being applied to decipher the regulatory logic in the genomes of both eukaryotes and bacteria (3, 4). Here, we put our understanding of how an individual bacterial transcription factor influences global gene expression to the test to explore the extent to which the binding sequence and the surrounding regulatory context tune promoter outputs.

Making precise predictions of genome-wide expression is hampered by an incomplete understanding of how regulatory information is encoded at different promoter regions (3, 5). We cannot accurately predict *a priori* what influence on gene expression a particular transcription factor will have at a given promoter. For the majority of transcription factors we only know where these factors are likely to bind and whether the transcription factor acts as an activator or repressor. Current models typically cannot quantitatively determine the magnitude of expression, nor predict how expression is modulated by the number of regulatory proteins per cell or the promoter architecture (defined as the sequence, orientation, location, and number of transcription factor

binding sites, and their proximity to the RNA polymerase binding site). Even for well-characterized regulators, we are often surprised by experimental results contradicting expected trends (6-8). A predictive understanding of how regulatory information is encoded in the genome would lead to more meaningful comparisons between genomes of related organisms, enhance our understanding of regulatory-genome changes associated with niche differentiation, and improve our ability to design regulatory networks.

We chose the ferrous iron [Fe(II)] responsive two-component system, BqsRS, from *Pseudomonas aeruginosa* (9) to test our ability to quantitatively predict the regulatory output for a given transcription factor. Most organisms require iron for essential cellular processes, including electron transfer steps in metabolism, but iron is a limiting nutrient in many environments (10). Because iron uptake and localization are critical for growth and function, the cellular response to iron is complex and tightly regulated. In the body, our immune system controls pathogen proliferation in part by sequestering iron through high-affinity ferric iron [Fe(III)]-binding molecules such as lactoferrin (11). In many environments, including the cystic fibrosis (CF) lung and soil, iron is present in both the Fe(II) and Fe(III) forms (12, 13). Furthermore, elevated Fe(II) in CF sputum correlates with severe disease states (12). BqsRS consists of a sensor histidine kinase (BqsS) and a response regulator (BqsR) that responds specifically to Fe(II) at low micromolar concentrations in the periplasmic space (9). BqsRS is known to be involved in rhamnolipid production and biofilm dispersal (14) and it mediates cellular defenses against cationic

stressors, including aminoglycoside and polymyxin antibiotics (15). The consensus BqsR DNA binding motif (15) is found in promoter regions of genes that are upregulated by BqsR, but it remains unclear whether BqsR binding alone is sufficient to elicit a regulatory response, or if the impact of BqsR depends on the context at each promoter.

One common model used to predict the influence of transcription factors on global gene expression is based on position weight matrices (PWM) (16, 17). Here, we used the previously derived PWM for the Fe(II) responsive transcription factor BqsR (15) to search for potential binding sites throughout the genome. Thousands of potential binding sites were found; RNA-Seq expression measurements revealed that the majority of genes containing potential BqsR operators were not strongly regulated by BqsR. This discrepancy suggests that either the PWM of BqsR does not accurately describe the interaction between BqsR and the genome, or that the surrounding operator context has a greater influence on promoter outputs than the local BqsR binding sequence. Through systematic measurements of how the architecture of BqsR containing promoters influences gene expression, we address how BqsR regulatory information is encoded in the genome and to what extent understanding the regulatory context is critical to develop a predictive understanding of the global effect of an individual transcription factor.

## **Results**

### *Operator sequence diversity throughout the genome*

To construct a quantitative model capable of predicting the magnitude of gene expression directly for any given promoter region, we first dissected how variability in the

BqsR operator modulates gene expression. BqsR is an activator that binds upstream of the gene transcription start site (Fig. 4-1A). Earlier work established that the BqsR binding sequence contained a pair of highly conserved pentamers (Fig. 4-1A) (15). Many potential BqsR operators were found within promoter regions throughout the genome (Fig. 4-1B). For this study, we defined the promoter region as 1000 bp upstream from the protein coding sequence. There were 330 operators with 2 or fewer mutations in the pair of consensus repeated pentamers (TTAAG(N)<sub>6</sub>TTAAG), and over 4000 potential operators containing 3 mutations. This frequency of binding sites is not expected to occur by chance in the genome (Supplemental Fig. 4-S1). Given the large quantity of close matches to the 10 most conserved bp (2 repeats of the pentamer), it is not obvious at which of these operators BqsR binds and has a regulatory consequence.

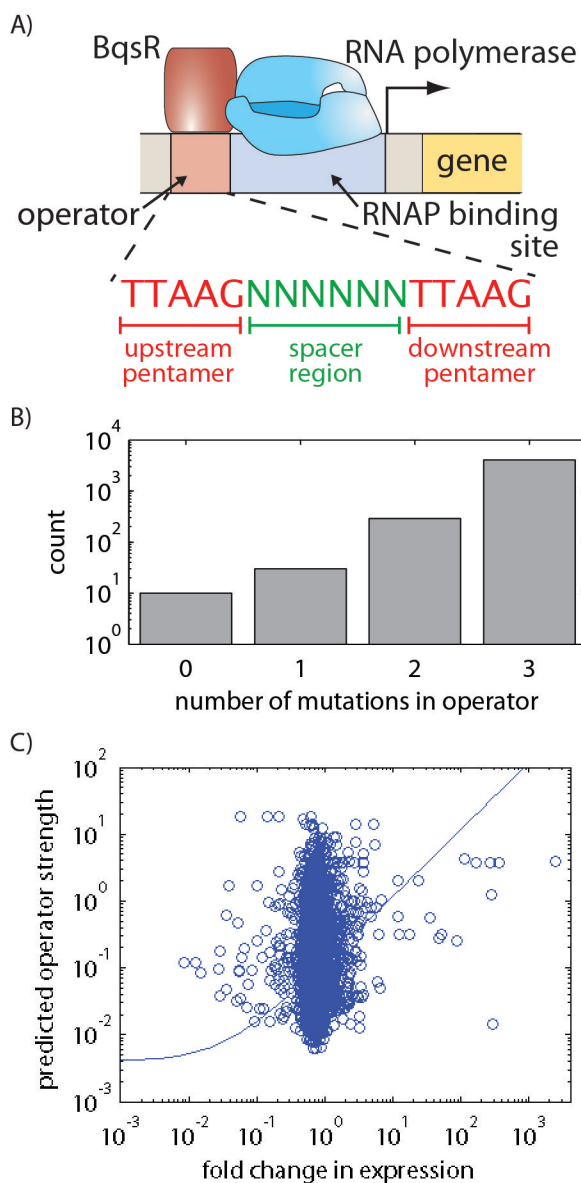


Figure 4-1: The BqsR binding motif in the genome and its impact on global expression. A) The response regulator BqsR activates gene expression. The BqsR operator sequence contains a repeated pentamer (TTAAG) separated by 6 bp. B) The number of potential BqsR operators in the genome of *P. aeruginosa* PA14 sorted by the number of mutations in the repeated pentamers. C) A comparison of the predicted operator strength with the observed experimental fold change in expression. The curve shows the best-fit linear relationship between the experimentally measured fold change in gene regulation due to iron shock and the predicted operator strength. For operator regions containing multiple BqsR binding sites, only the site with the highest score was plotted.

The magnitude of BqsR-mediated gene expression for each of these potential operators can be predicted using a PWM for the operator binding site. A PWM uses a set of promoter sequences to generate a DNA sequence motif reflecting the nucleotide frequency for each position in the sequence (16, 17). The PWM can then be used to rank order the regulatory strength of each potential operator. The operator strength is calculated using

$$\text{operator strength} = e^{-S}, \quad (\text{Equation 1})$$

in which  $S$  is the operator score calculated using the PWM (Supplemental Material for further details). Operator strength is assumed to be proportional to the affinity of the operator for BqsR. In Figure 4-1C, the scores were calculated using the PWM derived in (15). Here, scores are normalized to the score of the downstream BqsR binding site in the PA14\_04180 promoter region, a reference operator whose sequence is listed in Figure 4-2A.

Figure 4-1C compares the measured fold change in gene expression due to the BqsR operator strength calculated using Equation 1, revealing a poor correlation between the PWM predictions and the experimental expression. Typically it is assumed that the rate of gene expression at a given promoter is proportional to the affinity for the transcription factor to the operator, known as the occupancy hypothesis (6). The fact that several binding sites with high PWM scores were not induced by an Fe(II) shock raises several questions: how is regulator-binding specificity achieved; how does operator

sequence modulate promoter outputs; and to what extent does the surrounding promoter context influence BqsR-mediated regulatory responses?

*Influence of the pentamer sequence on BqsR-mediated regulation*

One potential cause of disagreement between model predictions and experimental measurements was that PWM did not accurately capture the relationship between operator sequence and binding affinity. To determine whether the PWM was missing key regulatory information, we experimentally dissected how the structure of the operator (*i.e.*, the sequence, length, and position) affected the level of gene expression. Although we analyzed nucleotide frequency in the genome to inform our experiments, direct experimental measurements were used to construct our model. First we looked at the operator diversity throughout the genome for clues as to which operator variations might impact BqsR-mediated expression. For this analysis FIMO (Find Individual Motif Occurrences), part of the MEME Suite (18), was used to identify potential BqsR binding sites in the genome. These potential operator sequences were further characterized by comparing the sequence, location, and orientation of the binding sites. The operator was split into three regions: the upstream and downstream pentamers and the spacer region (Fig. 4-1A). A small library of synthetic promoters was fused to the *lacZ* reporter gene and inserted into the genome at the *glmS* locus to quantify the influence of specific changes in promoter architecture on expression output. The synthetic promoter library was based on the BqsR binding sequence in the promoter for gene PA14\_04180, the gene most highly upregulated by BqsR.



Previously, two repeated pentamers were found to be highly conserved and sufficient for BqsR binding (15). The frequency and identity of mutations in the pentamer regions of potential BqsR binding sites were examined (Fig. 4-2B). The PWM of the pentamer regions was calculated using the 330 operator sequences from Figure 4-1B containing up to 2 mutations in the pentamer regions (Fig. 4-2B). The PWM score indicates the frequency of a given bp at each position in the sequence compared to the background distribution of bp in the genome [Supplemental Equations S1 and S2 and (16, 17)]. Bases with a high score indicate a particular base being favored at a given position, and bases with negative scores indicate bases that are rare at a given position. There is a non-uniform distribution of the scores, implying that some bases contribute more than others to the binding energy of BqsR to the operator.

A library of synthetic promoters based on a modified PA14\_04180 promoter was constructed to test the influence of the pentamer sequence on gene expression. Only the upstream pentamer was mutated because the downstream pentamer overlaps the -35 region of the RNA polymerase (RNAP) binding site (15). Additionally, the symmetry in the PWM scores in Figure 4-2B suggests the two pentamer sequences may interact similarly with BqsR. The reference DNA sequence of the BqsR operator is shown in Figure 4-2A. The library contains all possible single point mutations for the upstream pentamer sequence, 15 constructs in total. In Figure 4-2C, we report the gene expression level after Fe(II) shock for each mutant relative to the expression level of the wild-type promoter containing the binding site shown in Figure 4-2A. Expression analysis of the promoter

library revealed heterogeneity in the contribution of each position to the expression level. For example, the bases in position 5 have a weak influence on regulation, whereas mutating position 1 from T to C reduced expression nearly 10 fold. The large decrease in expression of this particular mutation may explain why a C in position 1 was rare in the potential binding sites found in the genome (Fig. 4-2B).

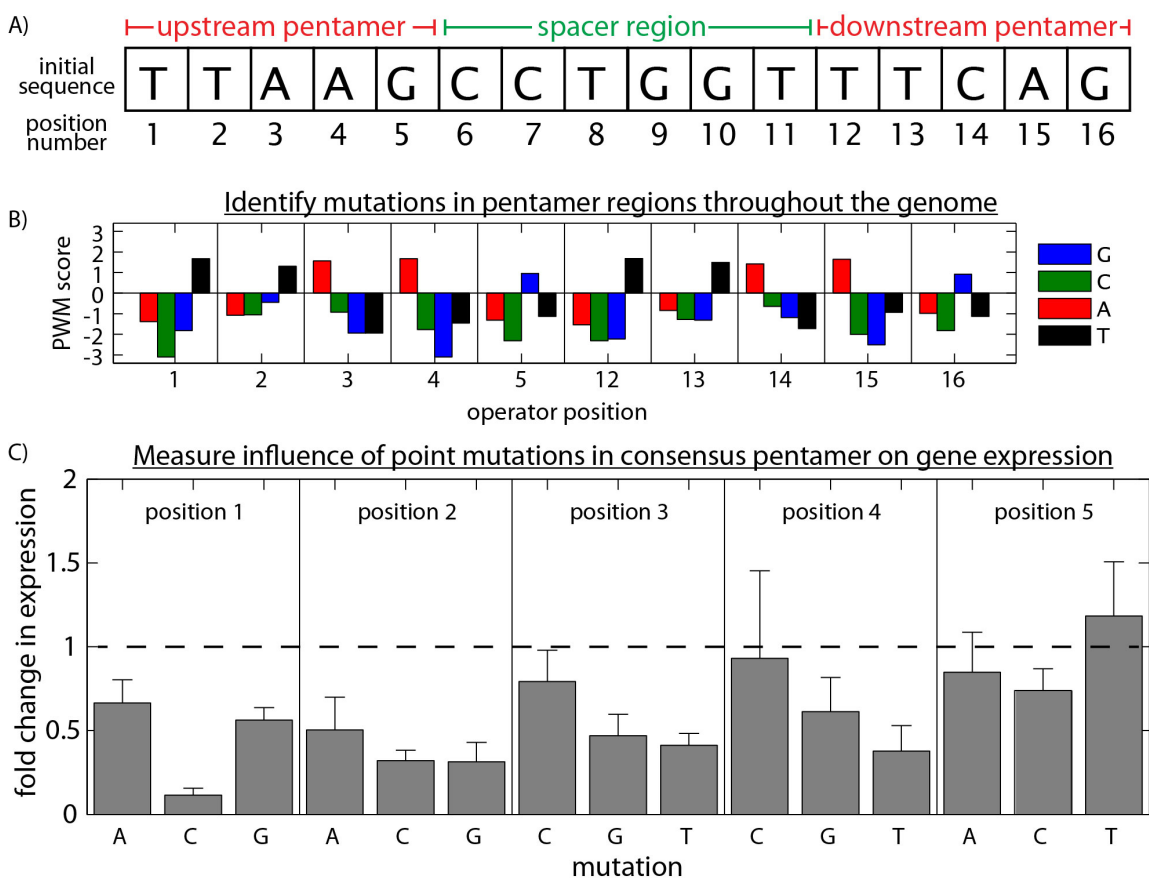


Figure 4-2: Variability of the BqsR operator's repeated pentamer sequence and its influence on gene regulation. A) The reference operator found in the PA14\_04180 promoter region contains two pentamers separated by a 6 bp spacer region. B) The graph shows the PWM for the upstream and downstream pentamer sequences. PWM scores were calculated using the operators from Figure 4-1 containing between 0 and 2 mutations in the pentamer regions. C) Expression measurements of synthetic constructs quantified the influence of each point mutation for all nucleotides in the upstream pentamer on gene expression. Error bars show standard error of biological triplicates.

*Influence of the spacer sequence on BqsR-dependent regulation*

In addition to the pentamer region, we also determined the natural variability and influence of the spacer region between the two pentamers on gene regulation. First, we examined the variability of spacer region lengths throughout the genome. All previously reported operator sequences had a spacer length of 6 bp, but it was unclear if spacer length modulated BqsR binding. Analyzing all the potential operators in the genome with spacer lengths of 5, 6, or 7 pairs shows that all the pentamer pairs with no mutations had 6 bp spacer regions, and that a spacer length of 6 bp was most common for operators with a single pentamer mutation (Fig. 4-3A). Indeed, spacer length is a key operator parameter. Spacers of length 5 or 7 lowered gene expression to less than 10%, an expression level similar to an operator for which the upstream pentamer has been deleted (control in Fig. 4-3B). Together these results suggest that a 6 bp spacer region is essential for operator binding. We also examined expression in operators where the orientation of the upstream operator was changed. In iron responsive regulatory elements such as Fur, repeated hexamers are known to appear in a variety of orientations such as tandem repeats and palindromic orientations (19). For BqsR, it was found that the reverse or reverse complement of the upstream pentamer reduced expression to background levels (Supplemental Fig. 4-S2).

Because the variation in the sequence of the spacer region differed from the background nucleotide frequency in *P. aeruginosa*, the sequence of the spacer region may have contributed to setting the overall expression level. Synthetic promoters were

created to quantify the influence of the spacer region on gene regulation. Because the number of potential spacer sequences was large ( $4^6 = 4096$ ), targeted changes were made to the spacer region based on analysis of the spacer regions from pentamers analyzed in Figure 4-2B. The spacer PWM scores shows that some bases occur frequently at specific positions, such as a C at position 11 (Fig. 4-3).

We constructed spacer regions with single mutations from the wild type sequence, position 11 T to C and position 10 G to A, which Figure 4-3C suggested would be strongly beneficial and weakly detrimental to expression, respectively. We also constructed two operators containing arbitrary spacer sequences, labeled random in Figure 4-3D, and two more containing more common or less common spacer sequences based on the results of Figure 4-3C. Figure 4-3D shows measurements of gene expression from these constructs, revealing the spacer region sequence has the potential to moderately influence expression levels. For the six sequences measured, gene expression levels were found to vary up to a factor of 2.

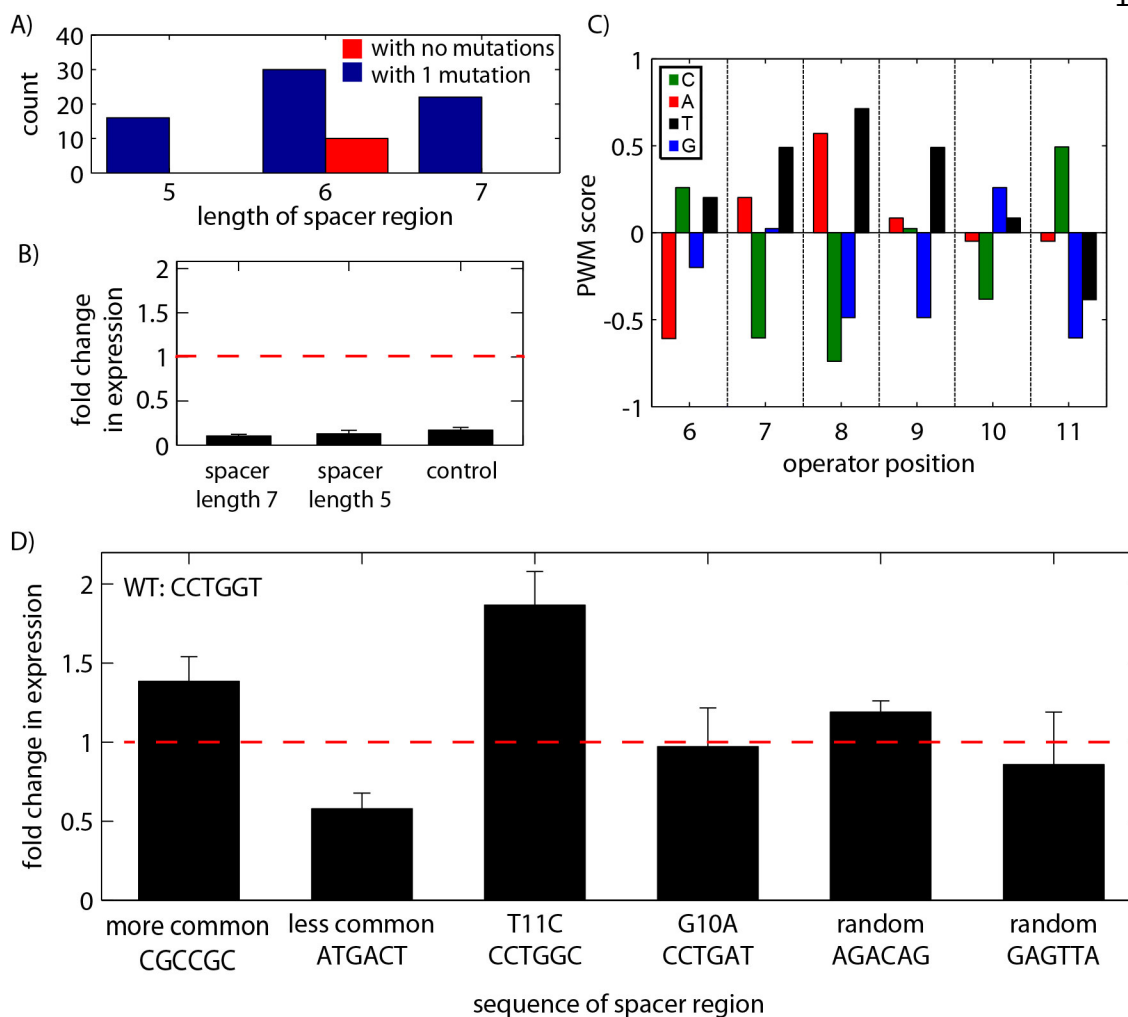


Figure 4-3: Variability of the spacer sequence and its influence on gene regulation. A) The number of operators containing spacers of length 5, 6, or 7 bp is plotted for potential operators containing either 0 or 1 mutations in the pentamer regions, revealing a preference for spacer regions of length 6 bp. B) Experimental measurements of synthetic operators confirm that a 6-bp spacer length is critical for BqsR-mediated regulation, as spacer lengths of 5 or 7 bp resulted in expression levels similar to the negative control in which the upstream pentamer was deleted. C) Operators from (A) containing a 6 bp spacer region were analyzed for sequence preference by calculating the PWM score. D) Gene expression measurements of 6 synthetic spacer sequences showed that the sequence of the spacer region modulates the level of gene regulation up to a factor of 2.

### *Binding site clusters*

Yet another aspect of promoter architecture is the number of binding sites in promoter region. We searched the genome for operator clusters with a 6 bp spacer region between pentamers (Fig. 4-3C) and only considered operators with a maximum of 3 pentamer mutations. The genome contained hundreds of binding sites with 2-5 copies of BqsR operators and promoters with up to 18 clustered potential operators (Fig. 4-4A). Supplemental Fig. 4-S3 shows that the quantity of large binding site clusters cannot be accounted for by a random distribution of binding sites. Some promoters contained overlapping operators that share a common pentamer sequence, creating multiple repeats of the pentamer sequence separated by a 6 bp spacer region with up to 6 repeated pentamers (Fig. 4-4B).

Because many promoter regions contain repeats of the BqsR operator sequence, we next experimentally dissected the role of operator clusters in the regulatory output. The native promoter region for gene PA14\_04180 contains 4 overlapping pentamers as shown in Fig. 4-4B (the truncated PA14\_04180 promoter used in Fig. 4-2 contained only the 2 downstream pentamers). The most downstream pentamer, labeled 4 in Figure 4-4B, is unique given that it overlaps the -35, RNAP binding site in the promoter. We created several synthetic promoters that lacked one or more pentamers, all of which retained pentamer 4 given its dual role in both BqsR and RNAP binding (Fig. 4-4D). A pentamer was deleted by mutating the pentamer sequence from TTAAG to ACTCA. Deleting pentamer 2 or both pentamers 1 and 2 resulted in a 50% decrease in expression

level. Deleting either 3 or 1 and 3 resulted in approximately 95% reduction in expression. This suggested BqsR binding to pentamers 3 and 4 was critical for strong expression. The supplemental binding sites 1 and 2 led to an additive increase in promoter output, but were not essential for expression. From these results, we inferred that for other repeated binding sites one pair of pentamers is sufficient to elicit a BqsR response, with additional operators further increasing expression.

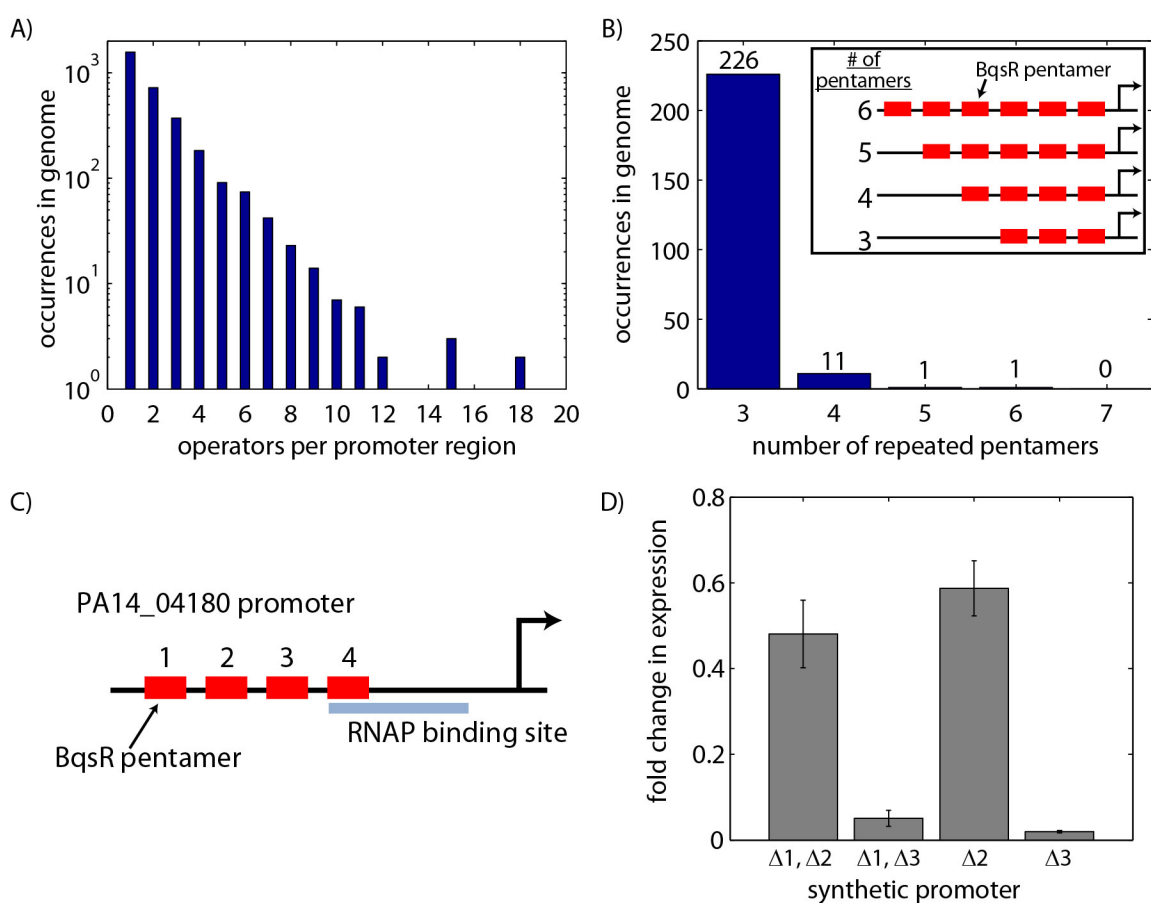


Figure 4-4: Promoters regions with multiple operators. (A) The number of operators within each promoter region of the genome. Operators are sequences 16 bp in length and contain up to 3 mutations in the pentamer regions. (B) Several of these clustered promoters are spaced by 6 bp, the same spacing between the two pentamer regions of an individual promoter, creating arrays of overlapping binding sites as shown in the schematic. (C) The promoter for PA14\_04180 is one such promoter containing 4 repeated

pentamer regions. (D) Gene expression measurements on synthetic versions of the PA14\_04180 promoter, in which individual or pairs of pentamer repeats were removed, revealed that each pentamer repeat contributes to the overall level of gene regulation, although not all the BqsR binding sites contribute equally to the regulatory output.

#### *Deriving a binding energy matrix for BqsR*

From the experimental dissection of BqsR-mediated gene expression, we constructed a BqsR binding energy matrix with a 6 bp spacer region to predict gene regulation for the entire genome. We assumed an additive contribution of operator clusters, operator occupancy was linearly proportional to the change in expression level, and stronger BqsR binding increased transcription. The energy matrix translates the operator nucleotide sequence to binding affinity (Fig. 4-2C). Such binding energy matrices have been developed previously for other transcription factors (20, 21), and in the case of the Lac repressor in *E. coli* have been shown to accurately predict promoter outputs (22).

For the pentamer region, measurements of the fold change in expression from Figure 4-2C were converted to a change in binding energy using

$$DE = -k_B T \ln(\text{fold change}), \quad (\text{Equation 2})$$

where  $k_B$  was Boltzmann's constant,  $T$  was the temperature, and  $DE$  was the change in the binding energy of BqsR to the operator (see Supplemental section "Predicting gene expression from operator sequence" and Figs. 4-S4 and 4-S5 for further details). For the spacer region, a "best fit" binding energy matrix was fit to the expression data for the 7 variants of the spacer sequence measured in Figure 4-3D (see Supplemental Materials section "Deriving the energy matrix for the spacer region" and Fig. 4-S6).



The final 16-mer energy matrix reported the change in binding energy for any operator sequence relative to the initial sequence (Fig. 4-5B and Supplemental section “Predicting gene expression from operator sequence”). Using this matrix we predicted the expression of a given promoter relative to a promoter containing the reference operator (Fig. 4-5A). For comparison, Figure 4-5C showed the binding energy matrix derived from the PWM. The PWM method of constructing an energy matrix differed from ours in that we did not assume that the consensus sequence in the genome has the strongest binding energy. Although we analyzed potential operators throughout the genome for clues as to what operator parameters might influence regulation, ultimately, our matrix was based on direct, quantitative experimental measurements of gene expression from a library of synthetic operators containing *lacZ*-promoter fusions.

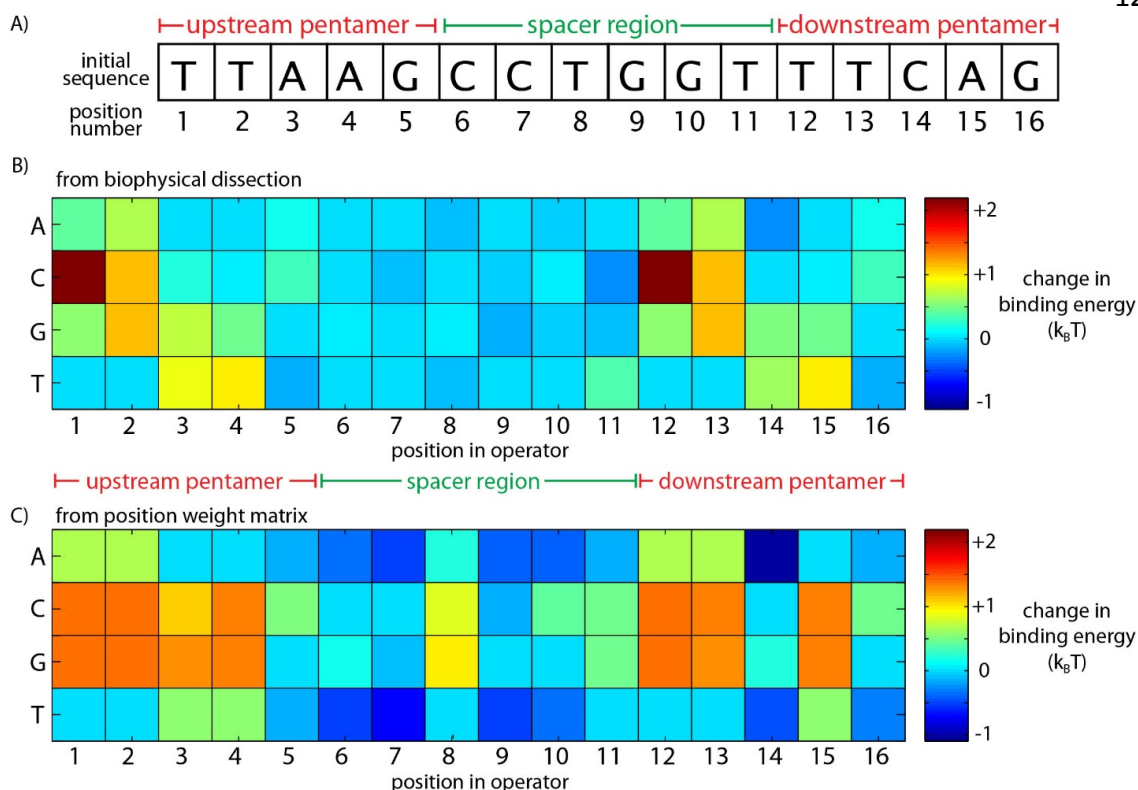


Figure 4-5: Creating a binding energy matrix for the BqsR operator sequence. A) The reference operator sequence from gene PA14\_04180. B) Energy matrix derived from the experimental measurements of regulation by synthetic operators. C) For comparison, the energy matrix derived from the position weight matrix used for prediction shown in Fig. 4-1C, which was constructed using qPCR expression data for 12 genes. Binding energies in  $k_B T$  units are relative to the binding strength to the reference operator in (A). Matrix positions in dark red indicate bp changes that cause large reductions in expression levels, whereas matrix positions in dark blue indicate bp that increase expression levels.

#### *Global prediction of BqsR-mediated gene regulation*

With these operator rules and the experimentally derived energy matrix (Fig. 4-5B), we made new predictions for BqsR-mediated gene regulation to test if the PWM missed key regulatory information encoded in the primary DNA sequence. If sequence information was missing from the PWM approach, predictions derived from Fig. 4-5B will significantly improve our predictive capability. The fold change in gene expression due to

each potential operator in the genome was calculated using Equation 2 (see Supplemental Fig. 4-S7 and 4-S8 for the distribution of predicted operator affinities). For promoter regions containing multiple BqsR binding sites, we assumed each operator acted independently and the total fold change was the sum of the fold change for each individual operator (see Supplemental section “Additive approximation for multi-operator promoters” for more details). Since it is known that the position of the operator relative to the transcription start site plays a role in regulatory outputs, we only included potential operators up to 1000 bp upstream from the protein coding sequence. Additionally, genes were predicted to be BqsR-regulated if they contain a maximum of 3 mutations in the 10 base pairs encoding the upstream and downstream pentamers, resulting in 103 predicted BqsR regulated transcriptional units.

These predictions were compared to RNA-Seq measurements of global BqsR-mediated gene expression reported previously (15). As a conservative annotation of BqsR-regulated genes we compare transcriptional units [regions which may contain several genes that are co-transcribed as defined by Wurtzel *et al.* (23)] whose expression was changed 2-fold or greater in WT compared to the  $\Delta bqsR$  mutant in response to Fe(II) shock. Figure 4-6A shows comparisons of the predicted fold change in expression for 103 transcriptional units, predicted using either our binding energy matrix (Fig. 4-5B) or the PWM derived binding energy matrix (Fig. 4-5C), to the experimentally measured fold change in expression. All predictions were normalized to the expression level of the reference gene (Fig. 4-2A). The predictions from our model and the model based on the

PWM were similar although not identical (Fig. 4-6B). Additionally, our model predictions were accurate for promoters giving a strong response, greater than 25 fold, to ferrous iron shock in experiments (Fig. 4-6C). These results suggested that the energy matrix accurately predicted operator occupancy. Our inability to predict the regulatory influence of BqsR at most promoters was not rooted in a misunderstanding of BqsR binding; instead, for most genes the surrounding regulatory context of each promoter was more important than operator occupancy in setting expression levels.

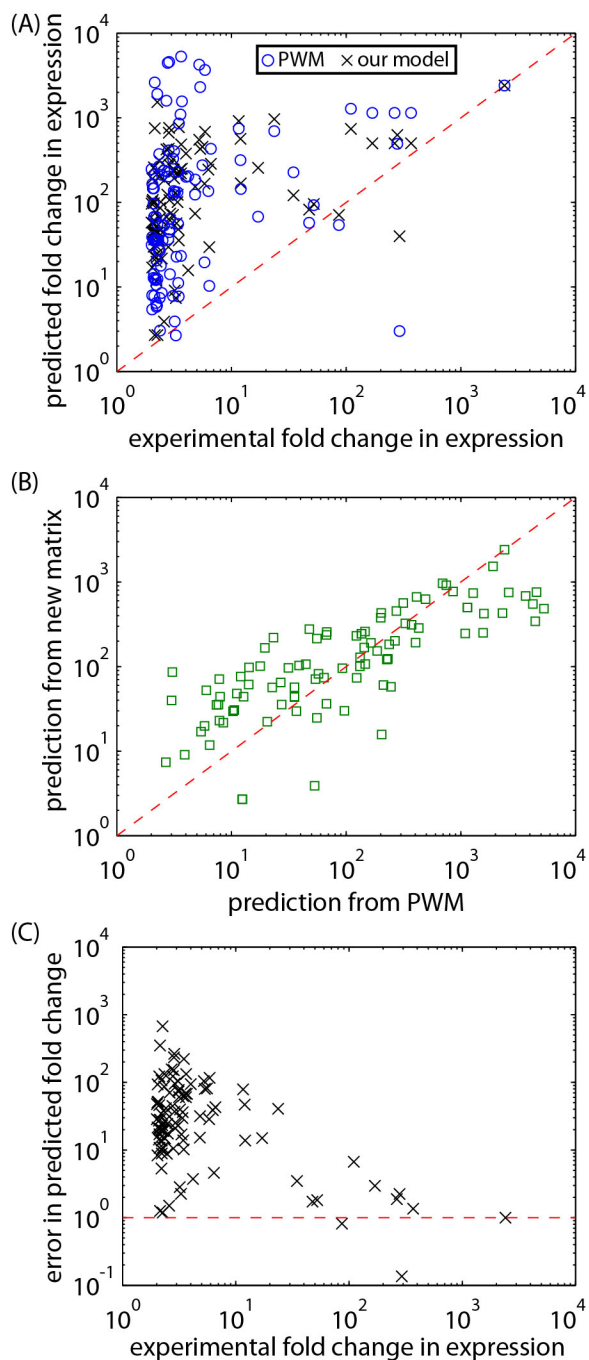


Figure 4-6: Comparison of global predictions of gene expression using the energy matrices in Figure 4-5 to gene expression measured using RNA-seq. A) Predicted vs. experimentally measured BqsR-mediated fold change in expression levels for 103 genes. Black x's show predicted expression based on the energy matrix shown in Figure 4-5B. Blue circles show predicted expression based on the PWM shown in Figure 4-5C. B) Both PWM model and our model made similar predictions of fold change in expression. C) Error in prediction,

defined as the ratio of predicted to experimentally measured fold change in gene expression, plotted as a function of measured fold change in expression.

*A closer look at genes that were poorly predicted*

Both energy matrices in Figure 4-5 poorly predicted BqsR-mediated regulation for the majority of the transcriptional units containing predicted BqsR operators. To explore why this might be, we examined the broader regulatory context of each promoter to determine whether the model overlooked key inputs into BqsR-mediated regulatory decisions. One parameter ignored in our initial predictions was the position of the BqsR operator relative to the gene transcription start site (TSS). The results shown in Figure 4-6 consider operator position relative to the protein coding sequence (CDS), only including BqsR binding sites within 1000 bp upstream of the protein coding sequence. However, the position of the BqsR binding site relative to TSS as opposed to the CDS was more relevant to transcriptional regulation. The location of a transcription factor binding site relative to the -35, RNAP binding site of the promoter influenced promoter output (6).

To gauge the influence of the spatial relationship between the TSS and the BqsR binding site on gene expression, the distance between the TSS and the BqsR operators was examined. Figure 4-7A showed the ratio of the predicted expression level to the experimentally measured expression level as a function of operator position. Only 75 of the 103 transcriptional units predicted in Figure 6 are plotted in Figure 4-7A, as the TSS for many transcriptional units could not be determined from the RNA-Seq data and were not previously reported (23). Operator positions were reported relative to the TSS. A ratio near 1 indicated an accurate prediction, with higher values signifying greater error.

Although it was interesting that most of the potential operators were found near the TSS, Figure 4-7A indicated predictability was not correlated with operator position.

Predictions for genes with operators located downstream of the TSS all had ratios of the predicted to experimentally measured change in gene expression greater than 45 (Fig. 4-7A), which may reflect an inability of BqsR to regulate gene expression from a downstream position. However, the sample size of 6 was too small to be conclusive.

In general we predicted a higher level of expression than experimentally observed, potentially caused by our assumption of additive contributions from multiple operators in the same promoter region. To analyze the impact of operator number on predictive ability, we examined the ratio of predicted to measured expression as a function of the number of putative operators in the promoter region (Fig. 4-7B). Each individual BqsR operator contained two pentamer regions separated by 6 bp with any number of nucleotides allowed between the 16-mer operators. The results reveal a trend of poor predictions for promoter regions containing multiple BqsR binding sites. For promoter regions containing up to three operators, predictive ability varied, but in general predictions were within 50 fold of measured values. For promoter regions containing 4 or more binding sites, predictions were typically higher than experimental measurements by a factor of 50 or more, with the poorest predictions occurring for the two promoters containing 8 and 9 operators. These results suggest that the assumption that operators behave independently may not be valid for promoter regions containing many operators, and that some operators in large clusters of binding sites may be nonfunctional. To

further analyze operator clusters we modeled a promoter region with two binding clusters. Considering the assumptions that operators act independently, the strongest binding operator dominates expression level, only the operator closest to the coding sequence changes has a regulatory consequence, and operator orientation relative to the direction of gene coding sequence did not significantly improve predictions (See Supplemental Figs. 4-S9 through 4-S12).

### **Role of additional transcription factors in regulation**

We next examined whether the poorly predicted transcriptional units had other known transcription factor binding sites in their promoter regions. To address this, we took a two-pronged approach. We used the PWMs reported for the 12 *P. aeruginosa* transcription factors annotated on the Prodoric database (24). However, due to our findings that PWMs vastly overpredict potential operators, we also compared our predicted regulon (defined as the genes under BqsR control) with 13 experimentally validated *P. aeruginosa* regulons (25-38).

Only two transcription factors, Anr (25) and PqsR regulons (32, 33), had experimentally determined regulons that had a statistically significant overlap with the transcriptional units predicted to be upregulated by BqsR but which were not upregulated in the RNA-Seq data (Fig. 4-7C “false positives” and Supplemental Table 4-S1). PqsR controls the *Pseudomonas* quorum sensing regulon that responds to PQS; Dong *et al.* (14) showed the concentration of PQS is reduced in the  $\Delta bqsR$  mutant. Cells were grown in anaerobic conditions to ensure Fe(II) remained stable, the condition where Anr is active.



Figure 4-7C “false negatives” also shows overlap of the RpoN (34) regulon with transcriptional units whose expression levels changed as a result of the iron shock, but were not predicted. RpoN encodes an alternate sigma factor induced in stationary phase (39) and under nitrogen limitation, and has been shown to influence quorum sensing regulation (40). The cells used in the RNA-Seq experiment were harvested in late stationary phase. All of the overlapping genes predicted showed upregulation by RpoN.

The PWMs from the Prodigic database were used to predict potential transcription factor binding sites throughout the *P. aeruginosa* PA14 genome. In an attempt to limit the error associated with PWM-based predictions, for each predicted transcription factor operator we assumed only the most probable 100 binding sites were capable of having a regulatory influence. Overall we identified potential secondary regulators in 36 of the 103 predicted transcriptional units (Fig. 4-7D). Some transcription factors occurred in several genes that contain BqsR binding sites, suggesting there could be significant crosstalk between these factors when making regulatory decisions. To gauge whether the presence of specific transcription factors led to poor predictions of gene expression, the average prediction to measurement ratio for all the genes containing each of the secondary transcription factor was calculated (Fig. 4-7D). Genes containing the transcription factors Fur, PsrA, and Vfr were all predicted well, despite the potential influence of these additional regulators. Expression measurements for genes containing Anr, AlgR, RcsB, RhIR, and ExsA in their promoter regions deviated most from predictions. Because the presence of a potential RhIR binding site in promoter regions caused the

most deviation between predictions and experimental measurements, the effect of

RhIR on BqsR-mediated gene expression was examined in further detail.

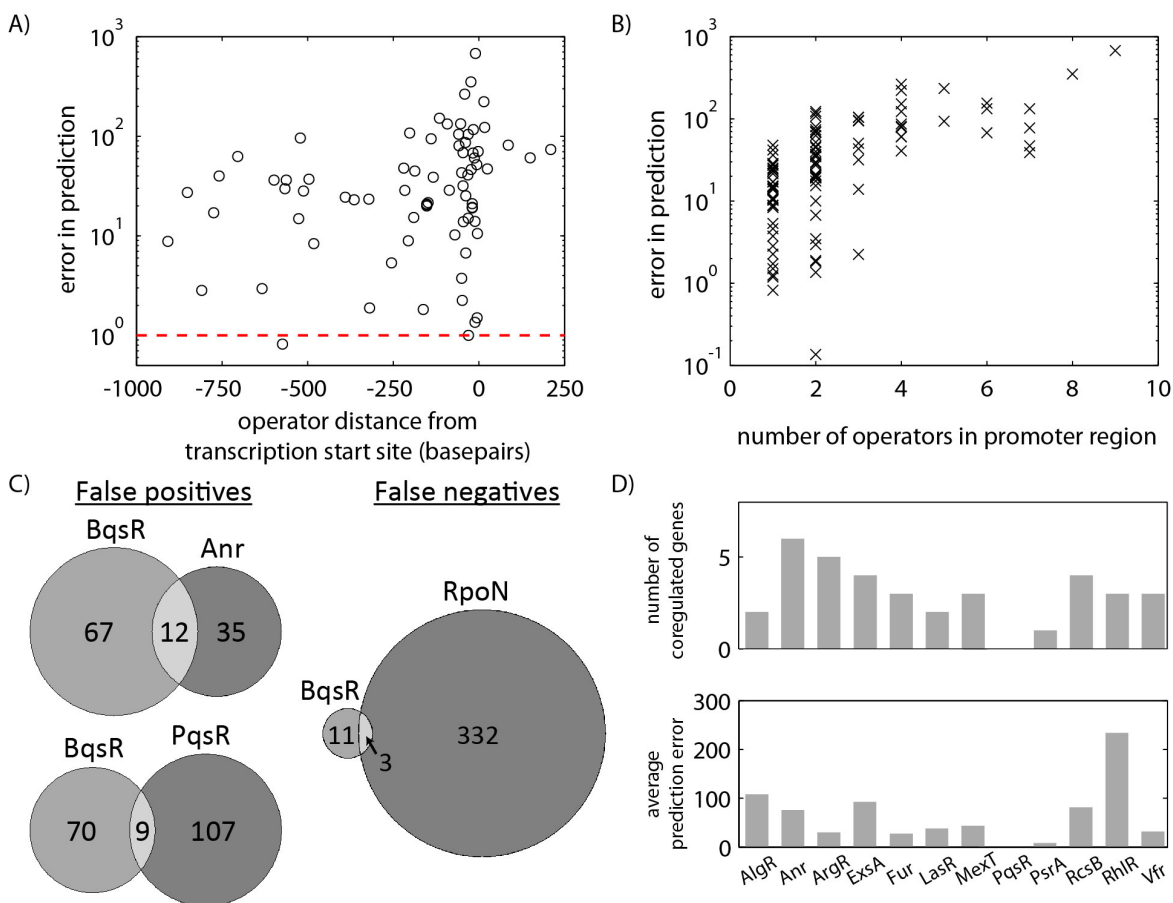


Figure 4-7: The influence of operator position and secondary transcription factors on the ability to predict BqsR-mediated gene regulation. A) For the genes predicted in Figure 4-6A, the error in prediction, defined as the ratio of predicted over experimentally measured expression levels, is shown as a function of operator positions relative to the transcription start site. A prediction error of 1 indicates the measurement exactly matched the prediction. B) Error in prediction plotted as a function of the number of BqsR binding sites found in the promoter region. C) Venn diagrams showing regulon overlaps between the predicted genes and selected transcriptional regulators with statistically significant overlap. See section “Comparison to other operons” in Supplemental Materials. D) Position weight matrices for 12 transcription factors were used to identify which of the predicted genes were likely to be coregulated by an additional transcription factor. The upper graph shows the number of promoter regions

containing a potential binding site for a second transcription factor. Bars on the bottom graph show the average prediction error for each set of genes containing a secondary transcription factor; the larger the ratio, the greater the error.

*Measuring gene expression from promoters coregulated by both BqsR and RhIR*

Prodoric was used to search promoter regions of BqsR-regulated genes for potential RhIR binding sites. Intriguingly, the RhIR binding motif overlapped with the BqsR binding site in many of the promoters. In the *bqs* promoter the RhIR binding site overlaps with the upstream BqsR pentamer and the downstream BqsR pentamer overlaps with the -35 RNAP binding site. In these overlapping promoters, because RhIR and BqsR cannot bind simultaneously, high levels of RhIR should competitively exclude BqsR and thus lower expression. Ordinarily the quorum sensing gene *rhIR* is upregulated in response to an autoinducer indicative of high cell density during stationary phase. RhIR can act as a repressor or an activator (41). An *rhIR* overexpressing strain was used to express RhIR in early exponential phase. To determine the effect of RhIR on *bqsR* expression, qPCR analysis was used. Figure 4-8b shows BqsR-dependent expression of genes predicted to have an RhIR binding site. For all genes, the RhIR overexpressing strain showed a statistically significant ( $p$ -value  $\leq 0.05$  by unpaired two-tailed t-test) decrease in expression compared to wildtype. For *oprH*, a gene containing a BqsR-responsive operator but not a putative RhIR binding site, expression did not significantly change when RhIR was overexpressed. RhIR significantly changes the regulatory influence of BqsR in promoter regions containing BqsR and RhIR operators.

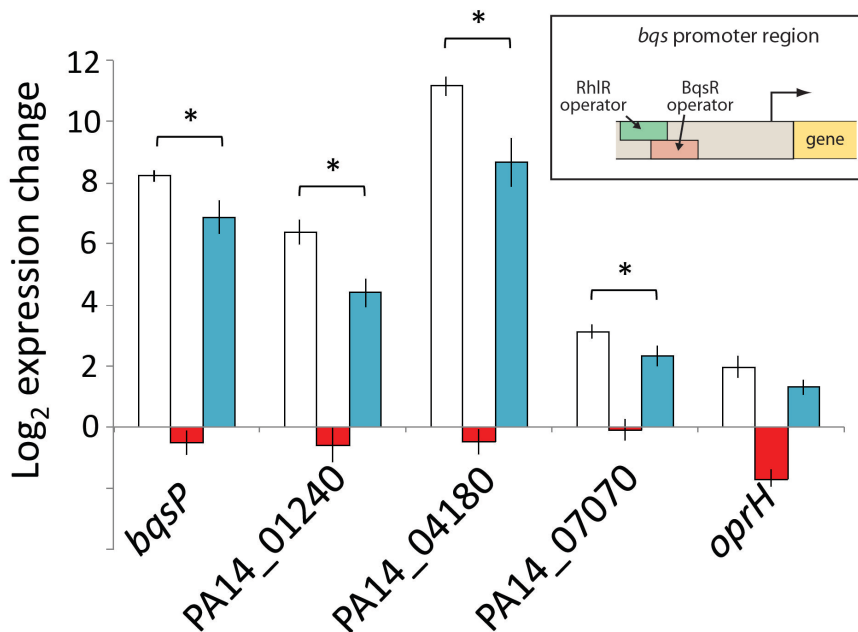


Figure 4-8: Fold change in expression due to ferrous iron shock for promoters coregulated by both BqsR and RhIR. For genes containing both BqsR and RhIR binding (*bqsP*, PA14\_01240, PA14\_04180, and PA14\_07000), the fold change in expression due to Fe(II) shock was decreased upon overexpression of RhIR, demonstrating the ability of RhIR to modulate the effect BqsR has on expression at these promoters. Expression from the *oprH* promoter, which does not contain an RhIR binding site, was not significantly influenced by RhIR overexpression. The inset shows the *bqs* promoter, which contains overlapping operators. WT response is shown in white,  $\Delta bqsR$  is shown in red, and WT overexpressing *rhIR* is shown in blue. \*indicates a p-value < 0.05.

## Discussion

Using the *P. aeruginosa* Fe(II) responsive regulator, BqsR, as a test case, we examined our ability to make quantitative predictions about the influence of an individual transcription factor on global gene expression levels. A PWM model did not accurately predict global gene expression patterns, leading us to hypothesize that either the PWM model did not capture how operator affinity was encoded in the operator sequence, or that the regulatory influence BqsR was dictated by the surrounding regulatory context of

each promoter. A detailed model of operator binding through a synthetic promoter library revealed that predicting transcription factor affinity alone was insufficient to predict the global expression levels. By examining the effect of operator number and distribution, proximity to the transcription start site, and the presence of secondary regulators, we found that secondary regulators were more important than operator occupancy in determining the influence of BqsR on expression levels at promoters throughout the genome.

While our model of operator occupancy outperformed the original PWM-derived model, the improvement was modest. Similarity between our binding energy matrix and the PWM supports the ability of PWM to describe operator affinity, at least when enough operator sequences can be identified to calculate an accurate binding matrix. However, our results underscore that caution should be used in relating operator strength to expression levels. Many genes highly upregulated in experiments were accurately predicted (within a factor of 3), but more than 150 operators with binding strengths predicted to be stronger than the most upregulated gene (PA14\_04180) had weak or immeasurable expression. This type of disagreement between experimental measurements and predictions highlights how difficult it is to make reliable predictions of gene expression directly from the genome sequence, and call attention to the need to more systematically study the influence of the broader regulatory context.

That the broader regulatory context may control the influence of individual transcription factors is not a new idea. Several studies have attempted to predictively

integrate inputs from multiple transcription factors at a single promoter; however, none of these studies used their findings to predict expression for all promoter contexts within the genome (42-45). The role of DNA structure in modulating promoter inputs has also been examined. Genome shape and mechanics, such as nucleosome wrapping and DNA loop formation, mediate both transcription factor binding and interactions (3, 8, 46). DNA can also mediate allosteric effects between adjacent transcription factors (47). Regulatory interactions have also been explored from a systems perspective (48), with one study reported that only 60% of the interactions between regulators could be accurately predicted in *E. coli* (4). Despite careful studies on many aspects of the broader regulatory context in several systems, it remains unclear to what extent a general framework can predict the influences of promoter context on regulation.

Moving forward, we can leverage the lessons learned here by examining in more detail remaining questions. For example, several promoters have large clusters of binding sites, unlikely to be present by pure chance, however our predictive ability decreased with increasing cluster size. Perhaps under the conditions measured, weaker operators have a low probability of occupancy and therefore do not contribute to regulation (49), although how weak an operator must be before it no longer modulates expression level is unclear, and may be context dependent (50). By examining a broader set of expression conditions, we may be able to develop a set of rules that predict when and how the number of operators in the clusters is significant. Additionally, we should transition from the bioinformatic analysis presented here to rigorous experimental quantification of the role

multiple transcription factors play in modulating promoter outputs. Future quantitative regulatory models should incorporate feedback in the dynamics of transcription factor levels. These types of experiments would enhance our understanding of how natural genomes encode regulatory outputs. Predicting signal integration at the promoter level remains a challenge. Bacteria respond to a wide variety of external stimuli, offering useful model systems in which to understand the logic and mechanisms of signal integration at the promoter level. Such work would complement ongoing efforts in synthetic and developmental biology (45, 48, 51).

## Methods

*Growth media and culturing conditions.* *P. aeruginosa* PA14 was grown both aerobically and anaerobically at 37 °C in MOPS minimal medium (MOMM) in acid washed glassware to ensure cells were Fe-limited. The basic MOMM is composed of 40 mM  $C_4H_4Na_2O_4 \cdot 6H_2O$ , 9.3 mM  $NH_4Cl$ , 2.2 mM  $KH_2PO_4$ , 25 mM  $KNO_3$ , 25 mM  $NaNO_3$ , 25 mM MOPS, 25 mM  $NaMOPS$  pH 7.2. Additionally, immediately prior to inoculation 100  $\mu M$   $CaCl_2$ , 1  $\mu M$   $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$ , 1 mM  $MgSO_4$ , and trace metals were added (52). Any composition changes are noted. All PA14 cultures were prepared by inoculation of MOMM media with the desired strains for 16 hours overnight shaking aerobically then grown anaerobically in a coy chamber with an atmosphere of 80%  $N_2$ , 15%  $CO_2$ , and 5%  $H_2$  at 37°C.

*Strain construction.* The strains used in this work were constructed from the wild type strain *P. aeruginosa* UCBPP-PA14. To monitor gene expression in these strains, the gene

construct containing the versions of the PA14\_04180 promoter attached to the *lacZ* reporter gene were inserted into the genome. Briefly, Gibson assembly was used to insert a gene construct containing 530 bases of the wildtype PA14\_04180 attached to *lacZ* between the transposon sites of the plasmid pUC18T-mini-Tn7t. The region between the transposon sites contained a selection marker for growth on gentamicin. This base construct was then mutated using site directed mutagenesis to create synthetic versions of the promoter. Gene constructs were transferred into the *glmS* locus of the genome of *P. aeruginosa* using triparental mating (53). Once inserted, constructs were verified using Sanger sequencing.

*Measuring gene expression of mutant library in response to iron shock.* The gene reporter *lacZ* quantified the change in gene expression in response to changes in ferrous iron. Anaerobic cultures grown to final OD<sub>500</sub> of 0.2-3 were uncapped inside of an anaerobic chamber and aliquoted to 1.7 ml tubes containing a final concentration of 400 mM ferrous iron (solutions of FeNH<sub>4</sub>SO<sub>4</sub>). Shocking with 400 mM Fe(II) more rapidly induced the maximal BqsR-mediated regulatory response observed with 200 mM Fe(II), which was used in previous RNA-Seq experiments (9) (Supplemental Fig. S13). Fe(II) has been measured approaching 300 μM in CF sputum (12). A ferrozine assay was used to confirm the concentration of the stock ferrous iron solution. The ferrous iron treatment was performed in triplicate for each strain measured. Cells were incubated in the anaerobic chamber at room temperature for 4 hours and then transferred to a 96 well plate for



measurement of gene expression. Each well received 150  $\mu$ L of ferrous iron treated cells with 50  $\mu$ L of a media containing 50 ng/ml of the fluorogenic LacZ indicator fluorescein di- $\beta$ -D-Galactopyranoside (FDG).  $OD_{550}$  and fluorescence with an excitation and emission of 490 nm and 520 nm, respectively, were measured in each well under anaerobic conditions every 5 min for 1 hour with a BioTek Synergy 4 plate reader. See Supplemental Fig. 4-S13 for control experiments regarding this procedure.

To calculate the change in gene expression after Fe(II) shock, the background corrected fluorescence measurements were divided by the background corrected absorbance measurement to quantify the gene expression per cell. To calculate the fold change in gene expression for a given strain, the gene expression measurement was divided by the expression from a strain containing the reference BqsR operator shown in Fig. 4-2A. The wildtype or initial operator sequence is shown in Figure 4-2A in the main text and is based on the downstream operator of the PA14\_04180 gene.

*RhIR and BqsR co-regulon prediction.* The position weight matrix for the RhIR DNA binding site (30) was input into FIMO (18), a tool which searches for a consensus sequence within a database. In this case, the database supplied was the 500 bp upstream from the translation start site for genes in the BqsR regulon.

*Effect of RhIR on Fe(II) shock conditions.* Aerobic cultures of *P. aeruginosa* wt-pMQ72 (54),  $\Delta bqsR$ -pMQ72, and wt-pMQ72-*rhIR* were grown in 3 ml MOMM supplemented with 100  $\mu$ g/ml gentamycin at 37  $^{\circ}$ C for 36 hours. Anaerobic cultures were grown in 20 ml MOMM

supplemented with 100 µg/ml gentamycin and 1% arabinose (to induce *rhIR* expression) with 1% inoculum from aerobic overnight. When the cells reached early exponential phase (Beckman spectrophotometer 20; OD<sub>500</sub> = 0.2), 9 ml of culture was removed. 4.5 ml culture was added to 9 ml of RNAprotect (Qiagen) before and after a 30 min 200 µM ferrous ammonium sulfate shock at room temperature. The cells were incubated with RNAprotect for 5 min and centrifuged for 10 min at 5000 x g. The supernatant was discarded and the pellets stored at -80 °C.

*mRNA isolation and qPCR data analysis.* mRNA was isolated from stored cell pellets using the RNeasy kit mini (Qiagen) with optional on-column DNA digestion according to the manufacturer's instructions. Subsequently, the RNA was treated with TURBO DNA-free (Applied Biosystems). cDNA was generated with iScript (Bio-Rad) random-primed reverse transcriptase reaction following the manufacturer's protocol. An mRNA genomic contamination control and cDNA was used as template for quantitative-reverse transcriptase-PCR (Real Time 7500 PCR Machine, Applied Biosystems) using iTaq Universal SYBR Green Supermix (Bio-Rad). Samples were assayed with 3-5 biological replicates. *recA* and *clpX* were used as endogenous controls (55). Fold changes were calculated using the  $\Delta\Delta C_t$  method (9). To ensure *recA* was constant in all conditions tested, the relative fold change for the internal control *clpX*, whose expression was also expected to remain constant across all our treatments. Only those samples with a *clpX* fold change between

0.5 – 2 were used.  $\log_2$  of the final fold change was reported. Results were compared with an unpaired 2-tailed t-test assuming unequal variances.

*Ferrozine assay.* This colorimetric assay measures Fe(II) concentration. The Stookey method (56) was modified for 96-well plate format. All measured Fe(II) concentrations were within 5% of reported value.

*Bioinformatics.* To analyze potential BqsR binding sites in the genome, occurrences of the motif (TTAAG(N)<sub>6</sub>TTAAG) were found in the genome of *Pseudomonas aeruginosa* PA14 using the program FIMO (Find Individual Motif Occurrences), part of the MEME Suite (18). Motif occurrences were then sorted and analyzed using Matlab. The same process was used to locate binding sites of other transcription factors using binding motifs listed in the ProDoric database (24).

*Comparison to the other operons.* From published microarray and RNA-Seq papers regulons (25-38) for other transcription factors were defined. The list of genes in the transcription factor regulons were converted to transcriptional units (23). Comparisons between transcriptional units (TU) in the regulon to two lists were made to discover the number of shared TUs: TUs in prediction but not observed in RNA-seq data and TUs in RNA-Seq data but not predicted. In R, a hypergeometric test assigned a p-value to the

overlapping regulons. For those regulons with significant overlap, whether the TUs were upregulated or downregulated was noted. See Supplemental Material for further details.

### Acknowledgments

*This work was supported by the National Institutes of Health, grant numbers DP1 OD000217, (Directors Pioneer Award), 5R01HL117328-03, R01 GM085286 and R01 GM085286B, and 1 U54 CA143869 (Northwestern PSOC Center); La Fondation Pierre Gilles de Gennes (R.P.); the Center for Microbial Interactions at Caltech (J.B.). D.K.N. is an HHMI Investigator.*

### Supplemental Material

#### *Position weight matrix calculations*

Position weight matrix (PWM) calculations are performed following (16,17). From a list of  $N$  operator sequences, the frequency of each bp at each operator position is calculated using

$$p_{xi} = \frac{n_{xi}+1}{N+4}, \quad (\text{Equation S1})$$

where  $p_{xi}$  is the estimated probability of having nucleotide  $x$  at position  $i$ ,  $n_{xi}$  is the number of sequences containing nucleotide  $x$  at position  $i$ . The +1 in the numerator ensure that  $p_{xi} > 0$ .

From these frequencies and given the background frequency of AT in the genome as 66% (57), the score ( $S$ ) of any operator sequence of length  $L$  is calculated using,

$$S = \sum_{i=1}^L \ln \frac{p_{xi}}{q_x}, \quad (\text{Equation S2})$$

where  $q_x$  is the background frequency of the nucleotide in position  $x$ . The score is proportional to the binding energy of BqsR to a given operator sequence, with greater scores indicating stronger binding. Score as defined in Equation S2 is used in Figures 4-1C, 4-2B, and 4-3C.

To determine how a change in the operator sequence influences the BqsR binding energy, the score of the mutated operator is subtracted from the score of the original operator.

#### *Bioinformatic analysis of the genome*

FIMO (Find Individual Motif Occurrences), part of the MEME Suite (18), was used to find potential BqsR binding sites in the genome of *Pseudomonas aeruginosa* strain UCBPP-PA14. An input sequence of TTAAG(N<sub>6</sub>)TTAAG was used with a threshold p-value of 0.1, resulting in over 80,000 output sequences. A large p-value threshold was used to include all potential sequences.

Custom Matlab code was used to sort through the FIMO output sequences. Each sequence was mapped onto the genome to determine the orientation and position of each sequence relative to the coding sequences of each gene. Sequences not falling within 0 to -1000 bp of a coding sequence were discarded. Mutations in the pentamer regions of each sequences were then identified, as compared to TTAAG(N<sub>6</sub>)TTAAG using a Matlab code. To find sequences with spacer lengths of 5 or 7, the input sequence to FIMO was adjusted to TTAAG(N<sub>5</sub>)TTAAG or TTAAG(N<sub>7</sub>)TTAAG, respectively.

### *Estimates of number of binding sites*

The number of binding sites expected within a genome of size 6,537,647 bp was calculated assuming the probability of a G or C is 33% at a given position and the probability of an A or T is 17% at a given position. Each binding sites has a specified number of mutations within the 10 bp of the repeated pentamer region and is separated by 6 random bp. These random binding sites can occur anywhere within the genome and have either orientation. For example, to calculate the probability of an operator with zero pentamer mutations,  $P_{\text{perfect operator}}$ , to occur (the identity of 10 bp is specified), we use

$$P_{\text{perfect operator}} = 2 (p_{\text{AT}}^8 p_{\text{CG}}^2), \quad (\text{Equation S3})$$

where  $p_{\text{AT}}$  is percent of A or T bp in the genome (16.7%) and  $p_{\text{CG}}$  is percent of C or G bp in the genome (33.3%). The factor of 2 accounts for the operator occurring in either orientation. Using the values for the *P. aeruginosa* genome, the probability of a zero mutant operator is approximately  $10^{-7}$ .

To estimate the number of zero mutant operators expected by chance in the genome, we multiple the probability of a zero mutant operator by the number of bp in the genome since each position in the genome is a potential starting sites for the operator sequence. Similar calculations estimate the number of 1, 2, and 3 mutation operators in the genome we should expect by chance. For these calculations we take into account that a mutation can occur in any of the 10 repeated pentamer positions. For example, to calculate the probability of an operator with 1 pentamer mutation, we use

$$P_{1 \text{ mutation operator}} = 2 ( 8(p_{\text{AT}}^7 p_{\text{CG}}^2) + 2(p_{\text{AT}}^8 p_{\text{CG}}) ) - P_{\text{perfect operator}} \quad (\text{Equation S4})$$

The factors of 8 and 2 inside the parentheses account for the number of ways in which a mutation in an A or T or a G could occur, respectively.  $P_{\text{perfect operator}}$  is subtracted because the first factor calculates the probability of an operator with either 0 or 1 mutation in the pentamer, *i.e.*, the changed base is just unspecified, meaning that it could either be a mutation or the original base.

As shown in Figure 4-S1, there are more pentamer regions without mutations than expected in the genome and fewer operators with 1, 2, or 3 pentamer mutations than would be expected by chance.

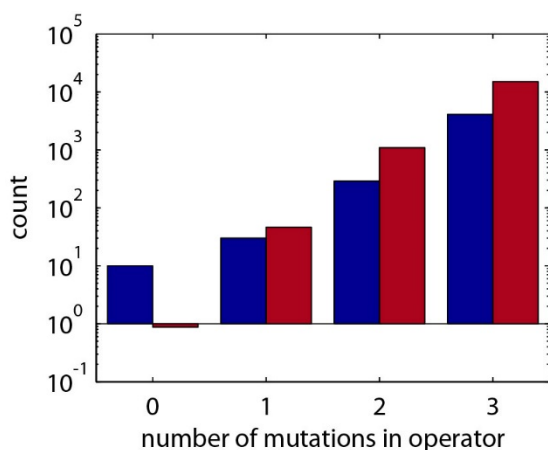


Figure 4-S1: The distribution of operators containing 0, 1, 2, and 3 mutations in the ten bp of the repeated pentamer. Blue bars show the actual distribution in the genome of *P. aeruginosa* and the red bars show the number of expected operators that would occur by chance in a genome of the same size with a GC content of 66%.

*Influence of pentamer orientation on BqsR-mediated gene regulation*

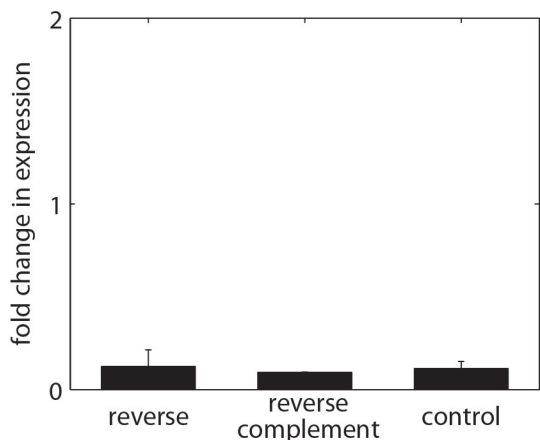


Figure 4-S2: Changing the upstream pentamer to the reverse (TTAAG to GAATT) or the reverse complement (TTAAG to CTAA) reduced the fold change in expression to the level of the negative control in which the upstream pentamer was deleted. Error bars show standard error.

#### *Distribution of Binding Sites in the Genome*

To put the distribution of repeated and clustered BqsR binding sites shown in Figure 4-4 into context, calculations estimated the occurrence of repeated and clustered binding sites. There are 6192 BqsR operators with up to 3 pentamer mutations throughout the genome of *P. aeruginosa*. To estimate the extent of operator clustering we might expect by chance, Matlab simulations were run to calculate the average cluster sizes when 6192 operators were inserted into a genome of size 6,537,647 bp. This genome contained genes in the same locations as the *P. aeruginosa*. The randomly inserted operators each operator have a unique starting position, but partial overlap of operator sequences was allowed. Each operator was assigned a random orientation. The operators could be inserted into any position within the genome. Five such random operator distributions were created and analyzed.



In these five random distributions, no repeated binding sites were observed, that is no repeats of more than two pentamers spaced by 6 bp with identical orientation were observed. Therefore, the repeats of up to 7 pentamers found in the actual genome, including 226 such repeats with 3 pentamers in a row, could not have happened by chance. On average these randomly placed 6192 BqsR operators ended up in the regulatory regions, from -1000 to 0 of the coding sequence, of 3,651 genes. This is slightly more than the 3,112 genes in the real genome containing BqsR operators. As shown in Figure 4-S3, the real genome has much larger clusters of operators within the same regulatory region. Cluster size is the number of operators within 1000 bp upstream of the gene coding sequence, regardless of operator orientation or relative spacing. Randomly placed operators only occasionally resulted in a cluster of 7 operators, i.e., a promoter region containing 7 total operators, whereas the real genome contains several clusters of more than 10 operators in the same regulatory region.

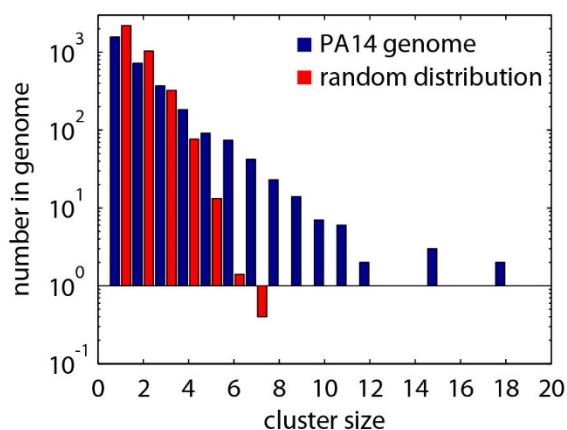


Figure 4-S3: Clusters of binding sites found within 1000 bp upstream of the start of gene coding sequences. The distribution of cluster sizes for the genome of *P. aeruginosa* (blue) compared to the number of clusters for a random redistribution of the same number of

operators throughout the genome. For the random distribution, a number in genome less than 1 indicates that not all random distributions measured contained a cluster of that size.

### *Predicting gene expression from operator sequence*

In our experiments we are measuring the fold change in gene expression upon mutating an operator sequence from its wild-type sequence. From this fold change we can calculate the relative change in the operator binding energy as a result of the change in the operator sequence.

The model used to analyze the data is derived as previously described for the promoters such as the *lac* promoter for *E. coli*. Briefly, in this model the rate of transcription is proportional to the probability of finding RNA polymerase bound to the promoter. The probability of finding the polymerase on the promoter is calculated using a thermodynamic quasi-equilibrium model. In the model the promoter has several possible microstates that are defined by the configuration of the RNA polymerase and the transcription factors on the promoter. The probability of a specific configuration is related to the copy numbers and binding energies of the proteins bound to the promoter region. These microstates and related Boltzmann terms are shown in Figure 4-S4 for the BqsR binding promoter we are using in this study. In these Boltzmann weights,  $P$  is the number of RNA polymerase molecules per cell,  $A$  is the number of BqsR transcription factors per cell,  $E_o$  is the binding energy of BqsR to its operator sequence,  $E_p$  is the binding energy of RNA polymerase to its promoter sequence,  $E_c$  accounts for cooperativity between BqsR

and RNA polymerase,  $N_{NS}$  is the number of nonspecific binding sites in the genome,  $k_B$

is the Boltzmann constant, and  $T$  is temperature.

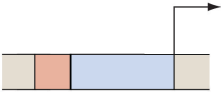
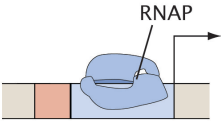
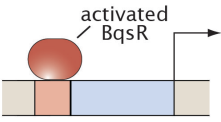
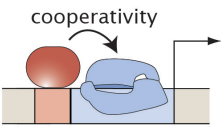
STATE	WEIGHT	PROMOTER ESCAPE RATE
	1	0
	$\frac{P}{N_{NS}} e^{-\frac{E_p}{k_B T}}$	$k_1$
	$\frac{A}{N_{NS}} e^{-\frac{E_a}{k_B T}}$	0
	$\frac{P}{N_{NS}} \frac{A}{N_{NS}} e^{-\frac{(E_p+E_a+E_c)}{k_B T}}$	$k_2$

Figure 4-S4: States and weights for BqsR-mediated gene regulation. In the bottom state there is cooperativity between RNA polymerase and BqsR, resulting in increased occupancy of the promoter by RNA polymerase in the presence of BqsR.

Another parameter which determines the production rate of mRNA is the rate constant for transcription of each promoter microstate. Any state in which the RNA polymerase is not bound has a transcriptional rate constant of zero. Each state in which the RNA polymerase is bound has a non-zero transcriptional rate constant, as listed in Figure 4-S4.

From the Boltzmann weights listed in Figure 4-S4, we can now derive the rate of transcription. The transcription rate is the sum of the probability of each state,  $p_i$ , times the rate of transcription,  $r_i$ . This gives,

$$\text{transcription rate} = \sum p_i r_i. \quad (\text{Equation S5})$$

The probability of a state is the Boltzmann weight for that state divided by the sum of the Boltzmann weights for every possible microstate:

$$p_i = \frac{w_i}{\sum w_i}. \quad (\text{Equation S6})$$

From these definitions we can derive the transcription rate for our particular system as,

$$\text{transcription rate} = \frac{k_1 \frac{P}{N_{NS}} e^{\frac{-E_p}{k_B T}} + k_2 \frac{P}{N_{NS}} \frac{A}{N_{NS}} e^{\frac{-(E_p + E_a + E_c)}{k_B T}}}{1 + \frac{P}{N_{NS}} e^{\frac{-E_p}{k_B T}} + \frac{P}{N_{NS}} \frac{A}{N_{NS}} e^{\frac{-(E_p + E_a + E_c)}{k_B T}} + \frac{A}{N_{NS}} e^{\frac{-E_a}{k_B T}}}. \quad (\text{Equation S7})$$

Now consider we make a change to the operator sequence to which the transcription factor BqsR binds. We assume that this change in the operator sequence only influences the binding energy of BqsR to its operator and has no influence on both the transcription rate constant and the cooperativity term from the RNA polymerase and BqsR bound microstate. The binding energy of BqsR to the mutated operator sequence we call  $E_m$ . From this definition we can calculate the fold change in the transcription rate upon mutating the BqsR operator as,

$$\text{fold change} = \frac{\frac{k_1 \frac{P}{N_{NS}} e^{\frac{-E_p}{k_B T}} + k_2 \frac{P}{N_{NS}} \frac{A}{N_{NS}} e^{\frac{-(E_p + E_a + E_c)}{k_B T}}}{1 + \frac{P}{N_{NS}} e^{\frac{-E_p}{k_B T}} + \frac{P}{N_{NS}} \frac{A}{N_{NS}} e^{\frac{-(E_p + E_a + E_c)}{k_B T}} + \frac{A}{N_{NS}} e^{\frac{-E_a}{k_B T}}}}{\frac{k_1 \frac{P}{N_{NS}} e^{\frac{-E_p}{k_B T}} + k_2 \frac{P}{N_{NS}} \frac{A}{N_{NS}} e^{\frac{-(E_p + E_m + E_c)}{k_B T}}}{1 + \frac{P}{N_{NS}} e^{\frac{-E_p}{k_B T}} + \frac{P}{N_{NS}} \frac{A}{N_{NS}} e^{\frac{-(E_p + E_m + E_c)}{k_B T}} + \frac{A}{N_{NS}} e^{\frac{-E_m}{k_B T}}}}}. \quad (\text{Equation S8})$$

Experimentally, this fold change is equivalent to the ratio of gene expression for cells containing the wild-type operator sequence to cells containing the mutated operator.

This equation can be simplified to

$$\begin{aligned}
 \text{fold change} &= \frac{\frac{\frac{P}{N_{NS}} e^{\frac{-E_p}{k_B T}} (1 + \frac{k_2}{k_1} \frac{A}{N_{NS}} e^{\frac{-(E_a + E_c)}{k_B T}})}{(1 + \frac{P}{N_{NS}} e^{\frac{-E_p}{k_B T}}) (1 + \frac{A}{N_{NS}} e^{\frac{-(E_a + E_c)}{k_B T}})}}{\frac{\frac{P}{N_{NS}} e^{\frac{-E_p}{k_B T}} (1 + \frac{k_2}{k_1} \frac{A}{N_{NS}} e^{\frac{-(E_m + E_c)}{k_B T}})}{(1 + \frac{P}{N_{NS}} e^{\frac{-E_p}{k_B T}}) (1 + \frac{A}{N_{NS}} e^{\frac{-(E_m + E_c)}{k_B T}})}}} = \frac{\frac{(1 + \frac{k_2}{k_1} \frac{A}{N_{NS}} e^{\frac{-(E_a + E_c)}{k_B T}})}{(1 + \frac{A}{N_{NS}} e^{\frac{-(E_a + E_c)}{k_B T}})}}{\frac{(1 + \frac{k_2}{k_1} \frac{A}{N_{NS}} e^{\frac{-(E_m + E_c)}{k_B T}})}{(1 + \frac{A}{N_{NS}} e^{\frac{-(E_m + E_c)}{k_B T}})}}}. \quad (\text{Equation S9})
 \end{aligned}$$

Making the approximation that the transcription rate in the BqsR bound case is much larger than the transcription rate when only the RNA polymerase is bound, i.e.,

$$\frac{k_2}{k_1} \frac{A}{N_{NS}} e^{\frac{-(E_a + E_c)}{k_B T}} \gg 1 \text{ and } \frac{k_2}{k_1} \frac{A}{N_{NS}} e^{\frac{-(E_m + E_c)}{k_B T}} \gg 1, \text{ we can further simplify the fold change}$$

expression,

$$\begin{aligned}
 \text{fold change} &= \frac{\frac{\frac{A}{N_{NS}} e^{\frac{-E_a}{k_B T}}}{(1 + \frac{A}{N_{NS}} e^{\frac{-E_a}{k_B T}})}}{\frac{\frac{A}{N_{NS}} e^{\frac{-E_m}{k_B T}}}{(1 + \frac{A}{N_{NS}} e^{\frac{-E_m}{k_B T}})}}}. \quad (\text{Equation S10})
 \end{aligned}$$

This approximation is well supported by the experimental data in Figure 4-S5, which shows the expression of the reporter gene LacZ is low the in the absence of strong BqsR activation.

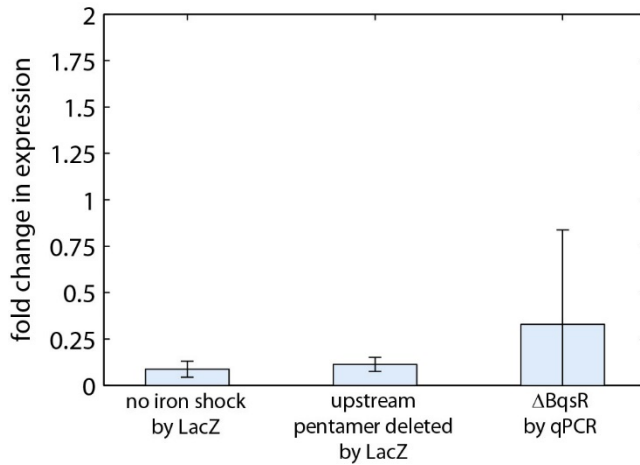


Figure 4-S5: Fold change in expression in the absence of strong BqsR activity. Without iron shock, no extra iron is added during the shock, the fold change in expression is  $9 \pm 4\%$ . Upon mutating all five bases of the upstream operator, the fold change in expression fell to  $11 \pm 4\%$ . Given the mutations made in the pentamer (TTAAG to ACTCA), we would calculate residual expression of 7%. These measurements were taken using the LacZ gene reporter for the mutated PA14\_04180 promoter region containing only two pentamers. The final bar shows the change in gene expression for the wild-type version of gene PA14\_04180 upon deleting BqsR from the genome, measured using qPCR. Together these indicate that the gene expression level of PA14\_04180 in the absence of BqsR-mediated activation should be negligible.

With this further simplification the transcription rate constants have now been removed from the fold change expression. The remaining expression has two interesting limits. In the limit that  $\frac{A}{N_{NS}} e^{\frac{-E_a}{k_B T}} \gg 1$  and  $\frac{A}{N_{NS}} e^{\frac{-E_m}{k_B T}} \gg 1$ , the fold change goes to 1. In this limit, the probability of BqsR being bound is very large, therefore it is not surprising that a change in the BqsR binding energy may not result in a change in the fold change.

We could call this the operator saturation limit. The more interesting limit is when

$\frac{A}{N_{NS}} e^{\frac{-E_a}{k_B T}} \ll 1$  and  $\frac{A}{N_{NS}} e^{\frac{-E_m}{k_B T}} \ll 1$ , when we would not expect to always find BqsR bound

to the operator. In other words, we still expect some “action” from changes in BqsR copy

number and binding energy. In the limit of not saturating the operator the fold change simplifies to,

$$\text{fold change} = e^{\frac{-(E_a - E_m)}{k_B T}} \quad (\text{Equation S11})$$

or

$$-\ln(\text{fold change}) = \frac{\Delta E_{mut}}{k_B T}, \quad (\text{Equation S12})$$

where  $\Delta E_{mut}$  is the change in the binding energy due to the mutation. From the fold change measurement we can calculate the change in the BqsR binding energy caused by the mutation in the operator.

In the reverse case, we can also then back calculate the fold change for a given operator if we know how each basepair mutation influences the binding energy. This assumes that each individual mutation in the operator sequence has an additive effect on the operator binding energy, a first approximation that has proven useful in previous examples (20,22).

To be more explicit, starting from the wild-type operator, if we have 3 point mutations in the operator we would expect the fold change to be,

$$\text{fold change} = e^{\frac{-(\Delta E_{m1} + \Delta E_{m2} + \Delta E_{m3})}{k_B T}} = e^{\frac{-\Delta E_{m1}}{k_B T}} e^{\frac{-\Delta E_{m2}}{k_B T}} e^{\frac{-\Delta E_{m3}}{k_B T}}. \quad (\text{Equation S13})$$

In Equation S13  $\Delta E_{mi}$  is the change in the binding energy due to the  $i^{\text{th}}$  mutation in the operator sequence. Of course we should keep in mind that this fold change is relative to the gene expression from the “wild-type” sequence and not an exact level of expression (unless of course we knew the precise level of gene expression in the wild-type case).

Another way of looking at this result is that you can multiple together the fold changes due to each change in the operator sequence. For example if mutation 1 has a fold change of 0.8 and mutation 2 has a fold change of 0.5, the predicted fold change for an operator with both mutations would be the product of the two fold changes or 0.4.

#### *Deriving the energy matrix for the spacer region*

To derive the “best-fit” matrix, first we assume the sequence of the spacer region does not influence expression. This is equivalent to saying all the bases of the spacer region contribute 0  $k_B T$  to the binding energy. From this starting matrix, we predict the expression level for synthetic promoters containing the 7 spacer sequences (Fig. 4-3D), and calculate the deviations of the predictions from the experimental measurements using Equation 2. The value of the matrix at a random position is then increased or decreased by 4%, and new deviations from the measured expression levels are calculated. If the prediction was more accurate, the change to the energy matrix was retained. If the prediction was less accurate, there was a 1% chance the change was kept. This process was iterated 10,000 times to generate a final matrix that is the best fit to the experimental data. To ensure that the fitting procedure did not get stuck at a local minimum, the whole process was then repeated from the all zeros spacer energy matrix 100 times and these 100 final matrices were averaged together. Supplemental Figure 4-S6 shows the final energy matrix after 10,000 iterations has reached an optimum.



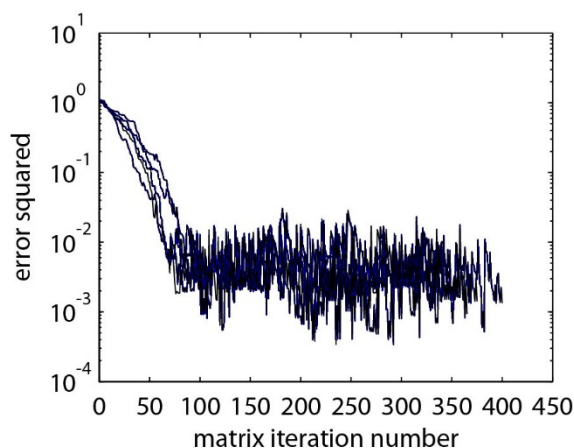


Figure 4-S6: In fitting the energy matrix for the middle 6 bp of the operator, the value of the binding energy matrix for a random basepair was increased or decreased by 4% and the resultant error in the prediction using this new energy matrix was calculated. As mutations accumulated, the square in the error quickly reduced to a level less than  $10^{-3}$ . The graph above shows the squared error as a function of the number of changes for 5 independent fits. The spacer region for the energy matrix reported in Figure 4-5B averages together 100 such replicate fits.

*Distribution of operator affinities and gene expression*

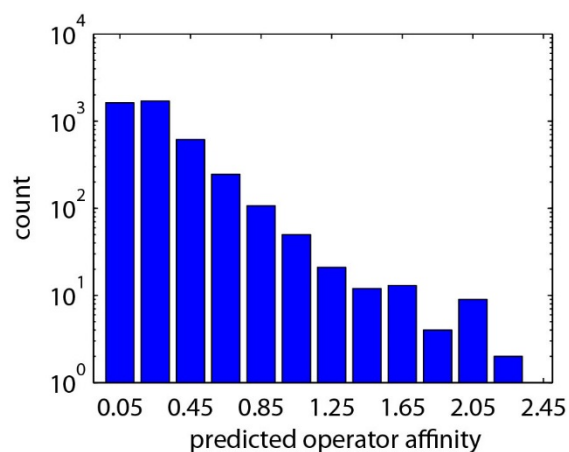


Figure 4-S7: The distribution of predicted operator affinities for all 4415 operators containing less than 3 pentamer region mutations found within 1000 bp upstream of the coding sequence. Operator affinity is the negative exponential of the change in binding energy divided by  $k_B T$ . Binding energies for each operator were calculated using the energy matrix in Figure 4-5B of the main text. The starting sequence shown in Figure 4-5A

has a predicted operator affinity of 1. The model assumes operator affinity is proportional to fold change in expression.

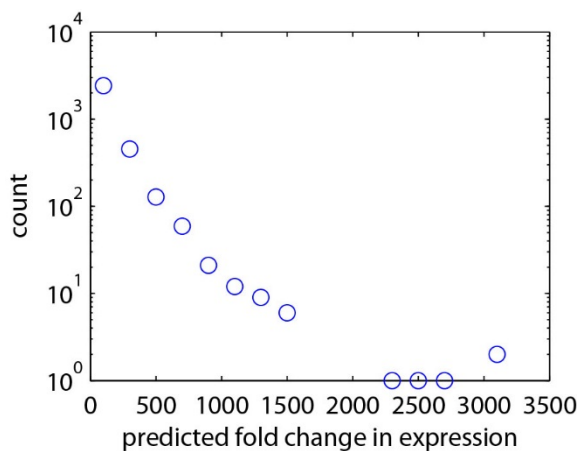


Figure 4-S8: The distribution of predicted fold change in gene expression for all genes containing a BqsR operator with fewer than 3 pentamer mutations within 1000 bp upstream of the coding sequence. Predictions are scaled such that the prediction for the reference gene PA14\_04180 matches the experimentally measured value.

#### *Additive approximation for multi-operator promoters*

To predict expression from promoter regions containing multiple BqsR binding sites, we used an additive model. From the results of Figure 4-4, we found that having multiple binding sites in the same promoter region leads to increased expression in the case of the PA14\_04180 promoter. We approximated the combined influence of multiple operators at the same promoter using an additive model in which

$$\text{fold change in expression}_{total} = \sum \text{fold change}_i. \quad (\text{Equation S14})$$

Fold change<sub>*i*</sub> is the fold change due to operator *i*, where fold change is calculated using Equation 2. This additive approximation is equivalent to assuming the two operators act independently on gene expression.

To access the consequences of this approximation, we next calculated the total fold change in gene expression for a promoter containing two operators when using the additive approximation versus a thermodynamic model, which incorporates the possibility of having both operators occupied simultaneously. For the additive case, the influence that each operator can be calculated using

$$\text{transcription rate} = \frac{k_1 \frac{P}{N_{NS}} e^{\frac{-E_p}{k_B T}} + k_2 \frac{P}{N_{NS}} \frac{A}{N_{NS}} e^{\frac{-(E_p + E_a + E_c)}{k_B T}}}{1 + \frac{P}{N_{NS}} e^{\frac{-E_p}{k_B T}} + \frac{P}{N_{NS}} \frac{A}{N_{NS}} e^{\frac{-(E_p + E_a + E_c)}{k_B T}} + \frac{A}{N_{NS}} e^{\frac{-E_a}{k_B T}}}. \quad (\text{Equation S15})$$

where  $P$  is the number of RNA polymerase molecules per cell,  $A$  is the number of BqsR transcription factors per cell,  $E_a$  is the binding energy of BqsR to its operator sequence,  $E_p$  is the binding energy of RNA polymerase to its promoter sequence,  $E_c$  accounts for cooperativity between BqsR and RNA polymerase,  $k_i$  is the rate of transcription from state  $i$ ,  $N_{NS}$  is the number of nonspecific binding sites in the genome,  $k_B$  is the Boltzmann constant, and  $T$  is temperature.

When not using the additive approximation, we derive the transcription rate from the states and weights shown in Figure 4-S9. As can be seen in these states, the two operators have different binding energies,  $E_{a1}$  and  $E_{a2}$ . The energy that accounts for cooperativity between BqsR and RNA polymerase is the regardless of whether operator 1, operator 2, or both operators are bound.

STATE	WEIGHT	PROMOTER ESCAPE RATE
	1	0
	$\frac{P}{N_{NS}} e^{-\frac{E_p}{k_B T}}$	$k_1$
	$\frac{A}{N_{NS}} e^{-\frac{E_{a1}}{k_B T}}$	0
	$\frac{A}{N_{NS}} e^{-\frac{E_{a2}}{k_B T}}$	0
	$\frac{A^2}{N_{NS}^2} e^{-\frac{(E_{a1}+E_{a2})}{k_B T}}$	0
	$\frac{P}{N_{NS}} \frac{A}{N_{NS}} e^{-\frac{(E_p+E_{a1}+E_c)}{k_B T}}$	$k_2$
	$\frac{P}{N_{NS}} \frac{A}{N_{NS}} e^{-\frac{(E_p+E_{a2}+E_c)}{k_B T}}$	$k_3$
	$\frac{P}{N_{NS}} \frac{A^2}{N_{NS}^2} e^{-\frac{(E_p+E_{a1}+E_{a2}+E_c)}{k_B T}}$	$k_4$

Figure 4-S9: The states and weights for a promoter containing two BqsR operators.

From the weights in Figure 4-S9, we derive the transcription rate for the case of two BqsR operators to be

*transcription rate* =

$$\frac{k_1 \frac{P}{N_{NS}} e^{-\frac{E_p}{k_B T}} + k_2 \frac{PA}{N_{NS}^2} e^{-\frac{(E_p+E_{a1}+E_c)}{k_B T}} + k_3 \frac{PA}{N_{NS}^2} e^{-\frac{(E_p+E_{a2}+E_c)}{k_B T}} + k_4 \frac{PA^2}{N_{NS}^3} e^{-\frac{(E_p+E_{a1}+E_{a2}+E_c)}{k_B T}}}{1 + \frac{P}{N_{NS}} e^{-\frac{E_p}{k_B T}} + \frac{A}{N_{NS}} e^{-\frac{E_{a1}}{k_B T}} + \frac{A}{N_{NS}} e^{-\frac{E_{a2}}{k_B T}} + \frac{A^2}{N_{NS}^2} e^{-\frac{(E_{a1}+E_{a2})}{k_B T}} + \frac{PA}{N_{NS}^2} e^{-\frac{(E_p+E_{a1}+E_c)}{k_B T}} + \frac{PA}{N_{NS}^2} e^{-\frac{(E_p+E_{a2}+E_c)}{k_B T}} + \frac{PA^2}{N_{NS}^3} e^{-\frac{(E_p+E_{a1}+E_{a2}+E_c)}{k_B T}}}$$

(Equation S16)

We calculated the ratios of transcription rates for the additive model, using Equations S14 and S15, over the full thermodynamic model, using Equation S16. The results are shown in Figure 4-S10.

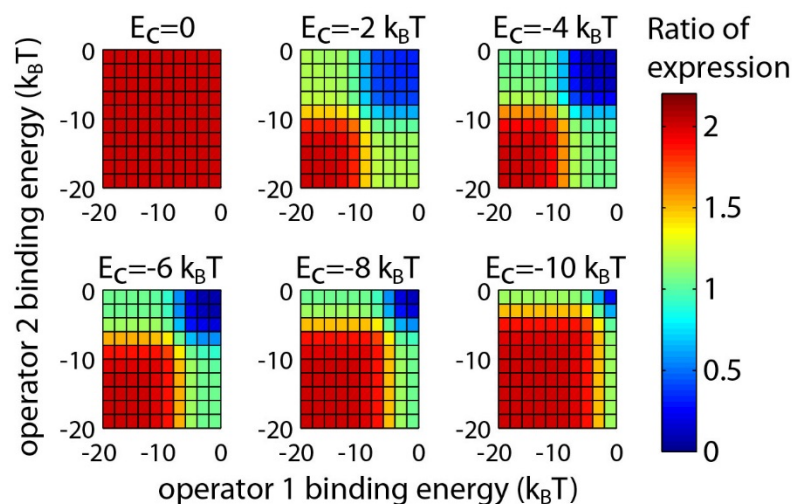


Figure 4-S10: The ratio of expression levels for a model which assumes operators act independently, the additive model using Equation S14, to a model which directly incorporates the states in which both operators are bound. Calculations are for a promoter containing two operators. The axes of each plot are the binding energies of the two operators. For these calculations the number of BqsR per cell was 100, the number of nonspecific binding sites was  $5 \times 10^6$ , the RNA polymerase binding energy was  $-3 k_B T$ , and the number of RNA polymerase per cell was 10,000. In these calculations we assumed the promoter escape rates,  $k_i$ 's, for all states were equal.

In Figure 4-S10, the interaction energies range from 0 to  $-10 k_B T$ , and the operator binding energies range from 0 to  $-20 k_B T$ . We found that when both operator binding energies are strong,  $< -10 k_B T$ , the prediction using the additive model will be within a factor of two of the prediction using the full thermodynamic model whose states are listed in Figure 4-S9. For intermediate interaction energies,  $-2 k_B T$  to  $-8 k_B T$ , the additive model is close to the full thermodynamic model when one of the operators had a strong

binding energy and the other had a weak binding energy. The calculations with the additive model were particularly poor when both operators had very weak binding, approximately less than  $5 k_B T$ . Given that typical binding energy are  $-5$  to  $-15 k_B T$ , and interaction energies with RNA polymerases have been observed to be  $0$  to  $-10 k_B T$ , we expect calculations made for a promoter containing two operators using the additive model to be within a factor of 2 of calculations using a full thermodynamic model. Therefore the additive model is unlikely to contribute significantly to the general overestimate of expression levels observed in Figure 4-6 of the main text.

#### *Single operator predictions*

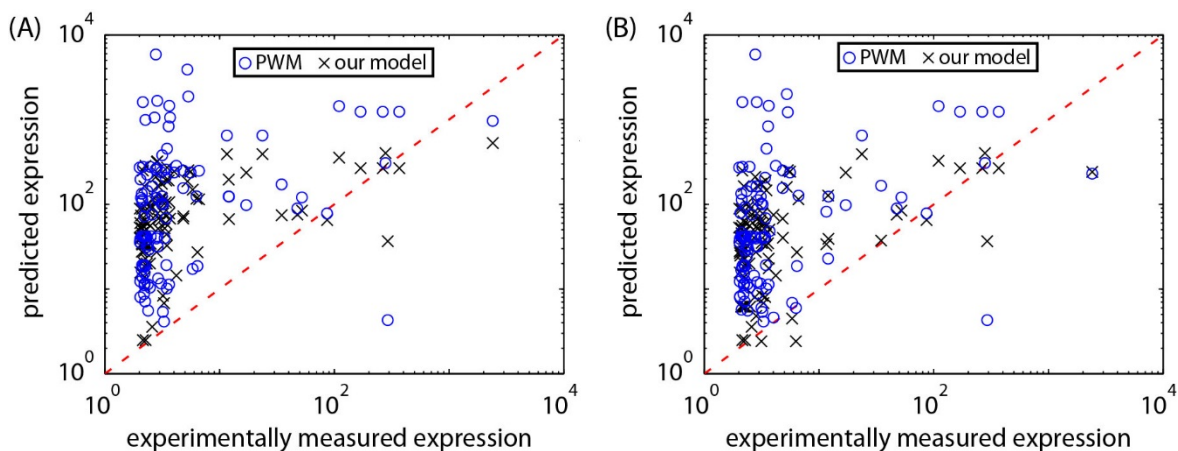


Figure 4-S11: Comparison of predictions to experimental measurements of gene expression. As in Figure 4-6 of the main text, the predictions use the BqsR binding matrices found in Figure 4-5. Unlike in Figure 4-6, in (A) we only use the strongest binding operator and in (B) we only use the closest operator to predict expression. All operators used contain less than 3 mutations and are found between  $-1000$  and  $0$  bases of the coding sequence.

Gene expression predictions taking into account operator orientation

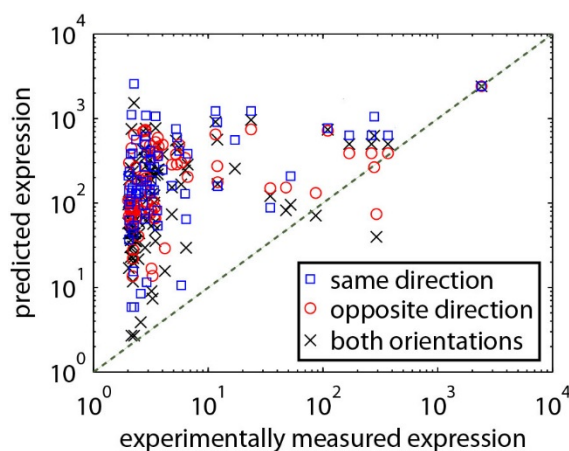


Figure 4-S12: Predictions of gene expression for the same set of genes shown in Figure 4-6 when taking operator direction into account. The black x's replot the original predictions from Figure 4-6B which allow the operators to be in either orientation relative to the direction of the regulated gene. The blue squares show the predictions for the same genes using only those operators that have the same orientation of the regulated gene, and the red circles show the predictions using only those operators that have the opposite orientation of the regulated gene. For some of the predicted genes, all of the operators face in the same direction; therefore, the red circles and blue squares each contain a subset of the predictions made using both orientations.

Measuring gene expression in response to ferrous iron shock

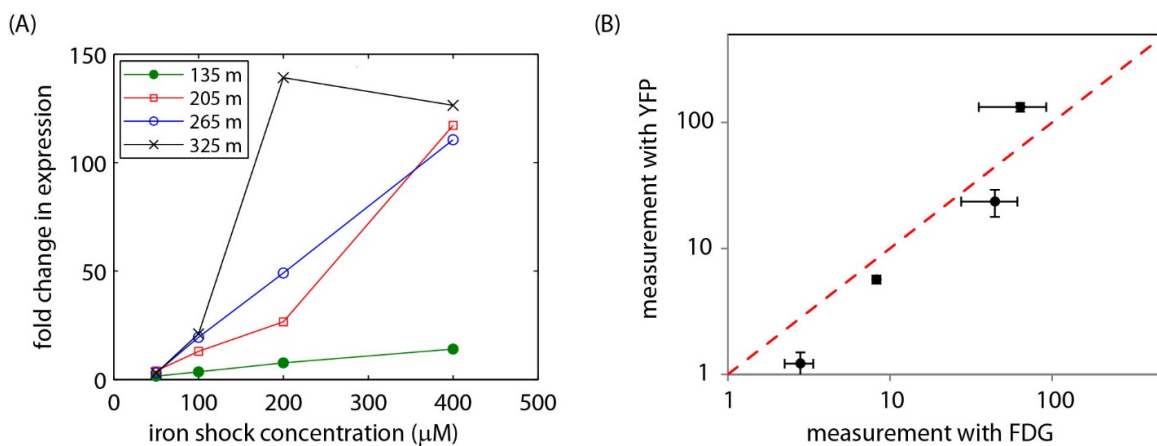


Figure 4-S13: (A) Cells containing a *lacZ* reporter attached to the promoter region of gene PA14\_04180 were shocked with ferrous iron concentrations between 50 and 400 mM for either 135 m, 205 m, 265 m, or 325 m (see legend). Following the shock, expression levels

of *lacZ* were quantified using the fluorogenic indicator FDG. It was found that upon shocking with 400 mM ferrous iron, the fold change reached a maximum in less than 205 min. (B) Comparison of the gene expression measured using the FDG assay and the LacZ reporter to a YFP reporter. Measurements were performed on aerobically grown *E. coli* strains containing a *lac* operator at +11 relative to the transcription start site and the wildtype number of Lac repressors per cells (58). The four data points correspond to construct containing the O1, O2, O3, and Oid *lac* operator.

### *Comparison to other operons*

We compare the genes experimentally found to be under BqsR control with 13 experimentally validated *P. aeruginosa* regulons from the literature. Most of these regulons did not significantly overlap with the BqsR regulon (see Table 4-S1). Figure 4-7C shows the results of a few of these comparisons. For some transcription factors, the regulon was reported more than once. Two published datasets for Fur were compared to the BqsR regulon. Only one dataset significantly overlapped with the BqsR regulon (28) and the expression of all of the overlapping TUs were only marginally perturbed by Fur. Two methods for finding the PqsR regulon were used (32,33) and both regulons overlapped with TUs we predicted to be BqsR regulated but did not appear in the RNAseq data with statistical significance. Anr represses all of the genes (25) that statistically significantly overlap with predicted BqsR TUs, but did not appear in the RNAseq data. RpoN activates many of the genes (34) that we observed in the experimental data but were not predicted using our binding energy matrix, referred to as the false negatives in the main text.



Table 4-S1. Known *P. aeruginosa* regulons comparison to predicted BqsR regulon.

\*p-value &lt; 0.05

regulon	total transcription units in regulon	overlap with false positive predictions	significance of overlap	overlap with false negative predictions	significance of overlap	Reference
Anr	47	12	*4.01E-12	0		25
Dnr	7	0		0		25
ArgR	22	1	7.56E-02	0		26
Fur 1	84	0		0		27
Fur 2	137	7	*7.06E-03	0		28
MexT	101	1	6.28E-01	0		29
LasR	55	1	3.18E-01	0		30
PqsR 1	78	8	*2.65E-05	0		32
PqsR 2	107	9	*5.90E-05	0		33
All quorum sensing	77	2	2.15E-01	0		30
RhlR	30	1	1.28E-01	0		30
RpoS exponential	49	1	2.72E-01	0		30
RpoS stationary	21	0		0		30
RsmA	375	10	1.52E-01	1	4.10E-01	36
Vfr	139	1	7.93E-01	1	9.10E-02	37
RpoN all	659	18	7.97E-02	5	*2.29E-02	34
RpoN induced	335	10	8.47E-02	3	*2.93E-02	34
RpoN repressed	327	8	2.38E-01	2	1.13E-01	34

## References

1. **Kung JT, Colognori D, Lee JT.** 2013. Long noncoding RNAs: past, present, and future. *Genetics* **193**:651-669.
2. **Meister G.** 2013. Argonaute proteins: functional insights and emerging roles. *Nat Rev Genet* **14**:447-459.
3. **Levo M, Segal E.** 2014. In pursuit of design principles of regulatory sequences. *Nat Rev Genet* **15**:453-468.
4. **Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS.** 2007. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol* **5**:e8.
5. **Spitz F, Furlong EE.** 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**:613-626.
6. **Garcia HG, Sanchez A, Boedicker JQ, Osborne M, Gelles J, Kondev J, Phillips R.** 2012. Operator sequence alters gene expression independently of transcription factor occupancy in bacteria. *Cell Reports* **2**:150-161.
7. **Mayo AE, Setty Y, Shavit S, Zaslaver A, Alon U.** 2006. Plasticity of the *cis*-regulatory input function of a gene. *PLoS Biology* **4**:e45.
8. **Lam FH, Steger DJ, O'Shea EK.** 2008. Chromatin decouples promoter threshold from dynamic range. *Nature* **453**:246-250.
9. **Kreamer NNK, Wilks JC, Marlow JJ, Coleman ML, Newman DK.** 2012. BqsR/BqsS constitute a two-component system that senses extracellular Fe(II) in *Pseudomonas aeruginosa*. *Journal of Bacteriology* **194**:1195-1204.
10. **Vasil ML, Ochsner UA.** 1999. The response of *Pseudomonas aeruginosa* to iron: genetics, biochemistry and virulence. *Molecular Microbiology* **34**:399-413.
11. **Singh PK, Parsek MR, Greenberg EP, Welsh MJ.** 2002. A component of innate immunity prevents bacterial biofilm development. *Nature* **417**:552-555.
12. **Hunter RC, Asfour F, Dingemans J, Osuna BL, Samad T, Malfroot A, Cornelis P, Newman DK.** 2013. Ferrous iron is a significant component of bioavailable iron in cystic fibrosis airways. *mBio* **4**.
13. **Castro AP, Fernandes GDR, Franco OL.** 2014. Insights into novel antimicrobial compounds and antibiotic resistance genes from soil metagenomes. *Frontiers in Microbiology* **5**:489.
14. **Dong Y-H, Zhang X-F, An S-W, Xu J-L, Zhang L-H.** 2008. A novel two-component system BqsS-BqsR modulates quorum sensing-dependent biofilm decay in *Pseudomonas aeruginosa*. *Commun Integr Biol* **1**:88-96.
15. **Kreamer NN, Costa F, Newman DK.** 2015. The ferrous iron-responsive BqsRS two-component system activates genes that promote cationic stress tolerance. *mBio* **6**.

16. **Hertz GZ, Stormo GD.** 1999. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* **15**:563-577.
17. **Vilar JMG.** 2010. Accurate prediction of gene expression by integration of DNA sequence statistics with detailed modeling of transcription regulation. *Biophysical Journal* **99**:2408-2413.
18. **Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS.** 2009. MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research* **37**:W202-W208.
19. **Lavrrar JL, McIntosh MA.** 2003. Architecture of a Fur binding site: a comparative analysis. *Journal of Bacteriology* **185**:2194-2202.
20. **Kinney JB, Murugan A, Callan CG, Jr., Cox EC.** 2010. Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proceedings of the National Academy of Sciences of the United States of America* **107**:9158-9163.
21. **Stauff DL, Bassler BL.** 2011. Quorum sensing in *Chromobacterium violaceum*: DNA recognition and gene regulation by the CviR receptor. *Journal of Bacteriology* **193**:3871-3878.
22. **Brewster RC, Jones DL, Phillips R.** 2012. Tuning promoter strength through RNA polymerase binding site design in *Escherichia coli*. *Plos Computational Biology* **8**:e1002811-e1002811.
23. **Wurtzel O, Yoder-Himes DR, Han K, Dandekar AA, Edelheit S, Greenberg EP, Sorek R, Lory S.** 2012. The single-nucleotide resolution transcriptome of *Pseudomonas aeruginosa* grown in body temperature. *PLoS Pathog* **8**:e1002945.
24. **Munch R, Hiller K, Barg H, Heldt D, Linz S, Wingender E, Jahn D.** 2003. PRODORIC: prokaryotic database of gene regulation. *Nucleic Acids Research* **31**:266-269.
25. **Trunk K, Benkert B, Quäck N, Münch R, Scheer M, Garbe J, Jänsch L, Trost M, Wehland J, Buer J, Jahn M, Schobert M, Jahn D.** 2010. Anaerobic adaptation in *Pseudomonas aeruginosa*: definition of the Anr and Dnr regulons. *Environmental Microbiology* **12**:1719-1733.
26. **Lu C-D, Yang Z, Li W.** 2004. Transcriptome analysis of the ArgR regulon in *Pseudomonas aeruginosa*. *Journal of Bacteriology* **186**:3855-3861.
27. **Ochsner UA, Wilderman PJ, Vasil AI, Vasil ML.** 2002. GeneChip® expression analysis of the iron starvation response in *Pseudomonas aeruginosa*: identification of novel pyoverdine biosynthesis genes. *Molecular Microbiology* **45**:1277-1287.
28. **Palma M, Worgall S, Quadri LN.** 2003. Transcriptome analysis of the *Pseudomonas aeruginosa* response to iron. *Arch Microbiol* **180**:374-379.
29. **Tian Z-X, Fargier E, Mac Aogáin M, Adams C, Wang Y-P, O'Gara F.** 2009. Transcriptome profiling defines a novel regulon modulated by the LysR-type

- transcriptional regulator MexT in *Pseudomonas aeruginosa*. *Nucleic Acids Research* **37**:7546-7559.
30. **Schuster M, Greenberg EP.** 2007. Early activation of quorum sensing in *Pseudomonas aeruginosa* reveals the architecture of a complex regulon. *BMC Genomics* **8**:1-11.
  31. **Wagner VE, Bushnell D, Passador L, Brooks AI, Iglewski BH.** 2003. Microarray analysis of *Pseudomonas aeruginosa* quorum-sensing regulons: effects of growth phase and environment. *Journal of Bacteriology* **185**:2080-2095.
  32. **Bredenbruch F, Geffers R, Nimtze M, Buer J, Häussler S.** 2006. The *Pseudomonas aeruginosa* quinolone signal (PQS) has an iron-chelating activity. *Environmental Microbiology* **8**:1318-1329.
  33. **Déziel E, Gopalan S, Tampakaki AP, Lépine F, Padfield KE, Saucier M, Xiao G, Rahme LG.** 2005. The contribution of MvfR to *Pseudomonas aeruginosa* pathogenesis and quorum sensing circuitry regulation: multiple quorum sensing-regulated genes are modulated without affecting *lasRI*, *rhIRI* or the production of *N*-acyl- L-homoserine lactones. *Molecular Microbiology* **55**:998-1014.
  34. **Damron FH, Owings JP, Okkotsu Y, Varga JJ, Schurr JR, Goldberg JB, Schurr MJ, Yu HD.** 2012. Analysis of the *Pseudomonas aeruginosa* regulon controlled by the sensor kinase KinB and sigma factor RpoN. *Journal of Bacteriology* **194**:1317-1330.
  35. **Schuster M, Hawkins AC, Harwood CS, Greenberg EP.** 2004. The *Pseudomonas aeruginosa* RpoS regulon and its relationship to quorum sensing. *Molecular Microbiology* **51**:973-985.
  36. **Burrowes E, Baysse C, Adams C, O'Gara F.** 2006. Influence of the regulatory protein RsmA on cellular functions in *Pseudomonas aeruginosa* PAO1, as revealed by transcriptome analysis. *Microbiology* **152**:405-418.
  37. **Wolfgang MC, Lee VT, Gilmore ME, Lory S.** 2003. Coordinate regulation of bacterial virulence genes by a novel adenylate cyclase-dependent signaling pathway. *Developmental Cell* **4**:253-263.
  38. **Hentzer M, Wu H, Andersen JB, Riedel K, Rasmussen TB, Bagge N, Kumar N, Schembri MA, Song Z, Kristoffersen P, Manefield M, Costerton JW, Molin S, Eberl L, Steinberg P, Kjelleberg S, Høiby N, Givskov M.** 2003. Attenuation of *Pseudomonas aeruginosa* virulence by quorum sensing inhibitors, vol. 22.
  39. **Farrell MJ, Finkel SE.** 2003. The growth advantage in stationary-phase phenotype conferred by *rpoS* mutations is dependent on the pH and nutrient environment. *Journal of Bacteriology* **185**:7044-7052.
  40. **Heurlier K, Dénervaud V, Pessi G, Reimann C, Haas D.** 2003. Negative control of quorum sensing by RpoN ( $\sigma$ 54) in *Pseudomonas aeruginosa* PAO1. *Journal of Bacteriology* **185**:2227-2235.
  41. **Medina G, Juárez K, Valderrama B, Soberón-Chávez G.** 2003. Mechanism of *Pseudomonas aeruginosa* RhIR Transcriptional Regulation of the *rhIAB* Promoter. *Journal of Bacteriology* **185**:5976-5983.

42. **Razo-Mejia M, Boedicker JQ, Jones D, DeLuna A, Kinney JB, Phillips R.** 2014. Comparison of the theoretical and real-world evolutionary potential of a genetic circuit. *Physical Biology* **11**:026005.
43. **Kuhlman T, Zhang Z, Saier MH, Jr., Hwa T.** 2007. Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **104**:6043-6048.
44. **Aow JS, Xue X, Run JQ, Lim GF, Goh WS, Clarke ND.** 2013. Differential binding of the related transcription factors Pho4 and Cbf1 can tune the sensitivity of promoters to different levels of an induction signal. *Nucleic Acids Res* **41**:4877-4887.
45. **Smith RP, Taher L, Patwardhan RP, Kim MJ, Inoue F, Shendure J, Ovcharenko I, Ahituv N.** 2013. Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nat Genet* **45**:1021-1028.
46. **Boedicker JQ, Garcia HG, Johnson S, Phillips R.** 2013. DNA sequence-dependent mechanics and protein-assisted bending in repressor-mediated loop formation. *Phys Biol* **10**:066005.
47. **Kim S, Brostroemer E, Xing D, Jin J, Chong S, Ge H, Wang S, Gu C, Yang L, Gao YQ, Su X-d, Sun Y, Xie XS.** 2013. Probing allostery through DNA. *Science* **339**:816-819.
48. **Segal E, Raveh-Sadka T, Schroeder M, Unnerstall U, Gaul U.** 2008. Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature* **451**:535-540.
49. **Brewster RC, Weinert FM, Garcia HG, Song D, Rydenfelt M, Phillips R.** 2014. The transcription factor titration effect dictates level of gene expression. *Cell* **156**:1312-1323.
50. **Tanay A.** 2006. Extensive low-affinity transcriptional interactions in the yeast genome. *Genome research* **16**:962-972.
51. **Kosuri S, Goodman DB, Cambray G, Mutalik VK, Gao Y, Arkin AP, Endy D, Church GM.** 2013. Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proceedings of the National Academy of Sciences* **110**:14024-14029.
52. **Kopf SH, Henny C, Newman DK.** 2013. Ligand-Enhanced Abiotic Iron Oxidation and the Effects of Chemical versus Biological Iron Cycling in Anoxic Environments. *Environmental Science & Technology* **47**:2602-2611.
53. **Choi K-H, Schweizer HP.** 2006. Mini-Tn7 insertion in bacteria with single attTn7 sites: example *Pseudomonas aeruginosa*. *Nature Protocols* **1**:153-161.
54. **Shanks RMQ, Caiazza NC, Hinsa SM, Toutain CM, O'Toole GA.** 2006. *Saccharomyces cerevisiae*-based molecular tool kit for manipulation of genes from gram-negative bacteria. *Applied and Environmental Microbiology* **72**:5027-5036.

55. **Dietrich LEP, Price-Whelan A, Petersen A, Whiteley M, Newman DK.** 2006. The phenazine pyocyanin is a terminal signalling factor in the quorum sensing network of *Pseudomonas aeruginosa*. *Molecular Microbiology* **61**:1308-1321.
56. **Stookey LL.** 1970. Ferrozine---a new spectrophotometric reagent for iron. *Analytical Chemistry* **42**:779-781.
57. **Winsor GL, Lam DKW, Fleming L, Lo R, Whiteside MD, Yu NY, Hancock REW, Brinkman FSL.** 2011. Pseudomonas Genome Database: improved comparative analysis and population genomics capability for *Pseudomonas* genomes. *Nucleic Acids Research* **39**:D596-D600.
58. **Garcia HG, Phillips R.** 2011. Quantitative dissection of the simple repression input-output function. *Proceedings of the National Academy of Sciences* **108**:12173-12178.

**Chapter V: Future Directions**

At the beginning of this project, I had the simple question: can bacteria distinguish between extracellular Fe(II) and Fe(III)? I found not only that *Pseudomonas aeruginosa* specifically sensed extracellular Fe(II) through BqsRS, but gained insight into how Fe(II) affects cellular processes. In the presence of Fe(II), BqsRS provided for spermidine homeostasis and protection against cationic stressors, including aminoglycoside and polymyxin drugs used to treat cystic fibrosis patients. Furthermore, many BqsR regulated genes were consistent with the hypothesis that Fe(II) served as a proxy for reducing and acidic environments, conditions where Fe(II) is stable. Although I discovered some mechanistic details of how BqsRS functions, much remains to be elucidated.

## **Future Experiments**

### *Metal binding experiments*

Previously, I determined that the RExxE motif in the periplasmic region of BqsS is critical for sensing Fe(II) *in vivo* (see chapter 4). In chapter 5, I construct a BqsR binding promoter-*lacZ* fusion and analyze the response with a fluorescent reporter. This makes experimental analysis of transcriptional response much faster than qPCR experiments. This system can be leveraged to further analyze the Fe(II)-BqsS interaction. Additionally, these types of experiments would be beneficial if BqsS proves to be too difficult to purify.

First, the metal specificity can be probed further. It would be interesting to test Co(II) because it similar in size and electronic properties to Fe(II). Although I previously



tested the transcriptional response to 5 mM  $\text{CaCl}_2$ , there has been some hints that BqsS responds to elevated Ca(II) concentrations. This includes a calcium signaling review referencing unpublished work (1) and that *bqsR* was approximately 2-fold repressed in a low calcium condition (2). However, the low calcium condition was induced by EGTA (2), which binds Fe(II) with a higher affinity than Ca(II), so this could be due to Fe(II) deprivation. Thus, it would be interesting to follow this literature and experimentally examine the transcriptional response to a range of Ca(II) concentrations. Furthermore, metal competition studies could be performed to assess if some metals are competitive inhibitors of the Fe(II) shock response.

The Fe(II)-sensing motif could be probed in more detail. The full sequence for the Fe(II)-sensing motif is REeaE. Testing variations of the RExxE sequence identified from a phylogenetic analysis (see Appendix B) would develop an understanding of the full Fe(II)-sensing potential of the observed motifs. For example, a RxExE motif is closely related to the Fe(II)-sensing motif. If RxExE could indeed bind Fe(II), then strains containing this motif are most likely Fe(II)-sensing bacteria. Additional mutations to examine include all combinations of single and double glutamate mutants. Also, changing the glutamates to aspartates should result in active Fe(II)-sensor and would serve as a positive control for the *in vivo* experiments (*i.e.* from mutants previously analyzed, it is possible that any mutation at this locus disrupts activity). Additionally E to D mutants would determine which of the D containing motif variants in the phylogenetic analysis are functional Fe(II) sensors. A more conservative substitution to generate a nonfunctional sensor would be E

to Q (rather than S or A used previously in chapter 3) as Q only substitutes an amino group for a carboxyl group. Also, a KExxE motif has been shown to bind Fe(II) (3), examining this change in BqsS would be interesting.

The Fe(II)-sensing motif remains to be biochemically characterized. Several attempts to purify the periplasmic domain of BqsS were performed. Beyond varying the standard induction parameters (time, temperature, and inducer concentrations) to obtain soluble protein, several affinity tags (His, Strep, and glutathione) were tested. Because the definition of the periplasmic region was based on predicted transmembrane helix cut-offs, multiple constructs were tested with slightly different edges (*i.e.* the total length of the periplasmic region was varied for the cut-off of the N and C terminal residues). New approaches to purification should include examining different constructs (see table S5-1), different expression strains, denaturing purification, or purifying the full length sensor [several similar full length sensor histidine kinases have been purified (4, 5)].

Once purified BqsS is obtained, metal binding studies could be performed with techniques such as isothermal calorimetry or differential scanning fluorimetry (6). These experiments could give the  $K_D$  not only for Fe(II), but also other metals such as Co(II) and Fe(III). The results of these experiments would give insight into the mechanism of metal selectivity. If the sensor only binds Fe(II), selectivity would be achieved solely through the sensor, whereas if the sensor binds several metals with similar strength, then the selectivity could be mediated through an additional factor, perhaps BqsP or BqsQ. Other metals could bind with higher affinity than Fe(II), but lock the sensor in an inactive

conformation. Metal competition studies would begin to address this; furthermore, the preference for Fe(II) in a more complex context would be elucidated. Metal competition studies would also be interesting to examine *in vivo*.

Ultimately, determining how Fe(II)-binding would be a vast improvement over the Fe(II)-sensing experiments completed. With purified BqsS, if the Fe(II) binding is strong enough, inorganic chemistry techniques such as  $^{57}\text{Fe}$ -Mössbauer spectroscopy could be used to understand the ligand environment in more detail. Additionally, x-ray crystallography of BqsS incubated with Fe(II) would show all residues in contact with Fe(II). Analysis of Fe(II) binding with site-directed mutant at key residues, including those in the RExxE motif, would further confirm the key Fe(II)-binding residues.

#### *Operon characterization*

In *P. aeruginosa* PA14, the *bqs* operon consists of two homologous single pass transmembrane proteins (*bqsP* and *bqsQ*), the response regulator (*bqsR*), the sensor histidine kinase (*bqsS*), and a hypothetical (*bqsT*) (see chapter 2 for further details). It would be interesting to study the function of *bqsP* and *bqsQ* because all strains with the RExxE motif within BqsS contain at least one copy of these genes (with an e-value < 10e-3 and 45% identity). These have homology to PePSY-type peptidases, which are known virulence factors (7). Also, they are homologous to Yycl, a short peptide known to modulate sensor kinase activity through its transmembrane domain; the function of the periplasmic domain is unknown (8, 9). Other single-pass membrane proteins have been

shown to modulate secondary two-component systems (10). BqsRS upregulates at least two other two-component systems; BqsP or BqsQ could bridge multiple systems. Performing transcriptional analysis of  $\Delta bqsPQ$ ,  $\Delta bqsP$ , and  $\Delta bqsQ$  (the first two clean deletion strains have already been made) would be interesting because it may modulate magnitude of the *bqs* operon expression. It would also be interesting to examine truncation mutants, both in the membrane domain and the periplasmic domain. The membrane domain may modulate the sensor's activity and the periplasmic domain could either act as a peptidase or help in Fe(II) sensing.

### *Regulon*

Several BqsR regulon members would be interesting to examine in greater detail. Compared with WT,  $\Delta bqsR/S$  have a severe growth defect that increases as the Fe(II) concentration increases (Fig. 5-1). I performed growth curves for several transposon mutants (annotation confirmed with PCR) from the PA14 transposon mutant library (11) in high Fe(II). The Fe(II) concentration required for a growth defect for most of these mutants was higher than for  $\Delta bqsR$  (Fig. 5-1). Since there is an observable phenotype associated with these genes, they are important for growth in high Fe(II) and would be interesting targets for further experiments.

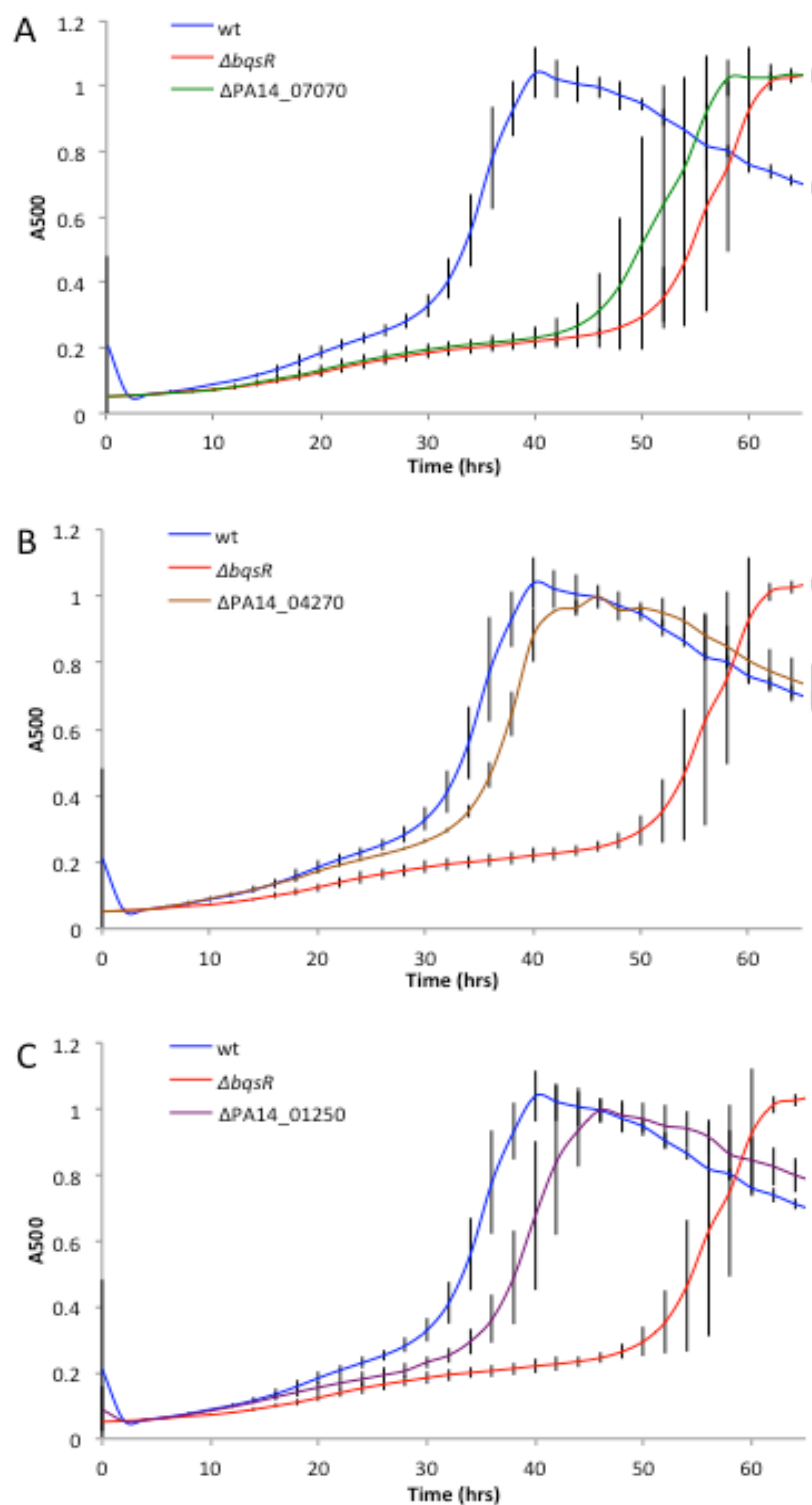


Figure 5-1. High Fe(II) growth defects for BqsR-regulon members. Cultures were grown in biological triplicate anaerobically at 37 °C in MOPS minimal media supplemented with 150

$\mu\text{M}$  Fe(II). Individual transposon mutant growth curves for PA14\_07070 (A), PA14\_04270 (B), and PA14\_01250 (C) are shown with WT and  $\Delta bqsR$  (see chapter 3 and appendix A for methods).

The mutant with the most severe growth defect in 150  $\mu\text{M}$  Fe(II) is PA14\_07070 and is almost as impaired as  $\Delta bqsR$  (Fig. 5-1). *Pseudomonas aeruginosa*, *P. stutzeri*, *P. chloritidismutans*, *P. xanthomarina*, and *Azotobacter* strains with the exact RExxE motif in BqsS (see Appendix B) contain a copy of this; the *Kushneria* strain with this motif contains a more distantly related homologue designated as ferredoxin (E-value =  $4\text{E}-7$ ). PA14\_07070 is a 48.2 kDa predicted ferric reductase, most likely containing a flavin (based on homology from pBLAST). This is an interesting target because in the context of high Fe(II), it most likely functions in the reverse direction oxidizing Fe(II) to Fe(III). This could serve as a defense against high Fe(II), which can be toxic anaerobically (12, 13), by effectively removing Fe from its environment, as Fe(III) is insoluble. The structure contains 6 transmembrane segments at the N-terminus with the majority of the soluble part of the protein facing the cytosol [predicted by DAS (14) and TMHMM (15)]. Purifying and verifying the Fe(II) oxidation capability of this protein would be interesting.

Another gene to examine in more detail is PA14\_04270, a predicted transcription factor. *Pseudomonas* and *Azotobacter* strains with the exact RExxE motif (see Appendix B) in BqsS contain a copy of this; the *Kushneria aurantia* strain does not (E-value <  $1\text{E}-3$ ). PA14\_04270 is a 36.2 kDa with 1 potential transmembrane domain predicted to be in the cytoplasmic membrane by TMHMM and contains a signal sequence detected by pSORTb (16). To examine its regulon, an RNA-Seq experiment comparing a time series after Fe(II)

shock with the strains  $\Delta$ PA14\_04270,  $\Delta$ bqsR, and WT could be used to distinguish what subset of the BqsR regulon is due to PA14\_04270. Also, a time course may identify other BqsR regulated genes. With some of the PA14\_04270 responsive genes identified, the consensus DNA binding motif could also be elucidated. It may also serve as mediator between different two-component systems, as suggested for BqsP/Q above.

PA14\_01250 may also be an interesting target for further study. *Pseudomonas*, *Azotobacter*, and *Kushneria aurantia* strains with the exact RExxE motif in BqsS (see Appendix B) contain a copy of this (E-value < 1E-4). It is a 54.8 kDa, 9-11 pass-transmembrane protein (TMHMM, DAS) predicted by pSORTb to be in the cytoplasmic membrane. Its predicted function is a sulfate permease. The upregulation and growth defect could be due to the counter-ion for Fe(II), because the Fe(II) source for most experiments has been ferrous ammonium sulfate. However, we did see it upregulated in the microarray where FeCl<sub>2</sub> was used (see chapter 2). It would be interesting to test this expression in a variety of Fe(II) sources with a different counter ion and see if it still is upregulated.

Finally, PA14\_04180 did not have a large growth defect (although the anaerobic chamber was not functioning properly and I have not been able to do the growth experiment at higher Fe(II) concentrations); however it is the most strongly upregulated gene, between 3000-6000 fold upregulated (chapters 2 and 3). Such a high level of upregulation makes this an interesting target. However, only the *Azotobacter* strain appears to have a homologue (E-value < 1E-3). The flanking genomic region on the left is

the polyamine permease (Pot operon) and a polyamine modification protein. From the genomic context, PA14\_04180 may be binding a polyamine precursor. It has N-terminal signal sequence and contains a poorly predicted transmembrane domain (DAS, TMHMM), and could therefore be in the cytoplasmic membrane, the periplasm, or extracellular with similar probability (pSORTb). It is a relatively small protein at 12.6 kDa that is a predicted bacterial OB-fold protein (BOF). BOFs contain a binding pocket for positively charged molecules. To examine this protein further, first determining if there is a growth defect in high Fe(II) would be important. Subsequently, the localization could be determined by affinity tagging the protein on the C-terminus with a Strep-tag (to avoid metal binding with His-tag and avoid interference with N-terminal the signal sequence). After an Fe(II) shock, the cell should be fractionated and labeled with antibodies to identify where PA14\_04180 expresses. Finally, the function of purified PA14\_04180 could be examined by testing Fe and amine binding. If PA14\_04180 does not bind Fe or amines, binding could be tested with a small molecule library and assessed with differential scanning fluorimetry (6).

### *Phenotypes*

The response mediated by BqsRS could be characterized further. The modifications to LPS and membranes could be biochemically assessed because transporter systems (PA14\_57830 - PA14\_57880, PA14\_01250, PA14\_04290), *oprH*, an outer membrane lipoprotein (PA14\_50740), and the LPS modifying *arn* operon (See



chapter 3 for full RNA-Seq results) are upregulated. It would be interesting to see if the surface charge of *P. aeruginosa* does indeed become more positive upon exposure to Fe(II). Additionally, since Fe(II) may serve as an environmental proxy for acidic and reducing conditions, it would be interesting to see if members of the BqsR regulon are upregulated in those conditions without Fe(II) present.

### **Conclusion**

By examining these research areas, much could be learned about Fe(II) sensing both on a biochemical level and on a physiological level. Understanding the nature of Fe(II) binding and how specificity is achieved would diversify the field of Fe chemistry because this type of Fe(II) binding is unique. A deep knowledge of how bacteria respond to their extracellular environment would allow for the development of biomarkers and therapeutic approaches to more effectively treat bacteria by targeting active processes. The BqsRS system is a good example of searching for an environmental parameter and then using laboratory techniques to find how a bacteria responds to it.

## References

1. **Dominguez DC, Guragain M, Patrauchan M.** 2014. Calcium binding proteins and calcium signaling in prokaryotes. *Cell calcium*. **57**:151-165.
2. **Wolfgang MC, Lee VT, Gilmore ME, Lory S.** Coordinate regulation of bacterial virulence genes by a novel adenylate cyclase-dependent signaling pathway. *Developmental Cell* **4**:253-263.
3. **Wedderhoff I, Kursula I, Groves MR, Ortiz de Orué Lucana D.** 2013. Iron binding at specific sites within the octameric HbpS protects *Streptomyces* from iron-mediated oxidative stress. *PLoS ONE* **8**:e71579.
4. **Türk M, Bierbaum G.** 2012. Purification and activity testing of the full-length YycFGHI proteins of *Staphylococcus aureus*. *PLoS ONE* **7**:e30403.
5. **Ma PK, Yuille HM, Blessie V, Gohring N, Igloi Z, Nishiguchi K, Nakayama J, Henderson PJF, Phillips-Jones MK.** 2008. Expression, purification and activities of the entire family of intact membrane sensor kinases from *Enterococcus faecalis*. *Molecular Membrane Biology* **25**:449-473.
6. **Niesen FH, Berglund H, Vedadi M.** 2007. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protocols* **2**:2212-2221.
7. **Yeats C, Rawlings ND, Bateman A.** 2004. The PepSY domain: a regulator of peptidase activity in the microbial environment? *Trends Biochem.Sci.* **29**:169-172.
8. **Szurmant H, Bu L, Brooks CL, Hoch JA.** 2008. An essential sensor histidine kinase controlled by transmembrane helix interactions with its auxiliary proteins. *Proceedings of the National Academy of Sciences* **105**:5891-5896.
9. **Szurmant H, Mohan MA, Imus PM, Hoch JA.** 2007. YycH and YycI interact to regulate the essential YycFG two-component system in *Bacillus subtilis*. *J. Bacteriol.* **189**:3280-3289.
10. **Buelow DR, Raivio TL.** 2010. Three (and more) component regulatory systems – auxiliary regulators of bacterial histidine kinases. *Molecular Microbiology* **75**:547-566.
11. **Liberati NT, Urbach JM, Miyata S, Lee DG, Drenkard E, Wu G, Villanueva J, Wei T, Ausubel FM.** 2006. An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proceedings of the National Academy of Sciences of the United States of America* **103**:2833-2838.
12. **Bird LJ, Coleman ML, Newman DK.** 2013. Iron and Copper Act Synergistically To Delay Anaerobic Growth of Bacteria. *Applied and Environmental Microbiology* **79**:3619-3627.

13. **Dunning JC, Ma Y, Marquis RE.** 1998. Anaerobic killing of oral *streptococci* by reduced, transition metal cations. *Applied and Environmental Microbiology* **64**:27-33.
14. **Cserzo M, Eisenhaber F, Eisenhaber B, Simon I.** 2002. On filtering false positive transmembrane protein predictions. *Protein Eng* **15**:745-752.
15. **Krogh A, Larsson B, von Heijne G, Sonnhammer EL.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**:567-580.
16. **Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FS.** 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**:1608-1615.

Table 5-S1. Suggestions for periplasmic BqsS constructs for purification. Strep-L30-A151 and GST-L30-R152 have been tested in *E. coli* BL21, with no condition yielding soluble protein, the other constructs remain to be tested in this strain. Other expression strains such as the ROSETTA strains should be tested since alternative strains yield dramatically different protein expression.

Construct	Test expression in <i>E. coli</i> BL21	Test different expression strain
Strep-L30-A151		X
*D32-R152	X	X
GST-L30-R152		X
D32-S149	X	X
F31-R152	X	X

**Appendix A:  $\Delta rhIR$  growth and  $\Delta bqsR$  rhamnolipid phenotype**

## Introduction

My work has shown that the two-component system (TCS), BqsRS, responds specifically to Fe(II) (chapter 2). I predicted regulon members and observed some associated phenotypes in high Fe(II) for the  $\Delta bqsR$  mutant including a severe growth defect, reduced spermidine production, and increased sensitivity to cationic antibiotics (chapter 3). Previously, the BqsRS two-component system has been implicated in modulating biofilm dispersal and quorum sensing (1). Dong *et al.* discovered this two-component system by noticing increased biofilm thickness in deletion mutants for this TCS (1). I have been unable to replicate this phenotype. However, Dong *et al.* also observed that  $\Delta bqsR/S$  produced less N-butyryl-L-homoserine lactone (C4-HSL), PQS, and rhamnolipids (1). *Pseudomonas aeruginosa* contained 3 quorum sensing systems: LasR/N-3-oxo-dodecanoyl-homoserine lactone (C12-HSL), RhIR/C4-HSL, and PqsR/PQS (2). One of the main products activated by the RhIR quorum sensing system was rhamnolipids (3). Rhamnolipids were associated with biofilm dispersal and swarming motility (4).

Since I was unable to reproduce the biofilm phenotype for  $\Delta bqsR/S$ , I wanted to ensure that I could reproduce at least one phenotype previously observed (1). I chose to compare the rhamnolipid production in  $\Delta bqsR$  and WT strains. If the deletion mutant showed less rhamnolipid production, then the corresponding quorum sensing deficiencies would most likely occur. Furthermore, because unlike Dong *et al.* I knew the BqsS ligand [Fe(II)], I could examine this phenomenon in greater detail. Biofilm studies were notoriously delicate and difficult to reproduce even within the same lab, whereas

measures of quorum sensing molecules and rhamnolipids were more consistent and quantitative. Since Dong *et al.* showed less 4C-HSL in  $\Delta bqsR$  (1), there may be a difference in the BqsR regulon expression when comparing expression in stationary phase between WT and a  $\Delta rhIR$  mutant.

My original intention was to compare wt,  $\Delta bqsR$ , and  $\Delta rhIR$  in both exponential (when quorum sensing is turned off) and stationary phase (when quorum sensing is turned on) to see how BqsR-mediated gene expression regulation interacts with the RhIR-based quorum sensing system. However, I found that the  $\Delta rhIR$  mutant grows much faster and to a greater density than WT and  $\Delta bqsR$  strains, making comparisons between strains impossible. Instead, I overexpressed RhIR in early exponential to compare with WT and  $\Delta bqsR$  (chapter 4). Here, I examine the  $\Delta rhIR$  mutant growth in more detail by examining growth different carbon sources and Fe conditions.

## Methods

*Growth conditions.* *P. aeruginosa* PA14 was grown both aerobically and anaerobically at 37 °C in MOPS minimal medium (MOMM) in acid washed glassware to ensure cells were Fe limited. The basic MOMM is composed of 40 mM succinate ( $C_4H_4Na_2O_4 \cdot 6H_2O$ ), 9.3 mM  $NH_4Cl$ , 2.2 mM  $KH_2PO_4$ , 25 mM  $KNO_3$ , 25 mM  $NaNO_3$ , 25 mM MOPS, and 25 mM NaMOPS pH 7.2. Additionally, immediately prior to inoculation 100  $\mu M$   $CaCl_2$ , 1  $\mu M$   $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$ , 1 mM  $MgSO_4$ , and trace metals were added (5). The Fe(II)

concentration used in each experiment is noted if at a different Fe(II) concentration.

Anaerobic work was performed in a Coy chamber with the 80% nitrogen, 15% carbon dioxide, and 5% hydrogen. For growth rhamnolipid extraction, cells were incubated aerobically in 1  $\mu$ M Fe(II) MoMM cultures for wt and  $\Delta bqsR$  overnight at 37 °C in quadruplicate. Inoculate 20 ml anaerobic 1  $\mu$ M Fe(II) MoMM cultures with 50 mM nitrate as the electron acceptor inoculated with 1% overnight for wt and  $\Delta bqsR$  in quadruplicate. Growth was monitored by Beckman Spec20. For growth comparing wt,  $\Delta bqsR$ , and  $\Delta rhIR$  in Fe-limited and Fe-replete conditions, strains were grown as above substituting 3.6  $\mu$ M Fe(II) to the overnights and anaerobic 5 ml cultures in Balch tubes for the Fe-replete condition in duplicate.

*Growth in various carbon sources.* WT,  $\Delta bqsR$ , and  $\Delta rhIR$  strains were grown in 96-well plates and the anaerobic BioTek Synergy 4 plate reader was used. All carbon sources were filter sterilized in 1 M stocks and added to MoMM base (prepared without a carbon-source) at 40 mM final concentration. The carbon sources were not normalized for carbon number because the cells were grown in Fe-limited conditions. The lack of Fe should affect growth before the concentration of carbon. The carbon sources tested were pyruvate, succinate, acetate, glucose, glutamate, and mannitol. The length of the overnights was varied so they reached the same OD at the time of anaerobic inoculation. Three replicates of overnights were grown in Fe-limited conditions in MoMM with acetate and mannitol for 36 hours, pyruvate and glucose for 24 hours, and succinate and



glutamate for 16 hours. Cultures were inoculated to  $OD_{500} = 0.02$  anaerobically in triplicate.

*Fe(II) shock for rhamnolipid extraction.* Cells were harvested between  $OD_{500} = 0.77 - 0.87$  and shocked with  $200 \mu\text{M Fe(II)}$  at room temperature. 1 ml samples were taken after 30 min and overnight for both Fe(II) shocked cells and unshocked cells. When cells were harvested they were centrifuged at max speed for 3 min and the supernatant saved for rhamnolipid analysis. Samples were stored at  $-20 \text{ C}$ .

*Rhamnolipid extraction and visualization.*  $500 \mu\text{l}$  of the frozen supernatant rhamnolipids were thawed and extracted 3 times with an equal volume of ethyl acetate in 1.7 ml eppendorf tubes. For the ethyl acetate extraction, the aqueous and organic phases were mixed well and centrifuged in the microfuge for 30 sec to separate. The organic phase, the top layer, from each extraction were pooled and evaporated to dryness with argon in glass GC vials. Samples were resuspended in  $100 \mu\text{l}$  of methanol.  $5 \mu\text{l}$  of each sample and  $5 \mu\text{l}$  of  $0.1 \text{ mg/ml}$  of rhamnolipid standard were spotted on TLC plates (silica 60). The running buffer was a mixture of chloroform, methanol, and acetic acid in a ratio of 65:15:2. To visualize the rhamnolipids on the TLC-plates, the plates were covered with a detection agent consisting of  $0.15 \text{ g}$  orcinol,  $8.2 \text{ mL}$  sulfuric acid (60%), and  $42 \text{ mL}$  deionized water. For preservation, dried plates were incubated at  $110^\circ\text{C}$  for 10 min.

## Results

### *Rhamnolipid analysis*

Previously most Fe(II) shock experiments were performed in early exponential phase, however rhamnolipids start being produced in stationary phase (3). First, to determine the optimal OD for rhamnolipid extraction, I extracted rhamnolipids at several ODs over a growth curve, which plateau between  $OD_{500} = 0.85-1.2$ . I harvested the cells at  $OD_{500} = 0.8$  because the cells consistently grew to this OD and rhamnolipids show increased concentration. The closest comparison between the Dong *et al.* (1) rhamnolipid extraction and my rhamnolipid extraction is comparing unshocked WT and  $\Delta bqsR$  cells (Fig. A-1). The difference between these extractions is that Dong *et al.* (1) performed the extraction from cells grown in aerobically LB and I performed the extraction anaerobically in MoMM. Figure A-1 shows that without Fe(II) shock WT has more rhamnolipids than  $\Delta bqsR$  after both 30 min (Fig. A-1A) and overnight (Fig. A-1B) at room temperature. Unshocked WT cells with 1  $\mu$ M Fe samples show slightly more rhamnolipid production when comparing the 30 min time point to the overnight time point at room temperature (Fig. A-2A). It appears as though the overnight  $\Delta bqsR$  also has more rhamnolipids than the 30 min time point, but it is a little difficult to discern because the TLC plate ran poorly (Fig. A-2B).

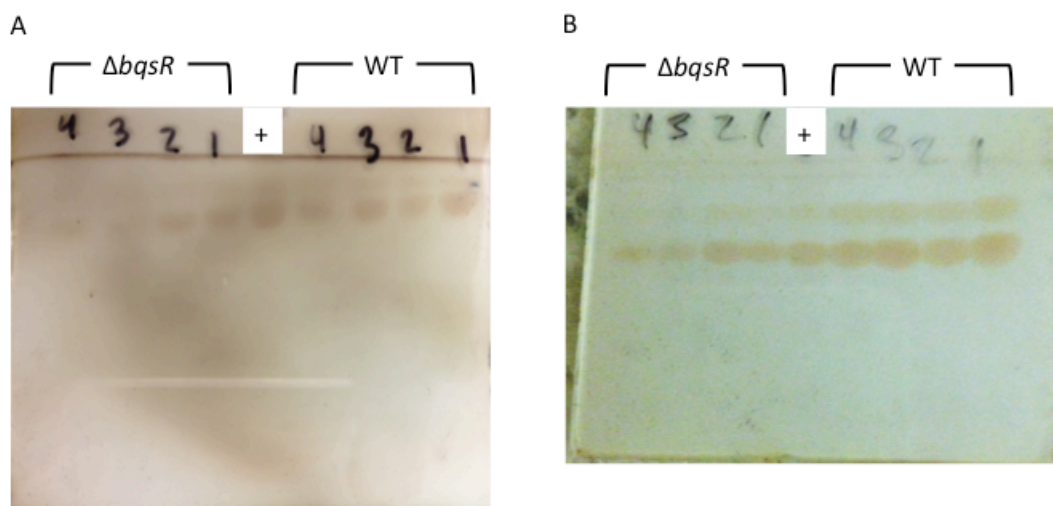


Figure A-1. Comparison of rhamnolipid production for unshocked cells comparing  $\Delta bqsR$  and WT. Lanes labeled 1-4 indicate biological replicates of that condition. A) Rhamnolipid extraction at the 30 min time point. B) Rhamnolipid extraction at the overnight time point.

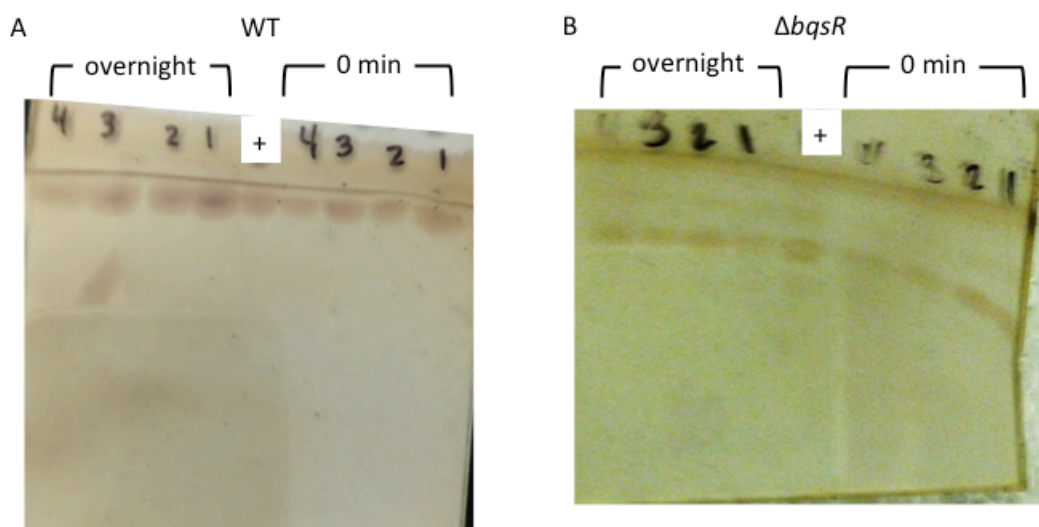


Figure A-2. Comparison of rhamnolipid production for unshocked cells comparing production over time for WT(A) and  $\Delta bqsR$  (B) strains. Lanes labeled 1-4 indicate biological replicates of that condition.

Since I know that Fe(II) activates BqsR, I further consider how Fe(II) shock affects rhamnolipid production in WT and  $\Delta bqsR$ . The concentration of rhamnolipids in the samples is examined at 30 min, the length of Fe(II) shock used for qPCR analysis.

Overnight samples are also taken because it may take longer for the cells to produce the correct enzymes and synthesize rhamnolipids than just the first step of eliciting a transcriptional response. There is no significant difference after 30 min of Fe(II) shock in WT or  $\Delta bqsR$  (Fig. A-3). It is difficult to discern because the TLC plate in Fig. A-4A ran poorly, but it looks like WT incubated overnight with Fe(II) has more rhamnolipids than unshocked cells. Overnight Fe(II) shocked  $\Delta bqsR$  samples also have more rhamnolipid than  $\Delta bqsR$  without Fe shock (Fig. A-4). Thus,  $\Delta bqsR$  has less rhamnolipid regardless of Fe(II) shock than WT, and Fe(II) shock increases rhamnolipid production in both WT and  $\Delta bqsR$ . These results indicate that there are both Fe(II) and BqsR-dependent and independent factors that affect production of rhamnolipids.

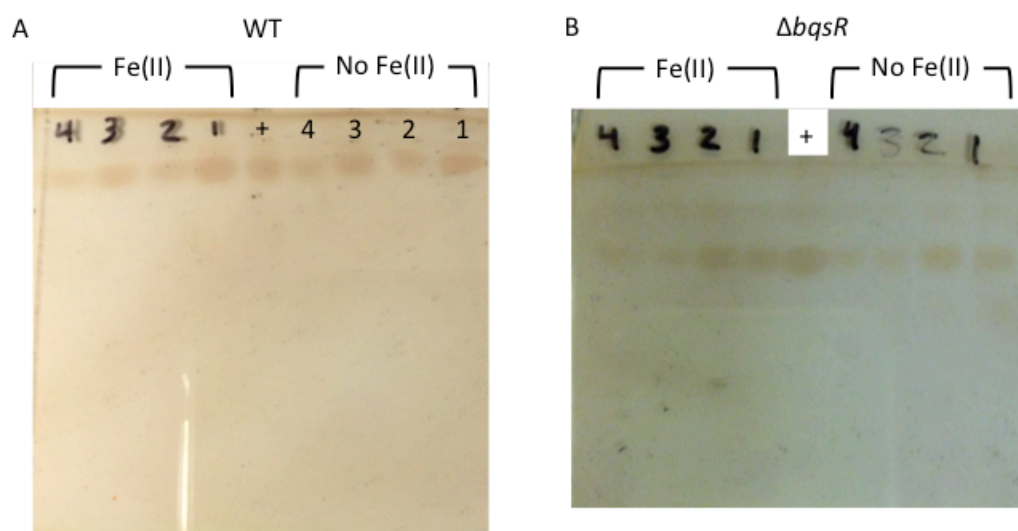


Figure A-3. Comparison of rhamnolipid production comparing Fe(II) shocked to unshocked cells at the 30 min time point for WT(A) and  $\Delta bqsR$  (B) strains. No significant difference in rhamnolipid concentration was observed between the unshocked and Fe(II) shocked cells. Lanes labeled 1-4 indicate biological replicates of that condition.

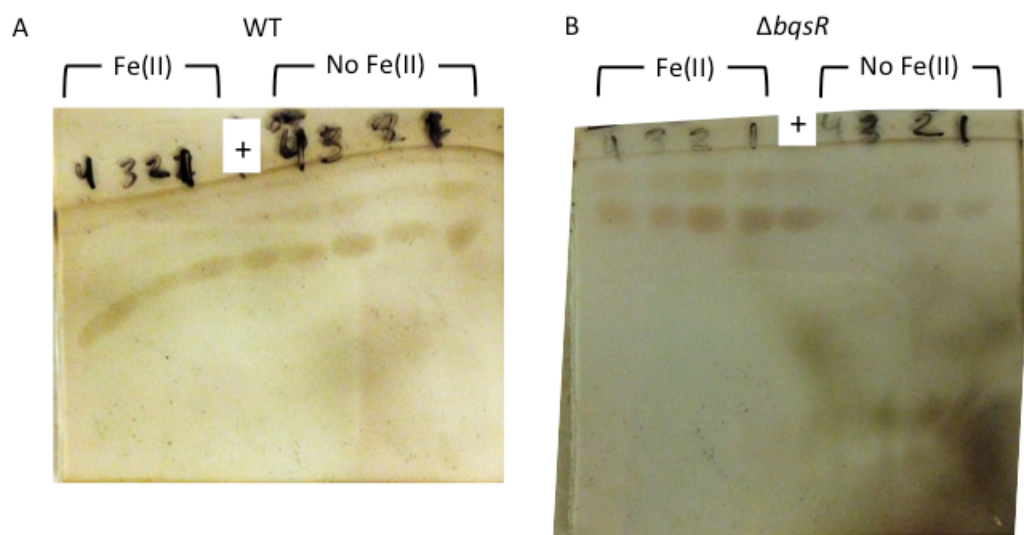


Figure A-4. Comparison of rhamnolipid production comparing Fe(II) shocked to unshocked cells at the overnight time point for WT(A) and  $\Delta bqsR$  (B) strains. Lanes labeled 1-4 indicate biological replicates of that condition.

#### *Growth in various carbon sources*

Once I was certain that rhamnolipids are produced, and that rhamnolipid production had a BqsR-dependent mechanism, I wanted to examine how  $\Delta rhIR$  affected the BqsR regulon. I was unable to examine the differences within the BqsR-regulon comparing WT,  $\Delta bqsR$ , and  $\Delta rhIR$  strains in stationary phase because  $\Delta rhIR$  grew much faster and to a much greater density than either WT or  $\Delta bqsR$ . Additionally, anaerobic  $\Delta rhIR$  at the end of a growth curve turned a bright yellow color that fluoresced white under UV, which might be consistent with the siderophore pyoverdine production. Therefore, the difference between  $\Delta rhIR$  and WT growth may be due to  $\Delta rhIR$ 's ability to obtain more Fe in the 1  $\mu\text{M}$  Fe-limited growth conditions. To test this I grew the cells in both Fe(II)-limited [1  $\mu\text{M}$  Fe(II)] and Fe(II)-replete [3.6  $\mu\text{M}$  Fe(II)]. While all strains grow to

a higher maximum OD in Fe(II)-replete conditions,  $\Delta rhIR$  outpaces all other strains regardless of Fe(II) concentration (Fig. A-5).

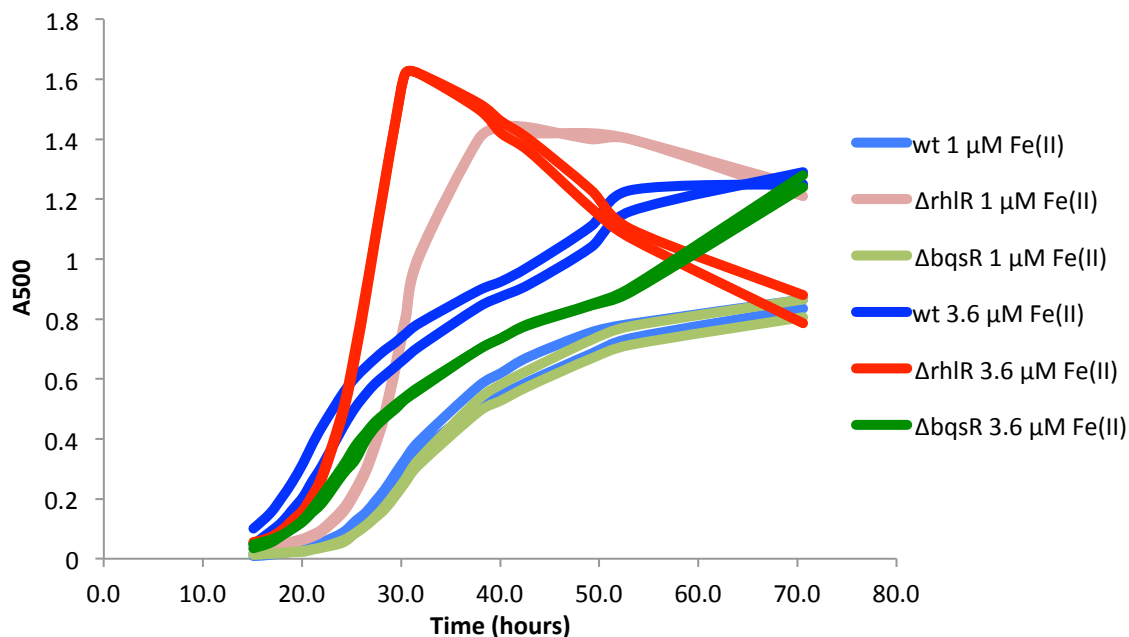


Figure A-5. Growth curves in MoMM with succinate as a carbon source in Fe(II)-limited (1  $\mu$ M) in dark colors and Fe(II)-replete (3.6  $\mu$ M) in pastel colors for WT,  $\Delta bqsR$ , and  $\Delta rhIR$ .

To examine the  $\Delta rhIR$  growth phenomenon in more detail, cells were grown in a variety of different carbon sources that spanned a range of carbon source preferences from very preferred TCA cycle intermediates (6) to mannitol, extensively used to derepress catabolite repression (6). This phenomenon could be confined to rapid growth conditions and not observed in poor carbon sources. Because  $\Delta rhIR$  cells grew faster and denser than WT in both Fe-limited and Fe-replete conditions, the growth curves were performed in Fe-limited conditions without normalization for carbon number. Figure A-6 shows representative graphs of biological triplicates replicates for growth curves in each carbon source. While the lag phase differed between 96-well plates grown on different

days, the trends between and within carbon sources for each strain remained constant. Generally WT and  $\Delta bqsR$  grow very similarly, and  $\Delta rhIR$  grows faster and to a higher final density. Since the cells were grown in Fe-limited conditions there is no growth defect in  $\Delta bqsR$  compared to WT growth as occurs during growth in high Fe(II) (see chapter 3).

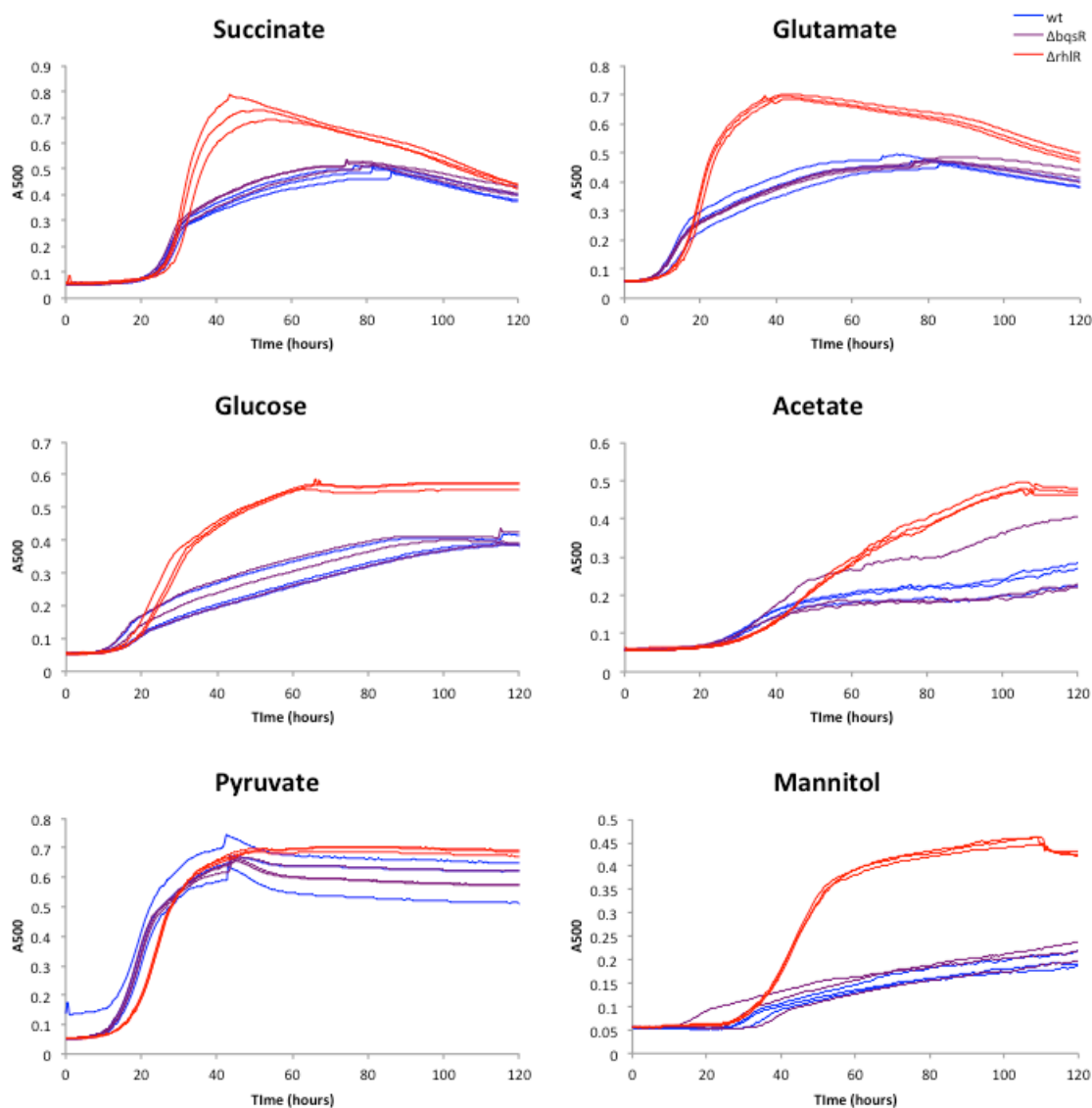


Figure A-6. Growth curves in MoMM with different carbon sources in Fe(II)-limited ( $1 \mu\text{M}$ ) for WT (blue),  $\Delta bqsR$  (purple), and  $\Delta rhIR$  (red).

The growth curves in succinate (Fig. A-6A) reveal that  $\Delta rhIR$  grows more quickly in exponential phase with a doubling time of  $3.6 \pm 0.2$  hours, whereas both WT and  $\Delta bqsR$  have a doubling time of  $4.4 \pm 0.2$  hours. The length of exponential phase is about twice as long for  $\Delta rhIR$  as for WT and  $\Delta bqsR$ . The maximum density between  $\Delta rhIR$  and WT/ $\Delta bqsR$  is different.  $\Delta rhIR$  achieves an OD of 0.8 whereas WT and  $\Delta bqsR$  only grow to an OD of 0.5. There is a severe die off in the  $\Delta rhIR$  strains that WT and  $\Delta bqsR$  don't seem to have. The features of the growth curves in glutamate (Fig. A-6B) grow similarly to succinate. The WT and  $\Delta bqsR$  strains have a slightly longer doubling time than  $\Delta rhIR$  (Table A-1).

	succinate doubling time
$\Delta rhIR$	$3.4 \pm 0.2$
$\Delta bqsR$	$4.2 \pm 0.2$
WT	$4.3 \pm 0.2$
	glutamate doubling time
$\Delta rhIR$	$3.9 \pm 0.4$
$\Delta bqsR$	$4.2 \pm 0.7$
WT	$4.3 \pm 0.7$
	glucose doubling time
$\Delta rhIR$	$6.9 \pm 0.6$
$\Delta bqsR$	$7.9 \pm 1.3$
WT	$7.6 \pm 1.0$
	acetate doubling time
$\Delta rhIR$	$13.3 \pm 0.9$
$\Delta bqsR$	$13.8 \pm 2.1$
WT	$12.8 \pm 1.6$
	pyruvate doubling time
$\Delta rhIR$	$4.5 \pm 0.4$
$\Delta bqsR$	$4.2 \pm 0.2$
WT	$4.6 \pm 0.8$
	mannitol doubling time
$\Delta rhIR$	$8.7 \pm 0.2$
$\Delta bqsR$	$23.5 \pm 22$
WT	$24.0 \pm 21$



Table A-1. Doubling times are from 3-6 biological replicates calculated from the period when the cells are growing exponentially.

In glucose  $\Delta rhIR$  grows exponentially for much longer, about 24 hours, rather than the approximately 7 hours for WT and  $\Delta bqsR$ . The rate of exponential growth in glucose is not significantly different between strains, rather  $\Delta rhIR$  continues to grow exponentially much longer (24 hours) and the highest density is OD of 0.5-0.6, whereas WT and  $\Delta bqsR$  grow to a maximum density of OD<sub>500</sub> = 0.4. Interestingly, the severe death for the  $\Delta rhIR$  strain observed in the more preferred C-sources, glutamate and succinate, does not occur in growth glucose. For cells grown in acetate in Fig. A-6 there is a large difference in final density between  $\Delta rhIR$  and the other strains, however there is a large range of final densities for  $\Delta bqsR$  and WT, which is seen in multiple trials. Similar to growth in glucose, the rate of exponential growth in acetate is not significantly different between strains, instead  $\Delta rhIR$  continues to grow in exponential phase longer. In contrast to all other carbon sources, in pyruvate there is no difference between any of the strains. In mannitol, none of the cells grow particularly well. However, the trends are more similar to glutamate and succinate, where  $\Delta rhIR$  doubles quicker (Table A-1) and with a longer duration in exponential phase and achieves a higher final density (OD<sub>500</sub> = 0.5). WT and  $\Delta bqsR$  only grow to an OD<sub>500</sub> = 0.2. Unlike in the more preferred carbon sources, there was no large and immediate  $\Delta rhIR$  die-off (Fig. A-6).

## Discussion

Here, I have confirmed that in *Pseudomonas aeruginosa* PA14  $\Delta bqsR$  produces less rhamnolipids than WT. Fe(II) shock increases rhamnolipid production in both WT and  $\Delta bqsR$ ; however, regardless of iron concentration WT always produces more rhamnolipids than  $\Delta bqsR$ . These results indicate that there are both Fe(II) dependent and independent factors that affect production of rhamnolipids in both strains. Multiple factors must integrate to determine the total amount of rhamnolipid produced. A more quantitative analysis is required to understand the mechanisms at play. However, this supports the hypothesis that RhIR and BqsR control a subset of overlapping genes.

Intriguingly,  $\Delta rhIR$  grows longer in exponential phase and to a higher peak density compared to WT in MoMM supplemented with most carbon sources assessed. To my knowledge, no similar growth phenotype has previously been observed for  $\Delta rhIR$ . In fact, in  $\Delta rhII$  mutant (produces the autoinducer, C4-HSL, that RhIR senses), without autoinducer supplementation (this should behave the same as a  $\Delta rhIR$  strain), had a slight growth defect anaerobically in LB + 100  $\mu$ M nitrate (7). I also grew the  $\Delta bqsR$  strain in all conditions, but since Fe(II) concentrations were low, there was no significant growth differences observed between  $\Delta bqsR$  and WT. When comparing growth between different carbon sources for the WT strain, the most preferred carbon sources are succinate and glutamate followed by glucose, acetate, and mannitol. Mannitol is clearly an extremely poor substrate for *P. aeruginosa*. In pyruvate, WT grows to the highest final density in the shortest time. In the literature, Catabolite repression is most strongly turned on for organic acids such as succinate and acetate and amino acids (e.g.,

glutamate) (6). Less strong catabolite repression is observed for glucose. Mannitol is the least preferred substrate; it is often used in catabolite repression studies to see when catabolite repression is completely turned off (6). Pyruvate does not exhibit catabolite repression and is used simultaneously in the presence of other carbon sources (8).

*ΔrhIR* grew strangely considering WT and the typical carbon source preferences observed in these experiments. While growth rates were highest for glutamate and succinate, once *ΔrhIR* reached its highest density, a large population die-off occurred. This die-off was not apparent for any of the less preferred carbon sources. The highest density was similar for all carbon sources, except acetate and mannitol had approximately 2/3 less density. Difference in carbon number does not account for these observed differences. Previously, it has been observed that different carbon sources affect rhamnolipid production (9), perhaps this can account for some of the growth differences observed. An alternative hypothesis is since Fe(II) was limiting – these pathways require more Fe-consumption. *ΔrhIR* develops a yellow color that fluoresces under UV which might be pyoverdine. Although I initially tested growth in both 1 μM Fe(II) (Fe-limited) and 3.6 μM Fe(II), which should be a Fe-replete condition, perhaps higher Fe concentrations would show less of a growth difference between strains. This may account for the growth phenotype, as *ΔrhIR* may obtain a larger quantity of Fe more quickly and thus may be able to grow in exponential phase longer and reach a higher final density more quickly. It would also be interesting to see how *ΔrhIR* grows in high Fe(II), where *ΔbqsR* has a growth defect, since BqsR and RhIR have overlapping regulons (see chapter 4).



## References

1. **Dong YH, Zhang XF, An SW, Xu JL, Zhang LH.** 2008. A novel two-component system BqsS-BqsR modulates quorum sensing-dependent biofilm decay in *Pseudomonas aeruginosa*. *Commun Integr Biol* **1**:88-96.
2. **Willcox MD, Zhu H, Conibear TC, Hume EB, Givskov M, Kjelleberg S, Rice SA.** 2008. Role of quorum sensing by *Pseudomonas aeruginosa* in microbial keratitis and cystic fibrosis. *Microbiology* **154**:2184-2194.
3. **van Ditmarsch D, Xavier J.** 2011. High-resolution time series of *Pseudomonas aeruginosa* gene expression and rhamnolipid secretion through growth curve synchronization. *BMC Microbiology* **11**:140.
4. **Caiazza NC, Shanks RMQ, O'Toole GA.** 2005. Rhamnolipids Modulate Swarming Motility Patterns of *Pseudomonas aeruginosa*. *Journal of Bacteriology* **187**:7351-7361.
5. **Kopf SH, Henny C, Newman DK.** 2013. Ligand-Enhanced Abiotic Iron Oxidation and the Effects of Chemical versus Biological Iron Cycling in Anoxic Environments. *Environmental Science & Technology* **47**:2602-2611.
6. **Rojo F.** 2010. Carbon catabolite repression in *Pseudomonas*: optimizing metabolic versatility and interactions with the environment **34**:658-684.
7. **Toyofuku M, Nomura N, Fujii T, Takaya N, Maseda H, Sawada I, Nakajima T, Uchiyama H.** 2007. Quorum sensing regulates denitrification in *Pseudomonas aeruginosa* PAO1. *Journal of Bacteriology* **189**:4969-4972.
8. **Collier DN, Hager PW, Phibbs Jr PV.** 1996. Catabolite repression control in the *Pseudomonads*. *Research in Microbiology* **147**:551-561.
9. **Santos A, Sampaio A, Vasquez G, Santa Anna L, Pereira N, Jr., Freire DG.** 2002. Evaluation of different carbon and nitrogen sources in production of rhamnolipids by a strain of *Pseudomonas aeruginosa*. *Appl Biochem Biotechnol* **98-100**:1025-1035.

**Appendix B: Phylogenetic analysis of BqsS**

## Introduction

*Pseudomonas aeruginosa* is best known as the dominant pathogen in cystic fibrosis (CF) patients. Fe(II) is present at elevated concentrations within CF patient sputum and Fe(II), not Fe(III), correlates with severe lung function decline (1). Fe(II) concentrations within CF sputum have been measured approaching 300  $\mu$ M Fe(II) (1). Therefore, coping with high concentrations of Fe(II) is critical for *P. aeruginosa* survival. In fact, deletion mutants for the Fe(II)-sensing two-component system, BqsRS, have a significant growth defect in high concentrations of Fe(II) (see chapter 3). Fe(II) sensing is mediated through the RExxE Fe(II)-sensing motif in the periplasmic domain of the sensor histidine kinase, BqsS (see chapter 3). It is somewhat surprising that Fe(II) would be bound by glutamates, as they tend to prefer Fe(III) ligands. We have not biochemically verified that the RExxE motif in BqsS directly binds iron. However, a crystal structure of an unrelated soluble protein from *Streptomyces reticuli*, HbpS, binds Fe(II) through K/RExxE motifs (2).

Examining how broadly distributed homologues for BqsS that contain the RExxE iron sensing motif, will give us insight into how common this Fe(II)-sensing strategy is. We could distinguish whether only *Pseudomonas* strains or most bacteria use this motif to sense Fe(II). Comparing whether pathogens and environmental strains contain a variant of the Fe(II)-binding motif will allow us to speculate if this sensor is a pathogenicity trait. Additionally, looking at the sequence diversity of potential Fe(II)-binding motifs will allow us to generate new hypothesis to broaden the definition of potential Fe(II)-binding motifs.

The only experimentally verified Fe(II)-sensing motif is the RExxE. In *P. aeruginosa* PA14, the full motif is REEAE. However, several variations of this motif may also bind Fe(II). For example, replacing the glutamates with aspartate residues should generate a mutant capable of interacting with Fe(II) because aspartate has the same carboxylic acid functional group and is only one carbon shorter. Additionally, glutamate in the third position may be able to replace the function of the glutamate at the second position (i.e., RxExE). Finally, since KExxE motif has previously been shown to bind Fe(II) (2), replacement of R to K should still allow Fe(II) binding.

## **Materials/Methods**

*Retrieve and align homologous sequences.* Using BqsS from *P. aeruginosa* (gi 116048575) as a query, protein BLAST (3) was used to search the non-redundant database, excluding uncultured/environmental sample sequences, and return 500 sequences for both the full length BqsS and the periplasmic BqsS domain. The resulting sequences were aligned using MAFFT version 7 (4) using the iterative refinement method, L-INS-i. The results from MAFFT are saved in the FASTA format. To convert to a format suitable for tree construction, Seaview (5) was used to convert to PHYLIP format. For all of the programs listed above, assume the default settings were applied to input unless otherwise indicated.



*Tree construction and visualization.* A tree was generated from the PHYLIP format of the aligned sequences with the program PhyML 3.0 (6) with the data type selected as amino-acids and the sequence file as interleaved. The default substitution model and tree searching parameters were used; the branch supports were aLRT. The tree was visualized and modified in iTOL (7). I manually annotated and categorized variants of the RExxE (REEAE in *Pseudomonas aeruginosa*) Fe(II)-sensing motif retrieved from BlastP. If the aligned sequence has no homology to the Fe(II)-sensing motif, white was assigned to that sequence (iTOL requires a color for all leaves in the tree). The number associated with color codes can be found on iTOL. I relabeled the leaves of the tree in iTOL, replacing the automated designators (GI number with some further information). The new leaf labels are the organism name and GI number (number only) to distinguish between multiple strains with the same name (e.g. multiple *Pseudomonas aeruginosa* strains). The approximately 150 sequences of the 500 retrieved sequences did not have a variant of the motif. These sequences fell in a sister clade to those that had the motif, suggesting that the motif evolved after these groups diverged. Since they were not of interest in this study they were removed from further analysis. The outgroup was defined as the clade sister to the last colored Fe(II)-binding motif, and the tree was rooted there. Clades were compressed for those which had many representatives of the same species.

*16S rRNA sequence retrieval and alignment.* 16S rRNA FASTA sequences of *Pseudomonas spp.*, *Cellvibrio spp.*, and *Azotobacter spp.* were obtained from JGI (8) and NCBI. If there were multiple 16S rRNA sequences for the same strain, only one was necessary because

all were identical. Unfortunately, only one full 16S rRNA sequence for *Azotobacter* species exists (*A. vinelandii* DJ at 1535 bp), so I took the largest partial sequences at 1537 bp for *Azotobacter* sp. Az15 and 1,528 bp for *Azotobacter chroococcum* strain 10006. MAFFT version 7 was used with the same parameters above to align the 16S rRNA sequences and Seaview used to convert alignment to PHYLIP format. I relabeled the leaves of the tree in iTOL, replacing the automated designators with the organism name. The *E. coli* 16S rRNA was used as the outgroup to root the tree.

## Results

### *Full length BqsS*

The most important result was that the RExxE Fe(II)-sensing motif was confined to the Pseudomonadaceae family. The vast majority of the strains that contained any variation of the Fe(II)-sensing motif are with the *Pseudomonas* genus with only 2 exceptions (Fig. B-1). Highlighted in red in Fig. B-1 are the other Pseudomonadaceae members, *Cellvibrio* and *Azotobacter* strains, containing a variant of the Fe(II) binding motif. In cyan (RxExE), there were many *Pseudomonas* strains, but also at the transition between *Pseudomonas* to other genus' *Cellvibrio japonicus* contained an RxExE and *Cellvibrio* sp. BR. did not have any conserved motif. The experimentally validated RExxE Fe(II)-sensing motif was confined to *Pseudomonas* strains, except where *Azotobacter vinelandii* was surrounded by *Pseudomonas* strains. Members of the same strain generally have the same Fe(II)-binding motif variant. Additionally, the variants of the Fe(II)-sensing motif clustered together without many gaps, e.g., the motif RxDxE or RxExD (shown in

dark yellow) cluster together. The motifs where the Fe(II) binding capability is the most uncertain [RxDxE or RxExD (dark yellow)] are also the furthest branching clades from the known Fe(II)-sensing motif RExxE motif (blue).

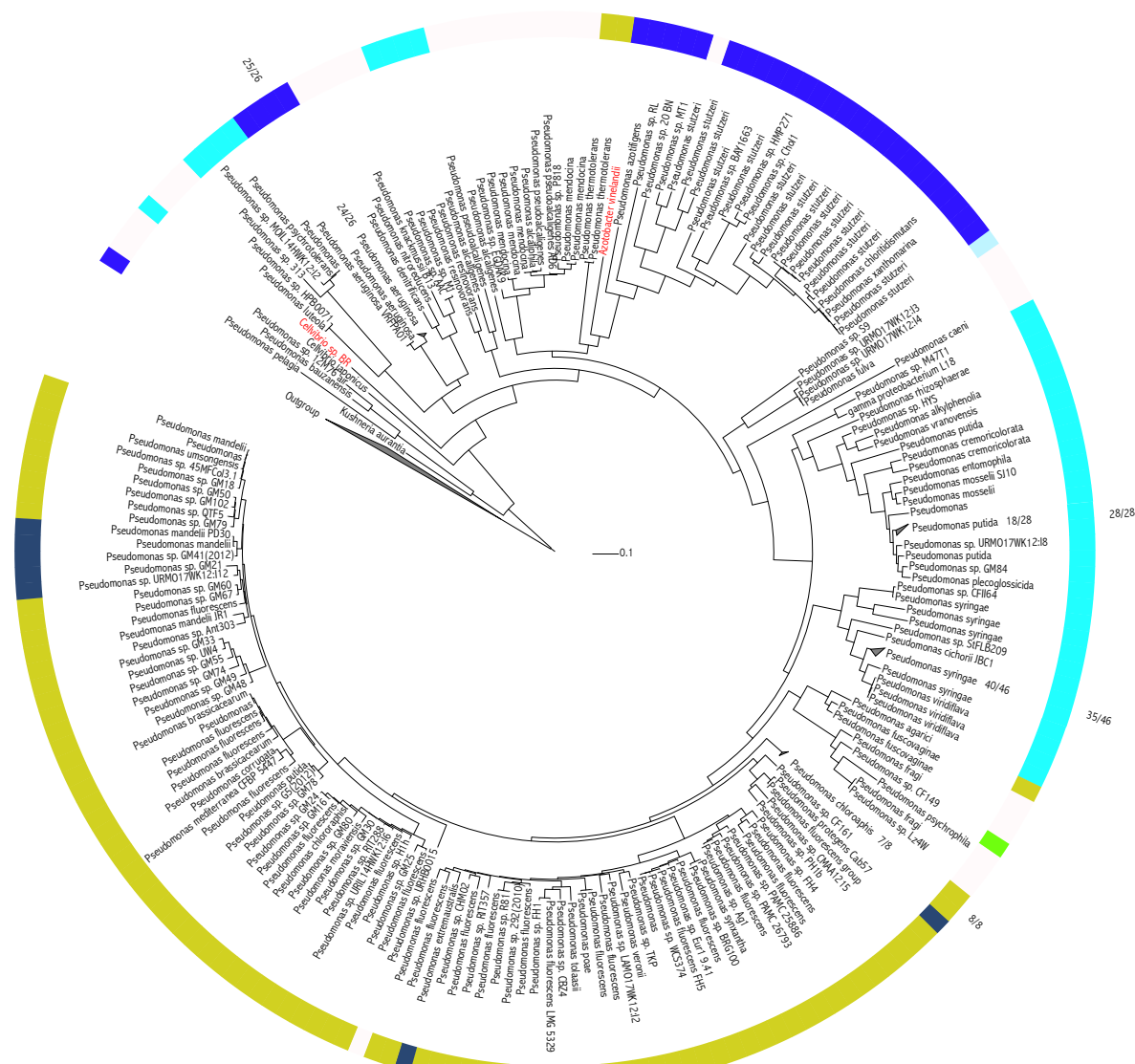


Figure B-1. The BqsS tree all sequences containing an Fe(II)-binding motif with strains without a motif collapsed into the outgroup, where the tree was rooted. The color indicates the variant of the Fe(II)-sensing motif: blue is RExxE, cyan is RxExE, dark yellow is RxDxE or RxExD, light cyan is KxExE, light green is KxDxE or KxExD, and navy is RDxxE or RExxD. The organisms with red text are the most distantly related phylogenetically.

Triangles represent collapsed clades. The fraction immediately following the strain name is the number of the named strain in that clade out of the total number of strains within the clade. The fraction outside of the color strip indicates the number of strains within the clade that contain the corresponding Fe(II)-binding motif out of the total number of strains within the collapsed clade. The scale bar represents 0.1 substitutions per site.

### 16s

Because it was unusual to observe a species clustering in the middle of different one (the *Azotobacter* strain surrounded by *Pseudomonas* strains) and that out of 2 *Cellvibrio* strains found on the tree only one contained an Fe(II)-sensing motif, a 16s tree was constructed with multiple *Pseudomonas*, *Cellvibrio*, and *Azotobacter* sequences (Fig. B-2). One possible explanation for the *Azotobacter* strain that appeared surrounded by many *Pseudomonas* sequences is that the *Azotobacter* was misannotated. The potentially misannotated sequence could actually be a *Pseudomonas* sequence. The *Azotobacter* strains clustered although they branched within the clade that contains *Pseudomonas psychrotolerans* L19. Therefore, they were more closely related to *Pseudomonas*. However, misannotation of *Azotobacter* cannot be ruled out because the *Azotobacter vinelandii* strain was the only complete sequence; alignments with partial sequences are not conclusive. The *Cellvibrio* strains clustered together and are clearly more distantly related to *Pseudomonas*. The *Pseudomonas* strains clustered together.

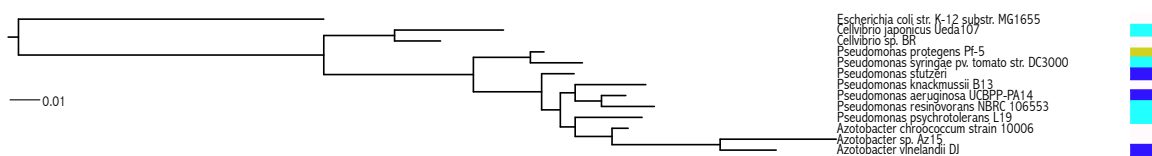


Figure B-2. The 16s tree constructed with multiple *Pseudomonas*, *Cellvibrio*, and *Azotobacter* sequences with *E. coli* as the outgroup. The color indicates the variant of the Fe(II)-sensing motif: blue is RExxE, cyan is RxExE, and dark yellow is RxDxE or RxExD. The scale bar represents 0.01 substitutions per site.

#### *Periplasmic region of BqsS*

The predicted periplasmic region of BqsS was examined in an attempt to capture more sequence diversity, perhaps pull out Fe(II) binding proteins that weren't two-component sensor histidine kinases. However, the BLAST results were all two-component systems. These results are very similar to the full length BqsS phylogenetic alignments, the only difference being that the *Azotobacter* is on the edge of the true RExxE motifs rather than in the middle of them (which makes more sense phylogenetically). But essentially both BLAST results show that the RExxE motif is restricted to within *Pseudomonas* strains and that similar Fe(II)-binding motifs cluster together. In fact, in this tree similar Fe(II)-binding motifs cluster together more tightly than the full length sensor tree (i.e., there is more distance between branches with different Fe(II)-binding motif variants).

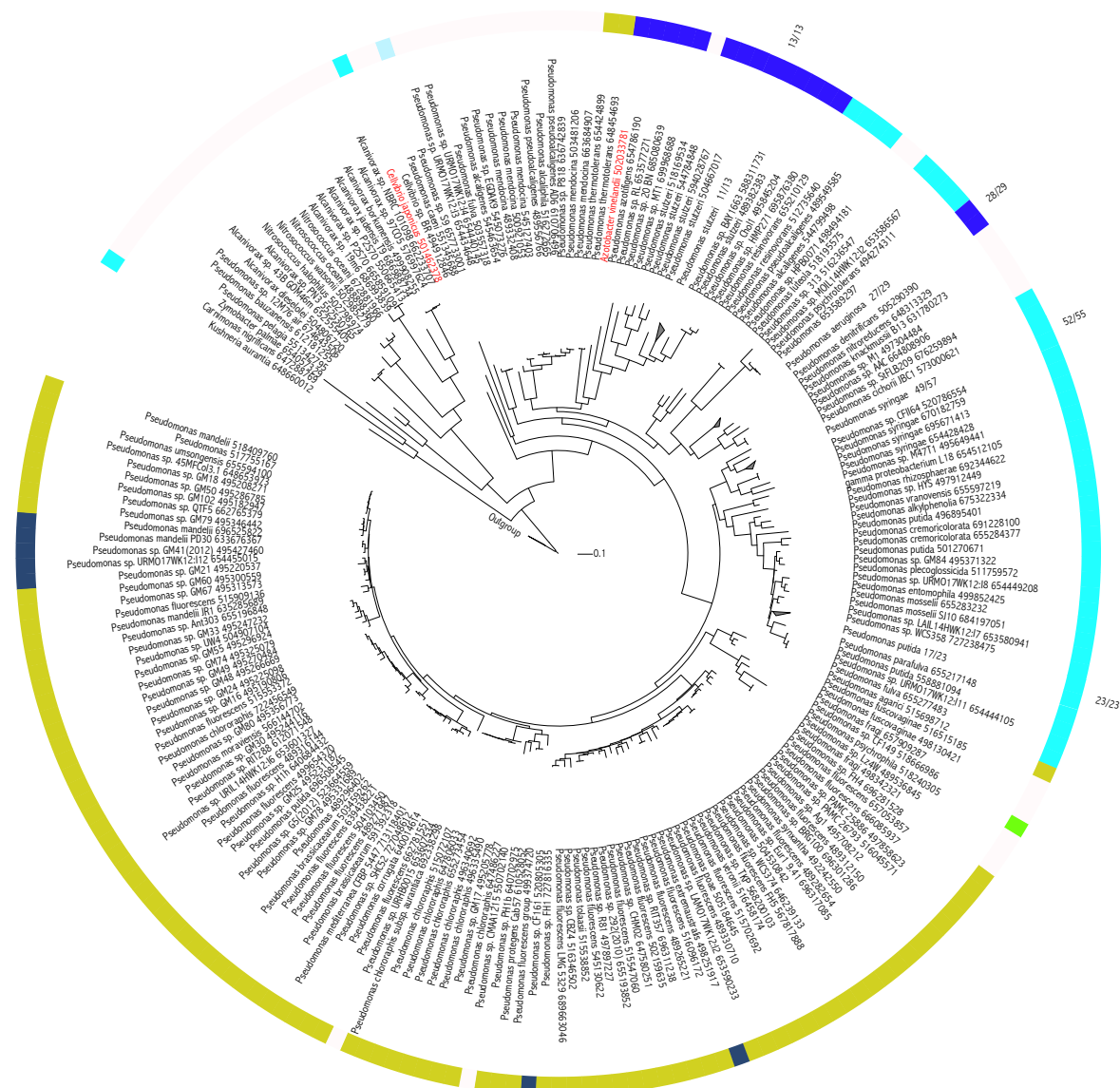


Figure B-3. The tree of the periplasmic region of BqsS with most divergent sequences that do not contain the Fe(II)-sensing motif removed. The color indicates the variant of the Fe(II)-sensing motif: blue is RExxE, cyan is RxxEx, dark yellow is RxDxE or RxxExD, light cyan is KxxEx, light green is KxDxE or KxxExD, and navy is RDxxE or RxxxD. The organisms with red text are the most distantly related phylogenetically. Triangles represent collapsed clades. The fraction immediately following the strain name is the number of the named strain in that clade out of the total number of strains within the clade. The fraction outside of the color strip indicates the number of strains within the clade that contain the

corresponding Fe(II)-binding motif out of the total number of strains within the collapsed clade. The scale bar represents 0.1 substitutions per site.

## Discussion

The RExxE Fe(II)-sensing motif both in the full length and the periplasmic domain BqsS alignment is confined to the Pseudomonadaceae family, with the vast majority of the sequences within the *Pseudomonas* genus. The same species clustered together as well as species within the same group (Fig. B-2). Sequences containing the same motif variant clustered together. The vast majority of the strains contain an R rather than an L as the first base. Arginine may be preferred because although Arg is highly positively charged, an Arg-Arg interaction could stabilize Fe(II) binding and BqsS dimerization rather than act as a repulsive force. Increased dimerization and Fe-binding stability could be mediated through the guanidinium moiety of each Arg participating in pi stacking interactions. Other proteins use this strategy (9) to stabilize interactions between domains.

Only the RExxE motif has experimentally verified Fe(II) sensing capability. The other potential Fe(II)-binding motifs should be verified both in the *in vivo* system used in Chapter 3 and biochemically with purified sensor (full length or periplasmic domain). Additionally, each glutamate should be individually mutated to see how each position contributes to Fe(II) binding. Furthermore, the sequence alignments of the experimentally validated Fe(II) sensing variants could be analyzed for other conserved residues in the periplasmic region to potentially reveal other residues critical for Fe(II) binding. For

example, in the *Salmonella* Fe(III) sensor histidine kinase, PmrB, a distal serine residue is also important for Fe(III) sensing (10).

Fe(II) sensing is not associated with pathogenicity because the motif occurs in a variety of pseudomonas strains. For example, the experimentally verified, RExxE motif is found in the human pathogenic species *P. aeruginosa* as well as the soil bacterium *P. stutzeri*. Other variants of the motif were found in the plant pathogenic species *P. syringae* as well as the typically non-pathogenic *P. fluorescens*. The Fe(II)-binding motif may have first evolved in soil because soils can frequently become anoxic (11) and soil environments can contain micromolar concentrations of Fe(II) (12). Unlike in the CF lung, where human immune systems are striving to limit free Fe (13), Fe in soil is more mobile (14). This hypothesis is supported by the phylogenetic result that Fe(II)-binding motifs are associated with pathogenic and non-pathogenic strains of *Pseudomonas*. To obtain a greater diversity of organism with this type of Fe(II) sensing, metagenomes from environmental samples rich in Fe(II) such as acid mine drainage microbial communities (15) should be examined.



## References

1. **Hunter RC, Asfour F, Dingemans J, Osuna BL, Samad T, Malfroot A, Cornelis P, Newman DK.** 2013. Ferrous iron is a significant component of bioavailable iron in cystic fibrosis airways. *mBio* **4**:e00557-13.
2. **Wedderhoff I, Kursula I, Groves MR, Ortiz de Orué Lucana D.** 2013. Iron binding at specific sites within the octameric HbpS protects *Streptomyces* from iron-mediated oxidative stress. *PLoS ONE* **8**:e71579.
3. **Gish W, States DJ.** 1993. Identification of protein coding regions by database similarity search. *Nat Genet* **3**:266-272.
4. **Katoh K, Standley DM.** 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**:772-780.
5. **Gouy M, Guindon S, Gascuel O.** 2010. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution* **27**:221-224.
6. **Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O.** 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* **59**:307-321.
7. **Letunic I, Bork P.** 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research* **39**:W475-W478.
8. **Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang J, Woyke T, Huntemann M, Anderson I, Billis K, Varghese N, Mavromatis K, Pati A, Ivanova NN, Kyripides NC.** 2014. IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Research* **42**:D560-D567.
9. **Magalhaes A, Maigret B, Hoflack J, Gomes JNF, Scheraga HA.** 1994. Contribution of unusual Arginine-Arginine short-range interactions to stabilization and recognition in proteins. *J Protein Chem* **13**:195-215.
10. **Wösten MMSM, Kox LFF, Chamnongpol S, Soncini FC, Groisman EA.** 2000. A signal transduction system that responds to extracellular iron. *Cell* **103**:113-125.
11. **Tiedje JM, Sexstone AJ, Parkin TB, Revsbech NP.** 1984. Anaerobic processes in soil. *Plant Soil* **76**:197-212.
12. **Gotoh S, Patrick WH.** 1974. Transformation of Iron in a Waterlogged Soil as Influenced by Redox Potential and pH1. *Soil Sci. Soc. Am. J.* **38**:66-71.
13. **Singh PK, Parsek MR, Greenberg EP, Welsh MJ.** 2002. A component of innate immunity prevents bacterial biofilm development. *Nature* **417**:552-555.
14. **Bongoua-Devisme AJ, Cebron A, Kassin KE, Yoro GR, Mustin C, Berthelin J.** 2012. Microbial Communities Involved in Fe Reduction and Mobility During Soil Organic Matter (SOM) Mineralization in Two Contrasted Paddy Soils. *Geomicrobiology Journal* **30**:347-361.
15. **Yelton A, Comolli L, Justice N, Castelle C, Denev V, Thomas B, Banfield J.** 2013. Comparative genomics in acid mine drainage biofilm communities reveals

metabolic and structural differentiation of co-occurring archaea. *BMC Genomics* **14**:485.